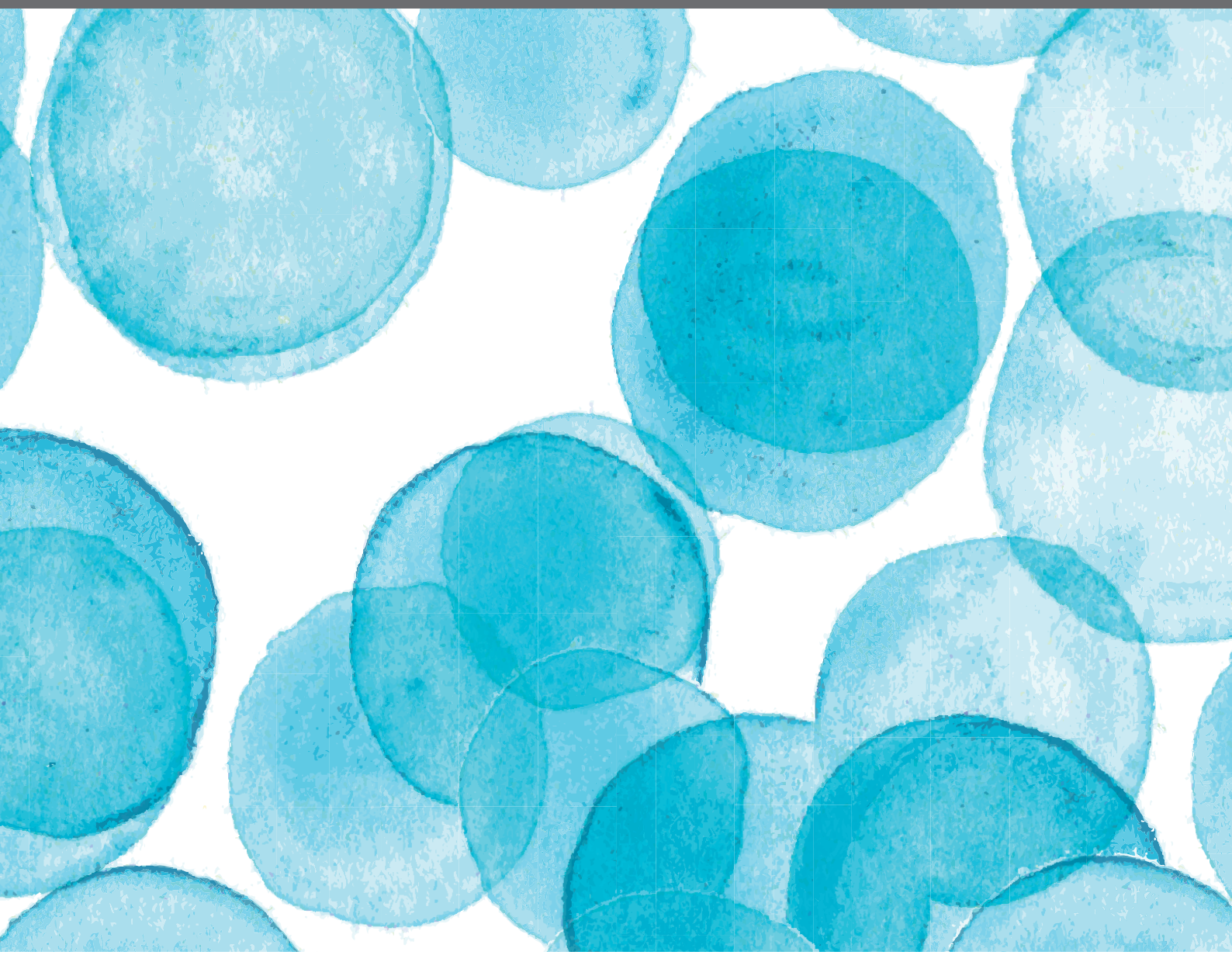
MINING, DESIGNING, MECHANISMS AND APPLICATIONS OF EXTREMOPHILIC ENZYMES

EDITED BY: Junpei Zhou, Massimiliano Fenice and Sunil Khare
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-875-7
DOI 10.3389/978-2-88971-875-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

MINING, DESIGNING, MECHANISMS AND APPLICATIONS OF EXTREMOPHILIC ENZYMES

Topic Editors:
Junpei Zhou, Yunnan Normal University, China
Massimiliano Fenice, University of Tuscia, Italy
Sunil Khare, Indian Institute of Technology Delhi, India

Citation: Zhou, J., Fenice, M., Khare, S., eds. (2021). Mining, Designing, Mechanisms and Applications of Extremophilic Enzymes. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-875-7

Table of Contents

frontiers in Microbiology

EDITORIAL
published: 25 October 2021
doi: 10.3389/fmicb.2021.709377

Editorial: Mining, Designing, Mechanisms and Applications of Extremophilic Enzymes

Massimiliano Fenice[1]*, *Sunil Kumar Khare*[2] *and Susanna Gorrasi*[1]

[1] *Dipartimento di Scienze Biologiche ed Ecologiche, University of Tuscia, Viterbo, Italy,* [2] *Department of Chemistry, Indian Institute of Technology Delhi, New Delhi, India*

Keywords: extremophiles, extremozymes, enzyme mining, enzyme purification, enzyme characterization, enzyme structure, biotechnology applications, protein engineering

Editorial on the Research Topic

Mining, Designing, Mechanisms and Applications of Extremophilic Enzymes

Enzymes play very important roles in industrial and environmental biotechnology. Their value has increased in recent decades mainly in food and detergent industries and is still growing, particularly in bioremediation (white biotechnology), and in medical, pulp/paper, textile, energy, and biosensor applications (Fernández-Lucas et al., 2017). Recent assessments have evaluated the global enzyme market at more than $7,000 million, and the business is estimated to rapidly exceed $10,000 million (Fernández-Lucas et al., 2017; Pasqualetti et al., 2019). Enzyme technologies provide economically viable and eco-friendly alternatives. The number of commercial enzymes continues to increase as we exploit, by traditional or molecular screening/selection methods, the astonishing microbial diversity to obtain new bio-catalysts and expand their range of application (Petruccioli et al., 1993; Xiao et al., 2015; Sysoev et al., 2021). Most extremophilic enzymes (extremozymes) come from microorganisms. Extremozymes remain active and stable in extreme conditions, such as high pH values, high salinity and hydrostatic pressure, and temperature extremes. They often have unique properties empowering them to satisfy the needs of many harsh industrial processes. From a commercial viewpoint, enzymes with high activity and adaptation to diverse extreme conditions show growing demand, since current applications of many existing enzymes are still constrained by activity, stability, and/or economic issues. One solution to this problem is searching for new enzymes from natural sources that show high activity/stability, uncommon specificity, and poly-extremophilic features (thermo/psychrophilic, acid/alkalophilic, and halophilic properties). However, not all extremozymes are isolated from extreme environments. Another winning strategy, which is expected to improve enzyme properties for biotechnological application, is protein engineering (Xiao et al., 2015; Liu et al., 2019). The objective of this Research Topic was to inspect new details in this growing research field, and to obtain a deeper understanding of the area, allowing us to outline its state-of-the-art: by showing enzymes that have high activity and stability in extreme or poly-extreme conditions; understanding their relative adaptation mechanisms; reporting successful mining, designing, characterization, and production of novel extremophilic enzymes also by bioinformatic tools. We cover 15 manuscripts that explore different aspects of this Research Topic.

To improve the thermostability of lipases from *Rhizopus chinensis*, Jiang et al. identified mutagenesis sites associated with enhanced flexibility based upon B-factor analysis and multiple sequence alignment. Two isoforms exhibited enhanced thermostability and improved residual activity. Novel highly thermostable mutant lipases for industrial applications can be predicted by B-factor analysis and constructed via site-directed mutagenesis.

OPEN ACCESS

Edited by:
Andreas Teske, University of North Carolina at Chapel Hill, United States

Reviewed by:
Ram Karan, King Abdullah University of Science and Technology, Saudi Arabia

Correspondence:
Massimiliano Fenice
fenice@unitus.it

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *13 May 2021*
Accepted: *29 September 2021*
Published: *25 October 2021*

Citation:
Fenice M, Khare SK and Gorrasi S (2021) Editorial: Mining, Designing, Mechanisms and Applications of Extremophilic Enzymes. Front. Microbiol. 12:709377. doi: 10.3389/fmicb.2021.709377

Yang et al. purified and characterized a thermostable laccase from *Trametes trogii* strain S0301. The enzyme matched the sequences of lcc3 in *T. trogii* BAFC463. It showed efficient dye decolorization at 60°C. This is the first report of a thermo-activated laccase from a thermophilic *Trametes trogii* strain having better properties among fungal laccases, with prospective applications in biotechnology.

Zhou et al. are reporting the heterologous expression of an invertase gene (GspInv) of *Gongronella* sp. w5 in *Komagataella pastoris*. The enzyme was characterized and shown to be an ideal candidate for high fructose syrup production with a high conversion efficiency (95%). When immobilized on cellulose, the protein retained the properties of invertase GspInv, suggesting that it could be a promising invertase for high fructose syrup preparation.

Perfumo et al. characterized enzymatic activities of a psychrophilic Antarctic strain of *Psychrobacter* sp., producing lipases and proteases. The sequence of an extracellular serine-protease was identified across *Psychrobacter* sp. genomes, and expressed in *Escherichia coli*. The purified enzyme was a cold-active alkaline protease that was found to be stable in presence of common inhibitors, compatible with detergents, and therefore suitable for new generation cold washing products.

A study by Yang et al. identified and characterized enzymes conferring salt tolerance to gut microbes. By screening the fecal metagenomic library, 10 out of the 48 salt-tolerant clones detected exhibited stronger tolerance and stability at different NaCl concentrations. High-throughput sequencing analysis showed that 91 genes encoded proteins involved in salt tolerance. Two trehalose-6-phosphate hydrolase genes were expressed in *Escherichia coli* and characterized. They were salt-tolerant and their activity increased with increasing NaCl concentration. The results indicate the existence of numerous salt-tolerant genes in gastrointestinal microbes, providing new insights into their salt-tolerance mechanisms.

Li et al. characterize the first cocrystal structure of sucrose phosphate synthase, a rate-limiting enzyme in sucrose synthesis, from *Thermosynechococcus elongatus* with UDP and sucrose-6-phosphate was characterized. It was observed that two sucrose phosphate synthase mutants lost all catalytic activity. Moreover, temperature gradient analysis shows that the enzyme exhibits the highest activity at 70°C, suggesting that it has potential uses in the industrial production of sucrose-6-phosphate.

Bhatt and Singh expressed an alkaline protease gene of *Bacillus lehensis* from a saline desert habitat in *Escherichia coli*. The recombinant protease (APrBL) belongs to the subtilase S8 protease family. After characterization, the APrBL protease turned out to be distinct from well-known commercial proteases and highly effective as a detergent additive and in whey protein hydrolysis.

The review by Zhang et al. highlights new aspects of DNA in hyperthermophilic Archaea that thrive in high-temperature environments. Although their genome is facing severe stability challenges due to the increased DNA damages by high temperature, hyperthermophilic Archaea display mutation frequencies similar to mesophilic microorganisms, indicating that the former must possess more efficient DNA repair systems to counteract the enhanced mutation rates under the harsher environment. This review is focused on advances in understanding the function of some HA proteins in DNA repair and proposes directions for future studies of an endonuclease family.

Vogler et al. characterized a novel γ-carbonic-anhydrase (γ-CA) from a polyextreme Red Sea brine pool by single-cell genome sequencing. The enzyme was expressed in *Halobacterium* sp. NRC-1 and characterized by X-ray crystallography and mutagenesis. Several possible structural determinants responsible for the enzyme's salt stability were highlighted. The study reveals insights into the halophilic γ-CA activity and its unique adaptations, providing bases for outlining strategies for salt adaptation, underlying protein evolution mechanisms, and yielding proteins with industrially valuable properties.

The review by Verma and Satyanarayana discusses xylanolytic enzymes sources, characteristics, and applications obtained through metagenomic approaches, and their amelioration by genetic engineering. Xylanolytic enzymes working under extreme conditions, which are prevalent in bioprocessing industries, are in high demand; however, their availability does not match industrial requirements and the needs of the market. Since DNA manipulations and protein-engineering are quite unsatisfactory in generating extremophilic xylanolytic enzymes, metagenomic approaches have been successfully employed to uncover hidden genes that were inaccessible by culture-dependent approaches.

The study by Shahraki et al. identified and characterized cellulases by high-throughput metagenomic data, based on optimum temperature and pH using a sequence similarity-based annotation and an ensemble of supervised learning algorithms. Two enzymes from cattle rumen were cloned, expressed, and characterized. This study highlights the strength of machine learning techniques to predict enzymatic properties based on their sequence.

Brandt et al. screened a Vietnamese fungal culture collection and identified 12 highly active xylan degraders. One of the best producers was an *Aspergillus sydowii* strain. Illumina sequencing was used for strain identification and to determine differences with the CBS reference strain. With activity based on in-gel zymography and subsequent mass-spec identification, three potential proteins responsible for xylan degradation were identified and characterized. The active site residues in both enzymes were confirmed by site-directed mutagenesis. The results justify the classification of both xylanases as highly interesting for further development.

Dong et al. identified a putative gene encoding superoxide dismutase was identified in a thermoacidophilic *Alicyclobacillus* strain and characterized for its activity in the presence of metal ions and optimal conditions of pH and temperature, which implies striking potentials as a food additive and for medical use.

In James et al., a novel transketolase has been reconstituted from two polypeptide chains encoded by a "split-gene" identified in the genome of the hyperthermophilic bacterium *Carboxydothermus hydrogenoformans*. The reconstituted enzyme was biochemically characterized. This is the first example of a reconstituted "split-gene" transketolase to be characterized, allowing its evaluation for industrial biocatalysis.

Ding et al. cloned the CDF gene of *Thermoactinomyces vulgaris* encoding a subtilisin-like protease (Als) and expressed it in *Escherichia coli*. The recombinant enzyme was secreted as a mature form (mAls), was purified and characterized, and showed high compatibility with commercial laundry detergents. Due to its polyextremotolerant properties and keratinolytic capacity, Als may find applications in various industries such as laundry detergents, food processing, non-aqueous biocatalysis, and feather processing.

In conclusion, the search for new extremozymes, together with the study of their peculiar characteristics, is still a subject of great topicality involving numerous scientists investigating the microbial biochemical diversity and looking for new or improved applications. In any case, these studies widely contribute to a deeper understanding of the microbial world and its enormous potentialities.

AUTHOR CONTRIBUTIONS

MF and SG wrote the Editorial. All authors edited and commented on it.

REFERENCES

Fernández-Lucas, J., Castañeda, D., and Hormigo, D. (2017). New trends for a classical enzyme: papain, a biotechnological success story in the food industry. *Trends Food Sci. Technol.* 68, 91–101. doi: 10.1016/j.tifs.2017.08.017

Liu, Q., Xun, G., and Feng, Y. (2019). The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnol. Adv.* 37, 530–537. doi: 10.1016/j.biotechadv.2018.10.011

Pasqualetti, M., Barghini, P., Giovannini, V., and Fenice, M. (2019). High production of chitinolytic activity in halophilic conditions by a new marine strain of *Clonostachys rosea*. *Molecules* 24:1880. doi: 10.3390/molecules24101880

Petruccioli, M., Fenice, M., and Piccioni, P. (1993). Distribution and typology of glucose oxidase activity in the genus *Penicillium*. *Lett. Appl. Microbiol.* 17, 285–288. doi: 10.1111/j.1472-765X.1993.tb01468.x

Sysoev, M., Grötzinger, S. W., Renn, D., Eppinger, J., Rueping, M., and Karan, R. (2021). Bioprospecting of novel extremozymes from prokaryotes—the advent of culture-independent methods. *Front. Microbiol.* 12:630013. doi: 10.3389/fmicb.2021.630013

Xiao, H., Bao, Z., and Zhao, H. (2015). High throughput screening and selection methods for directed enzyme evolution. *Indust. Eng. Chem. Res.* 54, 4011–4020. doi: 10.1021/ie503060a

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 19 February 2020
doi: 10.3389/fmicb.2020.00241

A Thermo-Active Laccase Isoenzyme From *Trametes trogii* and Its Potential for Dye Decolorization at High Temperature

*Xulei Yang[1], Yuanyuan Wu[1], Yu Zhang[1], En Yang[1], Yuan Qu[1], Huini Xu[1], Yuhui Chen[2], Chagan Irbis[1] and Jinping Yan[1]**

[1] Laboratory of Bioconversion, Life Science and Technology College, Kunming University of Science and Technology, Kunming, China, [2] College of Life Science, Southwest Forest University, Kunming, China

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Jing Si,
Beijing Forestry University, China
Xiaoyu Zhang,
Huazhong University of Science and Technology, China

***Correspondence:**
Jinping Yan
jpyan2019@163.com

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: 23 October 2019
Accepted: 31 January 2020
Published: 19 February 2020

Citation:
Yang X, Wu Y, Zhang Y, Yang E, Qu Y, Xu H, Chen Y, Irbis C and Yan J (2020) A Thermo-Active Laccase Isoenzyme From Trametes trogii and Its Potential for Dye Decolorization at High Temperature. Front. Microbiol. 11:241. doi: 10.3389/fmicb.2020.00241

A thermo-activation and thermostable laccase isoenzyme (Lac 37 II) produced by *Trametes trogii* S0301 at 37°C was purified to apparent homogeneity by anionic exchange chromatography and sephadex G-75 chromatography, with 12.3% of yeiled and a specific activity of 343.1 U mg^{-1}. The molecular weight of the purified Lac 37 II was estimated to be approximately 56 kDa in 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The optimal pH and temperature for the protein was 2.7 and 60°C, respectively. The purified Lac 37 II showed higher resistance to all tested metal ions and organic solvents except for Fe^{2+} and Cd^{2+} at 37°C and the activity of the purified Lac 37 was significantly enhanced by Cu^{2+} at 50 mM. The K_{cat}, K_m, and K_{cat}/K_m of Lac 37 II were 2.977 s^{-1}, 16.1 μM, and 184.9 s^{-1} μM^{-1}, respecively, in the condition of pH 2.7 and 60°C using ABTS as a substrate. Peptide-mass fingerprinting analysis showed that the Lac 37 II matched to the gene-deduced sequences of *lcc3* in *T. trogii* BAFC 463, other than *Lcc1*, *Lcc 2*, and *Lcc 4*. Compared with laccase prepared at 28°C, the onset of thermo-activation of Lac 37 II activity occurred at 30°C with an increase of 10%, and reached its maximum at the temperatures range of 40–60°C with an increase of about 40% of their original activity. Furthermore, Lac 37 II showed the efficient decolorization ability toward triphenylmethane dyes at 60°C, with decolorization rates of 100 and 99.1% for 25 mg L^{-1} malachite and crystal violet in 5 h, respectively, when hydroxybenzotriazole (HBT) was used as a mediator. In conclusion, it is the first time to report a thermo-activation laccase from a thermophilic *T. trogii* strain, which has a better enzyme property and higher decolorization ability among fungal laccases, and it also has a further application prospective in the field of biotechnology.

Keywords: *Trametes trogii*, thermoactive laccase, thermostable laccase, organic solvent tolerance, dye decolorization

INTRODUCTION

Laccases (EC1.10.3.2) are a group of copper-containing polyphenol oxidases that are known as "blue enzymes" for green chemistry due to their ability of oxidize diverse substrates which are similar to lignin or the degradation products of major lignin with molecular oxygen as the final electron acceptor. Due to their high catalytic efficiency and broad substrate specificity, laccases are used in various fields, including biopulping, delignification, biobleaching, environmental pollutants bioremediation, dye decolorization, etc. (Riva, 2006; Bertrand et al., 2017; Younes et al., 2019).

Fungi, especially white rot fungi, are the main laccase producers in nature, and the potential laccase producing strains be utilized in industrial application (Bertrand et al., 2017). Until now, many laccase-producing fungi have been studied and most of the fungi can produce several laccase isozymes (usually more than 10 isoenzymes in the same fungus strain) that showed different kinetic and physicochemical features, which makes it possible to seek new laccase isoenzymes and meets the demands in the industrial applications (Janusz et al., 2013; Zhuo et al., 2016; Zheng et al., 2017). However, the expression of different laccase isoenzymes in the same strain depends on many factors such as the presence of inducers (especially Cu^{2+} and phenolic compounds), the ratio of carbon and nitrogen, age of the culture and heat shock treatment (Baldrian, 2006; Piscitelli et al., 2011; Janusz et al., 2013; Zhuo et al., 2016; Bertrand et al., 2017). Untill now, few laccase isoenzymes have been isolated and characterized, usually one or two isoenzymes per fungus strain, and most of the laccases isoenzymes isolated so far are found sensitive to extreme conditions of temperature, pH, metal ions, etc. (Janusz et al., 2013; Fonseca et al., 2015; Jaiswal et al., 2015; Othman et al., 2018).

Laccase isoenzymes of thermophilic bacteria and fungi usually possess many attractive properties including high thermal stability, thermo-activation (stimulation of enzyme activity by pre-incubation), and tolerance to organic solvents and ionic concentrations (Hildén et al., 2007; Younes and Sayadi, 2011; Yan et al., 2014a,b), which are demanded biobleaching of pulp and treatment of colored industrial effluents (Wong et al., 2000; Asgher et al., 2008). Previous reports have found that increasing the temperature for laccase production (Tong et al., 2007) and heat shock treatment (Wang et al., 2012) in *Trametes* strains can induce the expression of different laccase isoenzymes and enhance laccase activity. In our experiment, we observed that *T. trogii* S0301 strain can grow at 37°C, but to date, only two native laccase isoenzymes (named Lcc1 and Lcc2) have been purified from this strain, and both of them were obtained from the supernatants cultured at 28°C (Colao et al., 2003; Yan et al., 2014a).

Thermo-active enzymes usually are more thermotolerant (Rathi et al., 2000; Hildén et al., 2007; Younes and Sayadi, 2011; Campos et al., 2016). Although the first thermo-active laccase was isolated as early as 1993, few thermo-active laccase isoenzymes have been isolated until now, such as laccase isoenzymes from *Fomes sclerodermeus*, *T. hirsutus*, *Coliolus zonatu*, *Marasmius quercophilus*, *Myceliophtora thermophile*, and *Scytalidium thermophilum* (Hildén et al., 2007; Younes and Sayadi, 2011). Thermo-activation has been observed in the heterogeneous expressed LCC3 of *T. trogii* BAFC 463 in *Pichia pastoris* (named the recombined LCC3) (Campos et al., 2016), but there are no reports about the thermo-active laccase isoenzyme originated from *Trametes* until now.

To explore the potential application of laccases in the thermotolerant *T. trogii* S0301 strain, laccase was obtained from this strain cultured at the temperature of 37°C. The main objectives of current study were (i) to purify and identify the laccase of *T. trogii* S0301 produced at 37°C; (ii) to characterize this laccase isoenzyme; and (iii) to assess the potential application of this laccase isoenzyme by dye decolorization experiments.

MATERIALS AND METHODS

Chemicals and Strain

2,2′-Azino bis (3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) and dyes (malachite green, bromophenol blue, and crystal violet) were purchased from Sigma-Aldrich and Merck, respectively. *T. trogii* S0301 strain employed in the present study was stored in the strain collection of Laboratory of Bioconversion of Life Science and Technology College, Kunming University of Science and Technology, and maintained on a GYP slant at 4°C (Yan et al., 2014a,b).

Laccase Production

Four 1-cm^2 plugs of the GYP plates incubated at 28°C for 4 days were excised with a sterilized cutter and added to each 250 mL Erlenmeyer flask containing 50 mL of GYP. After another 5 days incubation, the mycelia were homogenized with glass beads (0.3 mm in diameter) and transferred to GYP broth containing 2 mM Cu^{2+} with 10% (v/v) of the seed culture broths. The cultures were incubated in a rotary shaker at 200 rpm at 28 and 37°C, respectively. Ten-day-old liquid cultures were obtained by centrifuging (8000 rpm, 15 min) and the cell-free supernatants were designated as the crude enzyme for the further study.

Laccase Activity and Laccase Thermo-Activation Analysis

Laccase activity was determined with ABTS as the substrate. The 1.5 mL substrate solution includes 2 mM ABTS, 100 mM phosphate citrate buffer (pH 4.0), and 0.1 mL appropriately diluted crude or purified enzyme was used to determine the activity. The increase in absorbance was monitored at 420 nm for 3 min. One unit of the enzyme activity was defined as the amount of the enzyme that oxidized 1 μmol of the ABTS per minute according to the methods described by Yan et al. (2014b).

For thermo-activation analysis, the enzymes were pretreated at different temperatures (30–80°C) for 30 min in 100 mM phosphate citrate buffer (pH 4.0), and then thoroughly cooled on ice for another 30 min. Next, the residual laccase activity was determined. The same amount of enzymes that were not heat-treated but placed on ice as positive control, and the heat-denatured enzymes were served as the negative control. All the assays were carried out in triplicate.

Laccase Purification

The purification of the laccase from the crude enzyme of 10-day-old liquid cultures under 37°C with the addition of 2 mM Cu^{2+} was carried out as described in Yang et al. (2011). Briefly, the total protein was precipitated from the crude enzyme using ammonium sulfate (80% saturation). After dialysis, the enzyme solution was successively treated with a Q Sepharose^{TM} ion-exchange chromatography column (GE Healthcare) and a Sephadex G-75 Medium chromatography (Biotopped) column. The fraction containing laccase activity was collected and stored at −20°C for further studies. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and native sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Native-PAGE) were carried out according to Shi et al. (2014).

Laccase Identification

The purified laccase was further separated by Native-PAGE. After the electrophoresis, the gel was stained with citrate-phosphate buffer (100 mM, pH 4.0) containing 1.0 mM ABTS and the laccase band was collected for further study. MALDI-TOF/TOF-MS analysis used commercial service provided by Sangon Biotech on 4800 Plus MALDI TOF/TOFTM Analyzer (ABI, Foster City, CA, United States). Mass spectra were obtained in positive ions regime using reflectron. The program Mascot[1] was used for protein identification by "peptides fingerprints" and fragmentation spectra. The database NCBI was used for searching homology among proteins of all organisms and fungi with the accuracy mentioned taking into account possible methionine oxidation by atmospheric oxygen and possible modification of cysteine by acrylamide (Younes and Sayadi, 2011; Zheng et al., 2017).

pH and Temperature Effects on the Purified Laccase and Kinetic Analysis

Effects of pH on Laccase Activity and Laccase Stability

Characterization analysis of the purified laccase was carried out according to Yan et al. (2014b). To determine the optimum pH of the purified laccase, the laccase activity was assayed in 100 mM citrate-phosphate buffer adjusted to various pH values between 2.0 and 8.0, with 1.0 intervals at 30°C. Effect of pH on the stability of the purified laccase was studied by verifying the remaining activity after incubating the purified enzyme in the buffer solutions mentioned above at 30°C for 36 h. Next, the residual laccase activity of each treatment was compared with the control under the standard assay conditions. All the experiments were performed in triplicate.

Effects of Temperature on Laccase Activity and Laccase Stability

To determine the optimum pH of the purified laccase, enzymatic reaction was conducted at temperatures from 30 to 80°C with the enzyme in 100 mM citrate-phosphate buffer (pH 4.0), with 10°C intervals (Wu et al., 2010). For the thermostability analysis, the half-life at certain temperature ($T_{1/2}$) was determined after the purified laccase incubated at given temperature (60, 70, or 75°C) in phosphate citrate buffer (100 mM, pH 4.0) with different time intervals. The residual laccase activity was determined by the standard conditions and the activity of untreated enzyme was used to represent 100% relative activity. All the experiments were performed in triplicate.

Effects of Metal Ions on Laccase Activity

Metal ions including Na^{+}, Fe^{2+}, Cd^{2+}, Mn^{2+}, Zn^{2+}, Mg^{2+}, Co^{2+}, and Cu^{2+} were added to the standard laccase reaction mixture with a final concentration of 5 and 100 mM, and the residual activities were measured under the standard conditions (Shi et al., 2014). The laccase activity of the reaction mixture without metal ions was recorded as 100%. All the experiments were performed in triplicate.

Effects of Various Organic Solvents on Laccase Activity

To determine the effects of organic solvents on laccase activity, commonly used solvents including methanol, ethanol, and acetonitrile were added to the standard laccase reaction mixture with a final concentration of 1, 5, or 10% (v/v) and the residual activities were measured under the standard conditions (Shi et al., 2014). The laccase activity of the reaction mixture without organic solvent was recorded as 100%. All the experiments were performed in triplicate.

Kinetic Study

The kinetic constants (K_m and K_{cat}) were determined by using ABTS as substrate in series concentrations ranged from 0 to 2 mM at the optimal condition of the purified laccase in 100 mM citrate–phosphate buffer (pH 2.7) at 60°C. The laccase activity for each substrate concentration was determined three times. The K_m and K_{cat} values were evaluated by the Lineweaver–Burk plot using the Originpro 8 for Windows. All the experiments were performed in triplicate.

Dye Decolorization

Decolorization experiments were carried out according to Yan et al. (2014a,b). The dye decolorization reaction was conducted at 60°C with or without the addition of hydroxybenzotriazole (HBT) as a laccase mediator (Younes et al., 2007). Decolorization efficiency was calculated according to the following formula: decolorization (%) = $A_0 - A_t/A_0$ * 100%, where A_0 is the initial absorbance and A_t is the final absorbance (Younes et al., 2016).

Statistical Analysis

All data were presented as mean ± standard deviation (SD) for three replications for each sample. The ANOVA test using the software of Origin pro 8 for Windows used to analyze the experiment data. P-value < 0.05 was considered significant. In addition, all statistical charts were drawn by Origin pro 8 for Windows. Protein sequence alignment was analyzed by DNAMAN software.

[1] http://www.matrixscience.com

RESULTS AND DISCUSSION

Thermo-Activation Comparison of the Crude Enzymes From *T. trogii* S0301 Produced at 28 and 37°C

Previously, the thermo-active laccase isoenzymes have been observed in the thermophilic or thermotolerant strains, such as *Melanocarpus albomyces*, *M. thermophila*, and *S. thermophilum* (Kiiskinen et al., 2002; Younes and Sayadi, 2011). The original crude laccase activity of the culture filtrate produced at 37 and 28°C were 3330.62 and 4861.11 U L^{-1}, respectively. We observed that the crude laccase prepared at 37°C showed temperature-dependent activation after 30 min of thermal treatment under the different temperature conditions, while the crude laccase of this strain prepared at 28°C did not show obvious thermo-activation under the same experiment conditions (**Figure 1A**). The onset of thermo-activation of laccase activity occurred at 30°C with an increase of 10%, and reached its maximum at the temperatures range of 40–60°C with an increase of about 40% of their original activity (**Figure 1A**). Further raising temperatures caused the loss of the enzyme activity and the disappearance of thermo-activation. Until now, only two laccase isoenzymes (Lcc1 and Lcc2) have been purified from *T. trogii* strains and neither of them is thermo-active (Yan et al., 2014a; Campos et al., 2016). To further explore whether thermo-activation due to different laccase isoenzymes in the crude enzyme, laccase of this strain produced at 37°C was purified and identified.

Laccase Purification

Using Native-PAGE, two main laccase isoenzymes (Lac 37 I and Lac 37 II) were detected from the crude laccase of *T. trogii* S0301 which produced at 37°C in liquid GYP medium added with 2 mM Cu^{2+} (**Figure 1B**). Laccase which secreted in culture medium was purified after ammonium sulfate precipitation. By the first step of the purification, the specific lacase activity of crude laccase was changed from 55.3 to 96.0 U mg^{-1} (**Table 1**), followed by the anionic exchange chromatography column and the Sephadex G-75 chromatography column kept in tandem. Six protein peaks were obtained by anionic exchange chromatography column, and the first peak showed the laccase activity (111.9 U mg^{-1}) (**Table 1**). The protein solution from peak 1 was collected and then purified by Sephadex G-75 chromatography column. A final specific activity of 343.06 U mg^{-1} was achieved. A purification fold of 6.2 and a total enzyme yield of 9.6% were obtained (**Table 1**). A single laccase band detected on

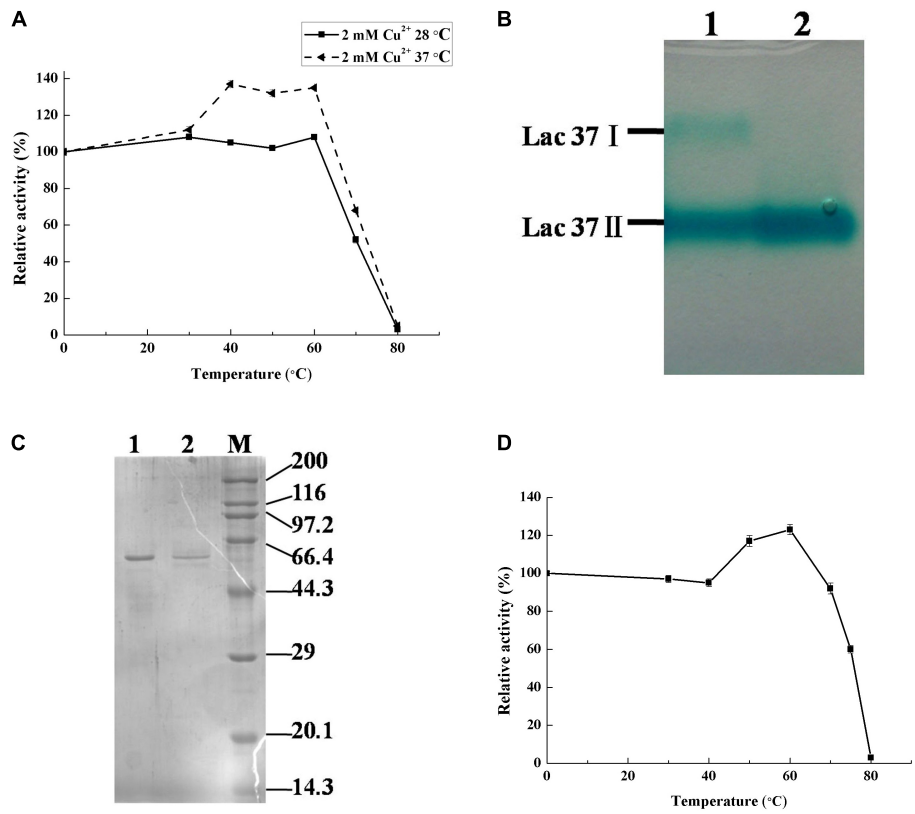


FIGURE 1 | Thermo-activated analysis and purification of laccase from *T. trogii* S0301. Native-page **(B)** and SDS-page **(C)** analysis. Lanes 1 and 2 are the crude and purified laccase produced at 37°C, respectively. The crude laccase activity of the culture filtrate produced at 37 and 28°C were 3330.62 and 4861.11 U L^{-1} **(A)**. Thermo-activated analysis of purified Lac 37 II from 37°C, which the original activity is 343.06 U mg^{-1} **(D)**. Activity is represented as % relative to the not heat-treated, assigned as 100%. Data points are the average of triplicate measurements, and the error bars represent the standard deviation.

TABLE 1 | Summary of laccase purification from *T. trogii* S0301.

Purificatin step	Total volume (mg)	Total activity (U)	Total protein (mg)	Specific activity (U mg^{-1})	Yield (%)	Purification fold
Crude laccase	1712	3330.62	59.6	55.31	100	1
Ammonium sulfate precipitation	59.57	735.03	11.137	95.99	22.07	1.74
Anionic exchange chromatography	4.83	440.5	4.3	111.91	13.22	2.02
Sephadex G-75 chromatography	3.4	321	1.02	343.06	9.6	6.2

TABLE 2 | Comparison of kinetic properties of the purified laccases mainly from strains of *Trametes* genus.

Strain	Specific activity (U mg^{-1})	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/K_m (s^{-1} μM^{-1})	Optimal condition (Tem/pH)	$T_{1/2}$ (min)	References
T. trogii S0301 at 37°C (Lac 37 II)	343.06	16.1	2977	184.96	60°C/2.7	>360 (at 60°C) 120 (at 70°C)	This study
T. trogii S0301 at 28°C	352.1	69	7958	115	45°C/3.0	180 (at 60°C)	Yan et al., 2014a
T. trogii BAFC 463[a]	–	–	–	–	50°C/4.4	>120 (at 60°C) <60 (at 70°C)	Grassi et al., 2011
T. trogii 201	152	30	3.3	0.11	–/3–3.5	–	Garzillo et al., 1998
T. trogii YDHSD		7.32	260	35.6	70°C/2.2	90 (at 60°C)	Ai et al., 2015
Recombined Lcc 1 of *T. trogii*	232	9.2	98.1	10.6	–/2.2	Lost 90% after 3 h at 60°C, pH 6	Colao et al., 2006
Recombined Lcc 2 of *T. trogii*	–	218	5.8	0.03	–/2.5	–	Colao et al., 2009
Recombined Lcc 3 of *T. trogii* BAFC 463	–	250	399	1.59	50°C/2.7	>180 (at 60°C) 45 (at 70°C)	Campos et al., 2016
T. trogi LacA	11.85	54.6			50°C/4.5	>240 (at 40°C) Unstable at 60°C	Guan et al., 2011
T. trogi LacB	4.52	17.7			60°C/4.0		
T. pubescens	18.543	105	876	8.34	50°C/5.0	120 (at 75°C)	Si et al., 2013
T. versicolor sdu-4	1320	47.5	284	99.7	–/2.2	132 (at 70°C)	Zhu et al., 2011
S. thermophilum at 42°C	139.4	260	1431	5.5	80°C/5.0	120 (at 65°C) 90 (at 70°C)	Younes et al., 2007
Cladosporium cladosporioides at 42°C	–	19.6	–	–	40–70°C/3.5	5 (at 70°C)	Halaburgi et al., 2011
Echinodontium taxodii 2538		41.4	–	–	60°C/3	>2 (at 50°C)	Shi et al., 2014

[a]*Laccase was purified but the data were based on the crude laccase. All data in this table were using ABTS as substrate.*

Native-PAGE indicated that only Lac 37 II was purified in this study. Lac 37 II showed a single band by SDS-PAGE, with a predicted molecular mass of approximately about 60 kDa (**Figure 1C**). Most studies showed that the molecular weight of fungal laccase monomer is between 50 and 100 kDa, which is consistent with the molecular weight of the purified Lac 37 II (Kunamneni et al., 2008). However, we failed to obtain the other laccase isoenzyme, Lac 37 I, in this study. In order to obtain the Lac 37 I, maybe we should change the purification conditions.

The purified Lac 37 II exhibited obvious temperature-dependent activation at temperatures from 40 to 60°C and reached its maximum at 60°C, with an increase of approximately 45% of the original activity (343.06 U mg^{-1}) (**Figure 1D**), which was similar to that of the crude enzyme of *T. trogii* S0301 produced at 37°C (**Figure 1A**). Thus, we suggest that Lac 37 II is the main source of the thermo-active laccase. In our study, the thermo-activation temperature range was agreed with other thermo-cativation laccase isoenzymes, but a highest increasing rate of laccase activity was observed in Lac 37 II, which have approximately 20% of increase compared to that of *Physisporinus rivulosus* and *S. thermophilum* (Hildén et al., 2007; Younes and Sayadi, 2011).

pH and Temperature Effects on Lac 37 II and Kinetic Analysis

According to the literature, the enzyme properties of many purified and recombined laccases mainly from *T. trogii* strains and some thermotolerant fungi are summarized in **Table 2**.

Generally, fungal laccases demonstrate their optimal pH of 2.0–6.0 using ABTS as substrate (Si et al., 2013; Shi et al., 2014; Ai et al., 2015). Lac 37 II, in this study, exhibited maximal activity at pH 2.7, which is in accordance with the recombined LCC3 and the purified laccases from *T. trogii* S0301 produced at 28°C with the optimum of pH 2.7 and 3.0 using ABTS as the substrate, respectively (**Figure 2A** and **Table 2**; Yan et al., 2014a; Campos et al., 2016). The original activity of Lac 37 II was stable, maintaining >80%, after incubation at pH 4 and 5 for 36 h (**Figure 2B**). When the pH was <3, the laccase activity was significantly inhibited, with an activity of 18%. These results are in

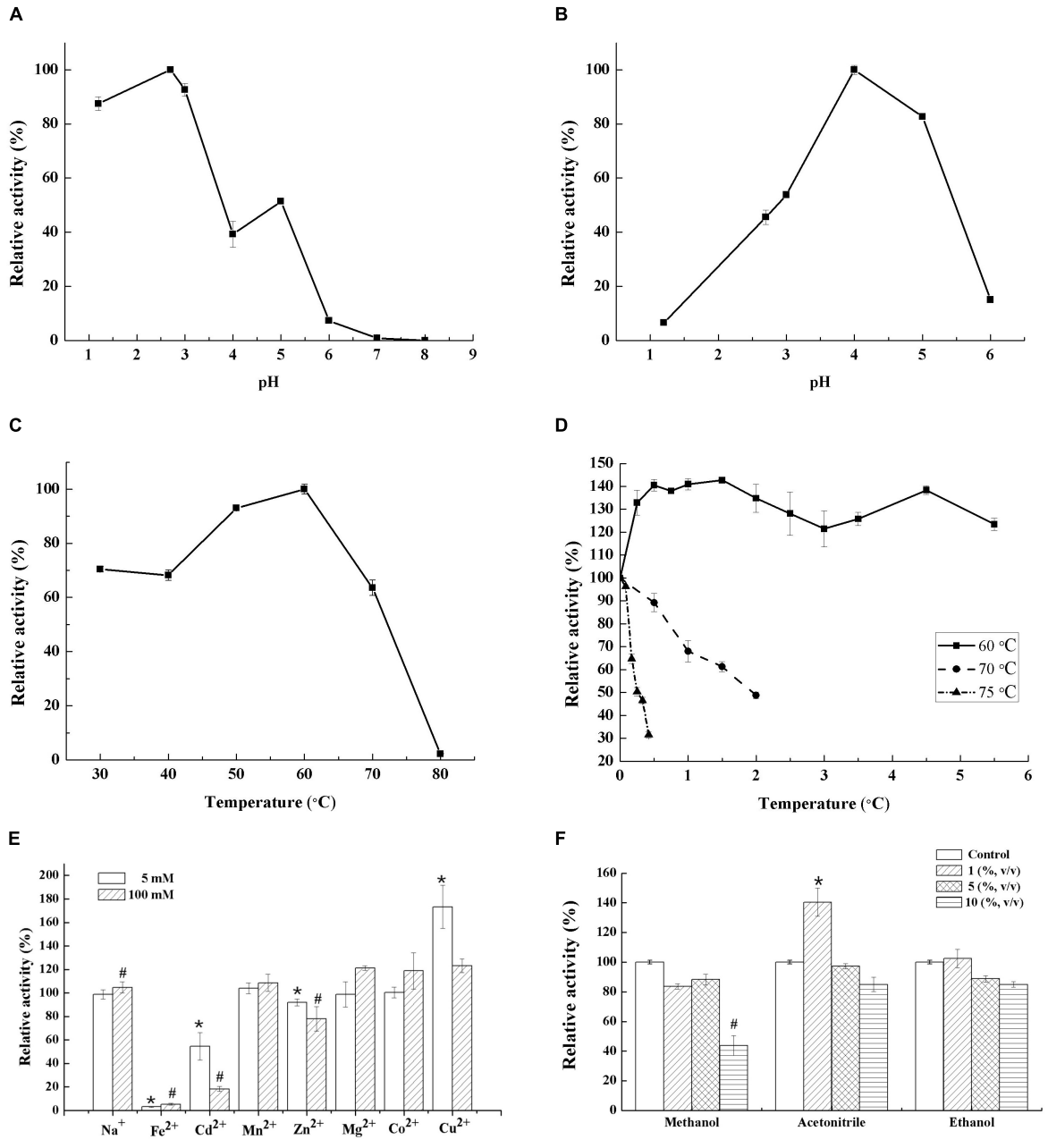


FIGURE 2 | Characterization of Lac 37 II. Effect of pH on laccase activity **(A)** and stability for 36 h **(B)**. Effect of temperature on laccase activity **(C)** and stability **(D)**. Effect of metal ions (5 or 100 mM) and organic solvent on laccase activity **(E,F)**. Laccase activity that without heat treatment, metal ions, and organic solvent at pH 4.8 was 343.06 U/mg. Activity is represented as % relative to laccase activity under the standard conditions, assigned as 100% for stability studies. The maximum laccase activity was recorded as 100% for the optimal temperature and pH studies. The data are presented as means from three independent measurements ± the standard deviations (indicated by the error bars). *p and $^{\#}p < 0.05$, as determined by one-way ANOVA.

accordance with other fungal laccases (Shi et al., 2014; Yan et al., 2014b).

The optimum temperature for Lac 37 II was 60°C with ABTS as a substrate, which was higher than that of the recombined LCC3 with the highest laccase activity at 50°C (**Figure 2C** and **Table 2**). The half-lives of enzymatic activity at various temperatures ($T_{1/2}$) at pH 4.0 were >12 h at 60°C, 2 h at 70°C, and 15 min at 75°C with ABTS as the substrate (**Figure 2D**). Based on the thermostability assays, Lac 37 II in this study exhibited a notable advantage over almost all laccases from the *Trametes* genus and other sources (**Table 2**), except for that of *T. versicolor* sdu-4 ($T_{1/2}$ of 132 min at 70°C) (Zhu et al., 2011) and *T. pubescens* ($T_{1/2}$ of 120 min at 75°C) (Si et al., 2013).

Kinetic analysis was carried out with ABTS as a substrate at optimal conditions (pH 2.7 and 60°C). The K_m, K_{cat}, and K_{cat}/K_m of Lac 37 II were 16.1 μM, 2977.8 s^{-1}, and 184.9 s^{-1} μM^{-1}, respectively (**Table 2**). Lac 37 II in this study possessed higher thermostability and catalytic efficiency, which makes the laccase isoenzyme have further prospective for the biotechnological applications.

Effect of Metal Ions on Activity of Lac 37 II

Metal ions are widely distributed in environmental pollutants, and laccases with higher resistance to metal ions are thus attractive (Younes and Sayadi, 2011). In this study, the effects of several metal ions (Na^+, Fe^{2+}, Cd^{2+}, Mn^{2+} Zn^{2+}, Mg^{2+}, Co^{2+}, or Cu^{2+}) on Lac 37 II were investigated. The purified Lac 37 II exhibited a high degree of resistance to some metal ions. When the concentration of metal ions was 5 and even 100 mM, Na^+, Mn^{2+}, Mg^{2+}, Co^{2+}, and Cu^{2+} had little effects on the laccase activity. However, the other metal ions such as Fe^{2+}, Cd^{2+}, and Zn^{2+} showed inhibitory effects on the activity of Lac 37 II, especially Fe^{2+}. Fe^{2+} completely inhibited the activity of Lac 37 II even at a low concentration (5 mM), and laccase activity decreased to 54.5 and 90.2% in the presence of Cd^{2+} and Zn^{2+} at 5 mM, respectively (**Figure 2E**). Some reports have shown that metal ions have some effects on laccase, most of which inhibit laccase activity (Hu et al., 2014; Zhuo et al., 2015). Previous studies have demonstrated that even at low concentrations, Fe^{2+} (1 or 10 mM) can strongly inhibit laccase activity in many strains, including *Pleurotus ferulae*, *Pycnoporus* sp., *T. trogii* YDHSD, *S. thermophilum*, *Trametes* sp. MA-X01, and *T. trogii* S0301 (Younes and Sayadi, 2011; Yan et al., 2014a; Ai et al., 2015; Wang et al., 2018). And the purified rLAC-EN3-1 from *P. pastoris* was also sensitive to Cd^{2+} with relative activities of 62% at 10 mM and 18% at 100 mM (Zhuo et al., 2015).

Cu^{2+}, by contrast, obviously enhanced the laccase activity of Lac 37 II with relative activities of 173.3 and 123.1% at 5 and 100 mM, respectively (**Figure 2E**), which could be due to the role of free copper ions as reducing agents in the solution and reducing the copper center in laccase (Qiao et al., 2017). Those results were in good agreement with the laccase of this strain produced at 28°C and of *T. pubescens* with a relative activity of 128% at 100 mM Cu^{2+} and 111.3% at 25 mM Cu^{2+}, respectively (Si et al., 2013; Yan et al., 2014b). Similarly, there was activation of laccase by 10 mM Cu^{2+} from *Sporothrix carnis* CPF-05. However, the laccases of *Bacillus subtilis* cjp3 and *T. trogii* YDHSD were sensitive to Cu^{2+} with relative activities of 14% (Qiao et al., 2017) and 80.9% at 10 mM (Ai et al., 2015), respectively. Younes and Sayadi (2011) reported that 100 mM Co^{2+} greatly inhibited laccase activity in *S. thermophilum* and *F. fomentarius*, compared with relative activities of 100.3 and 77.8% at 100 mM for laccases from *T. trogii* S0301 at 37 and 28°C, respectively.

Effect of Organic Solvents on Activity of Lac 37 II

Many substrates of laccases are organic pollutants that contain high concentrations of organic solvents used to enhance solubility (Maté et al., 2010). These will lead to undesirable side reactions of hydrolysis, which is not conducive to thermodynamic equilibrium and difficult to product recovery. The reaction of enzyme catalyst in various organic solvents is greatly limited (Klibanov, 2001). Similarly, the existence of organic solvents is also involved in the application of enzyme membrane immobilization, although it ensures the stability of enzyme to a certain extent (Liu et al., 2019). Thus, fungal laccases with organic cosolvent tolerance have practical uses. The effects of three common solvents (methanol, ethanol, and acetonitrile at a concentration of 1, 5, and 10%) on Lac 37 II activity was investigated. The purified Lac 37 II maintained >80% of its activity in buffer containing ethanol, even at high concentration (10%, v/v). Among all tested organic solvents, 1% (v/v) acetonitrile increased laccase activity by approximately 40.4%, while activity slightly declined to 97.3 and 85.0% of the control at 5 and 10% (v/v), respectively (**Figure 2F**). The tolerance of Lac 37 II to acetonitrile and ethanol in this study were similar to that of laccase-like enzyme from the marine sediment samples (Yang et al., 2018). The promotive effect of acetonitrile on laccase activity has been confirmed for the crude laccase of *T. trogii* LK13 (Yan et al., 2015). In addition, methanol at concentrations ranging from 1 to 5% (v/v) slightly lowered activity by 11.7–16.3%, and 10% ethanol (v/v) led to a 56.1% loss of activity. Similarly, the catalytic activity of laccase in *S. carnis* CPF-05 was almost lost when 10% of the organic solvents added (Olajuyigbe and Fatokun, 2017). In addition, the solvent tolerance of the enzyme is considered to be positively correlated with the thermal stability, which is also in line with the thermo-active and solvent tolerance of Lac 37 II in this study (Rasekh et al., 2014).

Laccase Identification

Using MALDI-TOF MS, five peptides of Lac 37 II were obtained and the sequences of them were dertermined as follows: KVIAPDGYPR, GPLVVYDPHDPHK, YSFVLEANQPK, ANPNHANFVGFNDGINSAILR, and SAGSSEYNYKNPVQR. These peptides from Lac 37 II accurately matched to the gene-deduced sequences of *lcc3* (GenBank KU055623) in *T. trogii* BAFC 463, but did not match other laccase isoenzymes of *T. trogii* strains (Lcc1, Lcc 2, Lcc 4, or the purified laccase of *T. trogii* S0301 at 28°C) (**Figure 3**). The theoretical protein molecular weight was 56 kDa, which is similar to the predicted molecular weight by SDS-PAGE. Previously, *lcc3* of *T. trogii* BAFC 463 has been expressed in *P. pastoris*, and the recombined LCC3 showed excellent thermostability and thermo-activation (Campos et al., 2016). The LCC3 was assumed to be due to the thermal stability observed in *T. trogii* BAFC 463 culture filtrates, but in their study, LCC3 did not purified from the fermentation supernate (Campos et al., 2016). Based on the results of laccase identification, we speculated that Lac 37 II purified in our study is the native LCC3, and it was the third laccase isoenzyme isolated from *T. trogii*.

We also observed that Lac 37 II showed certain great advantages over the recombined Lcc 3, such as smaller molecular

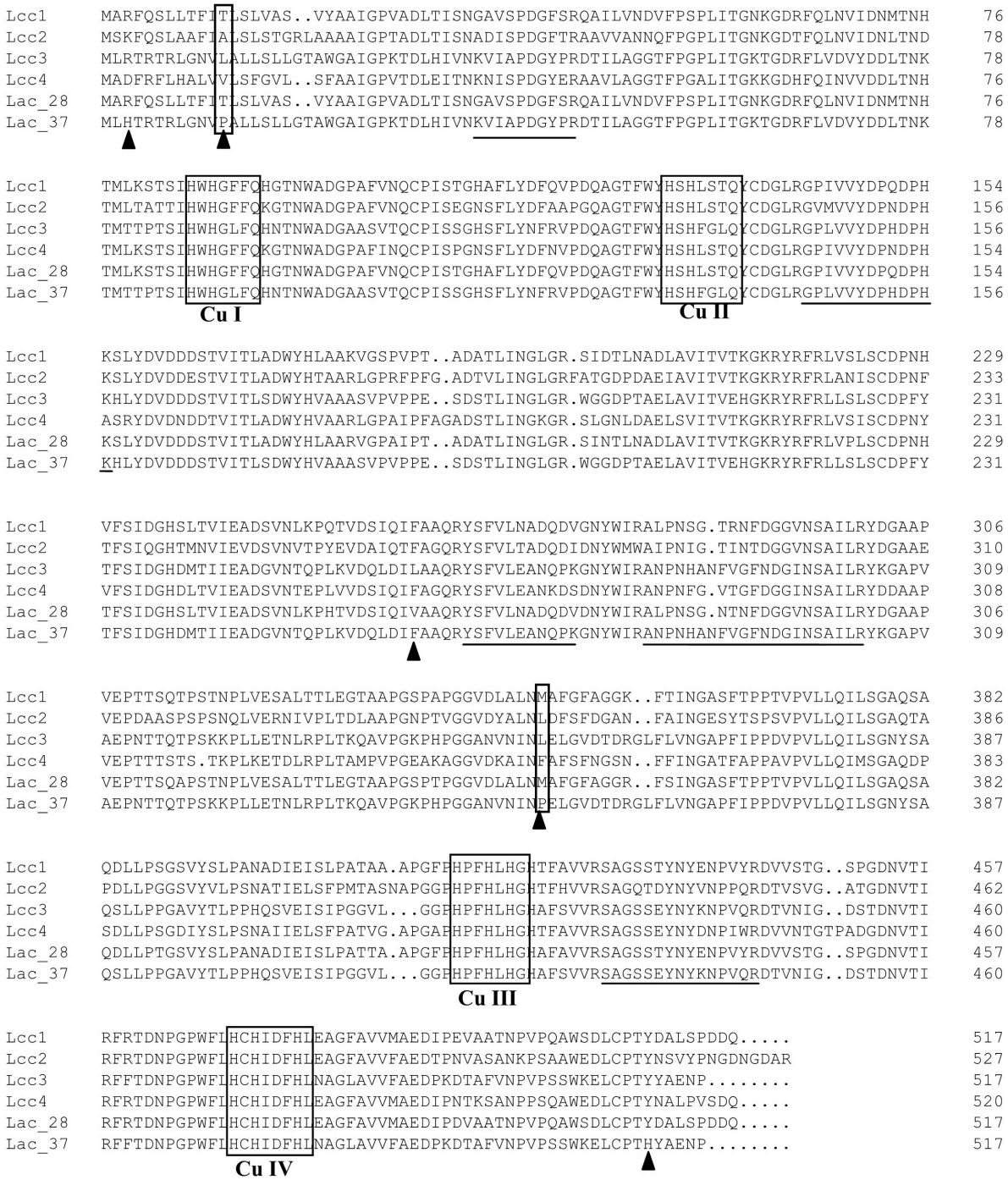


FIGURE 3 | Multiple amino acid sequences alignments of Lac 37 II with other fungal laccases of *T. trogii*. Four copper-binding conserved domains of typical laccase: CuI (HWHGFFQ), CuII (HSHLSTQ), CuIII (HPFHLHG), and CuIV (HCHIDFHL) were boxed in black. The underline showed the internal peptide sequences of Lac 37 II based on the results of MALDI-TOF MS. Lac_28 and Lac_37 were the purified laccase of *T. trogii* S0301 strain produced at 28°C and Lac 37 II at 37°C, respectively. ▲ indicated the five different amino acids in the gene-deduced sequences between Lac 37 II and Lcc 3.

weight and higher thermostability (**Figure 1**). In our study, Lac 37 II showed a smaller molecular mass than that of the recombined LCC3 in *T. trogii* BAFC 463. In addition, the optimum temperature, thermostability, and decolorizing efficiency of Lac 37 II were higher than those of the recombined LCC3 (**Table 2**).

The possible reasons for those differences are the post-translational modifications in different hosts (yeasts and

T. trogii), especially the glycosylation (Maestre-Reyna et al., 2015). Previous studies have proved that the recombined laccases in *P. pastoris* were always hyperglycosylated along with the changes of molecular mass and enzymatic properties, and the mechnism is that the glycosylation profile acts as the regulatory modules for substrate binding and turnover (Younes et al., 2007; Odón et al., 2009; Neha et al., 2012; Maestre-Reyna et al., 2015; Peter, 2016). In addition, we also found that the two amino acids (P13 and P351) of Lac37 II are different from lac3 (L13 and L351) (**Figure 3**). By predicting the protein structure of Lac 37 II, it was found that two prolines were located in the loop region. Suzuki et al. (1987) and Watanabe et al. (1991) considered that the proline may improve the thermal stability of protein by reducing the skeleton entropy of protein unfolding in the proper β-corner or random curl position. Two mutant lipases improved thermal stability by proline substitution mutagenesis, which were more stable than wild type (Mohammadi et al., 2016). Therefore, the difference of proline may be another factor that affects the thermal stability of Lac 37 II.

Dye Decolorization

The crude and purified laccase of *T. trogii* S0301 produced at 28°C and the recombined LCC3 all showed high efficiency toward decolorization of triphenylmethane dyes (Grassi et al., 2011; Yan et al., 2014a,b; Campos et al., 2016). Based on these results, two triphenylmethane dyes (malachite green and crystal violet) were chosen in this study to assess the potential application of this laccase isoenzyme. Due to the reasonably good thermostability of the purified laccase, dye decolorization analysis was carried out at 60°C in this study.

Without the addition of the laccase mediator, Lac 37 II was less effective against malachite green, and crystal violet, with maximum decolorization of 8.6% for 25 mg L^{-1} crystal violet and 16.0% for 10 mg L^{-1} malachite green in 15 h, which was similar to the recombined LCC3 (Campos et al., 2016), but much lower

TABLE 3 | Comparison of the decolorization ability of Lac 37 II with other *Trametes* laccases.

Strains	Laccase activity (U mL^{-1}) [a]	Dyes (mg L^{-1})[b]	Condition	Time (h)	Decolorization rate (%)	Laccase mediators[c]	References
T. trogii S0301 at 37°C	0.25	MG (10)	pH 4.0 at 60°C	1	100	+	This study
		MG (10)		20	16	–	
		MG (25)		3	100	+	
		CV (25)		5	100	+	
		CV (25)		20	8.6	–	
Recombined Lcc 3 of *T. trogii* BAFC 463	1	MG (18)	pH 4.5 at 60°C	24	0	–	Campos et al., 2016
		MG (18)		2	82.8	+	
T. trogii BAFC 463	6.5	MG (8)		24	98	+	Grassi et al., 2011
		MG (8)		24	25	–	
T. trogii S0301 at 28°C	0.25	MG (10)	pH 4.0 at 28°C	11	83.6	–	Yan et al., 2014a
		CV (25)		11	95.7	–	

[a]Laccase (U mL^{-1}) indicated the final laccase activity in the reaction mixture. [b]MG (malachite green), CV (crystal violet). [c]+/– indicated with or without the laccase mediators.

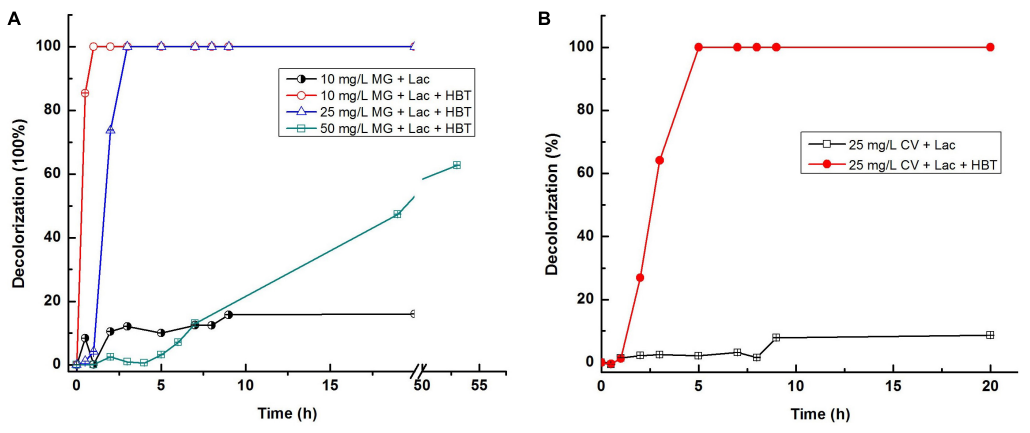


FIGURE 4 | Dye decolorization by Lac 37 II. Reactions were carried out using citrate-phosphate buffer (100 mM, pH 6.0) at 60°C. MG (malachite green) and CV (crystal violet). The degradation of malachite green **(A)** and crystal violet **(B)** were carried out using citrate-phosphate buffer (100 mM, pH 6.0) at 60°C. Each value is the mean value ± standard error of the mean of triplicates. Each value is the mean value ± standard error of the mean of triplicates.

than those of the purified laccase from *T. trogii* S0301 produced at 28°C with the maximum decolorization of 95.7% in 11 h for all dyes at the same concentration (Yan et al., 2014a) (**Table 3**).

Previous studies have indicated that the natural or artificial laccase mediators can increase decolorization of dyes by both the purified and crude laccases from many strains, including *T. trogii* and *T. villosa* (Grassi et al., 2011; Campos et al., 2016). To confirm whether laccase mediators can improve the decolorization ability of Lac 37 II, 1-HBT was chosed as a laccase mediator. When HBT was added at the concentration of 2 mM, the highest decolorization rate of Lac 37 II was detected as approximately 100% for 10 mg L^{-1} malachite green within 1 h (**Figure 4A**). Lac 37 II efficiently decolorized 25 mg L^{-1} malachite green, with maximum decolorization of 73.7 and 99.1% in 2 and 3 h, respectively (**Figure 4A**). To explore the decolorization ability of Lac 37 II at elevated concentrations of malachite green, 50 mg L^{-1} malachite green was used. Lac 37 II was able to decolorize with maximum decolorization of 47.3 and 62.7% in 20 and 53 h, respectively (**Figure 4A**). HBT also greatly enhanced the decolorization efficiency of Lac 37 II for crystal violet, with decolorization of 26.9, 64.1, and 99.1% in 2, 3, and 5 h, respectively, while Lac 37 II without HBT showed almost no effect on crystal violet even after 20 h of incubation, with decolorization of 9.1% at 20 h (**Figure 4B**).

In this study, HBT was chosen as the sole laccase mediator, and a greater increase in decolorization rate was obtained for all tested dyes. However, HBT has been demonstrated as a laccase mediator with a slight or no effect on dyes decolorization (Canas and Camarero, 2010; Campos et al., 2016). Thus, laccase mediators, especially more natural and effective ones, can be optimized in future studies to enhance the decolorization efficiency of Lac 37 II.

CONCLUSION

In this study, Lac 37 II, a novel native laccase isoenzyme of *Trametes trogii* S0301 was obtained by incubating this strain at 37°C, which is higher than the normal cultivation temperature of fungi. By purification and identification, we found that Lac 37 II is the third native laccase isoenzyme from *T. trogii* strains, and it is also the first thermo-active and the more thermostable isoenzyme of *Trametes* genus strains. With higher thermostability and catalytic efficiency, this laccase isoenzyme can efficiently decolorize triphenylmethane dyes with the addition of a laccase mediator, which makes Lac 37 II have further prospective for biotechnological applications.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

XY has carried out enzyme activity determination, thermal stability analysis, and laccase purification. YW was responsible for the effect of temperature and pH on the activity and stability of laccase. YZ was responsible for the effect of organic dissolution and metal ions on laccase activity. EY was responsible for the identification of laccase. YQ was responsible for the decolorization of dyes by laccase. HX was responsible for SDS-PAGE and Native-PAGE. YC was responsible for the preparation of laccase. CI has revised the original manuscript. JY has conceived the experiment plan, supervised the experiment process, and wrote the original manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 31560036).

ACKNOWLEDGMENTS

The authors gratefully thank Professor Zhou Tongxin and her team for their valuable help.

REFERENCES

Ai, M. Q., Wang, F. F., and Huang, F. (2015). Purification and characterization of a thermostable laccase from *Trametes trogii* and its ability in modification of kraft lignin. *J. Microbiol. Biotechnol.* 25, 1361–1370. doi: 10.4014/jmb.1502.02022

Asgher, M., Bhatti, H. N., Ashraf, M., and Legge, R. L. (2008). Recent developments in biodegradation of industrial pollutants by white rot fungi and their enzyme system. *Biodegradation* 19, 771–783. doi: 10.1007/s10532-008-9185-3

Baldrian, P. (2006). Fungal laccases - occurrence and properties. *FEMS Microbiol. Rev.* 30, 215–242. doi: 10.1111/j.1574-4976.2005.00010.x

Bertrand, B., Martínez-Morales, F., and Trejo-Hernández, M. R. (2017). Upgrading laccase production and biochemical properties: strategies and challenges. *Biotechnol. Prog.* 33, 1015–1034. doi: 10.1002/btpr.2482

Campos, P. A., Levin, L. N., and Wirth, S. A. (2016). Heterologous production, characterization and dye decolorization ability of a novel thermostable laccase isoenzyme from *Trametes trogii* BAFC 463. *Process Biochem.* 51, 895–903. doi: 10.1016/j.procbio.2016.03.015

Canas, A. I., and Camarero, S. (2010). Laccases and their natural mediators: biotechnological tools for sustainable eco-friendly processes. *Biotechnol. Adv.* 28, 694–705. doi: 10.1016/j.biotechadv.2010.05.002

Colao, M. C., Caporale, C., Silvestri, F., Ruzzi, M., and Buonocore, V. (2009). Modeling the 3-D structure of a re combinant laccase from *Trametes trogii* active at a pH close to neutrality. *Protein J.* 28, 375–383. doi: 10.1007/s10930-009-9204-1

Colao, M. C., Garzillo, A. M., Buonocore, V., Schiesser, A., and Ruzzi, M. (2003). Primary structure and transcription analysis of a laccase-encoding gene from the *Basidiomycetetrametes trogii*. *Appl. Microbiol. Biotechnol.* 63, 153–158. doi: 10.1007/s00253-003-1429-x

Colao, M. C., Lupino, S., Garzillo, A. M., Buonocore, V., and Ruzzi, M. (2006). Heterologous expression oflcc1gene from *Trametes trogii* in pichia pastoris and characterization of the recombinant enzyme. *Microb. Cell Fact.* 5, 31–40. doi: 10.1186/1475-2859-5-31

Fonseca, M. I., Farina, J. I., Sadanoski, M. A., D'Errico, R., Villalba, L. L., and Zapata, P. D. (2015). Decolorization of Kraft liquor effluents and biochemical

characterization of laccases from *Phlebia brevispora* BAFC 633. *Int. Biodeter. Biodegr.* 104, 443–451. doi: 10.1016/j.ibiod.2015.07.014

Garzillo, A. M. V., Colao, M. C., Caruso, C., Caporale, C., Celletti, D., and Buonocore, V. (1998). Laccase from the white-rot fungus *Trametes trogii*. *Appl. Microbiol. Biotechnol.* 49, 545–551. doi: 10.1007/s002530051211

Grassi, E., Scodeller, P., Filiel, N., Carballo, R., and Levin, L. (2011). Potential of *Trametes trogii* culture fluids and its purified laccase for the decolorization of different types of recalcitrant dyes without the addition of redox mediators. *Int. Biodeter. Biodegr.* 65, 635–643. doi: 10.1016/j.ibiod.2011.03.007

Guan, Y., Hua, S., Zhao, X. H., Zhang, Q. H., Zhong, L. J., Zhu, W. W., et al. (2011). Isolation, purification and enzymatic features of laccase from *Trametes trogii*. *J. Microbiol.* 31, 64–68. doi: 10.3724/SP.J.1077.2011.00073

Halaburgi, V. M., Sharma, S., Sinha, M., Singh, T. P., and Karegoudar, T. B. (2011). Purification and characterization of a thermostable laccase from the ascomycetes *Cladosporium cladosporioides* and its applications. *Process Biochem.* 46, 1146–1152. doi: 10.1016/j.procbio.2011.02.002

Hildén, K., Hakala, T. K., Maijala, P., Lundell, T. K., and Hatakka, A. (2007). Novel thermotolerant laccases produced by the white-rot fungus *Physisporinus rivulosus*. *Appl. Microbiol. Biotechnol.* 77, 301–309. doi: 10.1007/s00253-007-1155-x

Hu, X., Wang, C., Wang, L., Zhang, R., and Chen, H. (2014). Influence of temperature, pH and metal ions on guaiacol oxidation of purified laccase from *Leptographium qinlingensis*. *World J. Microbiol. Biotechnol.* 30, 1285–1290. doi: 10.1007/s11274-013-1554-3

Jaiswal, N., Pandey, V. P., and Dwivedi, U. N. (2015). Purification of a thermostable alkaline laccase from papaya (*Carica papaya*) using affinity chromatography. *Int. J. Biol. Macromol.* 72, 326–332. doi: 10.1016/j.ijbiomac.2014.08.032

Janusz, G., Kucharzyk, K. H., Pawlik, A., Staszczak, M., and Paszczynski, A. J. (2013). Fungal laccase, manganese peroxidase and lignin peroxidase: gene expression and regulation. *Enzyme Microb. Technol.* 52, 1–12. doi: 10.1016/j.enzmictec.2012.10.003

Kiiskinen, L. L., Viikari, L., and Kruus, K. (2002). Purification and characterisation of a novel laccase from the ascomycete *Melanocarpus albomyces*. *Appl. Microbiol. Biotechnol.* 59, 198–204. doi: 10.1007/s00253-002-1012-x

Klibanov, A. M. (2001). Improving enzymes by using them in organic solvents. *Nature* 409, 241–246. doi: 10.1038/35051719

Kunamneni, A., Plou, F. J., Ballesteros, A., and Alcalde, M. (2008). Laccases and their applications: a patent review. *Recent Pat. Biotechnol.* 2, 5–12. doi: 10.2174/187220808783330965

Liu, C., Saeki, D., Cheng, L., Luo, J., and Matsuyama, H. (2019). Polyketone-based membrane support improves the organic solvent resistance of laccase catalysis. *J. Colloid Interface Sci.* 544, 230–240. doi: 10.1016/j.jcis.2019.03.003

Maestre-Reyna, M., Liu, W. C., Jeng, W. Y., Lee, C. C., Hsu, C. A., Wen, T. N., et al. (2015). Structural and functional roles of glycosylation in fungal laccase from *Lentinus* sp. *PLoS One* 10:e0120601. doi: 10.1371/journal.pone.0120601

Maté, D., García-Burgos, C., García-Ruiz, E., Ballesteros, A. O., Camarero, S., and Alcalde, M. (2010). Laboratory evolution of high-redox potential laccases. *Chem. Biol.* 17, 1030–1041. doi: 10.1016/j.chembiol.2010.07.010

Mohammadi, M., Sepehrizadeh, Z., Ebrahim-Habibi, A., Shahverdi, A. R., Faramarzi, M. A., and Setayesh, N. (2016). Enhancing activity and thermostability of lipase a from *Serratia marcescens* by site-directed mutagenesis. *Enzyme Microb. Technol.* 9, 18–28. doi: 10.1016/j.enzmictec.2016.07.006

Neha, G., Nora, B., Tenzin, K., Meenu, C., Marion, A. S., and Saroj, M. (2012). Cloning, sequence analysis, expression of *Cyathus bulleri* laccase in *Pichia pastoris* and characterization of recombinant laccase. *BMC Biotechnol.* 12:75. doi: 10.1186/1472-6750-12-75

Odón, V.-V., Palomares, L. A., Dantán-González, E., Ayala-Castro, H. G., Martínez-Anaya, C., Valderrama, B., et al. (2009). The role of *N*-glycosylation on the enzymatic activity of a *Pycnoporus sanguineus* laccase. *Enzyme Microb. Technol.* 45, 233–239. doi: 10.1016/j.enzmictec.2009.05.007

Olajuyigbe, F. M., and Fatokun, C. O. (2017). Biochemical characterization of an extremely stable ph-versatile laccase from *Sporothrix carnis* CPF-05. *Int. J. Biol. Macromol.* 94, 535–543. doi: 10.1016/j.ijbiomac.2016.10.037

Othman, A. M., Elsayed, M. A., Elshafei, A. M., and Hassan, M. M. (2018). Purification and biochemical characterization of two isolated laccase isoforms from *Agaricus bisporus* CU13 and their potency in dye decolorization. *Int. J. Biol. Macromol.* 113, 1142–1148. doi: 10.1016/j.ijbiomac.2018.03.043

Peter, G. (2016). Effects of glycosylation on the enzymatic activity and mechanisms of proteases. *Int. J. Mol. Sci.* 17:1969. doi: 10.3390/ijms17121969

Piscitelli, A., Giardina, P., Lettera, V., Pezzella, C., Sannia, G., and Faraco, V. (2011). Induction and transcriptional regulation of laccases in fungi. *Curr. Genom.* 12, 104–112. doi: 10.2174/138920211795564331

Qiao, W., Chu, J., Ding, S., Song, X., and Yu, L. (2017). Characterization of a thermo-alkali-stable laccase from *Bacillus subtilis* cjp3 and its application in dyes decolorization. *J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng.* 52, 1–8. doi: 10.1080/10934529.2017.1301747

Rasekh, B., Khajeh, K., Ranjbar, B., Mollania, N., Almasinia, B., and Tirandaz, H. (2014). Protein engineering of laccase to enhance its activity and stability in the presence of organic solvents. *Eng. Life Sci.* 14, 442–448. doi: 10.1002/elsc.201300042

Rathi, P., Bradoo, S., Saxena, R. K., and Gupta, R. (2000). A hyper-thermostable, alkaline lipase from *Pseudomonas* sp. with the property of thermal activation. *Biotechnol. Lett.* 22, 495–498. doi: 10.1023/a:1005604617440

Riva, S. (2006). Laccases: blue enzymes for green chemistry. *Trends Biotechnol.* 24, 219–226. doi: 10.1016/j.tibtech.2006.03.006

Shi, L., Yu, H., Dong, T., Kong, W., Ke, M., Ma, F., et al. (2014). Biochemical and molecular characterization of a novel laccase from selective lignin-degrading white-rot fungus *Echinodontium taxodii* 2538. *Process Biochem.* 49, 1097–1106. doi: 10.1016/j.procbio.2014.03.028

Si, J., Peng, F., and Cui, B. (2013). Purification, biochemical characterization and dye decolorization capacity of an alkali-resistant and metal-tolerant laccase from *Trametes pubescens*. *Bioresour. Technol.* 128, 49–57. doi: 10.1016/j.biortech.2012.10.085

Suzuki, Y., Oishi, K., Nakano, H., and Nagayama, T. (1987). A strong correlation between the increase in number of proline residues and the rise in thermostability of five *Bacillus* oligo-1,6-glucosidase. *Appl. Microbiol. Biotechnol.* 26, 546–551. doi: 10.1007/BF00253030

Tong, P. G., Hong, Y. Z., Xiao, Y. Z., Zhang, M., Tu, X. M., and Cui, T. J. (2007). High production of laccase by a new basidiomycete, *Trametes* sp. *Biotechnol. Lett.* 29, 295–301. doi: 10.1007/s10529-006-9241-1

Wang, F., Guo, C., Wei, T., Zhang, T., and Liu, C. Z. (2012). Heat shock treatment improves *Trametes versicolor* laccase production. *Appl. Biochem. Biotechnol.* 168, 256–265. doi: 10.1007/s12010-012-9769-6

Wang, Q., Ding, L., and Zhu, C. (2018). Characterization of laccase from a novel isolated white-rot fungi *Trametes* sp. *MA-X*01 and its potential application in dye decolorization. *Biotechnol. Biotechnol. Equip.* 32, 1477–1485. doi: 10.1080/13102818.2018.1517028

Watanabe, K., Chishiro, K., Kitamura, K., and Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* kp1006. *J. Biol. Chem.* 266, 24287–24294. doi: 10.1016/S1567-4215(03)90519-4

Wong, K. K. Y., Richardson, J. D., and Mansfield, S. D. (2000). Enzymatic treatment of mechanical pulp fibers for improving papermaking properties. *Biotechnol. Prog.* 16, 1025–1029. doi: 10.1021/bp000064d

Wu, Y. R., Luo, Z. H., Chow, K. K., and Vrijmoed, L. L. P. (2010). Purification and characterization of an extracellular laccase from the anthracene-degrading fungus *Fusarium solani* MAS2. *Bioresour. Technol.* 101, 9772–9777. doi: 10.1016/j.biortech.2010.07.091

Yan, J., Chen, D., Yang, E., Niu, J., Chen, Y., and Chagan, I. (2014a). Purification and characterization of a thermotolerant laccase isoform in *Trametes trogii* strain and its potential in dye decolorization. *Int. Biodeter. Biodegr.* 93, 186–194. doi: 10.1016/j.ibiod.2014.06.001

Yan, J., Niu, J. Z., Chen, D. D., Chen, Y. H., and Chagan, I. (2014b). Screening of *Trametes* strains for efficient decolorization of malachite green at high temperatures and ionic concentrations. *Int. Biodeter. Biodegr.* 87, 109–115. doi: 10.1016/j.ibiod.2013.11.009

Yan, J. P., Chen, Y. H., Niu, J. Z., Chen, D. D., and Chagan, I. (2015). Laccase produced by a thermotolerant strain of *Trametes trogii*. *Braz. J. Microbiol.* 46, 59–65. doi: 10.1590/S1517-838246120130895

Yang, Q., Zhang, M., Zhang, M., Wang, C., Liu, Y., Fan, X., et al. (2018). Characterization of a novel, cold-adapted, and thermostable laccase-like enzyme with high tolerance for organic solvents and salt and potent dye decolorization ability, derived from a marine metagenomic library. *Front. Microbiol.* 9:2998. doi: 10.3389/fmicb.2018.02998

Yang, Y., Ma, F., Yu, H., Fan, F., Wan, X., Zhang, X., et al. (2011). Characterization of a laccase gene from the white-rot fungi *Trametes* sp. 5930 isolated from Shennongjia nature reserve in China and studying on the capability of decolorization of different synthetic dyes. *Biochem. Eng. J.* 57, 13–22. doi: 10.1016/j.bej.2011.07.006

Younes, S. B., Cherif, I., Dhouib, A., and Sayadi, S. (2016). *Trametes trogii*: a biologic powerful tool for dyes decolorization and detoxification. *Catal. Lett.* 146, 204–211. doi: 10.1007/s10562-015-1629-x

Younes, S. B., Khedher, S. B., Zhang, Y., Geissen, S. U., and Sayadi, S. (2019). Laccase from *Scytalidium thermophilum*: production improvement, catalytic behavior and detoxifying ability of diclofenac. *Catal. Lett.* 149, 1833–1844. doi: 10.1007/s10562-019-02771-1

Younes, S. B., Mechichi, T., and Sayadi, S. (2007). Purification and characterization of the laccase secreted by the white rot fungus *Perenniporia tephropora* and its role in the decolourization of synthetic dyes. *J. Appl. Microbiol.* 102, 1033–1042.

Younes, S. B., and Sayadi, S. (2011). Purification and characterization of a novel trimeric and thermotolerant laccase produced from the ascomycete *Scytalidium thermophilum* strain. *J. Mol. Catal. B Enzym.* 73, 35–42. doi: 10.1016/j.molcatb.2011.07.014

Zheng, F., An, Q., Meng, G., Wu, X. J., Dai, Y. C., Si, J., et al. (2017). A novel laccase from white rot fungus, *Trametes orientalis*: purification, characterization, and application. *Int. J. Biol. Macromol.* 102, 758–770. doi: 10.1016/j.ijbiomac.2017.04.089

Zhu, Y., Zhang, H., Cao, M., Wei, Z., Huang, F., and Gao, P. (2011). Production of a thermostable metal-tolerant laccase from *Trametes versicolor* and its application in dye decolorization. *Biotechnol. Bioprocess Eng.* 16, 1027–1035. doi: 10.1007/s12257-011-0129-0

Zhuo, R., He, F., Zhang, X., and Yang, Y. (2015). Characterization of a yeast recombinant laccase rLAC-EN3-1 and its application in decolorizing synthetic dye with the coexistence of metal ions and organic solvents. *Biochem. Eng. J.* 93, 63–72. doi: 10.1016/j.bej.2014.09.004

Zhuo, R., Yuan, P., Yang, Y., Zhang, S., Ma, F., and Zhang, X. (2016). Induction of laccase by metal ions and aromatic compounds in *Pleurotus ostreatus* HAUCC 162 and decolorization of different synthetic dyes by the extracellular laccase. *Biochem. Eng. J.* 117(Pt B), 62–72. doi: 10.1016/j.bej.2016.09.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 March 2020
doi: 10.3389/fmicb.2020.00346

Improving the Thermostability of *Rhizopus chinensis* Lipase Through Site-Directed Mutagenesis Based on B-Factor Analysis

Zhanbao Jiang[1†], Chengbo Zhang[2†], Minyuan Tang[1], Bo Xu[1,2,3,4], Lili Wang[5], Wen Qian[5], Jiandong He[5], Zhihong Zhao[5], Qian Wu[1,2,3,4], Yuelin Mu[1,2,3,4], Junmei Ding[1,2,3,4], Rui Zhang[1,2,3,4], Zunxi Huang[1,2,3,4] and Nanyu Han[1,2,3,4]**

[1] School of Life Sciences, Yunnan Normal University, Kunming, China, [2] Engineering Research Center of Sustainable Development and Utilization of Biomass Energy, Ministry of Education, Yunnan Normal University, Kunming, China, [3] Key Laboratory of Yunnan for Biomass Energy and Biotechnology of Environment, Yunnan Normal University, Kunming, China, [4] Key Laboratory of Enzyme Engineering, Yunnan Normal University, Kunming, China, [5] Yunnan Walvax Biotechnology Co., Ltd., Kunming, China

OPEN ACCESS

Edited by:
Sunil Khare,
Indian Institute of Technology Delhi, India

Reviewed by:
Rani Gupta,
University of Delhi, India
Likui Zhang,
Yangzhou University, China

***Correspondence:**
Zunxi Huang
huangzunxi@163.com
Nanyu Han
ha0001yu@e.ntu.edu.sg

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: 20 October 2019
Accepted: 17 February 2020
Published: 03 March 2020

Citation:
Jiang Z, Zhang C, Tang M, Xu B, Wang L, Qian W, He J, Zhao Z, Wu Q, Mu Y, Ding J, Zhang R, Huang Z and Han N (2020) Improving the Thermostability of Rhizopus chinensis Lipase Through Site-Directed Mutagenesis Based on B-Factor Analysis. Front. Microbiol. 11:346. doi: 10.3389/fmicb.2020.00346

In order to improve the thermostability of lipases derived from *Rhizopus chinensis,* we identified lipase (Lipr27RCL) mutagenesis sites that were associated with enhanced flexibility based upon B-factor analysis and multiple sequence alignment. We found that two mutated isoforms (Lipr27RCL-K64N and Lipr27RCL-K68T) exhibited enhanced thermostability and improved residual activity, with respective thermal activity retention values of 37.88% and 48.20% following a 2 h treatment at 50°C relative to wild type Lipr27RCL. In addition, these Lipr27RCL-K64N and Lipr27RCL-K68T isoforms exhibited 2.4- and 3.0-fold increases in enzymatic half-life following a 90 min incubation at 60°C. Together these results indicate that novel mutant lipases with enhanced thermostability useful for industrial applications can be predicted based upon B-factor analysis and constructed via site-directed mutagenesis.

Keywords: lipase, thermostability, B-factor analysis, multiple sequence alignment, site-directed mutagenesis

INTRODUCTION

Lipases (triacylglycerol acyl hydrolases, EC 3. 1. 1. 3) mediate the hydrolysis of triglycerides into monoglycerides, fatty acids, diglycerides, and glycerol at oil-water interfaces (Naik et al., 2010), and can further catalyze interesterification, esterification, alcoholysis, acidolysis, and aminolysis in non-aqueous environments (Pandey et al., 1999; Yu et al., 2014). Owing to their unique catalytic properties, lipases are utilized in a wide range of industrial contexts, such as in the production of food, leather, pharmaceuticals, and bioenergy (Persson et al., 2002; Ferreira-Dias et al., 2013).

Lipases are present in all forms of life, from microbes to mammals. Microbe-derived lipases exist in a wide variety of forms (Gupta et al., 2015). Compared with lipases from animals or plants, microbe-derived lipases can operate across a wider range of temperatures and pH values (Liu et al., 2017). *Rhizopus*-derived lipases have been shown to exhibit superior stability in acidic conditions, making them ideal for industrial applications. Dozens of commercial *Rhizopus* lipases have been produced to date, with *Rhizopus chinensis* CCTCC M201021 being the most commonly used strain for lipase production (Sha et al., 2013). However, lipases derived from *R. chinensis* are most active at

moderate temperatures, whereas the reactions catalyzed by lipases in the context of oil processing require higher temperatures (70–90°C) that can lead to the deactivation of lipases lacking heat resistance. As such, it is vital that heat-resistant lipases should be developed so that they may be utilized as superior biocatalysts (Yu et al., 2012a,b).

Many different approaches have been employed to improve lipase kinetic and thermodynamic stability (Bommarius and Paye, 2013; Stepankova et al., 2013; Su et al., 2014). Directed evolution and semi-rational design are the two most common protein engineering approaches used to generate thermostable mutants of target enzymes (Li et al., 2018). Directed evolution through error-prone PCR and DNA shuffling is an effective means of improving the performance of enzymes under high-temperature conditions (Suen et al., 2004; Goomber et al., 2016). Screening through the many colonies necessary to identify superior enzyme isoforms, however, is time-consuming and cannot be effectively performed in host species that grow slowly. An alternative approach to improving enzymatic thermal stability that has been implemented successfully is the B-factor (Xie et al., 2014; Han et al., 2017). This approach relies upon improving thermostability via increasing enzymatic rigidity at certain sites, with B-factors derived from X-ray data offering insight into atom fluctuations and rigidity relative to their equilibrium positions (Ringe and Petsko, 1986). In the present study, we normalized the B-factor values of the Lipr27RCL X-ray structure and thereby identified residues with pronounced flexibility. Following multiple sequence alignment, we then conducted site-directed mutagenesis to improve *R. chinensis* lipase thermostability.

MATERIALS AND METHODS

Materials

Pichia pastoris GS115 was from Invitrogen (Shanghai, China). A site-directed mutagenesis kit and DMT chemically competent cells were from TransGen (Beijing, China). The *R. chinensis* lipr27RCL gene cloned in the pPIC9K vector was deposited in our laboratory. We obtained 4-nitrophenol palmitate (4-NPP) from Sigma. Mutagenic primers were synthesized by Shuoqing (Kunming, China). All other chemicals were commercially available analytical grade materials.

Gene Cloning and Site-Directed Mutagenesis

For site-directed mutagenesis experiments, we utilized pPIC9K-*r27RCL* as a template to construct four single-mutant lipase constructs (*Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T,* and *Lipr27RCL-K68T*) via introduction of point mutations into *lipr27RCL* through site-directed mutagenesis with a Fast MultiSite Mutagenesis System based on provided directions. The primers for these mutations are compiled in **Table 1**. We transformed the resultant PCR products into *P. pastoris*, and DNA sequencing was used to confirm the identity of the resulting recombinant strains. PCR thermocycler settings were as follows: 5 min at 94°C, then 28 cycles of 30 s at 94°C, 2.5 min at 55°C, and 2 min at 72°C.

TABLE 1 | Oligonucleotide primers for Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T, and Lipr27RCL-K68T.

Primers	Primer sequences
r27RCL-K64N F	5′-GTCGTTCCAGGTACCAATTGGGACTGTAAG-3′
r27RCL-K64N R	5′-GAACATACTTGAGACATTGCTTACAGTCCCAA-3′
r27RCL-W65I F	5′-GTTCCAGGTACCAAGATTGACTGGGACTGT-3′
r27RCL-W65I R	5′-AGGAACATACTTGAGACATTGCTTACAGTCAAT-3′
r27RCL-D66T F	5′-ACCAAGTTGGACTGTACGCAATGTCTCAAG-3′
r27RCL-D66T R	5′-CTTACCATCAGGAACATACTTGAGACATTGCG-3′
r27RCL-K68T F	5′-ACCAAGTGGGACTGTACGCAATGTCTCAAG-3′
r27RCL-K68T R	5′-CTTACCATCAGGAACATACTTGAGACATTGCG-3′

Protein Expression and Purification

*Sal*I was used to linearize the WT pPIC9K/*Lipr27RCL* and mutant constructs, which were then individually transformed into *P. pastoris* GS115 via electroporation. Yeast extract peptone dextrose (YPD) medium supplemented with 200 μg/mL G-418 (Geneticin) was used to select for clones, which were then grown for 2 days at 30°C in a 50 mL volume of buffered glycerol-complex media. This media was then exchanged for 50 mL of buffered methanol-complex media to drive expression of the lipase proteins. Protein purification was conducted using an Amicon centrifugal filter device (cutoff 10.000). Proteins were then sequentially purified via Q Sepharose^{TM} Fast Flow anion exchange column chromatography and Phenyl-Sepharose 4 FF hydrophobic chromatography column chromatography, after which SDS-PAGE was used to confirm purification results (**Figure 1** and **Supplementary Figure S3**), and a protein quantification kit was used to measure enzyme levels.

Assessment of Enzymatic Activity

We defined one unit of enzymatic activity (1 U) as the quantity of enzyme necessary to mediate the release of 1 μmol p-nitrophenol (p-NP) per minute from 4-NPP (Li et al., 2011). Thermostability was analyzed through assessment of the residual enzymatic activity following incubation at 50°C for 5, 10, 15, 20, 30, 40, 50, 65, 80, 100, and 120 min, and at 60°C for 1, 3, 6, 10, 15, 20, 25, 30, 35, 45, 60, 75, and 90 min at the established optimal pH.

To establish key kinetic parameters for these purified lipases (K_m, V_{max}, and k_{cat}), enzymes were measured in a Tris–HCl (pH 9) buffer solution at 37°C with 4-PNN as a substrate provided across a range of concentrations (0.078125–10 mM), with reactions monitored based on p-NP production. Lineweaver–Burk plots were then used to fit the results and to determine these kinetic parameter values (Lineweaver and Burk, 1934).

Lipase r27RCL Temperature Factor Calculation

B-factor values for all Lipr27RCL Cα atoms were extracted from the PDB file, these B-factors were normalized such that they had a zero mean and unit variance distribution as follows: $B' = \frac{B - \langle B \rangle}{\sigma(B)}$, where the ⟨b⟩ is the average of all Cα atoms and σ(B) is the standard deviation of the

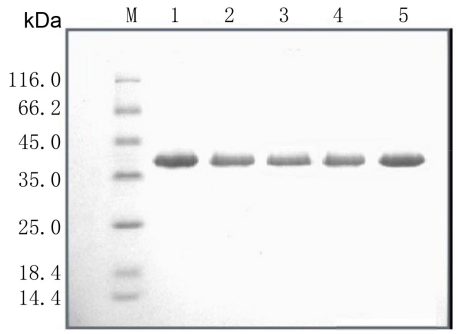


FIGURE 1 | SDS-PAGE analysis of the recombinant lipases. M: standard protein molecular mass markers; Lanes 1–5: Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65, Lipr27RCL-D66T, and Lipr27RCL-K68T, respectively.

B-factors for the individual protein (Yuan et al., 2003). We have successfully used these equations in previous studies (Han et al., 2017).

Circular Dichroism (CD)

Circular dichroism spectra were recorded using a Circular Dichroism (Model: Chirascan, Instrument: 0547) from Applied Photophysics Limited (United Kingdom). All spectra were recorded at 20°C. Conditions, including pathlength: 1 mm, time-per-point: 1 s (25us × 40000), step size: 1 nm, and bandwidth: 1 nm, were used for scanning the 200–250 nm spectra range. CD spectra were collected from the solution of 50 mM sodium phosphate buffer (PBS, pH 7.4) with the lipase concentration of 0.1 mg/ml. After three scans, the final spectrum was corrected by removing the recorded baseline of the PBS control medium.

MD Simulation Details

The X-ray crystal structure of Lipr27RCL was taken from PDB 6A0W, while structures for the four recombinant lipases Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T, and Lipr27RCL-K68T, were constructed using the SWISS-MODEL server (Arnold et al., 2006), using the default parameters. Normal MD simulations of Lipr27RCL, Lipr27RCL-K64N, and Lipr27RCL-K68T were performed at 60°C. After 1000-step energy minimization, all systems were first equilibrated for 5 ns in NVT ensemble and then equilibrated for 5 ns in NPT ensemble by restraining all heavy atoms, and each system was simulated for 30 ns. All systems were solvated with TIP3P waters in an octahedral box, and the minimal distance between each protein and edge of the box was set to 1.0 nm (Jorgensen et al., 1983). Sodium and chloride ions were added with a concentration of 100 nM to neutralize the systems. The GROMACS program suite version 4.5.7 and Amber ff99SB force field were applied in all simulations (Hornak et al., 2006; Hess et al., 2008). All simulations were performed in an isothermal-isobaric ensemble (60°C, 1 bar).

RESULTS

Prediction of Mutagenesis Sites Based on B-Factor Analysis

There is a linear relationship between B-factors determined based upon X-ray diffraction data and the mean square displacement of atoms relative to their average positions. As such, B-factors derived from protein crystal structures offer invaluable insights into the flexibility, stability, and dynamics of individual proteins. In the present study, we extracted and normalized the B-factor values of Cα atoms for Lipr27RCL from its crystal structure (**Figure 2A**). Residues from T63 to K68 (63-TKWDCK-68) of Lipr27RCL exhibited a high degree of flexibility. Multiple sequence alignment was then performed using 92 lipase sequences derived from thermophilic fungal lipases in the NCBI database, and beneficial sequences from residues 63 to 68 (63-TNITCT-68) were thereby discovered (**Figure 2B**). Based on the results from this normalized B-factor analysis and multi-sequence alignment, we thus identified K64N, W65I, D66T, and K68T as putative sites that may improve Lipr27RCL thermostability.

Construction and Characterization of Mutant Lipr27RCL Isoforms

Based on the results from B-factor analysis and multiple sequence alignment, we used site-directed mutagenesis to produce mutant forms of Lipr27RCL in which the 64K, 65W, 66D, and 68K residues had been mutated to 64N, 65I, 66T, and 68T, respectively. The resultant single mutant lipases (K64N, W65I, D66T, and K68T) were termed Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL- D66T, and Lipr27RCL-K68T for the purposes of this study. These lipases had the same molecular mass (39.50 kDa) as did Lipr27RCL when analyzed via SDS-PAGE (**Figure 1**). Following purification, we found the specific activities of Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T, and Lipr27RCL-K68T to be 2218.52, 2379.71, 1961.14, 2313.66, and 2449.40 U/mg, respectively.

We next assessed the changes in lipase thermostability as a result of this mutagenic campaign by assessing residual activity of these four Lipr27RCL enzymes and the WT isoform following a 90 min incubation at a range of temperatures. We found that all five of these enzymes functioned best at 40°C (**Figure 3**). Both Lipr27RCL-K64N and Lipr27RCL-K68T remained stable at 50°C, with residual activities of 66.12 and 76.44% after a 2 h incubation at this temperature, respectively. In contrast, the residual activities of the other lipase isoforms declines substantially under these conditions. Lipr27RCL and Lipr27RCL-D66T had half-lives of 65 min under these conditions, with 28.24 and 28.60% residual activities, respectively, after a 2 h treatment. Lipr27RCL-W65I had a half-life of just 40 min and a residual activity of 10.09% following this 2 h treatment (**Figure 4A**). At temperatures above 50°C, Lipr27RCL-K64N and Lipr27RCL-K68T were also more stable than the other tested enzymes. Lipr27RCL-K64N retained 49.23% activity following 1 h at 60°C, which was approximately the half-life of this enzyme, and it retained 29.84% activity following

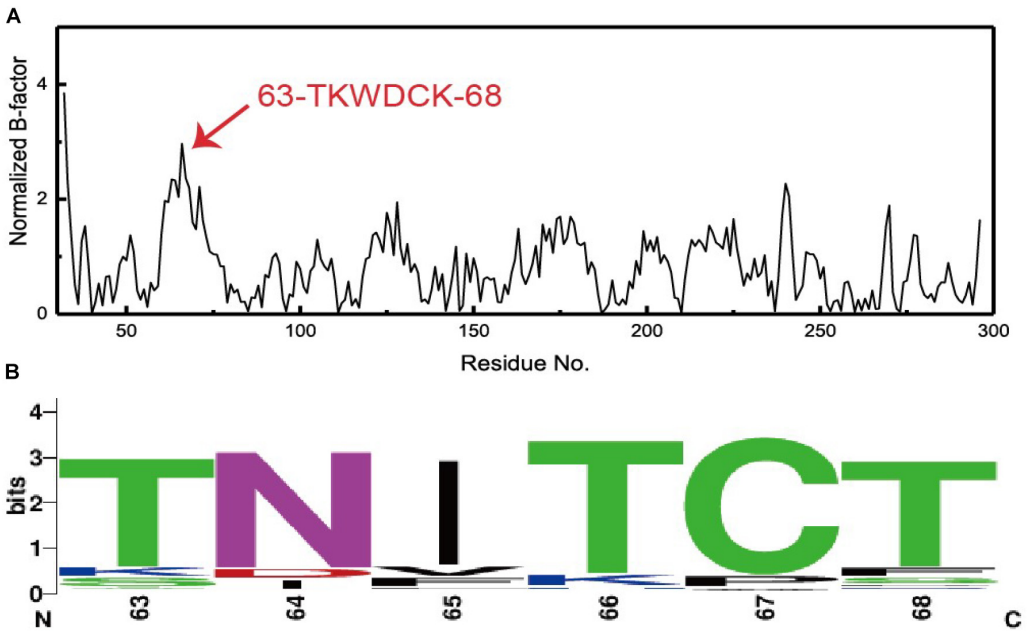


FIGURE 2 | (A) Normalized B-factor Lipr27RCL and **(B)** multiple sequence alignment based on 92 lipase sequences from different species of thermophilic fungi.

a 90 min incubation at this temperature. Lipr27RCL-K68T retained 51.44% activity after a 75 min incubation at 60°C, which was close to its half-life, and it retained 41.47% activity following a 90 min incubation at this temperature. In contrast, Lipr27RCL retained 53.28% activity after only 25 min at 60°C, with just 5.34% activity being retained following 90 min at this temperature. Similarly, Lipr27RCL-D66T retained 48.90% activity after 20 min at 60°C, with 4.44% activity having been retained after 90 min at this temperature. After just 10 min at 60°C, Lipr27RCL-W65I retained just 48.9% activity, with no activity at all remaining after 75 min at this temperature (**Figure 4B**). Relative to the WT lipase isoform, Lipr27RCL-K64N and Lipr27RCL-K68T exhibited residual activities which were increased by 37.88% and 48.20%, respectively, after 120 min at 50°C, and their respective half-lives had increased by 2.4- and 3.0-fold at 60°C for 90 min. Lipr27RCL-D66T performed nearly identically to WT Lipr27RCL, whereas Lipr27RCL-W65I had poorer thermostability than did WT Lipr27RCL. These findings thus suggested that two of the tested amino acid substitutions (K64N and K68T) were advantageous for Lipr27RCL thermostability.

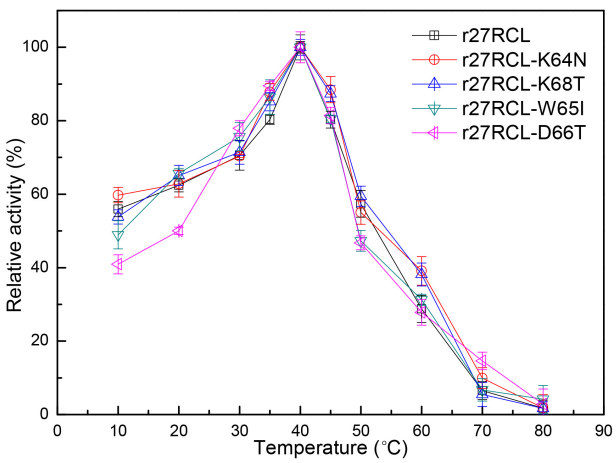


FIGURE 3 | The impact of temperature on recombinant Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65, Lipr27RCL-D66T, and Lipr27RCL-K68T.

Kinetic Analysis of Mutant Lipr27RCL Isoforms

We assessed the kinetic parameters of these five lipase isoforms at a pH of 9.0 at 37°C. The *p*-NP method was used to monitor these reactions, with purified proteins being combined with a range of 4-PNN concentrations (0.078125–10 mM). Kinetic measurements revealed that the apparent Michaelis constant (K_m) values for Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T, and Lipr27RCL-K68T were 0.36, 0.29, 0.45, 0.38, and 0.29 mM, respectively (**Table 2**). The smaller K_m values for Lipr27RCL-K64N and Lipr27RCL-K68T were indicative of the increased kinetic efficiency of these enzymes relative to WT Lipr27RCL. The catalytic efficiency (k_{cat}/K_m) of Lipr27RCL-K64N and Lipr27RCL-K68T was also increased (1.22-fold and 1.24-fold, respectively). The K_m and k_{cat}/K_m values for Lipr27RCL-D66T were similar to those for WT Lipr27RCL. However, the K_m of Lipr27RCL-W65I was 1.25-fold that of Lipr27RCL, with a clear decline in k_{cat}/K_m (**Table 2**). These kinetic analyses thus revealed that the K64N and K68T substitutions,

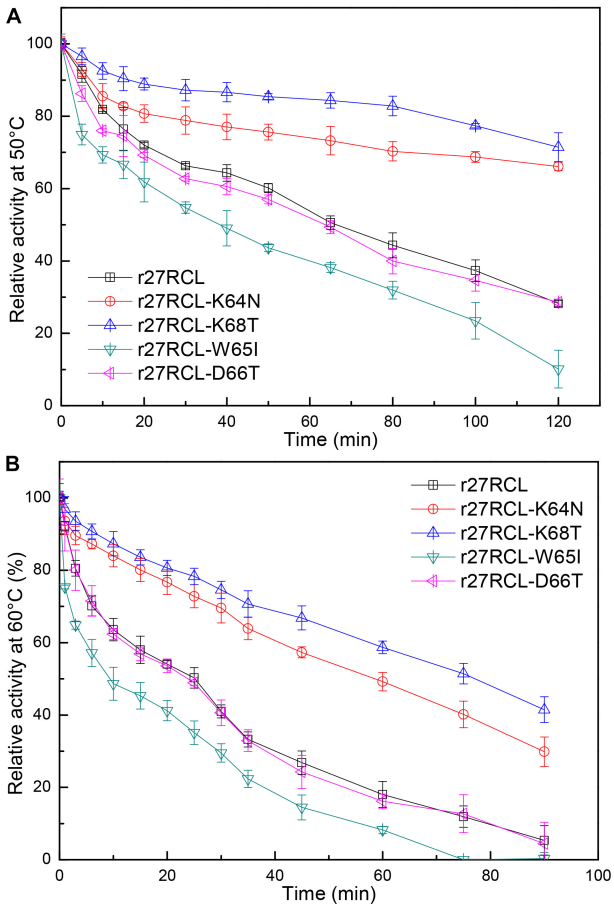


FIGURE 4 | The thermostability of Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65, Lipr27RCL-D66T, and Lipr27RCL-K68T at 50°C **(A)** and 60°C **(B)**.

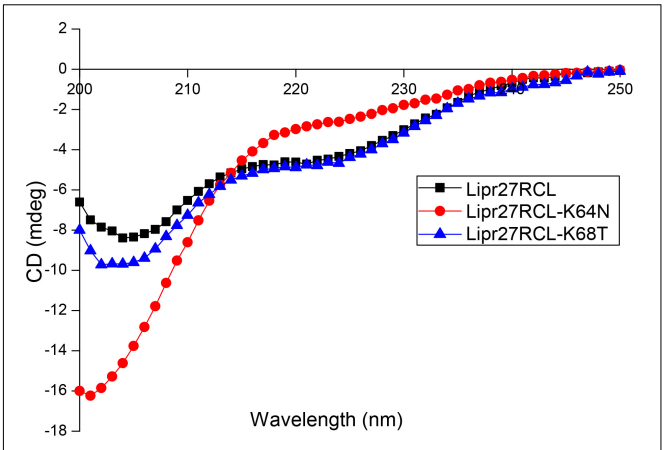


FIGURE 5 | Circular dichroism spectra (CD) of Lipr27RCL, Lipr27RCL-K64N, and Lipr27RCL-K68T.

although primarily intended to improve thermostability, also enhanced the catalytic efficiency and substrate binding of these lipases.

Circular Dichroism (CD) Spectroscopy

We selected Lipr27RCL-K64N and Lipr27RCL-K68T with improved thermal stability using CD spectroscopy to compare with Lipr27RCL to prove whether the single mutation caused changes in the secondary structure of Lipr27RCL. CD technology can give information under the protein conformation in solution through the dependence of the optical activity of a peptide chain, with almost no side chain interference (except aromatic amino acids) (Nascimento et al., 2019). The CD spectra of the three lipase isoforms (Lipr27RCL, Lipr27RCL-K64N, and Lipr27RCL-K68T) are shown in **Figure 5**. In addition, to investigate the thermal denaturation and renaturation of these three isoforms, we heat-treated them at 65°C for 5 min and placed them on ice, and then performed CD spectrum detection. At the same time, the renaturation samples were heat-treated at 65°C and placed on ice, and then recovered at 4°C for 2 days for CD spectrum scan (**Supplementary Figure S1**).

The CD spectra showed that compared with Lipr27RCL, the ellipticities of the α-helix and β-sheet of the Liper27RCL-K64N isoform changed obviously, while the ellipticity of α-helix of Lipr27RCL-K68T isoform had minor changes with almost no changes of β-sheet (**Figure 5**). These suggested that mutations at these two single sites might cause changes in the secondary structure of the Lipr27RCL isoform. In addition, the CD spectra of Liper27RCL-K64N and Liper27RCL-K68T were also different, indicating that there were some differences in the thermal stability of the two mutants. Moreover, the CD spectra of the three thermal denaturation isoforms showed that only the ellipticities of β-sheet of the Lipr27RCL-K64N and Lipr27RCL-K68T isoforms changed after heat treatment, while the ellipticities of α-helix and β-sheet of Lipr27RCL both changed (**Supplementary Figure S1**). These indicated that the thermal stability of Lipr27RCL-K64N and Lipr27RCL-K68T isoforms might be improved. However, the CD spectra of the three isoforms of thermal denaturation and renaturation showed that the ellipticities of the α-helix and β-sheet of these isoforms had little change before and after thermal denaturation and renaturation (**Supplementary Figure S1**). These indicated that it was difficult for them to recover enzyme activity after denaturation.

TABLE 2 | Kinetics of Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-W65, Lipr27RCL-D66T, and Lipr27RCL-K68T.

Enzymes	V_{max}	K_m	k_{cat}	k_{cat}/K_m
	(U/mg)	(mM)	(/s)	(/s/mM)
Lipr27RCL	714.29 ± 23.57	0.36 ± 0.06	188.10 ± 18.21	522.50
Lipr27RCL-K64N	588.24 ± 19.63	0.29 ± 0.09	184.86 ± 18.46	637.45
Lipr27RCL-W65	500.00 ± 23.02	0.45 ± 0.04	131.50 ± 19.38	292.22
Lipr27RCL-D66T	709.00 ± 19.03	0.38 ± 0.05	186.66 ± 16.28	491.21
Lipr27RCL-K68T	714.29 ± 20.21	0.29 ± 0.06	187.95 ± 13.65	648.10

Enhanced Thermal Tolerance of the Single Mutants Explored by MD Simulations

To fully understand the basis of enhanced thermostability in these recombinant lipases, MD simulations for Lipr27RCL and two single mutants (Lipr27RCL-K64N and Lipr27RCL-K68T) with improved thermostability were conducted at 60°C. To investigate the improved thermostability of mutants, novel interactions formed with the mutational sites were monitored during the simulations. It is discovered hydrogen bonding interactions between N64···S58 in Lipr27RCL-K64N was stronger than that in K64···S58 Lipr27RCL. Specifically, the hydrogen bond forming probability between K/N64···S58 were 70.6% and 62.5% in Lipr27RCL-K64N and Lipr27RCL, respectively (**Figures 6A,B**). Additionally, comparing the root mean square fluctuation (RMSF) values which reflect the flexibility of the mutated residues in the three lipases, we found that RMSF value of the mutated residue N64 in Lipr27RCL-K64N was smaller than that of K64 in Lipr27RCL, indicating improved stability of the mutated residue (**Table 3**). The side chain of N64 is smaller than that of K64, and as a result N64 exhibited increased local stability owing to its mild floating amplitude, and this coupled with the retention of normal hydrogen bond interactions led to the enhanced thermostability of Lipr27RCL-K64N.

Similarly, RMSF value of the mutated residue T68 in Lipr27RCL-K68T was smaller than that of K68 in Lipr27RCL, and the smaller RMSF value of T68 indicate the mutated site was relative stable during simulation (**Table 3**). T68 in Lipr27RCL-K68T also had a shorter side chain relative to that of K68, and the α-helix where T68 located was thus more stable than that of the wild type enzyme.

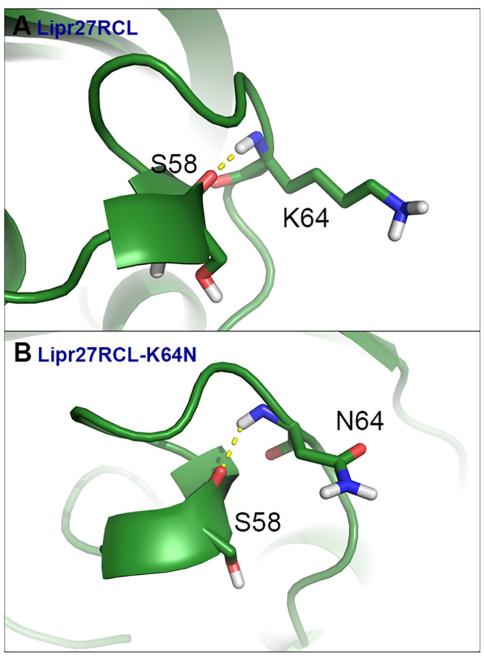


FIGURE 6 | MD simulations Of Lipr27RCL **(A)** and Lipr27RCL-K64N **(B)** conducted at 60°C.

TABLE 3 | RMSF for mutated residues over the whole simulation in Lipr27RCL and two mutants.

Enzymes	K/N64 (nm)	K/T68 (nm)
Lipr27RCL	0.1576	0.1995
Lipr27RCL-K64N	0.1187	0.1786
Lipr27RCL-K68T	0.1636	0.1478

DISCUSSION

The development of thermostable lipases has clear and practical potential to meet with industrial demands. Furthermore, generating such stable lipases may simultaneously improve their kinetic efficiency (Rogers and Bommarius, 2010; Han et al., 2017). In the present study, we utilized a B-factor comparison and multi-sequence alignment strategy to guide our efforts to design and produce thermostable lipase mutants. The resultant Lipr27RCL-K64N and Lipr27RCL-K68T proteins, which were designed via site-directed mutagenesis, exhibited both higher thermostability and superior catalytic efficiency.

Normalized B-factor analysis of Lipr27RCL was the key step in the design of thermostable recombinant isoforms of this enzyme, given that B-factors offer insight into protein fluctuations, which are indicative of atom rigidity relative to their corresponding positions (Ringe and Petsko, 1986), and previous work has shown that improved thermostability depends upon achieving a higher degree of rigidity (Podar and Reysenbach, 2006). As such, we analyzed normalized B-factor values from Lipr27RCL, and aligned this sequence with those of other thermophilic lipases using sequences from NCBI. We were thereby able to identify residues with pronounced degrees of flexibility, and we then constructed recombinant mutant lipases (Lipr27RCL-K64N, Lipr27RCL-W65I, Lipr27RCL-D66T, and Lipr27RCL-K68T) in which these flexible residues had been mutated. In subsequent thermostability and kinetic analyses, we found that Lipr27RCL-K64N and Lipr27RCL-K68T exhibited superior thermostability, substrate affinity, and catalytic efficiency. As such, these results confirm that B-factor analysis can facilitate efforts to selectively improve lipase thermostability profiles.

Based on sequence and structural analyses, we found further evidence suggesting that Lipr27RCL-K64N and Lipr27RCL-K68T may exhibit greater thermostability relative to other untested lipase isoforms, leading us to focus our in depth characterization efforts on these isoforms. We had also constructed a quadruple mutant lipase (K64N, W65I, D66T, and K68T) for this study, but found its thermostability to be 40% lower than the WT Lipr27RCL in a matter of minutes. This is likely at least in part due to the loss of hydrogen bond network interactions introduced by the I65 mutation, thereby adversely impacting the formation of hydrogen bond networks and the α-helix around N64 and T68. This loss of hydrogen bond networks in turn reduced the regional thermostability, markedly decreasing the thermostability of this quadruple mutant lipase. As such, other lipase isoforms

containing the I65 mutation will similarly not display improved thermostability owing to this loss of hydrogen bond network interactions. Similarly, lipase isoforms bearing T66 mutations are likely to exhibit limited thermostability due to the lack of stable interactions between this residue and other surrounding residues. As such, the characterization of lipase isoforms containing mutated I65 or T66 residues was not tested in this work. However, the double mutant lipase (Lipr27RCL-K64N/K68T) exhibited cumulative thermostability of Lipr27RCL-K64N and Lipr27RCL-K68T (**Supplementary Figure S2**), indicating that the role of the two mutation sites may be aggregate.

In summary, we found that we were able to produce novel mutant lipases (Lipr27RCL-K64N and Lipr27RCL-K68T) with improved thermostability through a combination of B factor analysis and site-directed mutagenesis, and these resultant enzymes represent attractive candidates for use in industrial applications.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

CZ performed the major experiments containing enzyme production, purification, and CD analysis. ZJ and MT carried out the experiments containing site-directed mutagenesis and characterization of lipases. LW, WQ, JH, and ZH helped with protein purification. NH analyzed the B-factor and MD simulation. QW and YM prepared experimental materials. BX, JD, and RZ dealt with the problems encountered in the experiment and coordinated the study. CZ wrote the manuscript. ZH and NH revised this manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (Grant Number 2017YFB0308401), the National Natural Science Foundation of China (Grant Numbers 31660240, 31660304, and 31960131), and the Yunling Scholar Fund and Yunling Technical Leader Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00346/full#supplementary-material

FIGURE S1 | Circular dichroism spectra (CD) of Lipr27RCL **(A)**, Lipr27RCL-K64N **(B)**, and Lipr27RCL-K68T **(C)** under the condition of thermal denaturation at 65°C for 5 min and renaturation at 4°C for 2d, respectively.

FIGURE S2 | The thermostability of Lipr27RCL, Lipr27RCL-K64N, Lipr27RCL-K68T, and Lipr27RCL-K64N/ K68T at 60°C.

FIGURE S3 | The original image of **Figure 1**.

REFERENCES

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770

Bommarius, A. S., and Paye, M. F. (2013). Stabilizing biocatalysts. *Chem. Soc. Rev.* 42, 6534–6565. doi: 10.1039/C3CS60137D

Ferreira-Dias, S., Sandoval, G., Plou, F., and Valero, F. (2013). The potential use of lipases in the production of fatty acid derivatives for the food and nutraceutical industries. *Electron. J. Biotechn.* 16, 1–38. doi: 10.2225/vol16-issue3-fulltext-5

Goomber, S., Kumar, R., Singh, R., Mishra, N., and Kaur, J. (2016). Point mutation Gln121-Arg increased temperature optima of *Bacillus lipase* (1.4 subfamily) by fifteen degrees. *Int. J. Biol. Macromol.* 88, 507–514. doi: 10.1016/j.ijbiomac.2016.04.022

Gupta, R., Kumari, A., Syal, P., and Singh, Y. (2015). Molecular and functional diversity of yeast and fungal lipases: their role in biotechnology and cellular physiology. *Prog. Lipid. Res.* 57, 40–54. doi: 10.1016/j.plipres.2014.12.001

Han, N., Miao, H., Ding, J., Li, J., Mu, Y., Zhou, J., et al. (2017). Improving the thermostability of a fungal GH11 xylanase via site-directed mutagenesis guided by sequence and structural analysis. *Biotechnol. Biofuels* 10:133. doi: 10.1186/s13068-017-0824-y

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. doi: 10.1002/prot.21123

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Li, G., Fang, X., Su, F., Chen, Y., Xu, L., and Yan, Y. (2018). Enhancing the thermostability of *Rhizomucor miehei* lipase with a limited screening library by rational-design point mutations and disulfide bonds. *Appl. Environ. Microbiol.* 84:e2129-17. doi: 10.1128/AEM.02129-17

Li, Z., Li, X., Wang, Y., Wang, Y., Wang, F., and Jiang, J. (2011). Expression and characterization of recombinant *Rhizopus oryzae* lipase for enzymatic biodiesel production. *Bioresour. Technol.* 102, 9810–9813. doi: 10.1016/j.biortech.2011.07.070

Lineweaver, H., and Burk, D. (1934). The determination of enzyme dissociation constants. *J. Am. Chem. Soc.* 56, 658–666. doi: 10.1021/ja01318a036

Liu, Y. H., Liu, H., Huang, L., Gui, S., Zheng, D., Jia, L. B., et al. (2017). Improvement in thermostability of an alkaline lipase I from *Penicillium cyclopium* by directed evolution. *RSC Adv.* 7, 38538–38548. doi: 10.1039/C7RA06307E

Naik, S., Basu, A., Saikia, R., Madan, B., Paul, P., Chaterjee, R., et al. (2010). Lipases for use in industrial biocatalysis: specificity of selected structural groups of lipases. *J. Mol. Catal. B Enzym.* 65, 18–23. doi: 10.1016/j.molcatb.2010.01.002

Nascimento, P. A. M., Pereira, J. F. B., and De Carvalho Santos-Ebinuma, V. (2019). Insights into the effect of imidazolium-based ionic liquids on chemical structure and hydrolytic activity of microbial lipase. *Bioprocess. Biosyst. Eng.* 42, 1235–1246. doi: 10.1007/s00449-019-02121-w

Pandey, A., Benjamin, S., Soccol, C. R., Nigam, P., Krieger, N., and Soccol, V. T. (1999). The realm of microbial lipases in biotechnology. *Biotechnol. Appl. Biochem.* 29, 119–131. doi: 10.1111/j.1470-8744.1999.tb00541.x

Persson, M., Costes, D., Wehtje, E., Adlercreutz, P. J. E., and Technology, M. (2002). Effects of solvent, water activity and temperature on lipase and

hydroxynitrile lyase enantioselectivity. *Enzyme Microb. Technol.* 30, 916–923. doi: 10.1016/S0141-0229(02)00033-9

Podar, M., and Reysenbach, A. L. (2006). New opportunities revealed by biotechnological explorations of extremophiles. *Curr. Opin. Biotechnol.* 17, 250–255. doi: 10.1016/j.copbio.2006.05.002

Ringe, D., and Petsko, G. A. (1986). Study of protein dynamics by X-ray diffraction. *Method Enzymol.* 131, 389–433. doi: 10.1016/0076-6879(86)31050-4

Rogers, T. A., and Bommarius, A. S. (2010). Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes. *Chem. Eng. Sci.* 65, 2118–2124. doi: 10.1016/j.ces.2009.12.005

Sha, C., Yu, X. W., Zhang, M., and Xu, Y. (2013). Efficient secretion of lipase r27RCL in *Pichia pastoris* by enhancing the disulfide bond formation pathway in the endoplasmic reticulum. *J. Ind. Microbiol. Biotechnol.* 40, 1241–1249. doi: 10.1007/s10295-013-1328-9

Stepankova, V., Bidmanova, S., Koudelakova, T., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2013). Strategies for stabilization of enzymes in organic solvents. *ACS Catal.* 3, 2823–2836. doi: 10.1021/cs400684x

Su, F., Li, G. L., Zhang, H. J., and Yan, Y. J. (2014). Enhanced performance of *Rhizopus oryzae* lipase immobilized on hydrophobic carriers and its application in biorefinery of rapeseed oil deodorizer distillate. *Bioenerg. Res.* 7, 935–945. doi: 10.1007/s12155-014-9415-y

Suen, W. C., Zhang, N. Y., Xiao, L., Madison, V., and Zaks, A. (2004). Improved activity and thermostability of *Candida antarctica* lipase B by DNA family shuffling. *Protein Eng. Des. Sel.* 17, 133–140. doi: 10.1093/protein/gzh017

Xie, Y., An, J., Yang, G., Wu, G., Zhang, Y., Cui, L., et al. (2014). Enhanced enzyme kinetic stability by increasing rigidity within the active site. *J. Biol. Chem.* 289, 7994–8006. doi: 10.1074/jbc.M113.536045

Yu, X. W., Tan, N. J., Xiao, R., and Xu, Y. (2012a). Engineering a disulfide bond in the lid hinge region of *Rhizopus chinensis* lipase: increased thermostability and altered acyl chain length specificity. *PLoS One* 7:e46388. doi: 10.1371/journal.pone.0046388

Yu, X. W., Wang, R., Zhang, M., Xu, Y., and Xiao, R. (2012b). Enhanced thermostability of a *Rhizopus chinensis* lipase by in vivo recombination in *Pichia pastoris*. *Microb. Cell. Fact.* 11:102. doi: 10.1186/1475-2859-11-102

Yu, X. W., Zhu, S. S., Xiao, R., and Xu, Y. (2014). Conversion of a *Rhizopus chinensis* lipase into an esterase by lid swapping. *J. Lipid. Res.* 55, 1044–1051. doi: 10.1194/jlr.M043950

Yuan, Z., Zhao, J., and Wang, Z. X. (2003). Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.* 16, 109–114. doi: 10.1093/proeng/gzg014

Conflict of Interest: LW, WQ, JH, and ZZ were employed by Yunnan Walvax Biotechnology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 April 2020
doi: 10.3389/fmicb.2020.00633

Identification and Immobilization of an Invertase With High Specific Activity and Sucrose Tolerance Ability of *Gongronella* sp. w5 for High Fructose Syrup Preparation

Gang Zhou[1,2,3], *Can Peng*[1,2,3], *Xiaosa Liu*[1,2,3], *Fei Chang*[1,2,3], *Yazhong Xiao*[1,2,3], *Juanjuan Liu*[1,2,3]* and *Zemin Fang*[1,2,3]*

[1] School of Life Sciences, Anhui University, Hefei, China, [2] Anhui Key Laboratory of Modern Biomanufacturing, Hefei, China, [3] Anhui Provincial Engineering Technology Research Center of Microorganisms and Biocatalysis, Hefei, China

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Hao Zhou,
Dalian University of Technology, China
Jose M. Bruno-Barcena,
North Carolina State University, United States

***Correspondence:**
Juanjuan Liu
liu_juan825@ahu.edu.cn
Zemin Fang
zemin_fang@ahu.edu.cn

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 03 December 2019
Accepted: 20 March 2020
Published: 09 April 2020

Citation:
Zhou G, Peng C, Liu X, Chang F, Xiao Y, Liu J and Fang Z (2020) Identification and Immobilization of an Invertase With High Specific Activity and Sucrose Tolerance Ability of Gongronella sp. w5 for High Fructose Syrup Preparation. Front. Microbiol. 11:633. doi: 10.3389/fmicb.2020.00633

Invertases catalyze the hydrolysis of sucrose into fructose and glucose and can be employed as an alternative in producing high fructose syrup. In this study, we reported the heterologous expression of an invertase gene (*GspInv*) of *Gongronella* sp. w5 in *Komagataella pastoris*. GspInv activity reached 147.6 ± 0.4 U/mL after 5 days of methanol induction. GspInv is invertase with a high specific activity of 2,776.1 ± 124.2 U/mg toward sucrose. GspInv showed high tolerance to sucrose (IC_{50} = 1.2 M), glucose (IC_{50} > 2 M), fructose (IC_{50} = 1.5 M), and a variety of metal ions that make it an ideal candidate for high fructose syrup production. A carbohydrate-binding module was sequence-optimized and fused to the N-terminus of GspInv. The fusion protein had the highest immobilization efficiency at room temperature within 1 h adsorption, with 1 g of cellulose absorption up to 8,000 U protein. The cellulose-immobilized fusion protein retained the unique properties of GspInv. When applied in high fructose syrup preparation by using 1 M sucrose as the substrate, the sucrose conversion efficiency of the fused protein remained at approximately 95% after 50 h of continuous hydrolysis on a packed bed reactor. The fused protein can also hydrolyze completely the sucrose in sugarcane molasses. Our results suggest that GspInv is an unusual invertase and a promising candidate for high fructose syrup preparation.

Keywords: invertase, expression, immobilization, high fructose syrup, *Gongronella* sp.

INTRODUCTION

High fructose syrup (HFS) is a major nutritive sweetener presented as a mixture of glucose and fructose (Singh et al., 2018a). In modern food and beverage industries, HFS has replaced sucrose as an alternative sweetener because of its many advantages (Rippe, 2014). For example, HFS is easier to store than sucrose because of its high osmotic pressure and anti-drying properties (Singh et al., 2018a). Furthermore, HFS is crystallization-controlled and easy to dissolve, making it flexible for many applications, such as seasoning and baking (Zhang et al., 2004). HFS can be divided into three categories according to fructose content: HFS-90 (90% fructose and 10% glucose), HFS-55

(55% fructose and 45% glucose), and HFS-42 (42% fructose and 58% glucose). In general, HFS-55 is more suitable for industrial applications than HFS-42 it has better flavor, sweetness, and other benefits (Fatourehchi et al., 2014; Neifar et al., 2019).

Commercially, HFS is obtained from cornstarch through multi-enzymatic hydrolysis and transformation in four steps, including (a) enzymatic liquefaction or partial hydrolysis of starch with α-amylase, (b) conversion of liquefied starch into dextrose hydrolyzate by employing amylo-glucosidase, (c) isomerization of dextrose to fructose using isomerase, and (d) refinement of the final fructose product (Singh et al., 2018a). Ultra-filtration and crystallization steps, such as activated carbon decolorization, ion exchange, chromatographic separation, and evaporation, are necessary to obtain high-purity fructose (Guthrie and Morton, 2000; Chen et al., 2016). All these complicated steps address related problems smoothly because the conventional HFS production approach is well-established commercially. However, certain drawbacks, including low conversion rate, low separation efficiency, labor-intensive preparation, poor enzyme technology, and low product yields, increase production costs (Wang et al., 2016). Hence, a convenient and cost-effective technique for HFS production needs to be developed.

Invertases (EC 3.2.1.26) are enzymes that catalyze the hydrolysis of sucrose into equimolar concentrations of glucose and fructose (inverted syrup) (Kotwal and Shankar, 2009). One of the most significant applications of invertase lies in the production of inverted syrup that can be used directly as HFS or as a substrate to obtain pure crystalline fructose (Kotwal and Shankar, 2009; Lima et al., 2011). However, the use of commercially available invertases to hydrolyze sucrose is costly because of the many drawbacks from substrate and enzyme aspects. For instance, the price of commercial invertase is high (Torres-Acosta et al., 2018). Meanwhile, most invertases are inhibited by the substrate sucrose and the end-product glucose and fructose (Sakakibara et al., 1996; Rashad and Nooman, 2009; Resa et al., 2009). For example, sucrose concentrations higher than 50 mM cause the yield to diminish sharply when commercially available invertase is used as the catalyst (Tomotani and Vitolo, 2007). Moreover, fructose and glucose are non-competitive inhibitors of invertase activities at high concentrations (Isla et al., 1999). Invert syrup production from sugarcane or beet sucrose is unavailable economically. Molasses (a by-product from sugarcane processing) (Deng et al., 2008) is used as the alternative substrate because it is about 10-times lower in price than the sucrose (Khatiwada et al., 2016; Gabisa et al., 2019) and is rich in sucrose and glucose (30–50%, v/v) and metal ions (Liu et al., 2008; Xia et al., 2016). Therefore, finding novel invertases that fit the characteristics of molasses, such as tolerance to high sucrose content and metal ions, and low-cost preparation is beneficial economically (Palai et al., 2014; Eskandarloo and Abbaspourrad, 2018; Mohammadi et al., 2018).

Gongronella sp. w5 is a soil-borne fungus that prefers sucrose as its carbon source (Hu et al., 2018). Previously, a deduced cytoplasmic glycoside hydrolase family 32 (GH32) invertase (named GspInv), with no significant sequence identity with well-characterized fungal invertases, was predicted in the genome (Dong et al., 2018). In the present study, GspInv was cloned and expressed in *Komagataella pastoris*. Our results showed GspInv is an unusual invertase suitable for HFS production. Furthermore, HFS production cost was reduced by fusing a sequence-optimized carbohydrate-binding module (CBM) (Oliveira et al., 2015) to the N-terminal of GspInv to facilitate the purification and immobilization of GspInv. The fused protein showed high efficiency in sucrose and molasses hydrolysis. As such, GspInv and its derivates have remarkable application potential in HFS production.

MATERIALS AND METHODS

Strains, Plasmid, and Reagents

Gongronella sp. w5 was obtained from China Center for Type Culture Collection (No. AF2012004) and cultured on potato dextrose agar slants at 4°C. *K. pastoris* GS115 and the expression vector pPIC9K were purchased from Invitrogen (Carlsbad, CA, United States). Yeast extract peptone dextrose medium, minimal dextrose medium (MD), buffered minimal glycerol, and buffered minimal glycerol yeast medium (BMGY) were prepared in accordance with the manual of the EasySelect *Komagataella* (*Pichia*) Expression kit (Invitrogen). All other chemicals and reagents were of analytical grade unless otherwise indicated.

Cloning, Expression, and Purification of Recombinant GspInv and Its Derivatives

Five fungal plugs of *Gongronella* sp. w5 with a diameter of 5 mm were grown at 37°C and 120 rpm in SAHX medium in accordance with the method of Hu et al. (2018). After growing for 2 days, the mycelia were collected and grounded with a mortar and pestle in the presence of liquid nitrogen. Total RNA was extracted using TRIzol reagent according to the manufacturer's instruction (TaKaRa, Dalian, China), followed by RNase-free DNase digestion (Promega, Beijing, China). The first-strand cDNA was synthesized using reverse transcriptase and a primer pair of *GspInvF* CCTAGGATGGTGCTTGCTGATCCT (*Bln*I site underlined) and *GspInvR* GCGGCCGCTCAAGGGCGATTGAACG (*Not*I site underlined) in accordance with the manufacturer's protocol (TaKaRa). GspInv cDNA was amplified by PCR by using the same primer pair of *GspInvF* and *GspInvR*. The cloned cDNA was digested further with *Eco*RI and *Not*I, ligated into the similarly digested expression plasmid pPIC9K, and linearized and transformed into *K. pastoris* GS115. The electroporated cells were plated onto MD agar plates to select the His^+ transformants. Some His^+ transformants were selected randomly and grown on the BMGY liquid medium at 28°C for 2 days.

GspInv was purified and immobilized onto cellulose through fusing a CBM_{24} from *Clostridium thermocellum* (GenBank No. HF912724.1) to GspInv (Chang et al., 2018). In brief, the CBM_{24} gene was codon-optimized in accordance with the *K. pastoris* codon bias and synthesized by Sangon Biotech (Shanghai, China). It was fused to the N- or C-terminus of *GspInv* by using overlap extension PCR (Chang et al., 2018). Given that CBM_{24} contains three potential glycosylation sites at 11, 65, and 121 in

the amino acid sequence, the amino acids at these sites were mutated to aspartic acid to obtain $CBM_{24}DG$ (DG stands for deglycosylation). The mutated sequence was fused further to the N- or C-terminus of the GspInv by using overlap extension PCR. Four plasmids, namely, CBM_{24}-GspInv, $CBM_{24}DG$-GspInv, GspInv-CBM_{24}, and GspInv-$CBM_{24}DG$, were constructed after digesting the four sequences with *Not*I and *Avr*II and inserting into the *K. pastoris* expression vector pPIC9K. They were transformed individually into *K. pastoris* cells and screened as described above.

The proteins were expressed by cultivating *K. pastoris* GS115 host cells carrying GspInv, CBM_{24}-GspInv, $CBM_{24}DG$-GspInv, GspInv-CBM_{24}, and GspInv-$CBM_{24}DG$ separately in 500 mL flasks containing 100 mL of BMGY medium at 28°C and shaken at 220 rpm. When the cell density of OD_{600} reached 2.0–6.0, cells were harvested and resuspended in 150 mL of BMMY medium in 500 mL flasks at a final cell density of OD_{600} = 1.0. The cultures were incubated at 28°C and shaken at 220 rpm with a daily addition of 0.5% (v/v) methanol to induce extracellular GspInv production. Samples were withdrawn every 24 h to determine the invertase activity and the biomass.

The aqueous solution was centrifuged at 20,000 × *g* for 30 min and concentrated at 4°C through ultrafiltration using a Minitan ultrafiltration system with a low-binding regenerated cellulose membrane (Millipore, Bedford, MA, United States). The heterologously expressed protein from *K. pastoris* GS115 culture broth was purified by centrifuging the concentrate at 20,000 × *g* for 30 min, and the supernatant was then dialyzed overnight against the citrate-phosphate buffer (50 mM, pH 5.5), followed by centrifugation as described previously. Then the supernatant was applied to a DEAE-Sepharose FF column (10 mm × 200 mm, Amersham Pharmacia, Uppsala, Sweden) pre-equilibrated with citrate-phosphate buffer. The column was eluted with a linear gradient of $(NH_4)_2SO_4$ (0–0.5 M in a citrate-phosphate buffer, with a flow rate of 1.0 mL min^{-1}).

Invertase Activity Assay

Protein samples were diluted in a suitable volume of citrate-phosphate buffer (50 mM, pH 5.0). Invertase activities were measured in 1 mL of reaction mixtures containing 20 μL purified enzyme, 50 mM citrate-phosphate buffer (pH 5.0), and 200 mM sucrose, and incubated at 45°C for 5 min. The reaction was terminated by heating the assay mixture at 100°C for 5 min. The released glucose and fructose were measured using the 3,5-dinitrosalicylic acid method (Ashwell, 1957). The unit (U) of invertase activity was defined as the amount of enzyme required to hydrolyze 1 μmol of sucrose per min under assay conditions.

Biochemical Characterization of Recombinant Proteins

The protein concentration was assayed using the Bradford method at 595 nm, with bovine serum albumin as the standard (Sangon Biotech). The homogeneity of GspInv and its derivatives was determined by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and stained with Coomassie brilliant blue R250. To verify that the protein in the gel was the recombinant GspInv, SDS-PAGE gel was washed with 50 mM citrate-phosphate buffer at pH 5.0 for 1 h to remove SDS. Then, it was incubated in an acetate-phosphate buffer (50 mM, pH 5.0) containing 200 mM sucrose at 45°C for 30 min, and actively stained using 100 mM NaOH solution containing 0.2% triphenyl tetrazolium chloride after the sucrose solution was removed (Van et al., 2013).

The effect of pH on enzymatic activity was determined at 45°C in 50 mM citrate-phosphate buffer (pH 3.5–6.5). The effect of temperature on the enzymatic activity was determined at pH 5.0 and temperatures ranging from 30–60°C. The enzyme stabilities against pH and temperature were determined by incubating proteins at various temperatures and different pH values. The residual activities were determined as mentioned above. All experiments were performed in triplicate.

The effects of metal ions including Mg^{2+}, Ba^{2+}, Ca^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, Cu^{2+}, and Fe^{3+}, on GspInv activity were investigated in the presence of each ion at different concentrations (1, 5, and 10 mM). The enzyme assays were carried out at pH 5.0 and 45°C.

Kinetic Analysis

The appropriate concentration of GspInv or its derivatives was utilized under the optimum conditions to determine the kinetic parameters (K_m, V_{max}, and k_{cat}/K_m). The reaction was carried out by incubating the enzyme in 50 mM citrate-phosphate buffer (pH 5.0) containing sucrose at a concentration range of 1 mM–2 M at 45°C for 5 min. Released glucose was quantified by using the glucose oxidase method (Rongsheng Biotech, Shanghai, China). The kinetic constants and their corresponding errors were calculated by fitting the measured rate to the Michaelis–Menten equation using the computer program Origin 8.0 (n = 9).

Effects of Mono-and Di-Sugars on Invertase Activity

The effect of sucrose on GspInv and its derivates activity was determined using 0.01–2 M sucrose as the substrate. The feedback inhibition of fructose and glucose on invertase activity was determined through the use of 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.5, and 2.0 M glucose or fructose added to the enzymatic reaction systems. The content of the glucose in the final product after the reaction was measured via the glucose oxidase method. The fructose concentration was determined using high-performance liquid chromatography (HPLC). Briefly, 20 μL of samples were withdrawn at different time intervals and analyzed at 50°C by using a TSKgel Amide-80 column (4.6 mm × 250 mm, 5 μm, Tosoh Corporation, Kyoto, Japan) and an evaporative light-scattering detector 2424 (Waters, United States). The eluting solution was acetonitrile: water (70:30, v/v).

Immobilization of Enzyme on Cellulose

The fermentation supernatant containing $CBM_{24}DG$-GspInv was withdrawn by centrifuging the cultures at 12,000 × *g* for 10 min and used directly for purification and immobilization. The microcrystalline cellulose was added into the supernatant at a ratio of 1:20 (g/mL) and incubated at 25°C for 4 h in a

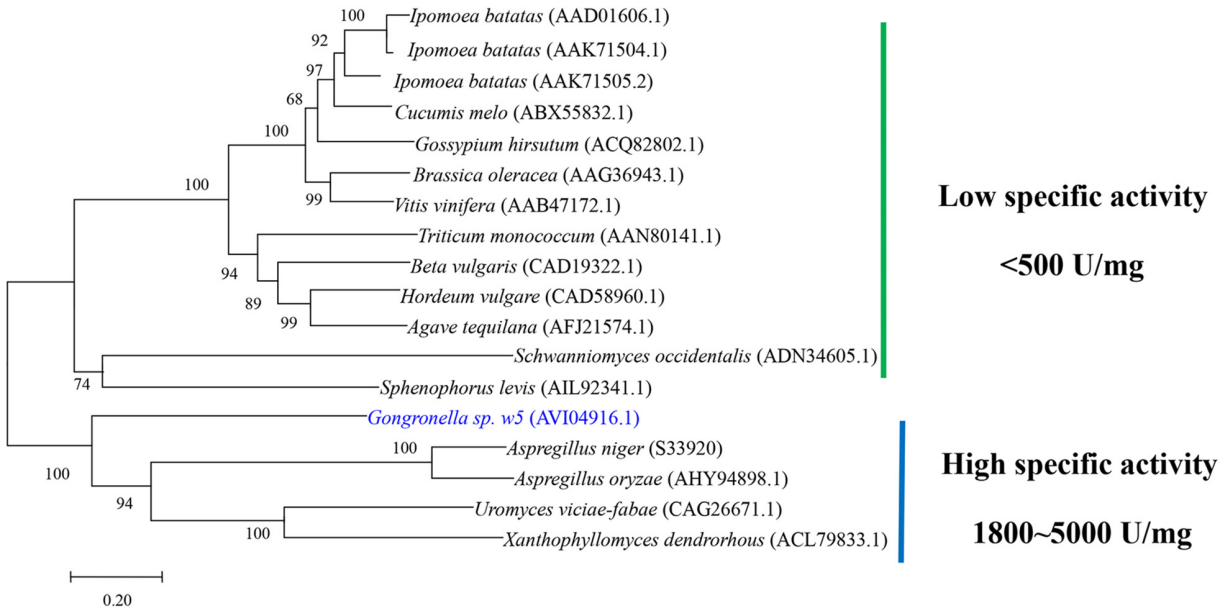


FIGURE 1 | Neighbor-joining phylogeny of GspInv with experimentally characterized fungal GH32 invertases. Sequences are retrieved from GenBank database. GenBank protein accession numbers are given in parentheses. Bootstrap support values based on 1000 replicates are given as percentages above individual branches.

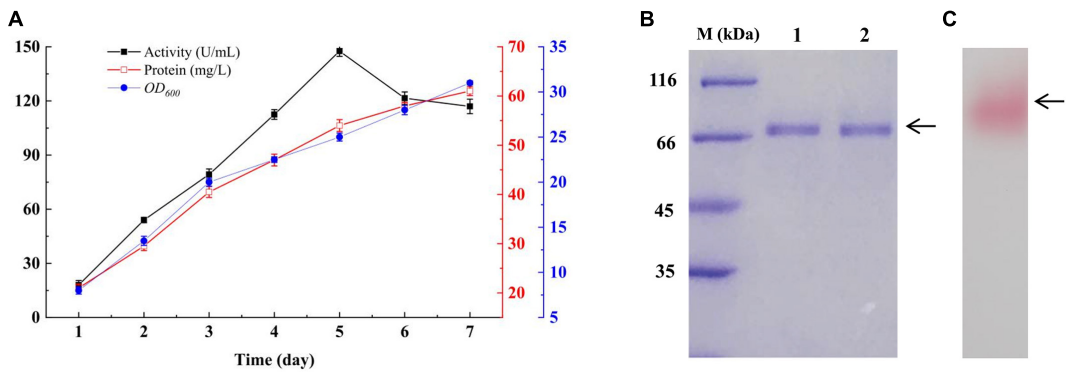


FIGURE 2 | Expression of GspInv in *K. pastoris*. **(A)** Time course of invertase activity, protein content, and *K. pastoris* concentration (OD_{600}) in the extracellular medium of *K. pastoris* cultures producing recombinant GspInv. The data presented are the average values from triplicate cultures (repeats of experiments gave similar results). **(B)** Coomassie brilliant blue R250 stained 15% SDS-PAGE gel of GspInv in crude culture supernatant (lane 1) and after ion-exchange chromatography (lane 2). **(C)** Zymogram gel of GspInv in crude culture supernatant.

shaker at 60 rpm/min. The samples were withdrawn at intervals of 0.5 h during the process and centrifuged at 5,000 × *g* to separate the supernatant and precipitate. The precipitate was then washed three times with an equal volume of citrate-phosphate buffer (50 mM, pH 5.0) and stored in this buffer. The residual invertase activity in the supernatant and the invertase activity on immobilized cellulose were separately tested. The immobilization efficiency of cellulose was calculated.

Reusability of Immobilized Invertase

For the continuous hydrolysis of sucrose, a packed bed reactor (PBR) was prepared by using 10 g of cellulose-CBM_{24}DG-GspInv (82,000 U) with a fixed bed height of 10 cm. The reactor was placed in a thermo-stated oven (Notting Scientific Equipment, Hangzhou, China) at 35°C to ensure constant reaction temperature. Then, the sucrose solution (1, 1.2, 1.5, or 2 M prepared in a citrate-phosphate buffer, pH 5.0) was fed into the base of the reactor by using a pump (Notting Scientific Equipment) at a flow rate of 2.5 mL/min for 5, 10, and 15 h, respectively. The eluent was withdrawn and the concentrations of sucrose, glucose, and fructose analyzed using HPLC as described above. The sucrose hydrolysis efficiency was then calculated.

The PBR containing cellulose-CBM_{24}DG-GspInv prepared as described above was loaded continuously with 1 M sucrose for 50 h at a flow rate of 2.5 mL/min to determine the reusability of cellulose-CBM_{24}DG-GspInv. The eluent was withdrawn

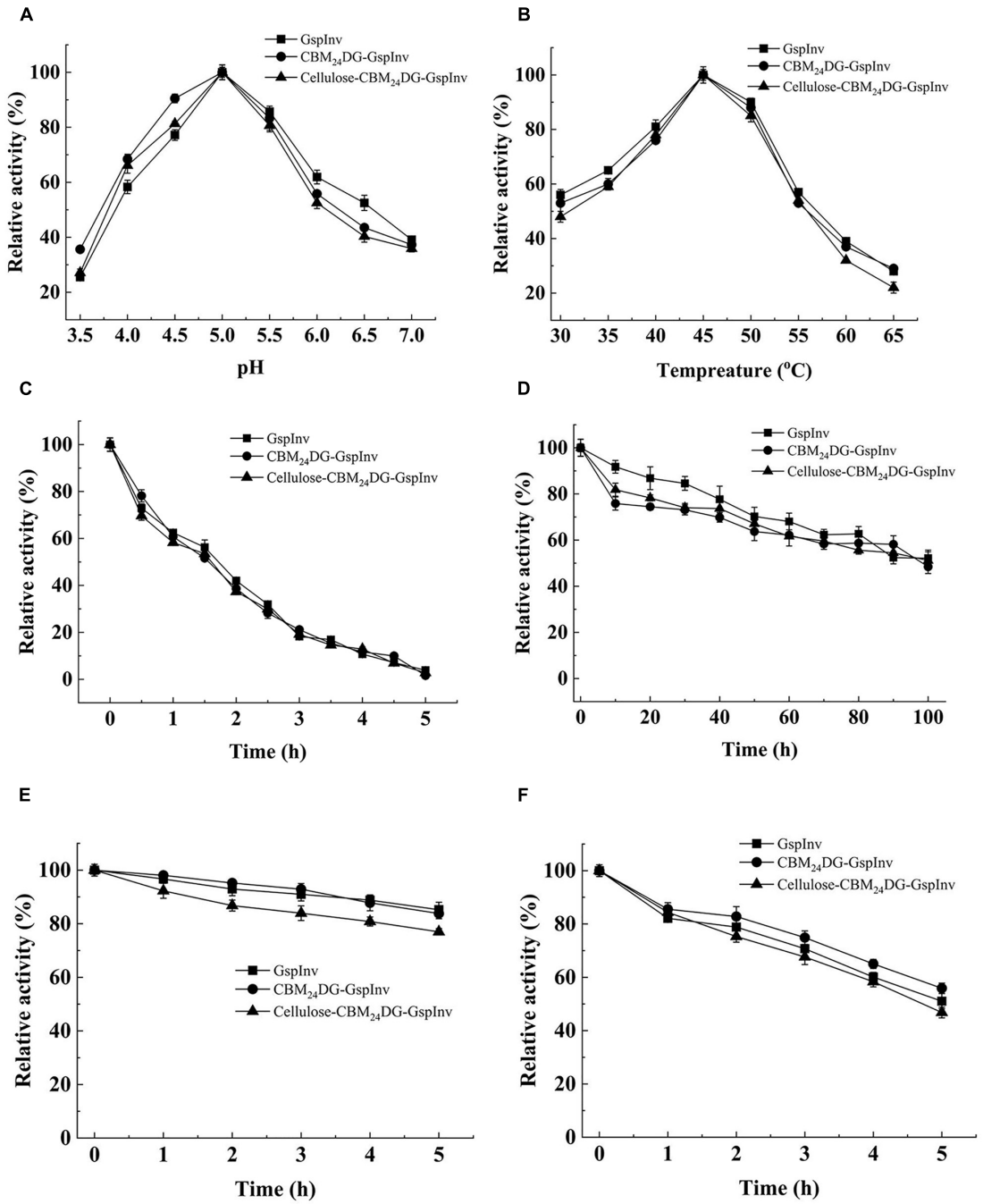


FIGURE 3 | Effects of pH and temperature on the activity and stability of GspInv, CBM_{24}DG-GspInv, and cellulose-CBM_{24}DG-GspInv. **(A)** pH optimum. Samples were incubated at 45°C. **(B)** Temperature optimum. Samples were incubated at pH 5.0. **(C,D)** Thermostability at 45°C and 40°C, respectively. Samples were incubated at pH 5.0. **(E,F)** pH stability at pH 5.5 and pH 6.0, respectively. Samples were incubated at 4°C. Standard deviations and values were calculated from triplicate technical repeats of measurements.

every 5 h, and the concentrations of sucrose, glucose, and fructose analyzed.

When molasses was employed as the substrate, the crude molasses (Xuanbo sugar-refinery, Guangxi, China) contains 40% (w/v) sucrose, 13.2% reduced sugars (glucose and fructose), 2.3% other carbohydrates, 9.1% ash, 3.9% salt, and 8.4% mineral substance was diluted with citrate-phosphate buffer (50 mM, pH 5.0) to obtain 20% sucrose concentration and centrifuged at 8,000 × *g* for 5 min to remove impurities. Then, 5 mg/L polyacrylamide was added into the liquid molasses, which was incubated at 85°C for 10 min, followed by centrifugation at 8,000 × *g* for 5 min to remove the precipitates. The supernatant

was loaded to the PBR for 5 h at a flow rate of 2.5 mL/min. The eluent was withdrawn every 1 h and the concentrations of sucrose, glucose, and fructose analyzed.

Bioinformatics Analysis

The sequence similarity search of GspInv was performed using Blast at UniProt[1]. The module structure of the enzyme was analyzed with a simple modular architecture research tool SMART[2]. Multiple sequence alignment of GspInv with other related invertase sequences was performed using Clustal X 2.0 and Phylogeny fr[3]. The phylogenetic tree was constructed using MEGA 7.

RESULTS AND DISCUSSION

GspInv Cloning and Sequence Analysis

The full-length GspInv cDNA was amplified successfully from the total RNA extracted from *Gongronella* sp. w5 mycelia using sucrose as the carbon source, with primers binding specifically to upstream and downstream of the cDNA coding sequence of GspInv deduced from the genome (Dong et al., 2018). The cloned GspInv cDNA was 1,761 bp in length. Compared to the 1,822 bp of its gene sequence, *GspInv* harbors only a 70 bp intron located between nucleotides 208 and 277. GspInv comprises 586 amino acids in length, with a theoretical molecular weight of 65.4 kDa and an isoelectric point of 4.89. The GspInv amino acid sequence shares the highest identities of 62.4 and 44% with a glycosyl hydrolase (ORZ22493.1) and a glycoside hydrolase family 32 protein (XP_018284693) deduced from the genome of *Absidia repens* and *Phycomyces blakesleeanus* NRRL 1555 and has not been characterized biochemically. Further analysis showed that GspInv has sequence identities of less than 37% to hypothetical proteins deduced from the genome sequences (**Supplementary Table S1**). These results suggested that GspInv is a novel protein with little information gained on its biochemical characteristics.

No signal peptide was predicted at the N-terminal of GspInv by using SingalP 5.0, which suggests that GspInv is an intracellular protein. This result is consistent with the invertase activity that can only be detected in the intracellular proteome of *Gongronella* sp. w5 (Hu et al., 2018). Module analysis suggested that GspInv is a GH32 protein carrying Pfam domain signatures, with accession numbers PF00251 (E value 3e-49) and PF08244 (E value 3.22e-25), which are characteristics of GH32 family protein (Finn et al., 2017). The alignment of the GspInv amino acid sequence with other characterized invertases revealed the GspInv contains conserved motifs of GH32 invertases (**Supplementary Figure S1**). Based on their alignments, residues D46, D173, and E251 of GspInv were identified as the nucleophile, transition-state stabilizer, and general acid/base catalyst, respectively (**Supplementary Figure S1**; Pons et al., 2004; Ramírez-Escudero et al., 2016). Trollope et al. (2015) reported that GH32 invertases can be divided into two groups, including that with low- and high-level fructooligosaccharide synthesis abilities based on the conserved motifs unique to each group. According to their prediction, GspInv belongs to the invertase with low-level fructooligosaccharide synthesis ability. Many GH32 members have been well characterized experimentally. In addition to sucrose hydrolysis ability, some invertases exhibit transferase activity at high substrate concentrations, resulting in the formation of fructooligosaccharides by transferring successive

[1]https://www.uniprot.org/

[2]http://smart.embl-heidelberg.de/

[3]http://www.phylogeny.fr/index.cgi

TABLE 1 | Effects of metal ions on GspInv activity.

Metal ions	1 mM	5 mM	10 mM
None	100.0	100.0	100.0
Mg^{2+}	102.6 ± 1.8	90.7 ± 8.5	82.2 ± 6.9
Ba^{2+}	106.7 ± 11.1	108.1 ± 3.9	109.7 ± 4.4
Ca^{2+}	111.9 ± 4.9	112.7 ± 3.0	120.9 ± 3.7
Co^{2+}	102.1 ± 6.2	85.1 ± 7.6	72.8 ± 11.1
Ni^{2+}	92.4 ± 1.4	78.7 ± 4.2	57.9 ± 2.3
Mn^{2+}	153.8 ± 3.6	143.2 ± 1.6	113.3 ± 17.2
Cu^{2+}	52.1 ± 2.3	22.3 ± 1.9	16.7 ± 1.4
Fe^{3+}	88.6 ± 4.6	89.2 ± 3.1	81.3 ± 2.4

Sulfates are used as the target mental ion donators. Invertase activity was determined at 45°C in citrate-phosphate buffer (50 mM, pH 5.0) using sucrose as the substrate. The data presented are the average values from triplicate technical repeats of measurements.

TABLE 2 | Comparison of the biochemical characteristics and kinetic constants of GspInv with characterized invertases.

Source	Optimum		Specific activity (U/mg)	Kinetic parameters		References
	pH	Temp (°C)		K_m (mM)	k_{cat} (s^{-1})	
A. niger	5.0–6.5	65–70	NR	4 ± 0.5	NR	Veana et al., 2014
S. cerevisiae	3.5–5.5	60	1590	NR	NR	Acosta et al., 2000
A. niger (SucB)	5.0	40	NR	2.0 ± 0.2	NR	Goosen et al., 2007
Bacillus sp. HJ14	8	30	2400	62.9	746.2	Zhou et al., 2016
Pichia anomala	4.0–6.5	38	1482	16	NR	Pérez et al., 2001
Elsholtzia haichowensis	5	70	NR	2.68 ± 0.1	NR	Cai et al., 2014
Aspergillus foetidus	5.6	37	257.2	NR	NR	Rehm et al., 1998
Gongronella sp. w5	5	45	2776 ± 124	8.7 ± 1.1	5100	This study

NR, not reported.

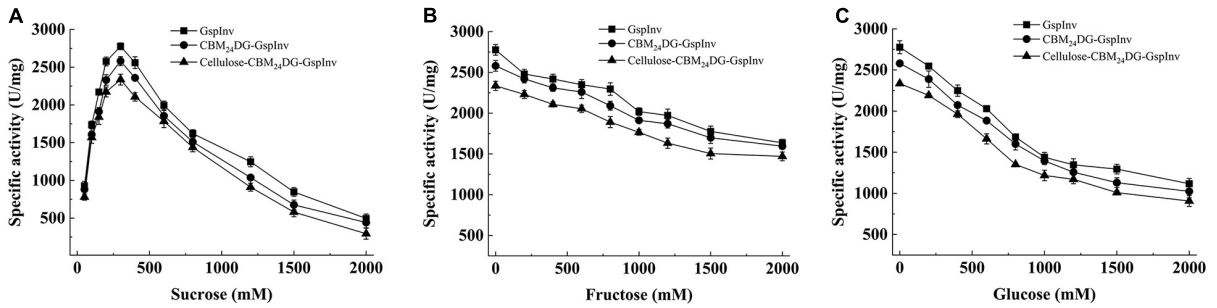


FIGURE 4 | Effects of sugars on GspInv, CBM_{24}DG-GspInv, and cellulose-CBM_{24}DG-GspInv activities. **(A)** Sucrose. **(B)** Glucose, **(C)** Fructose. Standard deviations and values were calculated from triplicate technical repeats of measurements.

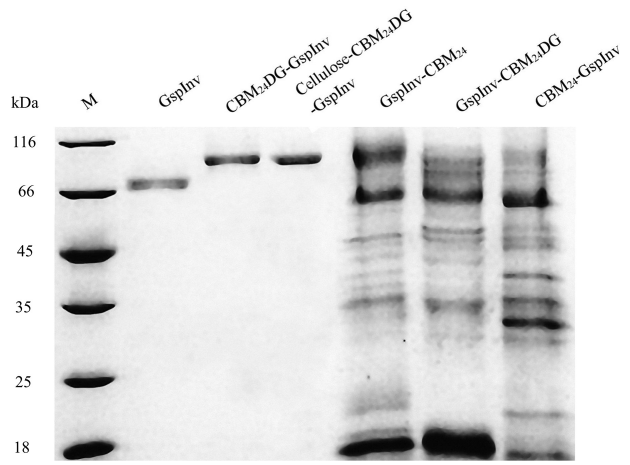


FIGURE 5 | Coomassie brilliant blue R250 stained 15% SDS-PAGE gel of GspInv (fermentation broth was 10-times concentrated), CBM_{24}DG-GspInv (10-times), cellulose-CBM_{24}DG-GspInv (10-times), GspInv-CBM_{24} (500-times), GspInv-CBM_{24}DG (500-times), and CBM_{24}-GspInv (500-times). M: standard protein marker.

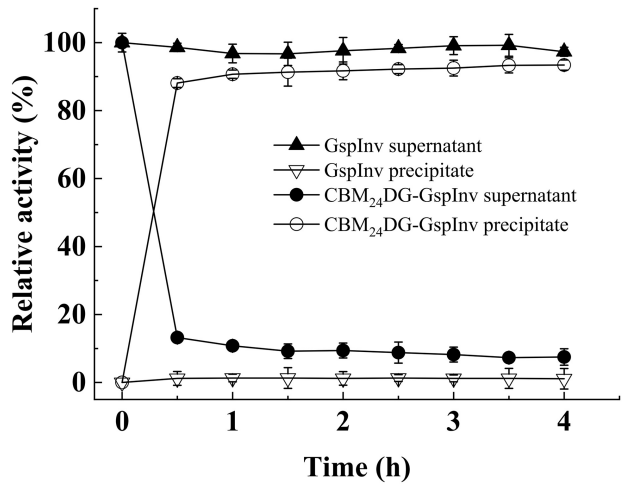


FIGURE 6 | Relative invertase activities in cell-free supernatant (crude enzyme) and precipitate after adding cellulose. The total invertase activities of GspInv and CBM_{24}DG-GspInv in supernatant at 0 h were set as 100% relative activity, respectively. Standard deviations and values were calculated from triplicate technical repeats of measurements.

fructose units to sucrose (Trollope et al., 2015). The invertase with low-level fructooligosaccharide synthesis ability will facilitate the preparation of HFS using high concentrations of sucrose as the substrate. Moreover, phylogenetic tree analysis suggested that GspInv belongs to the clade of invertases with high specific activity (**Figure 1**), and as such, GspInv may have application potential in modern industries.

Expression of GspInv in *K. pastoris* and Purification

GspInv was expressed in *K. pastoris* under the control of the inducible promoter AOX1 and with the α-factor signal sequence of the vector pPIC9K for secreted expression. The highest invertase activity of 147.6 ± 0.4 U/mL was obtained after induction with methanol for 5 days and using sucrose as the substrate. The extracellular soluble protein content was determined as 54.1 ± 0.7 mg/L (**Figure 2A**). Given that the specific activity of GspInv is 2,776.1 ± 124.2 U/mg as determined below, approximately 95% (>51 mg) of the total protein in the culture supernatant was GspInv on the 5th day. As such, only one band with an apparent molecular weight of approximately 67 kDa was presented on SDS-PAGE gel and the purity of the protein was similar to that after ion-exchange chromatography (**Figure 2B**). After active staining, the color of the band changed to pink, indicating that the protein was GspInv (**Figure 2C**; Van et al., 2013). MALDI-TOF-MS/MS identification of the peptide also suggested the only band on the gel was GspInv (**Supplementary Table S2**).

K. pastoris is considered one of the best hosts for invertase expression. Invertases from several strains, such as *Zymomonas mobilis* and *Aspergillus niger*, have been expressed heterologously in *K. pastoris* (Veana et al., 2014; Pérez de los Santos et al., 2016). The GspInv productivity was considerably higher than those of most invertases expressed heterologously in *K. pastoris* (**Supplementary Table S3**; Hsieh et al., 2006; Plascencia-Espinosa et al., 2014). Furthermore, GspInv is electrophoretically

homogenous on an SDS-PAGE gel after ultrafiltration, which means no additional purification step is necessary to purify GspInv and decrease production costs. This result is also different from most invertases with purification steps that require ultrafiltration and complicated chromatographic techniques that result in more than 50% activity loss (Huang et al., 2003; Wang et al., 2005; Hsieh et al., 2006; Veana et al., 2014; Pérez de los Santos et al., 2016; **Supplementary Table S3**). The complicated purification steps of these invertases cause the large-scale enzymatic hydrolyzing reactions to be impractical economically.

GspInv Characterization

The optimum pH of the purified GspInv was at pH 5.0. It showed more than 50% of its activity at a pH range of 4.0–6.5 (**Figure 3A**) and decreased sharply at pH 3.5 and above 6.5 when evaluated using sucrose as the substrate. The optimum temperature of GspInv was 45°C. More than 50% of the protein activity was shown when tested at temperatures between 30 and 55°C (**Figure 3B**). The optimum temperature and pH of GspInv were similar to most of the invertases from other fungi, with optimum temperature and pH of 40–70°C and 4.5–6.5, respectively (Acosta et al., 2000; Cai et al., 2014; Nadeem et al., 2015).

The thermostability of GspInv was assayed at the optimum pH 5.0. It retained approximately 65% of its original activity after incubation at 45°C for 1 h (**Figure 3C**). In comparison, GspInv retained 50% of the original activity after incubation at 40°C for 100 h (**Figure 3D**). GspInv was stable at pH 5.5. After incubation at pH 5.5 and 4°C for 5 h, the protein retained approximately 85% of its original activity (**Figure 3E**). GspInv retained 55% of its activity after 5 h at pH 6.0 (**Figure 3F**). pH and thermal stabilities are considered commercially as profitable features of an enzyme. Enzyme-catalyzed reactions operating at moderate temperatures and weak acidic pH reduce costs on energy and equipment (Singh et al., 2018b).

The effects of metal ions on GspInv activity were assayed. Invertase activity increased to 109.7 ± 4.4% (10 mM), 120.9 ± 3.7% (10 mM), and 153.8 ± 3.6% (1 mM) in the presence of Ba^{2+}, Ca^{2+}, and Mn^{2+}, respectively. The presence of metal ion Mn^{2+} increasing the enzymatic activity has been reported for several invertases. For example, it increased the INVA and INVB activities by 80 and 20%, respectively (Pérez de los Santos et al., 2016). This activating effect has been observed for the invertase from the yeast *Candida guilliermondii* (Plascencia-Espinosa et al., 2014). Mn^{2+} enhanced its activity by up to 277% for invertase from *Anthene phoenicis* (Rustiguel et al., 2015). GspInv retained 82.2 ± 6.9, 72.8 ± 11.1, and 81.3 ± 2.4% of its original activities in the presence of 10 mM Mg^{2+}, Co^{2+}, and Fe^{3+}, respectively. The tolerance of GspInv on these ions suggested it can hydrolyze substances such as molasses rich in sucrose but contain various ions. In comparison, ions including Ni^{2+} and Cu^{2+} showed inhibitory effects on the activity under all tested concentrations (1, 5, and 10 mM). GspInv activities were reduced to 57.9 ± 2.3 and 16.7 ± 1.4% at 10 mM of Ni^{2+} and Cu^{2+} (**Table 1**). This finding suggests the presence of thiol groups or His residues that are important for enzyme activity. Co^{2+}, Ni^{2+}, and Cu^{2+} may coordinate with His residues on protein groups and produce conformational changes in the protein structure (Pérez de los Santos et al., 2016). Furthermore, Cu^{2+} oxidizes cysteine residues in the protein and cause structural changes and alteration in the protein activity.

The substrate specificity and action mode of GspInv were investigated by incubating the enzyme with sucrose, trehalose, cellobiose, maltose, isomaltose, raffinose, melezitose, stachyose, and inulin at pH 5.0 and 45°C. Similar to most of yeast invertases, which are active only against sucrose and raffinose (Plascencia-Espinosa et al., 2014), GspInv released fructose from sucrose and raffinose and did not exhibit hydrolysis activity toward other substrates (**Supplementary Table S4**), suggesting that GspInv is invertase (Zhang et al., 2015). The kinetic constants of GspInv on sucrose were tested at optimal conditions. The values of the kinetic parameters K_m, k_{cat}, and k_{cat}/K_m were 8.7 ± 1.1 mM, 5,100 ± 10.8 s^{-1}, and 595 ± 76.6 mM^{-1} s^{-1}, respectively. The K_m value of GspInv falls at the lower end of the K_m values of 9.1–61.2 mM reported for most invertases (**Table 2**). This result suggests a relatively higher affinity for sucrose than most of invertases (Pérez et al., 2001; Nadeem et al., 2009; Zhou et al., 2016). In accordance with the phylogenetic analysis (**Figure 1**), the specific activities of GspInv on sucrose and raffinose were 2,776.1 ± 124.2 and 2,098.7 ± 123.6 U/mg, respectively (**Supplementary Table S4**), considerably higher than that of most invertases, such as that from *Aspergillus foetidus* (257.2 U/mg) (Rehm et al., 1998; **Table 2**).

The bioinformatics and biochemical data presented in this study support the conclusion that GspInv is invertase. Fungi are a rich source of GH32 invertases. However, no invertase from genus *Gongronella* has been reported, although previous studies have shown that fungi from genus *Gongronella*, such as *Gongronella butleri*, grow well when sucrose is used as carbon source (Kollerov et al., 2008). In the present study, we identified, cloned, and characterized an invertase GspInv from *Gongronella* sp. w5, representing the first report of invertase from this genus.

TABLE 3 | Kinetic parameters of GspInv, CBM_{24}DG-GspInv, and Cellulose-CBM_{24}DG-GspInv.

Enzyme	Specific activity (U/mg)	km (mM)	kcat (s^{-1})	kcat/km ($mM^{-1}s^{-1}$)
GspInv	2776	8.7 ± 1.1	5100 ± 10.8	595.9 ± 76.6
CBM_{24}DG-GspInv	2580	10.3 ± 1.2	4588 ± 7.9	451.4 ± 52.1
Cellulose-CBM_{24}DG-GspInv	2335	11.6 ± 1.9	4427 ± 11.2	393.4 ± 68.5

Invertase activity was determined at 45°C in citrate-phosphate buffer (50 mM, pH 5.0) using sucrose (1 mM–2 M) as the substrate. The data presented are the average values from triplicate technical repeats of measurements.

Effects of Mono- and Di-Sugars on GspInv Activity

GspInv activity was determined using varying concentrations of sucrose as the substrate to evaluate the application potential of

GspInv on inverted syrup and HFS production. GspInv activity increased with the increasing sucrose concentration from 10 to 300 mM. The highest invertase activity was observed at 300 mM sucrose, with a specific activity of 2,776.1 ± 124.2 U/mg protein. With the further increase in sucrose concentration, the specific activity of GspInv decreased gradually, with the IC_{50} (at which concentration GspInv retained 50% of the original activity) of 1.2 M. GspInv showed a specific activity of approximately 500 U/mg in the presence of 2 M sucrose (**Figure 4A**). Compare with sucrose, glucose and fructose inhibited GspInv activity. GspInv retained 70 and 50% of the original activities in the presence of 2 M glucose and 1.5 M fructose, respectively (**Figures 4B,C**).

Sucrose is regarded as one of the inhibitors that block the preparation of inverted sugar at high concentrations. Commercially used invertase from *Saccharomyces cerevisiae* is inhibited by 5% sucrose (∼146 mM) (Vásquez-Bahena et al., 2004). Another commercial yeast invertase (Bioinvert) is inhibited by 50 mM sucrose (Tomotani and Vitolo, 2007). For this reason, novel invertases tolerant to high-concentration sucrose should be explored for HFS preparation (Pérez de los Santos et al., 2016). Our data suggested the high sucrose tolerance ability of GspInv. As such, it may be a promising candidate for preparing HFS with a high sucrose concentration (>1 M). In comparison, the invertases from *A. niger* showed approximately 30% of its highest activity at 1 M sucrose (Goosen et al., 2007). Meanwhile, invertases possess high tolerance to hexoses, such as fructose and glucose, are also take advantages in HFS preparation when using alternative substrates such as molasses as the substrate. Molasses usually contains approximately 40%

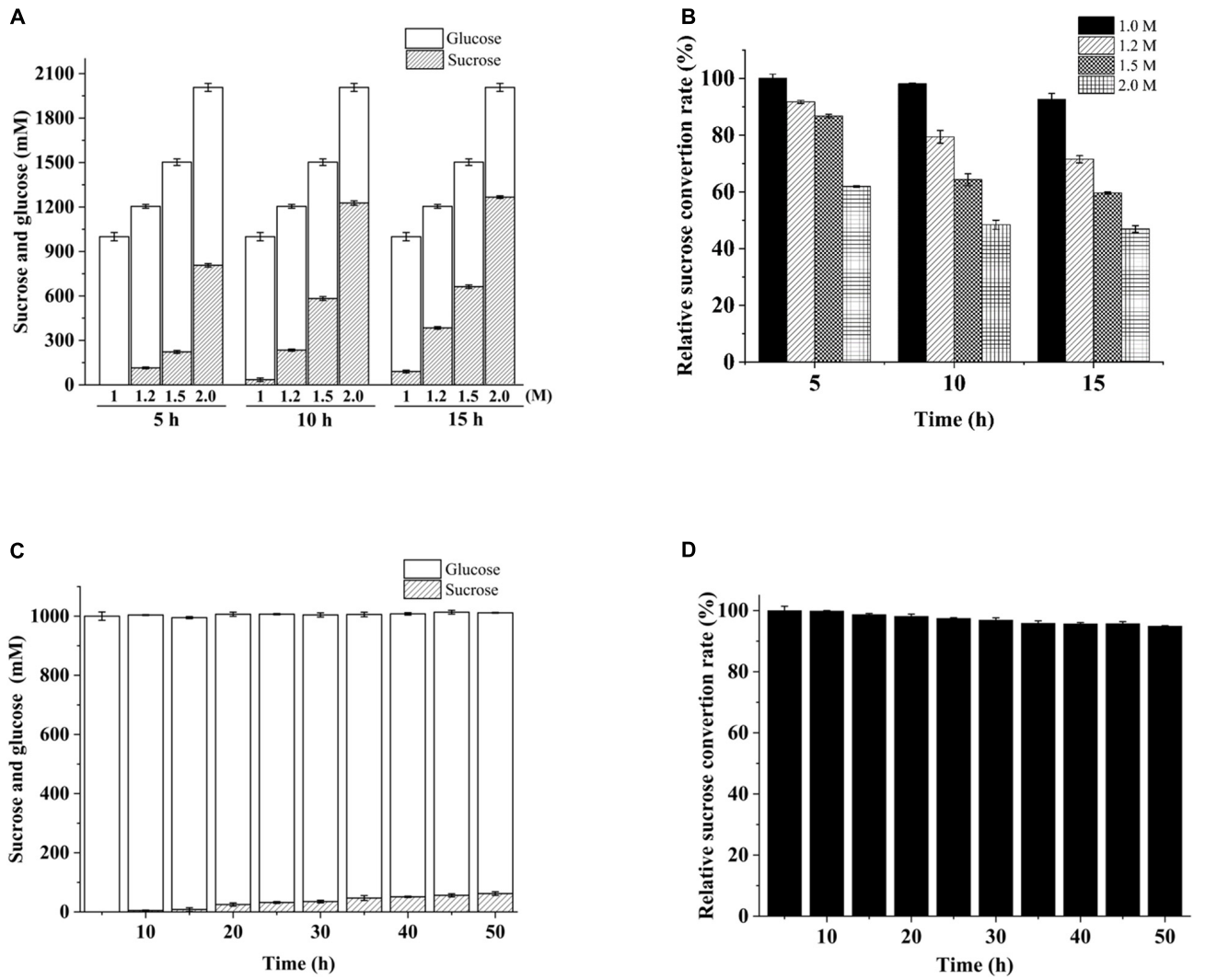


FIGURE 7 | Variation in conversion during continuous substrate hydrolysis catalyzed by immobilized cellulose-CBM_{24}DG-GspInv on PBR. **(A)** Sucrose and glucose concentration and **(B)** sucrose conversion rate after 5, 10, and 15 h continuous hydrolyzation using different concentrations of sucrose as the substrate. **(C)** Sucrose and glucose concentration and **(D)** sucrose conversion rate during 50 h continuous hydrolyzation in PBR using 1 M sucrose as the substrate. Standard deviations and values were calculated from triplicate technical repeats of measurements.

sucrose, 5–7% glucose, and 6–10% fructose (Najafpour and Shan, 2003). Because most invertases are inhibited by the end-product glucose and fructose (Ende and Laere, 1995; Lee and Sturm, 1996; Vorster and Botha, 1998; Hossain et al., 2011), invertases like GspInv that exhibit high hexose tolerance will facilitate the HFS preparation.

Construction GspInv Fusion Proteins and Their Immobilization on Cellulose

Immobilizing enzymes allow them to process large amounts of the substrate because they can be separated easily from the product; thus, the enzyme can be reused continuously. Extensive studies have been conducted on immobilizing invertases (Kotwal and Shankar, 2009). Results showed that the same kinetics could be obtained using a more conventional immobilizing matrix such as cellulose (Shoseyov and Warren, 1997; Boraston et al., 2002; Shoseyov et al., 2006; Lu et al., 2012). As a result, we immobilized GspInv on cellulose by employing CBM, without did control experiments by immobilizing GspInv on other matrixes. CBMs are small components of several cellulases that can bind specifically to cellulose (Boraston et al., 2004; Oliveira et al., 2015). Four GspInv derivates including CBM_{24}-GspInv, CBM_{24}DG-GspInv, GspInv-CBM_{24}, and GspInv-CBM_{24}DG were constructed and expressed in *K. pastoris* to facilitate the application of GspInv. Similar to GspInv, CBM_{24}DG-GspInv expressed and presented as the only one band on the SDS-PAGE gel (**Figure 5**). Trace activity (approximately 0.4 U/mL) was detected in the culture supernatant when CBM_{24}-GspInv was expressed by the strain, indicating that glycosylation may have a negative effect on CBM_{24}-GspInv activity (Wan et al., 2011). No target protein bands were observed in the culture supernatants of GspInv-CBM_{24} and GspInv-CBM_{24}DG (**Figure 5**), suggesting that CBM fused at the C-terminus of GspInv may affect the expression and/or secretion of proteins (Hussack et al., 2009; Tabuchi et al., 2010). As a result, CBM_{24}DG-GspInv was chosen for further research.

The crude fusion enzymes were absorbed onto cellulose directly by incubating CBM_{24}DG-GspInv with cellulose (designated as Cellulose-CBM_{24}DG-GspInv). Only one band was observed on SDS-PAGE gel after absorbing the CBM_{24}DG-GspInv onto cellulose (**Figure 5**). Based on the invertase activity determination, the immobilization efficiency of CBM_{24}DG-GspInv reached approximately 85% after incubation with cellulose for 0.5 h and reached saturation after approximately 1 h incubation (90%). The amount of CBM_{24}DG-GspInv that can adsorb to cellulose under optimal conditions was approximately 8,200 U/g cellulose, which transfer to approximately 3.5 mg of protein. The CBM_{24}DG-GspInv was adsorbed irreversibly with cellulose and approximately 2% of enzyme activity (120–200 U/g cellulose) was eluted from the cellulose during washing. Meanwhile, no GspInv was bound to cellulose (**Figure 6**). CBMs such as CBM_{24} from family 3 usually exhibit high absorption efficiencies toward cellulose because they usually have a flat hydrophobic binding surface comprising of aromatic residues (Boraston et al., 2004; Chang et al., 2018). The planar architecture of the binding sites is consistent with the flat surfaces of crystalline polysaccharides such as cellulose, facilitating their irreversible binding and reaching a plateau within a short time (Boraston et al., 2004).

TABLE 4 | Summary of data from studies of the continuous hydrolysis of sucrose by different immobilized invertases deployed in bioreactors.

Actual source (producer)	Method of immobilization	Sucrose concentration (g/L)	Operational conditions	Conversion factor	References
GspInv (this study)	CBM bound with cellulose	342	Continuous regime in PBR, 150 mL/h 50 h, 35°C, pH5	1–0.95	This study
Bioinvert (quest internatonal)	Adsorption on polystyrene Dowex beads	0.8	Continuous regime in membrane CSTR, 20 h, 30°C, pH 5.5	1	Tomotani and Vitolo, 2007
Baker yeast (Sigma Aldrich)	Covalent binding on activated montmorillonite	100	Continuous regime in PBR, 96 h, 30°C, pH6	1–0.75	Sanjay and Sugunan, 2006
S. cerevisiae (Fluka)	Covalent binding on chitosan and inclusion in alginate	68.4	Continuous regime, 50 h, 30°C, pH 4.6, 20 mL/h	1–0.8	Gómez et al., 2006
S. cerevisiae (Fluka)	Covalent binding on nylon micro-beads	273	Continuous regime, 38 h, 50°C, pH 5.5, 318 mL/h	0.97	Amaya-Delgado et al., 2006
S. cerevisiae Bento Gonc, Brazil)	Cross-linked enzyme aggregate methodology	100	Continuous regime, 40 h, 40°C, pH 6.0, 60 mL/h	0.75	Oliveira et al., 2018
Baker yeast (Sigma Aldrich)	Invertase immobilized on glutaraldehyde crosslinked chitosan beads	170	Continuous regime, 25 h, 30°C, pH 4.0, 18 mL/h	0.96	Koli and Gaikar, 2016

CBM_{24}DG-GspInv and Cellulose-CBM_{24}DG-GspInv Characterization and Comparison

The characteristics of CBM_{24}DG-GspInv and cellulose-CBM_{24}DG-GspInv were determined and compared to those of GspInv. **Figure 3** shows the optimum pH and temperature for CBM_{24}DG-GspInv and cellulose-CBM_{24}DG-GspInv were pH 5.0 and 45°C, respectively. They also shared similar pH and thermal stabilities with that of GspInv, with a half-life of approximately 100 h at 40°C and pH 5.0 (**Figure 3D**). Analysis of the apparent kinetic parameters showed a few differences in the K_m and k_{cat} values among GspInv, CBM_{24}DG-GspInv, and cellulose-CBM_{24}DG-GspInv. The K_m values of CBM_{24}DG-GspInv (10.3 ± 1.2 mM) and cellulose-CBM_{24}DG-GspInv (11.6 ± 1.9 mM) were slightly higher than those of GspInv (8.7 ± 1.1 mM) (**Table 3**). The specific activities of CBM_{24}DG-GspInv (2,580 U/mg) and cellulose-CBM_{24}DG-GspInv (2,335 U/mg) decreased when compared to that of GspInv (2,776 U/mg) (**Table 3**). This phenomenon may be due to the fusion of CBM_{24} with GspInv, which caused the conformational change of the invertase and the restriction of diffusion of the substrate into the enzyme in the solid matrix (Koli and Gaikar, 2016). Thus, substrate affinity was reduced and glucose and fructose could easily enter the substrate-binding pocket.

Reusability of the Immobilized Invertase During HFS Preparation

The repeated use of immobilized invertases is crucial to modern industries (Goradia et al., 2010). In our continuous flow experiment conducted on PBR, 1 M sucrose was hydrolyzed completely by cellulose-CBM_{24}DG-GspInv to an equivalent amount of glucose and fructose after 15 h of reaction. In comparison, approximately 1.0–1.3 M inverted sugar was obtained in the first 5 h when 1.2–2.0 M sucrose was used as the substrate (**Figure 7A**). The hydrolyzation rate decreased continuously in the reaction with 1.2–2.0 M sucrose as the substrate and decreased to 50% after 15 h reaction by using 2 M sucrose as the substrate (**Figure 7B**). As such, sucrose at concentrations lower than 1 M was suitable for HFS preparation.

PBR containing cellulose-CBM_{24}DG-GspInv was loaded continuously with 1 M sucrose for 50 h to determine further the reusability of cellulose-CBM_{24}DG-GspInv. The Results showed the hydrolyzation rate was maintained at 95% after 50 h reaction (**Figure 7C**). Repeated use of enzymes can reduce the cost of production significantly. Several groups have reported sucrose hydrolyzation using different invertases (**Table 4**). Compared with such invertases, GspInv has high sucrose concentration and conversion factor, which implies its application potential in producing inverted sugars (Amaya-Delgado et al., 2006; Gómez et al., 2006; Sanjay and Sugunan, 2006; Tomotani and Vitolo, 2007; Koli and Gaikar, 2016; Oliveira et al., 2018).

Commercially, invert syrup production by using sucrose is economically unavailable due to its high price (Piotr, 2019). Molasses is considered as an alternative of sucrose because it is rich in sucrose and glucose (30–50%, v/v) and lower in price (Deng et al., 2008). The application potential of GspInv in HFS production was evaluated by using molasses as the substrate. After dilution and pretreatment, the remained liquid molasses contained 20% sucrose, 2.7% fructose, and 3.2% glucose. GspInv hydrolyzed molasses efficiently. A final concentration of 142 g/L glucose was obtained after hydrolyzing the 20% molasses with 5,000 U/L GspInv or cellulose-CBM_{24}DG-GspInv for 20 min. However, quite different from that using sucrose as the substrate, the hydrolyzation rate decreased to 30% after 1 h hydrolyzation of 20% molasses on PBR containing cellulose-CBM_{24}DG-GspInv and decreased to zero within 2 h. The pretreated molasses still contains substantial amounts of mineral substance (4.2% in our case) and suspended colloids. These impurities may interact with GspInv and inhibit activity continuously.

CONCLUSION

In this study, GspInv from *Gongronella* sp. w5 was cloned and expressed in *K. pastoris*. Biochemical characterization indicated that GspInv is an unusual invertase with high specific activity and high tolerance to sucrose, glucose, fructose, and various metal ions. Its application potential in HFS production was explored by fusing a sequence optimized CBM to the N-terminal to purify and immobilize GspInv. The fused protein exhibited high efficiency in sucrose and molasses hydrolyzation. Our results suggest GspInv is an unusual invertase and a promising candidate for HFS preparation.

DATA AVAILABILITY STATEMENT

The sequence information of GspInv can be found in the GenBank with accession No. AVI04916.

AUTHOR CONTRIBUTIONS

JL and ZF conceived and designed the research. GZ, CP, XL, and FC organized and performed the experiments. JL, ZF, and YX analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Chinese National Natural Science Foundation grant numbers 31870098 and 31800051) and the Doctoral Research Start-up Funding of Anhui University (Y040418162).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00633/full#supplementary-material

REFERENCES

Acosta, N., Beldarraín, A., Rodríguez, L., and Alonso, Y. (2000). Characterization of recombinant invertase expressed in methylotrophic yeasts. *Biotechnol. Appl. Biochem.* 32, 179–187. doi: 10.1042/ba20000064

Amaya-Delgado, L., Hidalgo-Lara, M. E., and Montes-Horcasitas, M. C. (2006). Hydrolysis of sucrose by invertase immobilized on nylon-6 microbeads. *Food Chem.* 99, 299–304. doi: 10.1016/j.foodchem.2005.07.048

Ashwell, G. (1957). Colorimetric analysis of sugars. *Methods Enzymol.* 3, 73–105. doi: 10.1016/s0076-6879(57)03350-9

Boraston, A. B., Bolam, D., Gilbert, H., and Davies, G. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781. doi: 10.1042/BJ20040892

Boraston, A. B., McLean, B. W., Kavoosi, M., Haynes, C. A., and Kilburn, D. G. (2002). "Cellulose-binding fusion proteins," in *Methods and Tools in Biosciences and Medicine*, ed. M. N. Gupta (Basel: Birkhäuser), 148–162. doi: 10.1007/978-3-0348-8127-2_8

Cai, S., Xiong, Z., Li, L., Li, M., Zhang, L., Liu, C., et al. (2014). Differential responses of root growth, acid invertase activity and transcript level to copper stress in two contrasting populations of Elsholtzia haichowensis. *Ecotoxicology* 23, 76–91. doi: 10.1007/s10646-013-1153-y

Chang, F., Xue, S., Xie, X., Fang, W., Fang, Z., and Xiao, Y. (2018). Carbohydrate-binding module assisted purification and immobilization of β-glucosidase onto cellulose and application in hydrolysis of soybean isoflavone glycosides. *J. Biosci. Bioeng.* 125, 185–191. doi: 10.1016/j.jbiosc.2017.09.001

Chen, G. J., Yang, J. K., Peng, X. B., and He, J. R. (2016). High-level secretory expression of Aspergillus exo-inulinase and its use in the preparation of fructose syrup from inulin. *J. Mol. Catal. B. Enzym.* 133, 543–551. doi: 10.1016/j.molcatb.2017.09.001

Deng, Z. N., Li, N., Qin, Y. L., Chen, C. L., and Liang, Z. Q. (2008). Production of polyunsaturated fatty acids by *Mucor recurvus* sp. with sugarcane molasses as the carbon source. *Food Technol. Biotechnol.* 46, 73–79. doi: 10.1080/08905430802470235

Dong, Y., Sun, Q., Zhang, Y., Wang, X., Liu, P., Xiao, Y., et al. (2018). Complete genome of *Gongronella* sp. w5 provides insight into its relationship with plant. *J. Biotechnol.* 286, 1–4. doi: 10.1016/j.jbiotec.2018.08.022

Ende, W., and Laere, A. (1995). Purification and properties of a neutral invertase from the roots of *Cichorium intybus*. *Physiol. Plant.* 93, 241–248. doi: 10.1111/j.1399-3054.1995.tb02223.x

Eskandarloo, H., and Abbaspourrad, A. (2018). Production of galacto-oligosaccharides from whey permeate using β-galactosidase immobilized on functionalized glass beads. *Food Chem.* 251, 115–124. doi: 10.1016/j.foodchem.2018.01.068

Fatourehchi, N., Sohrabi, M., Dabir, B., and Royaee, S. J. (2014). Application of a novel type impinging streams reactor in glucose conversion to fructose using glucose isomerase enzyme. *J. Chem. Technol. Biotechnol.* 89, 1918–1923. doi: 10.1002/jctb.4276

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). Interpro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 45, 190–199. doi: 10.1093/nar/gkw1107

Gabisa, E. W., Bessou, C., and Gheewala, S. H. (2019). Life cycle environmental performance and energy balance of ethanol production based on sugarcane molasses in Ethiopia. *J. Clean. Prod.* 234, 43–53. doi: 10.1016/j.jclepro.2019.06.199

Gómez, L., Ramírez, H. L., Villalonga, M. L., Hernández, J., and Villalonga, R. (2006). Immobilization of chitosan-modified invertase on alginate-coated chitin support via polyelectrolyte complex formation. *Enzyme Microb. Technol.* 38, 22–27. doi: 10.1016/j.enzmictec.2004.10.008

Goosen, C., Yuan, X. L., van Munster, J. M., Ram, A. F. J., van der Maarel, M. J. E. C., and Dijkhuizen, L. (2007). Molecular and biochemical characterization of a novel intracellular invertase from *Aspergillus niger* with transfructosylating activity. *Eukaryot. Cell* 6, 674–681. doi: 10.1128/EC.00361-06

Goradia, D., Cooney, J., Hodnett, B. K., and Magner, E. (2010). Characteristics of a mesoporous silicate immobilized trypsin bioreactor in organic media. *Biotechnol. Prog.* 22, 1125–1131. doi: 10.1021/bp050334y

Guthrie, J. F., and Morton, J. F. (2000). Food sources of added sweeteners in the diets of Americans. *J. Am. Diet. Assoc.* 100, 43–51. doi: 10.1016/S0002-8223(00)00018-3

Hossain, M. M., Pervin, F., and Absar, N. (2011). Purification, characterization and n-terminal sequence analysis of betel leaf (Piper betle) invertase. *J. Chin. Chem. Soc.* 58, 389–397. doi: 10.1002/jccs.201190042

Hsieh, C. W., Liu, L. K., Yeh, S. H., Chen, C. F., Lin, H. I., Sung, H. Y., et al. (2006). Molecular cloning and functional identification of invertase isozymes from green bamboo *Bambusa oldhamii*. *J. Agric. Food Chem.* 54, 3101–3107. doi: 10.1021/jf052711s

Hu, J., Zhang, Y., Xu, Y., Sun, Q., Liu, J., Fang, W., et al. (2018). *Gongronella* sp. w5 elevates *Coprinopsis cinerea* laccase production by carbon source syntrophism and secondary metabolite induction. *Appl. Microbiol. Biotechnol.* 103, 411–425. doi: 10.1007/s00253-018-9469-4

Huang, W. C., Wang, A. Y., Wang, L. T., and Sung, H. Y. (2003). Expression and characterization of sweet potato invertase in *Pichia pastoris*. *J. Agric. Food Chem.* 51, 1494–1499. doi: 10.1021/jf026032i

Hussack, G., Luo, Y., Veldhuis, L., Hall, J. C., Tanha, J., and Mackenzie, R. (2009). Multivalent anchoring and oriented display of single-domain antibodies on cellulose. *Sensors* 9, 5351–5367. doi: 10.3390/s90705351

Isla, M. I., Vattuone, M. A., Ordóñez, R. M., and Sampietro, A. R. (1999). Invertase activity associated with the walls of *Solanum tuberosum* tubers. *Phytochemistry* 50, 525–534. doi: 10.1016/s0031-9422(98)00474-9

Khatiwada, D., Venkata, B. K., Silveira, S., and Johnson, F. X. (2016). Energy and GHG balances of ethanol production from cane molasses in Indonesia. *Appl. Energy* 164, 756–768. doi: 10.1016/j.apenergy

Koli, A. C., and Gaikar, V. G. (2016). Continuous cane sugar inversion process using immobilized invertase. *J. Chem. Technol. Biotechnol.* 4, 787–792. doi: 10.1002/jctb.5060

Kollerov, V. V., Shutov, A. A., Fokina, V. V., Sukhodol'skaya, G. V., and Donova, M. V. (2008). Biotransformation of 3-keto-androstanes by Gongronella butleri VKM F-1033. *J. Mol. Catal. B Enzym.* 55, 61–66. doi: 10.1016/j.molcatb.2008.01.009

Kotwal, S. M., and Shankar, V. (2009). Immobilized invertase. *Biotechnol. Adv.* 273, 11–322. doi: 10.1016/j.biotechadv.2009.01.009

Lee, H. S., and Sturm, A. (1996). Purification and characterization of neutral and alkaline invertase from carrot. *Plant Physiol.* 112, 1513–1522. doi: 10.1104/pp.112.4.1513

Lima, D. M., Fernandes, P., Nascimento, D. S., Ribeiro, R. D., and Assis, S. A. (2011). Fructose syrup: a biotechnology asset. *Food Technol. Biotechnol.* 49, 424–434. doi: 10.1016/j.ijfoodmicro.2011.07.006

Liu, Y. P., Zheng, P., Sun, Z. H., Ni, Y., Dong, J. J., and Zhu, L. L. (2008). Economical succinic acid production from cane molasses by Actinobacillus succinogenes. *Bioresour. Technol.* 99, 1736–1742. doi: 10.1016/j.biortech.2007.03.044

Lu, L., Xu, S., Zhao, R., Zhang, D., Li, Z., and Li, Y. (2012). Synthesis of galactooligosaccharides by CBD fusion β-galactosidase immobilized on cellulose. *Bioresour. Technol.* 116, 327–333. doi: 10.1016/j.biortech.2012.03.108

Mohammadi, M., Mokarram, R. R., Ghorbani, M., and Hamishehkar, H. (2018). Inulinase immobilized gold-magnetic nanoparticles as a magnetically recyclable biocatalyst for facial and efficient inulin biotransformation to high fructose syrup. *Int. J. Biol. Macromol.* 123, 846–855. doi: 10.1016/j.ijbiomac.2018.11.160

Nadeem, H., Rashid, M., Riaz, M., Asma, B., Javed, M., and Perveen, R. (2009). Invertase from hyper producer strain of *Aspergillus niger*: physiochemical properties, thermodynamics and active site residues heat of ionization. *Protein Pept. Lett.* 16, 1098–1105. doi: 10.2174/092986609789055322

Nadeem, H., Rashid, M. H., Siddique, M. H., Azeem, F., Muzammil, S., Javed, M. R., et al. (2015). Microbial invertases: a review on kinetics, thermodynamics, physiochemical properties. *Process Biochem.* 50, 1202–1210. doi: 10.1016/j.procbio.2015.04.015

Najafpour, G. D., and Shan, C. P. (2003). Enzymatic hydrolysis of molasses. *Bioresour. Technol.* 86, 91–94. doi: 10.1016/S0960-8524(02)00103-7

Neifar, S., Hlima, H. B., Mhiri, S., Mezghani, M., Bouacem, K., Ibrahim, A. H., et al. (2019). A novel thermostable and efficient class II glucose isomerase from the thermophilic *Caldicoprobacter algeriensis*: biochemical characterization, molecular investigation, and application in high fructose syrup production. *Int. J. Biol. Macromol.* 129, 31–40. doi: 10.1016/j.ijbiomac.2019.01.150

Oliveira, C., Carvalho, V., Domingues, L., and Gama, F. M. (2015). Recombinant CBM-fusion technology-applications overview. *Biotechnol. Adv.* 33, 358–369. doi: 10.1016/j.biotechadv.2015.02.006

Oliveira, M. A. C., Bontorin, B. M., Ulrich, L. G., Camargo, G. R. D. L., Arruda, R. M. P. D., and Waldir, T. P. (2018). Combined cleas of invertase and soy

protein for economically feasible conversion of sucrose in a fed-batch reactor. *Food Bioprod. Process.* 110, 145–157. doi: 10.1016/j.fbp.2018.05.006

Palai, T., Singh, A. K., and Bhattacharya, P. K. (2014). Enzyme, β-galactosidase immobilized on membrane surface for galacto-oligosaccharides formation from lactose: kinetic study with feed flow under recirculation loop. *Biochem. Eng. J.* 88, 68–76. doi: 10.1016/j.bej.2014.03.017

Pérez, J. A., Rodríguez, J., Ruiz, T., and Rodríguez, L. (2001). Expression of Pichia anomala INV1 gene in Saccharomyces cerevisiae results in two different active forms of hypoglycosylated invertase. *Arch. Microbiol.* 175, 189–197. doi: 10.1007/s002030100253

Pérez de los Santos, A. I., Cayetano-Cruz, M., Gutiérrez-Antón, M. A., Santiago-Hernández, A., Plascencia-Espinosa, M., Farrés, A., et al. (2016). Improvement of catalytical properties of two invertases highly tolerant to sucrose after expression in *Pichia pastoris*. Effect of glycosylation on enzyme properties. *Enzyme Microb. Technol.* 83, 48–56. doi: 10.1016/j.enzmictec.2015.11.008

Piotr, Z. (2019). Decolorisation of methylene blue with sodium persulfate activated with visible light in the presence of glucose and sucrose. *Water Air Soil Pollut.* 230, 313–330. doi: 10.1007/s11270-019-4372-x

Plascencia-Espinosa, M., Santiago-Hernández, A., Pavón-Orozco, P., Vallejo-Becerra, V., Trejo-Estrada, S., Sosa-Peinado, A., et al. (2014). Effect of deglycosylation on the properties of thermophilic invertase purified from the yeast *Candida guilliermondii* MpIIIa. *Process Biochem.* 49, 1480–1487. doi: 10.1016/j.procbio.2014.05.022

Pons, A. M., Delalande, F., Duarte, M., Benoit, S., Lanneluc, I., Sable, S., et al. (2004). Genetic analysis and complete primary structure of microcin L. *Antimicrob. Agents Chemother.* 48, 505–513. doi: 10.1128/AAC.48.2.505-513.2004

Ramírez-Escudero, M., Del Pozo, M. V., Marín-Navarro, J., González, B., Golyshin, P. N., Polaina, J., et al. (2016). Structural and functional characterization of a ruminal β-glycosidase defines a novel subfamily of glycosyl hydrolase family 3 with permuted domain topology. *J. Biol. Chem.* 291, 24200–24214. doi: 10.1074/jbc.M116.747527

Rashad, M. M., and Nooman, M. U. (2009). Production, purification and characterization of extracellular invertase from *Saccharomyses* cerevisiae NRRL Y-12632 by solid-state fermentation of red carrot residue. *Aust. J. Basic Appl. Sci.* 3, 1910–1919.

Rehm, J., Willmitzer, L., and Heyer, A. G. (1998). Production of 1-kestose in transgenic yeast expressing a fructosyltransferase from *Aspergillus foetidus*. *J. Bacteriol.* 180, 1305–1310. doi: 10.1128/jb.180.5.1305-1310.1998

Resa, P., Elvira, L., Sierra, C., and de Espinosa, F. M. (2009). Ultrasonic velocity assay of extracellular invertase in living yeasts. *Anal. Biochem.* 384, 68–73. doi: 10.1016/j.ab.2008.09.025

Rippe, J. M. (2014). *Fructose, High Fructose Corn Syrup, Sucrose, and Health: Modern Scientific Understandings*. New York, NY: Springer, 3–12. doi: 10.1007/978-1-4899-8077-9_1

Rustiguel, C. B., Jorge, J. A., and Guimarães, L. H. S. (2015). Characterization of a thermo-tolerant mycelial β-fructofuranosidase from *Aspergillus phoenicis* under submerged fermentation using wheat bran as carbon source. *Biocatal. Agric. Biotechnol.* 4, 362–369. doi: 10.1016/j.bcab.2015.05.004

Sakakibara, M., Wang, D., Takahashi, R., Takahashi, K., and Mori, S. (1996). Influence of ultrasound irradiation on hydrolysis of sucrose catalyzed by invertase. *Enzyme Microb. Technol.* 18, 444–448. doi: 10.1016/0141-0229(95)00128-x

Sanjay, G., and Sugunan, S. (2006). Fixed bed reactor performance of invertase immobilized on montmorillonite. *Catal. Commun.* 7, 1005–1011. doi: 10.1016/j.catcom.2005.12.029

Shoseyov, O., Shani, Z., and Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283–295. doi: 10.1128/MMBR.00028-05

Shoseyov, O., and Warren, R. A. J. (1997). Cellulose binding domains-a novel fusion technology for efficient, low cost purification and immobilization of recombinant proteins. *Novagen Novations Newsl.* 7, 1–3.

Singh, R. S., Chauhan, K., Pandey, A., and Larroche, C. (2018a). Biocatalytic strategies for the production of high fructose syrup from inulin. *Bioresour. Technol.* 260, 395–403. doi: 10.1016/j.biortech.2018.03.127

Singh, R. S., Chauhan, K., Pandey, A., Larroche, C., and Kennedy, J. F. (2018b). Purification and characterization of two isoforms of exoinulinase from *Penicillium oxalicum* BGPUP-4 for the preparation of high fructose syrup from inulin. *Int. J. Biol. Macromol.* 118, 1974–1983. doi: 10.1016/j.ijbiomac.2018.07.040

Tabuchi, S., Ito, J., Adachi, T., Ishida, H., Hata, Y., and Okazaki, F. (2010). Display of both N- and C-terminal target fusion proteins on the *Aspergillus oryzae* cell surface using a chitin-binding module. *Appl. Microbiol. Biotechnol.* 87, 1783–1789. doi: 10.1007/s00253-010-2664-6

Tomotani, E. J., and Vitolo, M. (2007). Production of high-fructose syrup using immobilized invertase in a membrane reactor. *J. Food Eng.* 80, 662–667. doi: 10.1016/j.jfoodeng.2006.07.002

Torres-Acosta, M. A., Morales-Guzman, S. I., Federico, R. R., Patricia, V. V., Willson, R. C., and Marco, R. P. (2018). Monte Carlo economic analysis of baker's yeast invertase purification using two- and three-phase partitioning. *J. Chem. Technol. Biotechnol.* 93, 2511–2517. doi: 10.1002/jctb.5730

Trollope, K. M., Gorgens, J. F., and Volschenk, H. (2015). Semirational directed evolution of loop regions in *Aspergillus japonicus* beta-fructofuranosidase for improved fructooligosaccharide production. *Appl. Microbiol. Biotechnol.* 81, 7319–7329. doi: 10.1128/aem.02134-15

Van, W. N., Trollope, K. M., Steenkamp, E. T., Wingfield, B. D., and Volschenk, H. (2013). Identification of the gene for beta-fructofuranosidase from Ceratocystis moniliformis CMW 10134 and characterization of the enzyme expressed in Saccharomyces cerevisiae. *BMC Biotechnol.* 13:100. doi: 10.1186/1472-6750-13-100

Vásquez-Bahena, J., Montes-Horcasitas, M. C., Ortega-López, J., Magaña-Plaza, I., and BFlores-Cotera, L. (2004). Multiple steady states in a continuous stirred tank reactor: an experimental case study for hydrolysis of sucrose by invertase. *Process Biochem.* 39, 2179–2182. doi: 10.1016/j.procbio.2003.11.007

Veana, F., Fuentes-Garibay, J. A., Aguilar, C. N., Rodríguez-Herrera, R., Guerrero-Olazarán, M., and Viader-Salvadó, J. M. (2014). Gene encoding a novel invertase from a xerophilic *Aspergillus niger* strain and production of the enzyme in *Pichia pastoris*. *Enzyme Microb. Technol.* 63, 28–33. doi: 10.1016/j.enzmictec.2014.05.001

Vorster, D. J., and Botha, F. C. (1998). Partial purification and characterisation of sugarcane neutral invertase. *Phytochemistry* 49, 651–655. doi: 10.1016/S0031-9422(98)00204-0

Wan, W., Wang, D., Gao, X., and Hong, J. (2011). Expression of family 3 cellulose-binding module (CBM3) as an affinity tag for recombinant proteins in yeast. *Appl. Microbiol. Biotechnol.* 91, 789–798. doi: 10.1007/s00253-011-3373-5

Wang, L. T., Wang, A. Y., Hsieh, C. W., Chen, C. Y., and Sung, H. Y. (2005). Vacuolar invertases in sweet potato: molecular cloning, characterization, and analysis of gene expression. *J. Agric. Food Chem.* 53, 3672–3678. doi: 10.1021/jf0480851

Wang, Y. J., Jiang, X. W., Liu, Z. Q., Jin, L. Q., Liao, C. J., Cheng, X. P., et al. (2016). Isolation of fructose from high-fructose corn syrup with calcium immobilized strong acid cation exchanger: isotherms, kinetics, and fixed-bed chromatography study. *Can. J. Chem. Eng.* 94, 537–546. doi: 10.1002/cjce.22418

Xia, J., Xu, J., Hu, L., and Liu, X. (2016). Enhanced poly (L-malic acid) production from pretreated cane molasses by Aureobasidium pullulans in fed-batch fermentation. *Prep. Biochem. Biotechnol.* 46, 798–802. doi: 10.1080/10826068.2015.1135464

Zhang, Y., Hidajat, K., and Ray, A. K. (2004). Optimal design and operation of SMB bioreactor: production of high fructose syrup by isomerization of glucose. *Biochem. Eng. J.* 21, 111–121. doi: 10.1016/j.bej.2004.05.007

Zhang, Y., Zhu, H., Wang, J., Zhou, X., Xu, W., and Shi, H. (2015). Isolation and identification of an inulinase-producing strain and the optimization of its fermentation condition. *Adv. Appl. Biotechnol.* 332, 93–107. doi: 10.1007/978-3-662-45657-6_11

Zhou, J., He, L., Gao, Y., Han, N., Zhang, R., and Wu, Q. (2016). Characterization of a novel low-temperature-active, alkaline and sucrose-tolerant invertase. *Sci. Rep.* 61, 233–242. doi: 10.1007/s12223-015-0430-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 April 2020
doi: 10.3389/fmicb.2020.00742

Crystal Structure and Active Site Engineering of a Halophilic γ-Carbonic Anhydrase

Malvina Vogler[1,2†], Ram Karan[1†], Dominik Renn[1], Alexandra Vancea[1], Marie-Theres Vielberg[2], Stefan W. Grötzinger[1], Priya DasSarma[3], Shiladitya DasSarma[3], Jörg Eppinger[1], Michael Groll[2*] and Magnus Rueping[1*]*

[1] KAUST Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] Center for Integrated Protein Science Munich, Department of Chemistry, Technische Universität München, Garching, Germany, [3] Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, Baltimore, MD, United States

OPEN ACCESS

Edited by:
Sunil Khare,
Indian Institute of Technology Delhi, India

Reviewed by:
Amy Michele Grunden,
North Carolina State University, United States
Mohammad Ali Amoozegar,
University of Tehran, Iran

***Correspondence:**
Ram Karan
ram.karan@kaust.edu.sa
Michael Groll
michael.groll@tum.de
Magnus Rueping
magnus.rueping@kaust.edu.sa

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: 18 February 2020
Accepted: 30 March 2020
Published: 28 April 2020

Citation:
Vogler M, Karan R, Renn D, Vancea A, Vielberg M-T, Grötzinger SW, DasSarma P, DasSarma S, Eppinger J, Groll M and Rueping M (2020) Crystal Structure and Active Site Engineering of a Halophilic γ-Carbonic Anhydrase. Front. Microbiol. 11:742. doi: 10.3389/fmicb.2020.00742

Environments previously thought to be uninhabitable offer a tremendous wealth of unexplored microorganisms and enzymes. In this paper, we present the discovery and characterization of a novel γ-carbonic anhydrase (γ-CA) from the polyextreme Red Sea brine pool Discovery Deep (2141 m depth, 44.8°C, 26.2% salt) by single-cell genome sequencing. The extensive analysis of the selected gene helps demonstrate the potential of this culture-independent method. The enzyme was expressed in the bioengineered haloarchaeon *Halobacterium* sp. NRC-1 and characterized by X-ray crystallography and mutagenesis. The 2.6 Å crystal structure of the protein shows a trimeric arrangement. Within the γ-CA, several possible structural determinants responsible for the enzyme's salt stability could be highlighted. Moreover, the amino acid composition on the protein surface and the intra- and intermolecular interactions within the protein differ significantly from those of its close homologs. To gain further insights into the catalytic residues of the γ-CA enzyme, we created a library of variants around the active site residues and successfully improved the enzyme activity by 17-fold. As several γ-CAs have been reported without measurable activity, this provides further clues as to critical residues. Our study reveals insights into the halophilic γ-CA activity and its unique adaptations. The study of the polyextremophilic carbonic anhydrase provides a basis for outlining insights into strategies for salt adaptation, yielding enzymes with industrially valuable properties, and the underlying mechanisms of protein evolution.

Keywords: extremophiles, halophiles, thermophiles, extremozyme, salt adaptation, mutagenesis, gamma-carbonic anhydrase

INTRODUCTION

Recent years have seen the discoveries of extremophiles in environments previously considered uninhabitable (Madigan and Marrs, 1997; Cavicchioli et al., 2011; Antunes et al., 2017; Jorquera et al., 2019; Merino et al., 2019). To date, enzymes from extremophiles gained increasing attention because they have adapted their structure and retained their function under harsh conditions, where their mesophilic homologs are non-functional (Persidis, 1998; Akal et al., 2019). In particular, these

proteins are attractive for biotechnological and chemical industries keen on replacing traditional catalysts with enzymes. As a result, these extremozymes provide a high stereoselectivity with fewer side reactions, and lower the burden on the environment, concomitantly accelerating reaction rates (Littlechild, 2017). However, many processes require high temperatures, use of salts, organic solvents, or other demanding conditions that are incompatible with the stability and function of most proteins (Littlechild, 2015; Amoozegar et al., 2019). Conversely, extremophilic organisms, which are naturally adapted to withstand harsh conditions, provide a perspective for optimization and rational protein-engineering approaches (Liszka et al., 2012).

The Red Sea constitutes a unique habitat of several anoxic deep-sea brine pools (Gurvich, 2006; Behzad et al., 2016) and, therefore, promises the discovery of a vast number of new extremophiles and enzymes. The anoxic environment is polyextremophilic, being filled with water, a high salt concentration, metal content, and elevated temperatures. Additionally, their increased density prevents mixing with the overlying seawater (Gurvich, 2006; Antunes et al., 2011). For example, the Discovery Deep brine pool below 2038 m is characterized by a salinity close to saturation (26.2%, w/v) with a temperature of 44.7°C (Hunt et al., 1967). Therefore, Discovery Deep's microorganisms have only been scarcely investigated; nonetheless, a recent surge in interest has begun to provide glimpses of the wealth of new information waiting to be explored (Antunes et al., 2011; Mwirichia et al., 2016; Grotzinger et al., 2018).

The limited accessibility and uncultivability of the aforementioned microorganisms with current laboratory techniques hamper the investigation of these exceptional species (Stewart, 2012). Consequently, methods independent of cultivation and their further development, such as the use of Single Amplified Genomes (SAG), are required (Kvist et al., 2007; Rashid and Stingl, 2015). Here, DNA from a single cell is amplified using the Multiple Displacement Amplification (MDA) techniques (Dean et al., 2001) to generate sufficient DNA for sequencing and to avoid the need for cultivation of the respected organisms. To improve the assembly and annotation accuracy for SAG-derived samples, our group developed the Profile and Pattern Matching (PPM) algorithm method (Grotzinger et al., 2014). In this work, we use data from SAG analysis to investigate an extremophilic γ-carbonic anhydrase (γ-CA) from the Red Sea Discovery Deep brine pool. CAs (E.C. 4.2.1.1) are ubiquitous metalloenzymes that catalyze the reversible hydration of carbon dioxide to bicarbonate ($CO_2 + H_2O \leftrightharpoons HCO_3^- + H^+$) (Supuran, 2008). To date, seven classes have been described: α-, β-, γ-, δ-, ζ-, η-, and θ-CAs. These hydrolases differ significantly in both sequence and structure, whilst all catalyzing the same reaction (Ferry, 2013; Del Prete et al., 2016; Supuran and Capasso, 2017). The α-CA class was the first to be isolated and described (Meldrum and Roughton, 1933), whereas the γ-carbonic anhydrase from the thermophilic archaeon *Methanosarcina thermophila* (Cam) was discovered in 1994 (Alber and Ferry, 1994). It was categorized as a new class showing a left-handed parallel β-helix fold, and central metal coordination in the active site by three histidine residues (Kisker et al., 1996). Subsequently, further γ-CAs have been investigated; however, half of them did not show measurable activity (Park et al., 2012; Herrou and Crosson, 2013), and the underlying molecular mechanisms are still controversial.

Here, we report on the first crystal structure of a halophilic γ-class CA (CA_D). The gene was derived from SAG analysis of an uncultured archaeon from the Red Sea Discovery Deep brine pool (Alam et al., 2013; Mwirichia et al., 2016) and was identified using the PPM algorithm (Grotzinger et al., 2014). The gene was expressed in the bioengineered haloarchaeon *Halobacterium* sp. NRC-1. We demonstrate that CA_D indeed encodes a γ-carbonic anhydrase. Activity analysis of structure-driven designed CA_D variants provided insights into the residues constituting the catalytic site.

MATERIALS AND METHODS

Chemicals and Reagents

Restriction enzymes, T4 DNA ligase, and DNA polymerase were purchased from New England Biolabs (Beverly, MA, United States). Chemicals were purchased from Sigma (St. Louis, MO, United States). Water was desalted and purified using a Milli-Q® Academic system (Merck, Darmstadt, Germany).

SAG Sampling Sites, Sample Preparation, and Genome Annotation

Samples were collected from the Discovery Deep brine pool in the Red Sea (21° 16.98′/38° 03.18′). Cells were sorted using fluorescence-activated cell sorting (FACS), lysed, the whole genome amplified and sequenced (Mwirichia et al., 2016). Genes were annotated using the INDIGO data warehouse system in combination with the profile pattern matching algorithm (PPMA) (Alam et al., 2013; Grotzinger et al., 2014).

Strains, Plasmids, Media, and Culture Conditions

Escherichia coli One Shot TOP10® chemical competent cells were purchased from Invitrogen (Carlsbad, United States). *E. coli* was grown at 37°C in Luria-Bertani (LB) medium supplemented with 100 μg/ml ampicillin. *Halobacterium* strains were cultured in CM^+ medium containing 4.3 M NaCl and trace metals at 42°C with shaking as previously described (DasSarma et al., 1995). For solid media, 2% (w/v) agar was added. Stock cultures were maintained in glycerol at -80°C. For short-term use, purified cultures were maintained on stock plates at 4°C.

Construction of the *Halobacterium* Carbonic Anhydrase Knockout Strain

To eliminate background carbonic anhydrase production, *icf* A was knocked out via the *ura*3-based gene deletion method for *Halobacterium* sp. NRC-1. Approximately 500 bp regions flanking the carbonic anhydrase gene (*icf* A) were amplified by crossover PCR (using primers shown in **Supplementary Table S5**). The resulting amplified crossover PCR fragment was

cloned into the suicide vector, pBB400 using flanking *Hin*dIII and *Eco*RI sites incorporated in the primers (**Supplementary Table S5**) (Berquist et al., 2006). The resulting plasmid, pBB400Δ*icf*A was transformed into *Halobacterium* sp. NRC-1Δ*ura*3 using the standard PEG-EDTA method (DasSarma et al., 1995). pBB400Δ*icf*A transformants were selected by plating on CM^+ uracil dropout media (HURA), colonies picked and grown in liquid HURA media, and integrant candidates were plated onto 5-FOA-CM^+ media plates. Knockout candidates were identified by DNA extraction and PCR using flanking primers listed in **Supplementary Table S5** (DasSarma et al., 1995; Berquist et al., 2006).

Construction of the Expression Plasmids

Synthetic genes were codon-optimized using the java codon adaptation online tool JCat (Grote et al., 2005) for *Halobacterium* sp. (strain NRC-1/ATCC 700922/JCM 11081). The optimized genes were ordered from GeneArt (Regensburg, Germany) and cloned into pRK42, which harbors an N-terminal His_6-tag, *csp*D2 promoter, origins of replication for *E. coli* and *Halobacterium*, and genes for ampicillin and mevinolin resistance for selection in *E. coli* and *Halobacterium*, respectively.

Expression of the Carbonic Anhydrase Genes in *Halobacterium* sp. NRC-1Δ*ura*3Δ*icf*A and Purification of the Encoded Proteins

Carbonic anhydrase genes containing vectors were transformed into the *Halobacterium* sp. NRC-1 Δ*ura*3Δ*icf*A strain using the PEG/EDTA method (DasSarma et al., 1995; Karan et al., 2013) and transformants were selected by plating on CM^+ agar plates using mevinolin resistance. For protein production, cells were grown to late log phase ($OD_{600\,nm}$ of 0.9–1.0) at 42°C in CM^+ medium supplemented with 20 μg/ml mevinolin. To induce carbonic anhydrase expression, the cultures were further incubated at 15°C for 72 h.

Cells were harvested by centrifugation (6,000 × *g*, 4°C, 10 min) in a 5430R centrifuge (Eppendorf, Germany) and disrupted in binding buffer (20 mM HEPES buffer pH 7.4 containing 2.0 M NaCl, 10% v/v glycerol, protease inhibitor cocktail, cOmplete from Roche, Germany and 30 mM imidazole) using a sonicator (Model Q500, QSONICA, Newtown, CT, United States) with a 1.9 cm probe (Thermo Scientific, Waltham, United States). Cell debris were removed by centrifugation (25,000 × *g*, 4°C, 10 min) in an Avanti J-26 XP centrifuge (Beckman Coulter, High Wycombe, United Kingdom) and the resulting crude extract was filtered through a 0.2 μm Nalgene membrane filter (Thermo Scientific, Rockford, IL, United States). The supernatant was loaded at a flow rate of 1.0 ml/min onto a 5-ml HiTrap Ni^{2+} chelating column (GE Healthcare Life Sciences, Piscataway, NJ, United States) pre-equilibrated with binding buffer. The column was washed with binding buffer, and the protein was eluted by increasing concentration of imidazole (30–300 mM) in binding buffer. The purified active fractions were combined and further purified and concentrated with Amicon® Ultra-4 Centrifugal Filter Units, 10 kDa (Cat no. UFC803024, Merck Millipore, MA, United States). Protein concentration was determined using the NanoDrop 2000c (Thermo Fisher Scientific, Pittsburgh, PA, United States) with an absorption coefficient of 9,970 M^{-1} cm^{-1} (Gasteiger et al., 2005).

Polyacrylamide Gel Electrophoresis

The SDS-PAGE analysis was performed by using the precast Novex® Tris-glycine gels (4–20%, Invitrogen, Carlsbad, CA, United States). The gels were stained with Imperial Protein Stain and Invision His-Tag In-gel stain (Thermo Fisher, Rockford, IL, United States).

Identification of Purified Protein by LC-MS/MS Analysis

The pure samples (10 μg) were digested with trypsin/LysC mix (Promega, Madison, WI, United States) using the FASP protocol (Wiśniewski et al., 2009). Peptides were measured using a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States) and analyzed using MASCOT v2.3 (Matrix Sciences Ltd, United Kingdom).

Activity Measurements

Carbonic anhydrase activity was measured by the SX20 Stopped-Flow Spectrometer (Applied Photophysics, Leatherhead, United Kingdom) using the pH indicator dye phenol red as described previously (Alber et al., 1999). Briefly, chamber A contained 100 μM phenol red in 20 mM MOPS buffer pH 9.8 containing 3 M KCl, with and without protein for catalyzed and uncatalyzed reaction, respectively. Chamber B contained CO_2-saturated water prepared by bubbling CO_2 into distilled deionized water at 25°C. The reaction was monitored spectrophotometrically by measuring the increase in absorbance at 557 nm. All slit widths were set at 0.5 mm. An attached water bath regulated the temperature as indicated. A carbonic anhydrase (0.5 μM) from bovine erythrocytes (Sigma, St. Louis, MO, United States) was used as a positive control. As a negative control, the crude lysate of untransformed *Halobacterium* sp. NRC-1Δ*ura*3Δ*icf*A cells was collected, purified using a Ni-NTA column, and the activity was measured for both the crude lysate and the purified fractions.

The specific activity was calculated via the Wilbur-Anderson unit (WAU) per 1 mg of protein, with one unit of activity being defined as $(T_0 - T)/T$, where T_0 (uncatalyzed reaction) and T (catalyzed reaction) are recorded as the time (sec.) required for the pH to drop from 9.8 to the transition point of the dye in a buffer control and in presence of enzyme or positive/negative control, respectively. Therefore, the reaction time until the activity plateau was reached was measured for the background reaction and all measurements were averaged and analyzed via the Excel solver.

X-ray Crystallography

For crystallization, the protein was dialyzed against 50 mM Tris-HCl buffer pH 8.0, containing 300 mM NaCl and concentrated to 17 mg/ml with Amicon® Ultra-4 Centrifugal Filter Units, 10 kDa (Cat no. UFC803024, Merck Millipore, MA, United States). Protein crystals were grown at 20°C using the hanging drop vapor diffusion method. The protein solution was mixed in a 1:1 ratio with the reservoir buffer, 0.1 M HEPES pH 7.5 containing 0.05 M

cadmium sulfate and 0.8 M sodium acetate. Crystals selected for measurement were flash-frozen in liquid nitrogen after soaking in cryobuffer (70%, v/v reservoir buffer and 30%, v/v glycerol).

Data Collection, Structure Solution and Refinement

Native diffraction datasets were collected at 2.6 Å resolution using synchrotron radiation at the X06SA-beamline, SLS, Villigen, Switzerland (see **Supplementary Table S1**). Recorded reflections were processed with XDS (Kabsch, 1993). CA_D crystallized in the cubic space group F432 with α = 362.6 Å, indicating five γ-CA-subunits in the asymmetric unit and a solvent content of 74%. Phases were obtained by Patterson search algorithms using the coordinates 1V3W as starting model (Jeyakanthan et al., 2008). The primary sequence was placed into the $2F_o$-F_c electron density map using COOT (Emsley and Cowtan, 2004) and refined with REFMAC5 (Murshudov et al., 1997). The model was completed in iterative rounds where temperature factors were anisotropically refined by translation/libration/screw motion-parameters, yielding crystallographic values of R_{cryst} = 0.177 and R_{free} = 0.204 (see **Supplementary Table S1**). Coordinates were confirmed to have adequate stereochemistry in the Ramachandran plot with 98.0% of residues in most favored, 1.8% in additionally allowed, and 0.2% in outlier regions. The crystal structure was deposited at the RCSB Protein Data Bank under accession codes 6SC4.

Structural Analysis

Crystal structure analysis and visualization were performed using programs PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC)[1] and Yasara (Krieger and Vriend, 2014), and SwissPDB Viewer 4.1 (Guex and Peitsch, 1997). Homolog searches were performed using the DALI server (Holm and Rosenstrom, 2010). For comparison of CA_D to homolog structures, all duplicates or structure variants belonging to the same homolog protein were neglected. Interactions between protein residues were calculated using Yasara (Krieger and Vriend, 2014), except for salt bridges which were calculated using VMD (Humphrey et al., 1996) with an oxygen-nitrogen cut-off of 4 Å (between residues Arg/Lys/His and Asp/Glu) and hydrogen bonds being estimated using Chimera (Pettersen et al., 2004).

Electrostatic surface potential calculations were performed using the PDB2PQR server (Dolinsky et al., 2004) along with the PyMOL plugin APBS (Baker et al., 2001). Surface-exposed amino acids were determined using Swiss PDB viewer 4.1 (Guex and Peitsch, 1997) (≥10% surface accessibility). An estimation of the oligomeric assembly was performed using the program PISA (Krissinel and Henrick, 2007).

Mutation Library

The QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, United States) was used according to manufacturer's instructions in combination with the primers listed in the **Supplementary Table S5**. pRK.CAD vector containing the CA_D gene was used as a DNA template.

[1]https://bioinformaticschap.wordpress.com/tools/visulization-tools/pymol/

RESULTS AND DISCUSSION

The remarkable stability of extremozymes and potential applications under harsh operational conditions has gained increased interest (Littlechild, 2017). Nonetheless, to gain a better understanding of the halophilicity in salt adapted proteins and the underlying molecular mechanisms of halophilic enzymes, additional studies of model proteins are required.

Generation of the CA_D Protein From SAG Analysis of a Discovery Deep Sample

The CA_D gene originated from SAG analysis was identified and annotated as a γ-carbonic anhydrase (γ-CA) using the PPM algorithm (Grotzinger et al., 2014). We used the genetically modifiable extreme haloarchaeon *Halobacterium* sp. NRC-1 (Karan et al., 2013) as an expression system. *Haloarchaea* contain an internal salt concentration of 4-5 M and maintain an isoosmotic balance of ion concentrations in the cytosol with the surrounding medium. *Halobacterium* sp. NRC-1 harbors a carbonic anhydrase gene, *icf*A, located in the chromosome, location 638911 ← 639570 (**Supplementary Figure S1**). To eliminate background carbonic anhydrase production, *icf*A was knocked out (Berquist et al., 2006). The resulting *Halobacterium* sp. NRC-1Δ*ura*3Δ*icf*A deletion strain was used for the CA_D expression. The enzyme was purified to homogeneity, with a total yield of 5 mg of protein/liter of culture (**Supplementary Figure S2A**). The identity of CA_D was confirmed by tryptic digest and LC-MS/MS analysis (89% coverage, **Supplementary Figure S2B**).

Crystal Structure of a Halophilic γ-Carbonic Anhydrase From the Discovery Deep Brine Pool

The crystal structure of CA_D was solved successfully to a resolution of 2.6 Å (PDB ID: 6SC4). Data collection and crystallographic quality statistics are shown in Table S1. The monomer contains a seven-turn, left-handed β-helix connected to an α-helix running antiparallel to the β-helix axis (**Figure 1A**). The majority of the structure comprises β-sheet (54.5%), while only a minimal amount consists of α-helix (13.5%), turn (12.9%), and coil structure (19.1%). The overall structure is highly conserved compared to reported γ-Cas (**Figure 1B**). Whereas the protein core is common to published γ-CA structures, differences in the connection of the helix motif is conspicuous. For example, CA_D contains a connecting β-sheet, while Cam consists of an additional α-helix instead. However, the comparison of CA_D with all 10 crystallized γ-CA structures revealed that the majority contained a β-sheet with the thermophilic Cam and carboxysomal CcmM being an exception. Since this characteristic feature did not correspond to all investigated thermophilic homologs, we aimed to analyze whether the ß-sheet has a role in the stability and rigidity of the overall architecture. Furthermore, it is likewise unclear if this plays a prominent role for the enzyme function.

Notably, CA_D is organized into trimers (**Figure 1C**), resembling the reported active conformation for γ-CAs, where the active site includes residues from the adjacent monomers

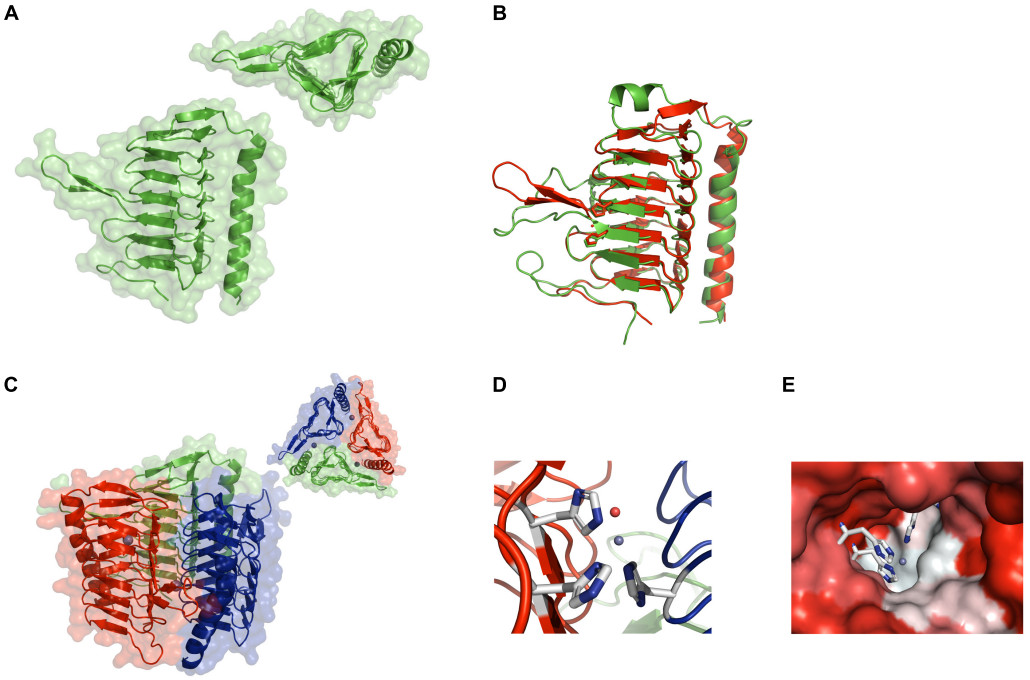


FIGURE 1 | Crystal structure of CA_D. **(A)** Ribbon structure of the CA_D monomer with the background showing the water-accessible surface. The inset shows a top view of the monomer depicting the triangular shape formed by the seven-turn, left-handed β-helix. **(B)** Structural overlay of the CA_D monomer (green) and the γ-CA Cam from *Methanosarcina thermophila* (red) (PDB ID 1qrg). **(C)** Ribbon structure of the CA_D trimer along with the water-accessible surface in the background. The inset contains a top view of the trimer. The three zinc ions are depicted as gray spheres. **(D)** The CA_D active center made up of three histidine residues (white-and-blue sticks) from two adjacent monomers. The zinc ion and a water molecule are depicted as gray and red spheres, respectively. **(E)** The active site cavity is colored according to hydrophobicity (red: hydrophilic, white: hydrophobic). The coordination residues are represented as sticks.

(Ferry, 2010). Closer inspection of the active site reveals a zinc ion, coordinated by three histidine residues (His64 chain A, His89 chain B, and His94 chain A). A well-defined water molecule is coordinated to the zinc ion, which acts as the nucleophile in the reaction mechanism of these enzymes (Supuran, 2016) (**Figure 1D**). Interestingly, one half of the active site pocket exhibits a more hydrophobic character whereas the facing side is more hydrophilic (**Figure 1E**) (Supuran and De Simone, 2015). Taken together, the overall structure of CA_D reveals a strong conservation to reported γ-CAs, supporting the γ-CA class gene annotation. Contrary to the sequence conservation, where the alignment of the CA_D sequence with structural homologs shows a sequence identity of less than 40% for alignments with a query coverage larger than 95%, the structure is, consequently, well conserved (**Supplementary Table S2**).

Halophilic Adaptation of CA_D Compared to Other Non-halophilic γ-CAs

While the overall CA_D structure is conserved to other known γ-CAs, differences to non-halophilic γ-CAs must provide the observed stability under high salt concentrations. Thus, the structural elucidation provides interesting insights into halophilic adaptation by comparing CA_D with meso- or thermophilic γ-CAs (**Supplementary Table S2**).

Comparison of Stabilizing Interactions

Compared to the average values for meso- and thermophilic CAs, the CA_D monomer contained an increased number of salt bridges (12 vs. an average of 10 and 9 for meso- and thermophilic homologs, respectively) (**Supplementary Table S3**), which is seen as a characteristic of haloadaptation (Frolow et al., 1996; Britton et al., 2006; Karan et al., 2012a). Despite the higher average number of salt bridges for the CA_D monomer, several homologs showed a comparable or increased amount of salt bridges. In fact, the discrepancy between the homologs within the group is large. Conversely, thermophilic homologs averaged a higher number of pi–pi interactions and an increased number of hydrogen bonds and hydrophobic interactions within the monomer that displayed an increased rigidity to maintain their structure (calculated numbers are presented in **Supplementary Table S3**). Oligomerization is expected to have a stabilizing effect and the calculation of the interactions within the trimers for CA_D and homologs demonstrates higher amount of interactions compared to monomers alone. Interestingly, while the CA_D monomer does not show a statistically significant increase in interactions compared to single homologs, the trimerization of CA_D stabilizes the enzyme by additional interactions on the monomer interfaces to a higher extent than for homologs. The CA_D trimer displayed an increased number of hydrogen bonds and salt bridges compared to all of the individual meso- and thermophilic homologs being additionally added to the enzyme

upon trimerization (hydrogen bonds: 69 vs. homolog average of 42; salt bridges: 10 vs. a homolog average of 5 or 5.5). As the – CA_D trimer constitutes the active form, this increased stabilization is critical to assemble the active enzyme under high salt conditions.

Comparison of the Surface-Charge

Halophilic proteins are typically characterized by a highly negative charged electrostatic surface (DasSarma and DasSarma, 2015). Interestingly, while the CA_D monomer shows both positive and negative charges on its surface, the overall change is slightly more negative (pI ∼6.8). The highest negative charge is found on the surface of the flexible β10–β11 loop extending from the β-helix (**Figure 2A**). The presence and sequence of this loop differs between homolog γ-CAs (Park et al., 2012). For CA_D it shows a high acidic surface charge. A highly acidic surface is an important and common trait of halophilic proteins, enabling protein hydration under high salt conditions (Frolow et al., 1996; Britton et al., 2006; Grotzinger et al., 2018).

The CA_D trimer contains an overall negative surface potential, while a positive charge is located only in two concentrated locations. The positive charge is possibly involved in the fast release of the bicarbonate product. The overall negative charge indicates the burying of positive patches found on the monomer surface (**Figure 2B**). Compared to mesophilic homologs, CA_D showed a slightly higher negative surface charge except for *E. coli* and *C. difficile* γ-CAs (PDB ID: 3tio, 4mfg) (**Figure 2C** and **Supplementary Figure S3**). Still, this negative surface-charge is lower than often described for halophilic proteins and possibly explains the flexibility that enables stability even at comparatively low salt concentrations.

Comparison of the Surface-Exposed Amino Acids

The most pronounced difference found for CA_D compared to homologs is the surface amino acid composition. The CA_D trimer protein surface contains a decreased number of hydrophobic and polar amino acid residues (**Supplementary Figure S4**) whilst showing an increased number of charged amino acid residues on the surface as compared to homologs. 56% of the amino acids on the CA_D protein surface are charged, compared to 38 and 39% in meso- and thermophilic homologs, respectively (**Figure 3A**). This decrease in hydrophobic amino acids leads to a decrease of the hydrophobic surface patch that facilitates the aggregation of the protein in CA_D (hydrophobic amino acids constitute 29% of the surface amino acids compared to an average of 36 and 39% for meso- and thermophilic homologs, respectively). An increased amount of charged amino acids is observed that form a stable hydration shell. This is essential for stability under high salt conditions. A closer look at positively and negatively charged amino acids on the trimer surface of CA_D reveals the dominance of glutamate (23.6 vs. 10.5 and 14.4% on the surface of mesophilic and thermophilic homologs, respectively) (**Figure 3B**). Interestingly, CA_D also contains a comparatively high amount of lysines.

The increase of surface glutamic acid residues described for CA_D is a characteristic hallmark of halophilic enzymes (DasSarma et al., 2006), as these residues have a high water-binding capacity via interaction with Na^+ or K^+ ions, and thereby attract the bound hydrating water (Trevino et al., 2007). This explains how the strong increase of glutamic acid on the surface of CA_D has such a pronounced effect on maintaining stability and activity at high salt concentrations. However, contradictory to reported halo-adaptation strategies, CA_D also shows increased surface lysine residues, compared to homologs. This explains why the negative surface charge is not as pronounced as for other reported halophilic proteins (Premkumar et al., 2005; Karan et al., 2012a; DasSarma et al., 2013). Lysine tends to disrupt the formation of an ordered hydration shell under elevated salt concentrations (Britton et al., 2006; Esclapez et al., 2007). At low ionic strength, the higher amount of glutamic acid on the surface causes electrostatic repulsion and, therefore, destabilizes the protein (Kohn et al., 1995; Frolow et al., 1996; Britton et al., 2006). Moreover, the long hydrophobic part of the lysine residues potentially plays a role in attracting hydrophobic CO_2 substrate for the catalytic reaction. Noteworthy, the negatively charged glutamate residues are located around the trimer surface, except for the top and bottom view of the multimer where positive patches stem from a network of arginine residues which are possibly initiating the rapid release of the formed product (Smith et al., 2002; Tripp et al., 2002).

CA_D Variant Library Design

Since the discovery of the first γ-CA in 1994 (Alber and Ferry, 1994), several γ-CAs have been characterized: some with a high reported activity and some with no detectable activity, raising the question whether essential residues are missing or an alternative function is appropriate (Macauley et al., 2009). We bioengineered the CA_D active center based on the CA_D crystal structure. To investigate the role of selected residues in the active site of CA_D, several variants were expressed, purified, and assayed for enzymatic activity. The selection of residues for mutagenesis was performed based on structural comparisons to γ-CA homologs as well as literature reports based on presumptions of conserved residues of γ-Cas (Smith et al., 1999; Iverson et al., 2000; Jeyakanthan et al., 2008; Ferry, 2010; Pena et al., 2010; Park et al., 2012; Ragunathan et al., 2013; Frost and McKenna, 2014). Thereby, the main comparison was focused on the Cam structure (Kisker et al., 1996). A structural comparison of the active center of CA_D (**Figure 4A**) with a simplified view of the Cam homolog (**Figure 4B**) revealed that CA_D residue I46 corresponds to Cam residue E62. This part of the enzyme plays an important role in product release, relaying protons during hydroxide formation from the zinc-bound water as well as forming hydrogen bonds with the bicarbonate. CA_D residues K58 and H166 are substituted by Q75 and N202 in Cam. These residues presumably orchestrate the orientation of the carbon dioxide for the nucleophilic attack in Cam. Moreover, Cam N202 together with E62 is thought to form hydrogen bonds with the product bicarbonate (Ferry, 2010; Pena et al., 2010).

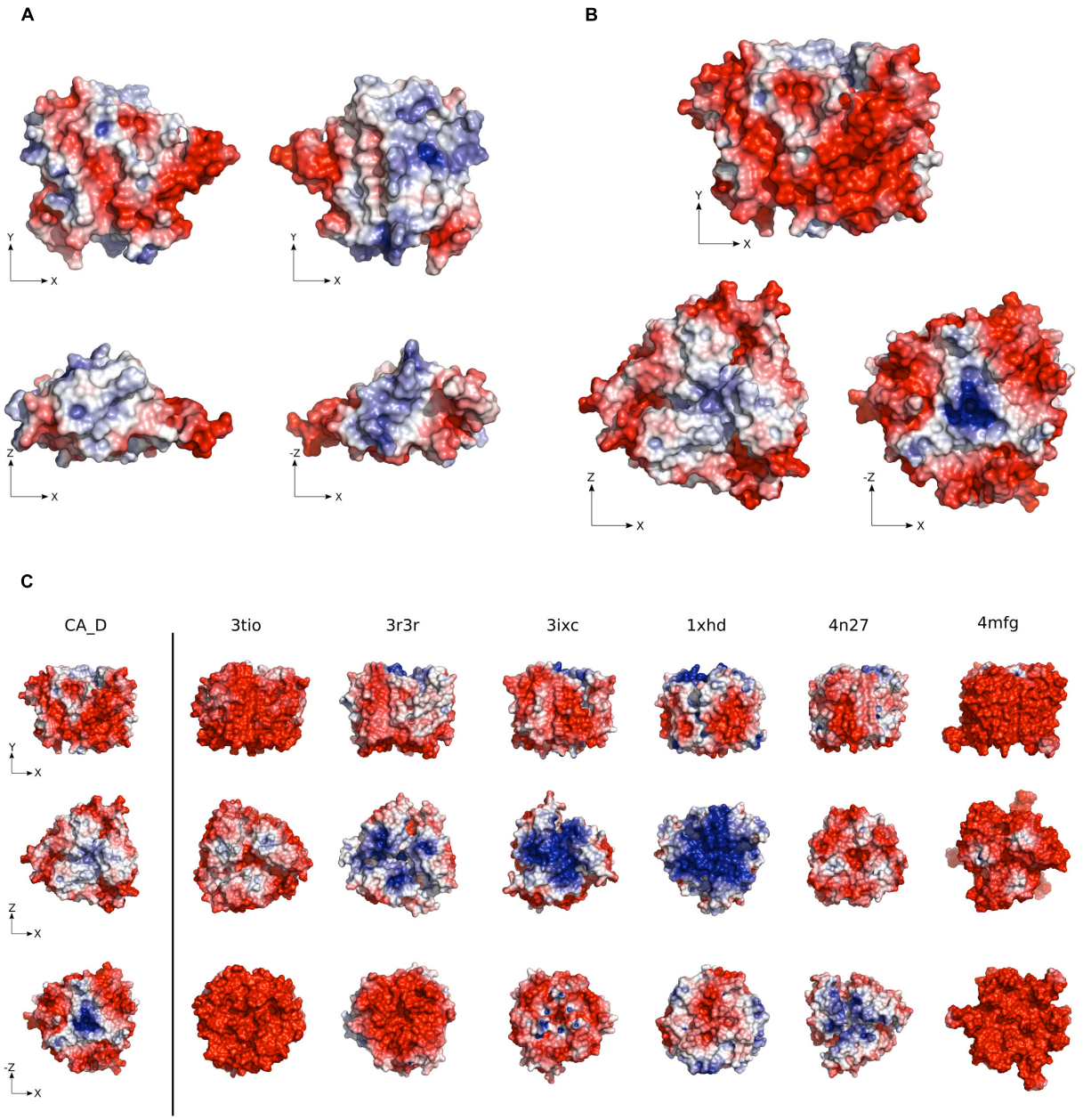


FIGURE 2 | Electrostatic surface potential of CA_D. **(A)** CA_D Monomer and **(B)** CA_D Trimer electrostatic surface potential color-coded from red (negative potential) to blue (positive potential). **(C)** The surface potential of CA_D compared to mesophilic γ-CA homologs (*Escherichia coli,* 3tio; *Salmonella enterica,* 3r3r; *Anaplasma phagocytophilum,* 3ixc; *Bacillus cereus,* 1xhd; *Brucella abortus,* 4n27; *Clostridium difficile,* 4mfg). Unit: –5 to +5 k_bT/e (k_b as the Boltzmann constant, T as the temperature in Kelvin and e as the charge of an electron).

The strict conservation of the E84 position is a matter of debate, as the CamH subclass has lost this residue (Ferry, 2010) and a mutagenesis study exhibits activity for the D84 and H84 substituted Cam variants (Tripp and Ferry, 2000). The presence of D residues instead of E is explained by the different abundance of aspartic and glutamic acid in thermophilic proteins, compared to mesophilic proteins (Lee et al., 2004). Thus, the selected point mutations were I46E, K58Q, H166N, I46E-K58Q, K58Q-H166N, I46E-H166N as well as a triple CA_D* (I46E-K58Q-H166N) (**Figure 4C**) and a quadruple CA_D*-D67E (I46E-K58Q-H166N-D67E) variant (**Table 1** and **Supplementary Figures S5**, **S6**).

Carbonic Anhydrase Activity Measurement of CA_D and Variants

To evaluate the potential impact of the mutations, we examined the CA_D variants for their activities expressed

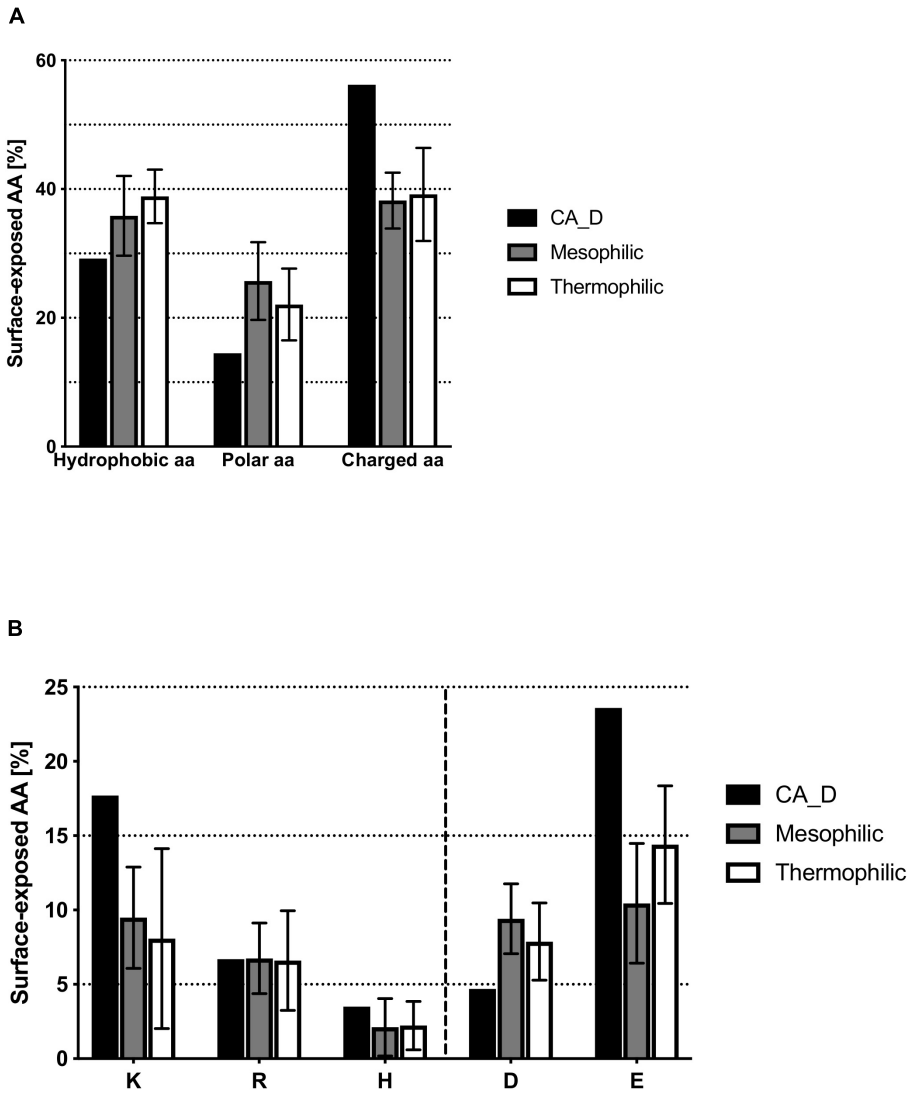


FIGURE 3 | Comparison of surface-exposed amino acids between CA_D and meso- and thermophilic homologs. **(A)** hydrophobic, polar or charged amino acids residues, **(B)** depicts the amount per individual charged amino acid.

in Wilbur-Anderson Unit per 1 mg of protein (WAU/mg) (Wilbur and Anderson, 1948; Hou et al., 2019). The colorimetric carbonic anhydrase activity assay by Wilbur-Anderson measures the time required for a saturated CO_2 solution to lower the pH of a specific buffer. Respective controls were: (i) a carbonic anhydrase from bovine erythrocytes as a positive control; and (ii) the crude lysate of untransformed *Halobacterium* sp. NRC-1Δ*ura3*Δ*icf*A cells as negative controls (**Supplementary Figure S7**). Purified lysate and the crude lysate did not show any enzyme activity. Therefore, these measurements confirmed that the observed activity resulted from the purified CA_D variants. As CA_D is from an uncultured archaeon from the Discovery Deep brine pool, the enzyme activity was measured at 40°C in a solution containing 3 M KCl.

The activity assays displayed a distinct profile for CA_D with a rather low activity of ∼33 WAU/mg. In contrast, the engineered variants showed a decreased activity for I46E, K58Q, and H166N, and undetectable activity for the double variants, I46E-K58Q, K58Q-H166N, and I46E-H166N (**Figure 5** and **Supplementary Table S4**). However, the triple variant CA_D* (I46E-K58Q-H166N) displayed a 17-fold increased activity of 566 WAU/mg (**Figure 5** and **Supplementary Table S4**) compared to CA_D.

These results are in close alignment with previous findings of related γ-carbonic anhydrases (γ-CAs). In the most prominent γ-CA Cam, residue E84 has been proposed as part of a proton shuttling network along with E62, E88, and E89 on the acidic loop (Tripp and Ferry, 2000). Conversely, when fully mimicking the expected proton shuttling network, activity of the quadruple

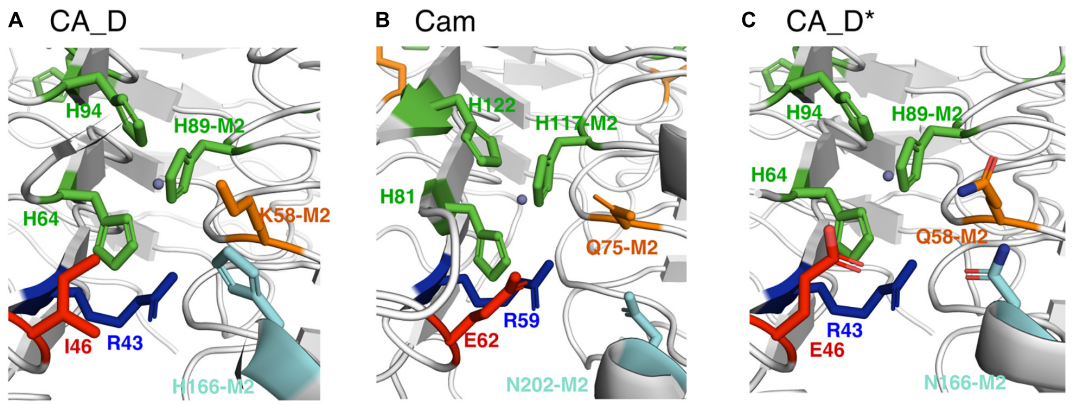


FIGURE 4 | Comparison of selected active site residues. **(A)** CA_D (crystal structure), **(B)** Cam (crystal structure) from *M. thermophila* (PDB ID: 1qrg), and **(C)** CA_D* including selected mutations such as I46E, K58Q, H166N (M2: indicates that residue belongs to the adjacent monomer). The central zinc ion is depicted as a gray sphere.

TABLE 1 | Rationalization of CA_D variants.

Constructs	Residues							
1QRG (Cam)	R59	E62	Q75	E84	N202	H81	H117	H122
CA_D	R43	**I46**	**K58**	D67	**H166**	H64	H89	H94
I46E	R43	**E46**	K58	D67	H166	H64	H89	H94
K58Q	R43	I46	**Q58**	D67	H166	H64	H89	H94
H166N	R43	I46	K58	D67	**N166**	H64	H89	H94
I46E-K58Q	R43	**E46**	**Q58**	D67	H166	H64	H89	H94
K58Q-H166N	R43	I46	**Q58**	D67	**N166**	H64	H89	H94
I46E-H166N	R43	**E46**	K58	D67	**N166**	H64	H89	H94
CA_D* (I46E-K58Q-H166N)	R43	**E46**	**Q58**	D67	**N166**	H64	H89	H94
CA_D*-D67E	R43	**E46**	**Q58**	**E67**	**N166**	H64	H89	H94

The residues in bold were modified based on the alignment with Cam shown in the first row.

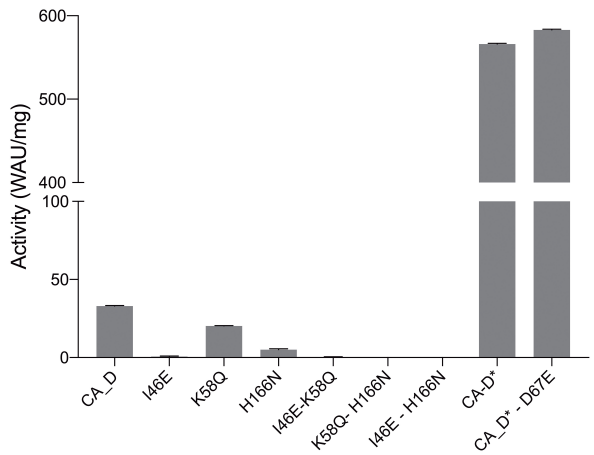


FIGURE 5 | Carbonic anhydrase activity assessment. CA_D WT and variants measured at 3 M KCl and 40°C.

variant CA_D*-D67E (**Figure 5**) is slightly higher as for the triple variant CA_D*, which is in agreement with the observations made in γ-CA Cam (Tripp and Ferry, 2000; Ferry, 2010).

Our findings support the proposed mechanism for Cam, in which the probed amino acids play a decisive role. Some conclusions, as to why CA_D is lacking activity could be drawn by comparing the altered residues in CA_D* to γ-CA Cam (Ferry, 2010). With the CA_D mutation I46E, a hydrophobic residue was replaced with a negatively charged one, the corresponding residue in the Cam homolog presumably is essential in water activation to enable the reaction (Ferry, 2010). K58Q and H166N are further substitutions based on Cam active residues and facilitate the orientation of the carbon dioxide, while the H166N substitution facilitates hydrogen bonding and release of the bicarbonate (Ferry, 2010). We, therefore, assume that CA_D is regulated by a similar proton shuttling network such as Cam (Ferry, 2010), but due to its halophilic nature CA_D behaves slightly differently.

CONCLUSION

High cytoplasmic salt concentrations critically affect the folding and activity of proteins and other macromolecules as they may induce protein aggregation due to enhanced hydrophobic interactions, increased hydration of ions, decreased

availability of unbound water molecules, and prevention of intra- and intermolecular electrostatic interactions (Karan et al., 2012a,b). Halophilic proteins are adapted to maintain their native conformation under high salt concentrations. They are functionally active in the presence of high salt concentrations, following halo-adaptation strategies such as high acidic amino acid content on the surface, low hydrophobicity at the core of the protein, and an increased number of salt bridges (Madern et al., 2000; Fukuchi et al., 2003; Bolhuis et al., 2008; Tadeo et al., 2009; DasSarma et al., 2013).

We used single amplified genomes to resuscitate a γ-carbonic anhydrase (γ-CA) from an uncharacterized haloarchaeon collected from a brine pool at the bottom of the Red Sea. The detailed structural analysis and comparison with previously solved structures of mesophilic γ-CAs revealed the molecular features of its extremophilic nature, caused by the unique habitat. The most prominent features of extremophilicity are the increased charged residues on the protein surface and an increased number of hydrogen bonds as well as salt bridges. Investigation of CA_D, therefore, demonstrates potential for further development and implementation of SAG analysis to generate extremozymes from previously inaccessible environments. This approach has to date only been scarcely used to characterize specific proteins of interest (Grotzinger et al., 2018). Further, mutation analysis provided an interesting insight into active site residue conservation for γ-CAs and enabled us to increase the CA_D function by 17-fold. Moreover, the applied combination of mechanistic insights from the thermophilic Cam into the scaffold of the halophilic CA_D, resulting in the CA_D* variant which demonstrated high activity and stability, underlines the potential of protein evolution for extremophilic proteins for industrial applications and the design of novel catalysts for industry.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the PDB ID: 6SC4.

AUTHOR CONTRIBUTIONS

MR and JE conceived and supervised the study. MV and RK designed and performed the experiments with the help of DR and AV. SG, PD, and SD provided plasmid and support. M-TV and MG solved the crystal structure. MV and RK wrote the manuscript. All the authors contributed to and commented on this manuscript.

FUNDING

The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST), Kingdom of Saudi Arabia. The staff of the Beamline X06SA at the Paul Scherrer Institute, SLS, Villigen, Switzerland, is acknowledged for assistance during data collection. Work in the DasSarma laboratory was supported by grant 80NSSC19K0463.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00742/full#supplementary-material

REFERENCES

Akal, A. L., Karan, R., Hohl, A., Alam, I., Vogler, M., Grötzinger, S. W., et al. (2019). A polyextremophilic alcohol dehydrogenase from the Atlantis II Deep Red Sea brine pool. *FEBS Open Biol.* 9, 194–205. doi: 10.1002/2211-5463.12557

Alam, I., Antunes, A., Kamau, A. A., Ba Alawi, W., Kalkatawi, M., Stingl, U., et al. (2013). INDIGO – INtegrated data warehouse of microbial genomes with examples from the red sea extremophiles. *PLoS One* 8:e82210. doi: 10.1371/journal.pone.0082210

Alber, B. E., Colangelo, C. M., Dong, J., Stalhandske, C. M., Baird, T. T., Tu, C., et al. (1999). Kinetic and spectroscopic characterization of the gamma-carbonic anhydrase from the methanoarchaeon *Methanosarcina thermophila*. *Biochemistry* 38, 13119–13128. doi: 10.1021/bi9828876

Alber, B. E., and Ferry, J. G. (1994). A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6909–6913. doi: 10.1073/pnas.91.15.6909

Amoozegar, M. A., Safarpour, A., Akbari Noghabi, K., Bakhtiary, T., and Ventosa, A. (2019). Halophiles and their vast potential in biofuel production. *Front. Microbiol.* 10:1895. doi: 10.3389/fmicb.2019.01895

Antunes, A., Ngugi, D. K., and Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* 3, 416–433. doi: 10.1111/j.1758-2229.2011.00264.x

Antunes, A., Simões, M. F., Grötzinger, S. W., Eppinger, J., Bragança, J., and Bajic, V. B. (2017). "Bioprospecting archaea: focus on extreme halophiles," in *Bioprospecting*, eds R. Paterson and N. Lima (Cham: Springer), 81–112. doi: 10.1007/978-3-319-47935-4_5

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10037–10041. doi: 10.1073/pnas.181342398

Behzad, H., Ibarra, M. A., Mineta, K., and Gojobori, T. (2016). Metagenomic studies of the Red Sea. *Gene* 576(2 Pt 1), 717–723. doi: 10.1016/j.gene.2015.10.034

Berquist, B. R., Müller, J. A., and DasSarma, S. (2006). 27 genetic systems for *Haloferax volcanii*. *Methods Microbiol.* 35, 649–680. doi: 10.1007/s00792-015-0794-6

Bolhuis, A., Kwan, D., and Thomas, J. (2008). "Halophilic adaptations of proteins," in *Protein Adaptation in Extremophiles*, eds K. S. Siddiqui, T. Thomas (New York: Nova Science Publishers Inc).

Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., et al. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4846–4851. doi: 10.1073/pnas.0508854103

Cavicchioli, R., Amils, R., Wagner, D., and McGenity, T. (2011). Life and applications of extremophiles. *Environ. Microbiol.* 13, 1903–1907. doi: 10.1111/j.1462-2920.2011.02512.x

DasSarma, S., Berquist, B. R., Coker, J. A., DasSarma, P., and Muller, J. A. (2006). Post-genomics of the model haloarchaeon *Halobacterium* sp. *NRC*-1. *Saline Syst.* 2:3. doi: 10.1186/1746-1448-2-3

DasSarma, S., Capes, M. D., Karan, R., and DasSarma, P. (2013). Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from antarctica. *PLoS One* 8:e58587. doi: 10.1371/journal.pone.0058587

DasSarma, S., and DasSarma, P. (2015). Halophiles and their enzymes: negativity put to good use. *Curr. Opin. Microbiol.* 25, 120–126. doi: 10.1016/j.mib.2015.05.009

DasSarma, S., Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., and Fleischmann, E. M. (1995). *Archaea: A Laboratory Manual – Halophiles*. Cham: Springer.

Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* 11, 1095–1099. doi: 10.1101/gr.180501

Del Prete, S., Vullo, D., De Luca, V., Carginale, V., Osman, S. M., AlOthman, Z., et al. (2016). Cloning, expression, purification and sulfonamide inhibition profile of the complete domain of the η -carbonic anhydrase from *Plasmodium falciparum*. *Bioorg. Med. Chem. Lett.* 26, 4184–4190. doi: 10.1016/j.bmcl.2016.07.060

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* (Suppl._2), W665–W667.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2126–2132. doi: 10.1107/S0907444904019158

Esclapez, J., Pire, C., Bautista, V., Martínez-Espinosa, R., Ferrer, J., and Bonete, M. (2007). Analysis of acidic surface of Haloferax mediterranei glucose dehydrogenase by site-directed mutagenesis. *FEBS Lett.* 581, 837–842. doi: 10.1016/j.febslet.2007.01.054

Ferry, J. G. (2010). The gamma class of carbonic anhydrases. *Biochim. Biophys. Acta* 1804, 374–381. doi: 10.1016/j.bbapap.2009.08.026

Ferry, J. G. (2013). Carbonic anhydrases of anaerobic microbes. *Bioorg. Med. Chem.* 21, 1392–1395. doi: 10.1016/j.bmc.2012.12.005

Frolow, F., Harell, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.* 3:452. doi: 10.1038/nsb0596-452

Frost, S. C., and McKenna, R. (2014). *Carbonic Anhydrase Mechanism, Regulation, Links to Disease, and Industrial Applications*. Cham: Springer.

Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327, 347–357. doi: 10.1016/s0022-2836(03)00150-5

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-890-0:571

Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., et al. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–W531. doi: 10.1093/nar/gki376

Grotzinger, S. W., Alam, I., Ba Alawi, W., Bajic, V. B., Stingl, U., and Eppinger, J. (2014). Mining a database of single amplified genomes from Red Sea brine pool extremophiles-improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA). *Front. Microbiol.* 5:134. doi: 10.3389/fmicb.2014.00134

Grotzinger, S. W., Karan, R., Strillinger, E., Bader, S., Frank, A., Al Rowaihi, I. S., et al. (2018). Identification and experimental characterization of an extremophilic brine pool alcohol dehydrogenase from single amplified genomes. *ACS Chem. Biol.* 13, 161–170. doi: 10.1021/acschembio.7b00792

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. doi: 10.1002/elps.1150181505

Gurvich, E. G. (2006). *Metalliferous Sediments of the World Ocean*. Berlin: Springer.

Herrou, J., and Crosson, S. (2013). Molecular structure of the Brucella abortus metalloprotein RicA, a Rab2-binding virulence effector. *Biochemistry* 52, 9020–9028. doi: 10.1021/bi401373r

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–W549. doi: 10.1093/nar/gkq366

Hou, J., Li, X., Kaczmarek, M. B., Chen, P., Li, K., Jin, P., et al. (2019). Accelerated CO_2 hydration with thermostable sulfurihydrogenibium azorense carbonic anhydrase-chitin binding domain fusion protein immobilised on chitin support. *Int. J. Mol. Sci.* 20:1494. doi: 10.3390/ijms20061494

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Hunt, J. M., Hays, E. E., Degens, E. T., and Ross, D. A. (1967). Red sea: detailed survey of hot-brine areas. *Science* 156, 514–516. doi: 10.1126/science.156.3774.514

Iverson, T. M., Alber, B. E., Kisker, C., Ferry, J. G., and Rees, D. C. (2000). A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39, 9222–9231. doi: 10.1021/bi000204s

Jeyakanthan, J., Rangarajan, S., Mridula, P., Kanaujia, S. P., Shiro, Y., Kuramitsu, S., et al. (2008). Observation of a calcium-binding site in the gamma-class carbonic anhydrase from *Pyrococcus horikoshii*. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 10), 1012–1019. doi: 10.1107/S0907444908024323

Jorquera, M. A., Graether, S. P., and Maruyama, F. (2019). Bioprospecting and biotechnology of extremophiles. *Front. Bioeng. Biotechnol.* 7:204.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26, 795–800. doi: 10.1107/s0021889893005588

Karan, R., Capes, M. D., DasSarma, P., and DasSarma, S. (2013). Cloning, overexpression, purification, and characterization of a polyextremophilic beta-galactosidase from the Antarctic haloarchaeon *Halorubrum lacusprofundi*. *BMC Biotechnol.* 13:3. doi: 10.1186/1472-6750-13-3

Karan, R., Capes, M. D., and DasSarma, S. (2012a). Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* 8:4. doi: 10.1186/2046-9063-8-4

Karan, R., Kumar, S., Sinha, R., and Khare, S. K. (2012b). *Halophilic Microorganisms as Sources of Novel Enzymes*. Cham: Springer, 555–579.

Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996). A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J.* 15, 2323–2330. doi: 10.1002/j.1460-2075.1996.tb00588.x

Kohn, W. D., Kay, C. M., and Hodges, R. S. (1995). Protein destabilization by electrostatic repulsions in the two-stranded α-helical coiled-coil/leucine zipper. *Protein Sci.* 4, 237–250. doi: 10.1002/pro.5560040210

Krieger, E., and Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* 30, 2981–2982. doi: 10.1093/bioinformatics/btu426

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022

Kvist, T., Ahring, B. K., Lasken, R. S., and Westermann, P. (2007). Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* 74, 926–935. doi: 10.1007/s00253-006-0725-727

Lee, D. Y., Kim, K.-A., Yu, Y. G., and Kim, K.-S. (2004). Substitution of aspartic acid with glutamic acid increases the unfolding transition temperature of a protein. *Biochem. Biophys. Res. Commun.* 320, 900–906. doi: 10.1016/j.bbrc.2004.06.031

Liszka, M. J., Clark, M. E., Schneider, E., and Clark, D. S. (2012). Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* 3, 77–102. doi: 10.1146/annurev-chembioeng-061010-114239

Littlechild, J. A. (2015). Archaeal enzymes and applications in industrial biocatalysts. *Archaea* 2015:147671. doi: 10.1155/2015/147671

Littlechild, J. A. (2017). Improving the 'tool box' for robust industrial enzymes. *J. Ind. Microbiol. Biotechnol.* 44, 711–720. doi: 10.1007/s10295-017-1920-1925

Macauley, S. R., Zimmerman, S. A., Apolinario, E. E., Evilia, C., Hou, Y. M., Ferry, J. G., et al. (2009). The archetype gamma-class carbonic anhydrase (Cam) contains iron when synthesized in vivo. *Biochemistry* 48, 817–819. doi: 10.1021/bi802246s

Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142

Madigan, M. T., and Marrs, B. L. (1997). Extremophiles. *Sci. Am.* 276, 82–87. doi: 10.1038/scientificamerican0497-82

Meldrum, N. U., and Roughton, F. J. (1933). Carbonic anhydrase. Its preparation and properties. *J. Physiol.* 80, 113–142. doi: 10.1113/jphysiol.1933.sp003077

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53(Pt 3), 240–255. doi: 10.1107/S0907444996012255

Mwirichia, R., Alam, I., Rashid, M., Vinu, M., Ba-Alawi, W., Anthony Kamau, A., et al. (2016). Metabolic traits of an uncultured archaeal lineage–MSBL1–from brine pools of the Red Sea. *Sci. Rep.* 6:19181. doi: 10.1038/srep19181

Park, H. M., Park, J. H., Choi, J. W., Lee, J., Kim, B. Y., Jung, C. H., et al. (2012). Structures of the gamma-class carbonic anhydrase homologue YrdA suggest a possible allosteric switch. *Acta Crystallogr. D Biol. Crystallogr.* 68(Pt 8), 920–926. doi: 10.1107/S0907444912017210

Pena, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2455–2460. doi: 10.1073/pnas.0910866107

Persidis, A. (1998). Extremophiles. *Nat. Biotechnol.* 16, 593–594. doi: 10.1038/nbt0698-593

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Premkumar, L., Greenblatt, H. M., Bageshwar, U. K., Savchenko, T., Gokhman, I., Sussman, J. L., et al. (2005). Three-dimensional structure of a halotolerant algal carbonic anhydrase predicts halotolerance of a mammalian homolog. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7493–7498. doi: 10.1073/pnas.0502829102

Ragunathan, P., Raghunath, G., Kuramitsu, S., Yokoyama, S., Kumarevel, T., and Ponnuraj, K. (2013). Crystallization, characterization and preliminary X-ray crystallographic analysis of GK2848, a putative carbonic anhydrase of *Geobacillus kaustophilus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 69(Pt 2), 162–164. doi: 10.1107/S1744309112051913

Rashid, M., and Stingl, U. (2015). Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnol. Adv.* 33, 1755–1773. doi: 10.1016/j.biotechadv.2015.09.005

Smith, K. S., Ingram-Smith, C., and Ferry, J. G. (2002). Roles of the conserved aspartate and arginine in the catalytic mechanism of an archaeal β-class carbonic anhydrase. *J. Bacteriol.* 184, 4240–4245. doi: 10.1128/jb.184.15.4240-4245.2002

Smith, K. S., Jakubzick, C., Whittam, T. S., and Ferry, J. G. (1999). Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 15184–15189. doi: 10.1073/pnas.96.26.15184

Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160. doi: 10.1128/JB.00345-312

Supuran, C. T. (2008). Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* 7, 168–181. doi: 10.1038/nrd2467

Supuran, C. T. (2016). Structure and function of carbonic anhydrases. *Biochem. J.* 473, 2023–2032. doi: 10.1042/BCJ20160115

Supuran, C. T., and Capasso, C. (2017). Carbonic anhydrase from porphyromonas gingivalis as a drug target. *Pathogens* 6:30. doi: 10.3390/pathogens6030030

Supuran, C. T., and De Simone, G. (eds). (2015). "Carbonicanhydrases: an overview," in *Carbonic Anhydrases as Biocatalysts* (Cham: Springer), 3–13.

Tadeo, X., Lopez-Mendez, B., Trigueros, T., Lain, A., Castano, D., and Millet, O. (2009). Structural basis for the aminoacid composition of proteins from halophilic archea. *PLoS Biol.* 7:e1000257. doi: 10.1371/journal.pbio.1000257

Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* 366, 449–460. doi: 10.1016/j.jmb.2006.10.026

Tripp, B. C., and Ferry, J. G. (2000). A structure-function study of a proton transport pathway in the gamma-class carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39, 9232–9240. doi: 10.1021/bi0001877

Tripp, B. C., Tu, C., and Ferry, J. G. (2002). Role of arginine 59 in the γ-class carbonic anhydrases. *Biochemistry* 41, 669–678. doi: 10.1021/bi010768b

Wilbur, K. M., and Anderson, N. G. (1948). Electrometric and colorimetric determination of carbonic anhydrase. *J. Biol. Chem.* 176, 147–154.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 May 2020
doi: 10.3389/fmicb.2020.00881

Discovery and Characterization of a New Cold-Active Protease From an Extremophilic Bacterium via Comparative Genome Analysis and *in vitro* Expression

Amedea Perfumo[1,2], Georg Johannes Freiherr von Sass[3], Eva-Lena Nordmann[1,4], Nediljko Budisa[3,5] and Dirk Wagner[1,6]*

[1] GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Section Geomicrobiology, Potsdam, Germany, [2] Polar Terrestrial Environmental System Division, Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Potsdam, Germany, [3] Institute of Chemistry, Technische Universität Berlin, Berlin, Germany, [4] Institute of Chemistry and Biology of the Marine Environment, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Germany, [5] Institute of Chemistry, University of Manitoba, Winnipeg, MB, Canada, [6] Institute of Geosciences, University of Potsdam, Potsdam, Germany

OPEN ACCESS

Edited by:
Massimiliano Fenice, University of Tuscia, Italy

Reviewed by:
Viktoria Shcherbakova, Institute of Biochemistry and Physiology of Microorganisms (RAS), Russia
Anna Zanfardino, University of Naples Federico II, Italy

***Correspondence:**
Amedea Perfumo
amedea.perfumo@awi.de

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *08 December 2019*
Accepted: *16 April 2020*
Published: *13 May 2020*

Citation:
Perfumo A, Freiherr von Sass GJ, Nordmann E-L, Budisa N and Wagner D (2020) Discovery and Characterization of a New Cold-Active Protease From an Extremophilic Bacterium via Comparative Genome Analysis and in vitro Expression. Front. Microbiol. 11:881. doi: 10.3389/fmicb.2020.00881

Following a screening of Antarctic glacier forefield-bacteria for novel cold-active enzymes, a psychrophilic strain *Psychrobacter* sp. 94-6PB was selected for further characterization of enzymatic activities. The strain produced lipases and proteases in the temperature range of 4–18°C. The coding sequence of an extracellular serine-protease was then identified via comparative analysis across *Psychrobacter* sp. genomes, PCR-amplified in our strain 94-6PB and expressed in the heterologous host *E. coli*. The purified enzyme (80 kDa) resulted to be a cold-active alkaline protease, performing best at temperatures of 20–30°C and pH 7-9. It was stable in presence of common inhibitors [β-mercaptoethanol (β-ME), dithiothreitol (DTT), urea, phenylmethylsulfonyl fluoride (PMSF) and ethylenediaminetetraacetic acid (EDTA)] and compatible with detergents and surfactants (Tween 20, Tween 80, hydrogen peroxide and Triton X-100). Because of these properties, the P94-6PB protease may be suitable for use in a new generation of laundry products for cold washing. Furthermore, we assessed the microdiversity of this enzyme in *Psychrobacter* organisms from different cold habitats and found several gene clusters that correlated with specific ecological niches. We then discussed the role of habitat specialization in shaping the biodiversity of proteins and enzymes and anticipate far-reaching implications for the search of novel variants of biotechnological products.

Keywords: bioprospecting, extremophilic bacteria, cold-active enzymes, genome mining, heterologous protein expression, microdiversity

INTRODUCTION

Cold habitats are an extraordinary reservoir of biotechnological molecules such as enzymes (Cavicchioli et al., 2011), antimicrobials (Borchert et al., 2017), and biosurfactants (Perfumo et al., 2018) that can function under extreme conditions. In particular, cold-active enzymes, having catalytic activity below 25°C, can be used in energy saving bioprocesses. Catalysis at moderate

temperature is an advantage also for the production of heat-sensitive bioproducts (e.g., drugs and therapeutics), the risk of contamination with mesophilic organisms is minimized as well as unwanted secondary products, and finally the bioprocess can be easily inactivated by raising the temperature (Cavicchioli et al., 2011; Santiago et al., 2016). Cold-adapted enzymes are also highly efficient because they can catalyze their reactions with chemical rates comparable to, and often exceeding, those of their warm-active counterparts (Struvay and Feller, 2012; Isaksen et al., 2016).

In this context, cold-active enzymes of microbial origin have tremendous biotechnological potential in a large range of markets such as detergents, food and beverages, textiles and can be used for specialty applications in molecular biology/research, pharmaceuticals, and diagnostics. Commercially available cold-active enzymes comprise mostly proteases and lipases, and to lesser extent amylases, cellulases, pectinases, mannanases, and others (Sarmiento et al., 2015). This market is expected to grow further in the coming years pushed forward by the rapid development of new technologies that both lead to the discovery of novel enzymes (e.g., genomics and metagenomics) and enable fine modifications of their chemical composition, structure and functional properties (e.g., genetic engineering and strain optimization or synthetic biology approaches) (Hoesl et al., 2011; Agostini et al., 2017).

The latest trend in the present omics-era is to carry out research on enzyme bioprospecting through environmental metagenomics (Berini et al., 2017) and functional genomics (Gerlt, 2017). Metagenome mining, especially if applied to microbial communities in extreme environments, holds great promise to discover entirely new classes of enzymes by-passing the technical challenges of culturing extremophilic microorganisms. However, the effective identification of promising enzymes from environmental metagenomes still suffers from low success rates, as highlighted by Ferrer et al. (2016). Therefore, a valid alternative is the functional genome mining approach. With the advances of high-throughput sequencing techniques, the number of bacterial genomes made publicly available has increased enormously, being at present in the order of tens of thousands (Ziemert et al., 2016). They obviously represent an extraordinary treasure trove of microbial biodiversity and functionality, where to search for genes coding for biomolecules and bioproducts of interest. In particular, comparative analysis of genome sequences is a relatively straightforward approach that allows the identification of target molecules across microorganisms, which differ for phylogeny, habitat or lifestyle. While this procedure, requiring known sequences of homologous enzymes, cannot lead to the discovery of entirely new molecules, it is still well suited to capture the molecular microdiversity that may arise from the adaptation to extreme environmental niches.

With this motivation, we have applied comparative genomics within the *Psychrobacter* genus to specifically identify and clone a gene encoding a cold-active protease in our isolate *Psychrobacter* strain 94-6PB, and further characterize the physicochemical properties of the expressed enzyme via *in vitro* analyses. *Psychrobacter* is a bacterium traditionally associated with low temperature environments (e.g., Siberian permafrost, Antarctic soil, seawater and sea-ice, deep-sea, glacial mud etc.) and is considered a model organism for studies on cold adaptation. Members of this genus have been shown to greatly differ in terms of both cold-adaptive traits and genome content (Zhang et al., 2017), and this can have important implications not only for the ecology of this organism but also for its use in industrial applications.

The biotechnological potential of *Psychrobacter* has been attracting increasing attention, for example for bioremediation treatments but also as a source of cold-active enzymes (e.g., lipases/esterases, proteases, β-lactamases, amylases, DNases) that can be used as catalysts in industrial bioprocesses (Dang et al., 2009). In this context, most of the research so far has specifically targeted lipases (most recently Novototskaya-Vlasova et al., 2012; Zhang et al., 2018), while in comparison little is known about proteases, the second industrially relevant enzyme. To contribute to closing the knowledge gap in this area, in this work we provide a description of a new cold-active protease from an Antarctic *Psychrobacter* together with a demonstration of an easy and effective experimental approach that can be also applied to the search and characterization of other types of microbial biomolecules.

MATERIALS AND METHODS

Microorganisms and Cultivation Conditions

A collection of 33 bacterial strains previously isolated from a glacier forefield soil in the Larsemann Hills, East Antarctica (Bajerski and Wagner, 2013) was used in the present study (**Supplementary Table S1**). All microorganisms were routinely cultivated on R2A medium at a temperature of 10°C. *Psychrobacter* strain 94-6PB was selected for further characterization with regard to growth temperature range by measuring the optical density ($OD_{600\,nm}$) of cultures in marine broth medium (Difco) at 0, 4, 10, 15, 22, and 30°C. To obtain an accurate taxonomic identification, the 16S rRNA gene of bacterial strains was sequenced in almost full length with primers 27F and 1492R following standard protocols.

On Plate Screening for Cold-Active Enzymes

Bacteria were screened for the synthesis of extracellular proteases and lipases at temperatures of 4, 10, and 18°C. Synthesis of proteases was assessed based on the original method by Frazier and Rupp (1928) by placing a volume of 10 μl of a pre-grown culture onto a calcium caseinate agar plate (Sigma-Aldrich, St. Louis, United States) and monitoring the formation of a clear halo around the colony during incubation for up to 4 weeks. As for lipases, bacterial strains were first screened for a general lipolytic activity by placing a volume of 10 μl of a pre-grown culture onto a R2A agar plate supplemented with sunflower oil (1% v/w) and rhodamine B (0.001% v/w) and monitoring

under UV irradiation the formation of a fluorescent orange-pink halo around the colony (Beisson et al., 2000). Positive strains were then further screened for the synthesis of either esterases (EC 3.1.1.1) capable to degrade short chain fatty acids using tributyrin (C4) as substrate (1% v/w) or lipases (EC 3.1.1.3) capable to degrade long chain fatty acids using triolein (C18) as substrate (1% v/w).

Comparative Genomic Analysis

Genome analysis and all further downstream analyses were performed on a selected bacterium, *Psychrobacter* strain 94-6PB. All genomes of *Psychrobacter* strains available at the NCBI database were individually searched for protease-coding genes (EC 3.4) and the retrieved sequences were screened with the program SignalP 4.1 (Petersen et al., 2011) for the presence of signal peptide cleavage sites to predict extracellularly released proteins. The resulting homologous sequences (**Supplementary Table S2**), including 500 bp-long untranslated 5′ and 3′ regions (UTR), were used for alignment with Clustal Omega and primers were designed with the program Primaclade (Gadberry et al., 2005) targeting conserved regions both within the coding sequence and in the UTRs. The complete list of the primers used for PCR amplification of the protease gene is reported in **Supplementary Table S3**.

Gene PCR-Amplification and Sequencing

The target gene coding an extracellular protease was first amplified and sequenced in overlapping fragments, which were used to assemble the full sequence. Then, a specific pair of primers (P*prot*-F 5′-GCTTAACTAGTATCAACACTGCTG-3′ and P*prot*-R 5′-TCGTGCCGTACAGGTATAATCG-3′) was designed based on the obtained sequence and used to amplify the gene as single PCR amplicon. The PCR reaction mix consisted of 0.025U HotStart Plus Taq Polymerase, 0.2 mM dNTPs, 20 μM each primer, 3 mM $MgCl_2$, 10 ng of DNA template, 2.5 μl of 10x PCR buffer and PCR grade water to final volume of 25 μl. Touch-Down PCR was performed as follows: initial denaturation at 95°C for 10 min, 11 cycles of denaturation at 95°C for 1 min, annealing starting at primer specific T_a (**Supplementary Table S3**) increased by 10°C for 1 min and then lowering by 1°C at each cycle, elongation at 72°C for 1 min, followed by 24 cycles of denaturation at 95°C for 1 min, annealing at primer specific T_a for 1 min, elongation at 72°C for 1 min, and final elongation at 72°C for 5 min. PCR products were checked for purity and molecular mass by gel electrophoresis.

Gene cloning was further applied in cases where it was not possible to obtain clear separation of the target band from unspecific products. Cloning of the gel-excised target band was performed using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, United States) with the cloning vector pCRTM2.1-TOPO vector Mach1^{TM}-T1R according to the manufacturer's instructions. Positive clones were picked from the plate and used for a screening PCR (initial denaturation at 96°C for 10 min, 35 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, elongation at 72°C for 1.5 min, and final elongation at 72°C for 10 min). PCR products were purified by agarose gel band excision and sequenced using the M13 reverse primer. Sequencing was done at GATC Biotech (Köln, Germany).

Gene Sequence Analysis and Protein Structure Prediction

The nucleotide sequence of the protease gene identified in our *Psychrobacter* strain 94-6PB (*prot94-6PB*) was compared with homolog sequences in other *Psychrobacter* organisms by multiple sequence alignment using Clustal Omega. The nucleotide sequence was translated into the corresponding amino acid sequence using the ExPASy online translation tool and the conserved domains were identified using NCBI's conserved domain database (CDD). The 3-dimensional structure was then modeled using the SWISS-MODEL online tool and the PDB entry 6IQR with 35.74% sequence identity to P94-6PB_SP as template. For visualization, the PyMOL software (PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) and CAVER 3.0 PyMOL plugin were used (DeLano, 2002; Chovancová et al., 2012).

Heterologous Gene Expression and Purification of the Recombinant Protease

The protease-coding gene was PCR amplified both with (*prot94-6PB_SP*) and without (*prot94-6PB)* the signal peptide using the primer pairs prot-for1 (5′-GGTACATATGATGAAAAAACAAC CAGCAC-3′ with NdeI restriction site underlined) and prot-rev1 (5′-ATTTGTCTCGAGCAACTTGTCTTCTGGGCTAAC-3′ with XhoI restriction site underlined) and protSP-for1 (5′-GG CACACATATGGCAAACACTGATACTGAAGGC-3′ with NdeI restriction site underlined) and prot-rev1, respectively. The PCR reaction mix contained 0.5 μl (1 U) of Phusion DNA Polymerase (NEB), 0.25 μM dNTPs, 0.5 μM each primer, 25 ng of genomic DNA, 1.5 μl of DMSO, 10 μl of 5x HF buffer and PCR grade water to a final volume of 50 μl. PCR was performed with an initial denaturation at 98°C for 30 sec, followed by 30 cycles at 98°C for 10 sec, 60°C for 30 sec and 72°C for 1 min, and final elongation at 72°C for 10 min. Gel-purified PCR products as well as pET30b expression vector were digested at room temperature for 3 h with Fast Digest NdeI and XhoI restriction enzymes in FD green buffer (Thermo Fisher Scientific). Digested PCR products and expression vector pET30b were ligated using T4 DNA ligase and T4 DNA ligase buffer. Ligation was performed at 16°C for 18 h and the recombinant plasmids were transformed into chemically competent *E. coli* TOP10 cells (Thermo Fisher Scientific). Obtained cell colonies were screened for successful ligation by colony PCR using the primers T7 promoter-f (5′-TAATACGACTCACTATAGGG-3′) and T7 terminator-r (5′-GCTAGTTATTGCTCAGCGG-3′) and standard PCR conditions, and plasmids were sequenced to confirm the inserted protease target gene into pET30b expression vector.

Recombinant plasmids pET30b/*prot94-6PB* and pET30b/*prot94-6PB_SP* were transformed into *E. coli* BL21 (DE3) and *E. coli* ArcticExpress competent cells (Agilent Technologies). Overnight cultures of transformed cells were inoculated into ZYP5052 autoinduction medium (Studier, 2005)

supplied with 1 mM kanamycin. Cells were grown at 37°C up to an optical density ($OD_{600\,nm}$) of 0.6 and afterward incubated at 13°C for 24 h. Cells were harvested (5000 × *g* at 4°C for 20 min), resuspended in 15 ml of wash buffer (50 mM Na_2HPO_4/NaH_2PO_4, 300 mM NaCl, pH 7.5) and incubated with lysozyme (300 μg/ml), DNaseI and RNase (10 μg/ml each) for 1 h on ice. Cells were disrupted using an M-110L Microfluidizer processor (Microfluidics, Westwood, MA, United States), yielding the crude extract. Cell debris was then removed by centrifugation (15,000 × *g* at 4°C for 60 min) and the supernatant was filtered (0.45 μm) to recover the cell lysate. P94-6PB protein purification was performed on a ÄKTA pure system (GE Healthcare, Munich, Germany) using Ni-NTA columns (GE Healthcare). The hexahistidin-tagged proteins were eluted with an imidazole gradient (0–500 mM) and protein-containing fractions were dialyzed against the storage buffer (15 mM Na_2HPO_4/NaH_2PO_4, 137 mM NaCl, pH 7.5). During cell disruption and purification, protein samples were analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie blue staining. Finally, protein concentration was determined using the Bradford assay and bovine serum albumin (BSA) as standard. Correct protein mass was verified by electrospray ionization mass spectrometry (LC-ESI-TOF-MS, Agilent 1260 HPLC coupled with Agilent 6530 Accurate-Mass Q-TOF mass spectrometer; Agilent, Santa Clara, CA, United States).

Protease Activity Assay

Activity assay was performed in a 100 μl volume using azocasein (Sigma-Aldrich) as substrate following the method described by Tomarelli et al. (1949). Briefly, hydrolysis of the casein releases the azo dye into the medium where it is detected by absorbance at 440 nm. Therefore, azocasein (final concentration 10 mg/ml) was preheated/cooled to the assay temperature and the reaction was started with the addition of the P94-6PB solution (final concentration 1 mg/ml). The mixture was shaken for 30 min and stopped with 100 μl Trichloroacetic acid (TCA) [10 % (w/v)]. In order to ensure a quantitative separation of precipitate and supernatant, the solution was centrifuged at 16,000 × g, for 60 min at 4°C, and 100 μl of the supernatant was mixed with 100 μl 2 M NaOH. The activity was determined by absorption measurements at 440 nm. For each condition, a blank measurement was performed. Here, 100 μl TCA [10% (w/v)] were added prior enzyme addition. All measurements were performed in triplicates.

Temperature- and pH-Range Activity Profile

For pH-dependent measurements, the protease activity assay was performed at 37°C. Reaction buffer was either Tris-buffer (0.2 M Tris, 0.1 M acetic acid, 0.1 M 2-N-Morpholino Ethane sulfonic acid, 0.01 M $CaCl_2$, pH 5–8) or Glycine buffer (0.2 M Glycine, 0.01 M $CaCl_2$, pH 9-11). For temperature-dependent measurements, the protease activity assay was performed at pH 9.0 using a Tris-buffer (0.1 M Tris, 0.1 M Glycin, 0.1 M NaCl, 0.01 M $CaCl_2$) at different temperatures ranging from 5 to 50°C.

Irreversible Thermal Inactivation

Enzyme solution (2 mg/mL) was incubated at different temperatures ranging from 5 to 50°C for 1 h and subsequently stored on ice for 30 min. Precipitated protein was spinned down and the supernatant was used for the protease activity assay as described above at 30°C and pH 9.0.

Effect of Enzyme Inhibitors

The effect of enzyme inhibitors was investigated at 30°C and pH 9.0 using a Tris-reaction buffer (0.1 M Tris, 0.1 M Glycine, 0.1 M NaCl, and 0.01 M $CaCl_2$). Urea, Dithiothreitol (DTT), β-mercaptoethanol (β-ME) and Ethylenediaminetetraacetic acid (EDTA) were dissolved in the reaction buffer, Phenylmethylsulfonyl fluoride (PMSF) was dissolved in methanol (0.5 M each). Protease solution was incubated with 2 or 5 mM of the respective inhibitor for 30 min on ice to minimize any temperature-related structural instability. As a control, the protease was incubated with the same volume of buffer/methanol. Reaction was started with the addition of the pre-incubated enzyme solution to the azocasein solution and the assay was performed as described above.

Effect of Surfactants and Detergents

The effect of surfactants and detergents was investigated at 30°C and pH 9.0 using a Tris reaction buffer (0.1 M Tris, 0.1 M Glycine, 0.1 M NaCl, and 0.01 M $CaCl_2$). Tween 20 (Sigma-Aldrich), Tween 80 (Sigma Aldrich), sodium dodecylsulfate (SDS; Roth), Triton X-100 (Roth) and the oxidizing agent H_2O_2 (Roth) were prepared as 20% stock solutions in the reaction buffer. Protease solution was incubated with 1 or 5% of the respective chemical for 30 min on ice. As a control, the protease was incubated with the same volume of buffer. Reaction was started with the addition of the pre-incubated enzyme solution to the azocasein solution and the assay was performed as described above.

Phylogenetic and Pairwise Identity Analysis

Percentage pairwise identity was computed with EBI's MUSCLE tool and visualized as Heat-Map with Morpheus software[1]. Evolutionary relationships were determined by using the Maximum Likelihood method based on the Tamura-Nei model. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016).

Sequence Accession Number

The nucleotide sequence of the gene *prot94-6PB* has been submitted to GenBank database under the accession number MN606318.

[1] https://software.broadinstitute.org/morpheus

RESULTS

Abundance and Biodiversity of Bacteria With Enzymatic Activity in Glacier Forefields

The 33 bacterial isolates from glacier forefield soils in the Larsemann Hills (East Antarctica) used in this study were largely represented by Actinobacteria (17/33; 52%), particularly of the *Arthrobacter* genus, followed by Gamma-proteobacteria (7/33; 21%), Alpha-proteobacteria (5/33; 15%), Bacteroidetes (3/33; 9%) and Deinococcus (1/33; 3%) (**Figure 1A**). Among these isolates, 85% (28/33) screened positive for enzymatic activity, while 15% (5/33) showed no activity. All active strains were found to produce lipases and 33% (11/33) produced both lipases and proteases. The latter ones comprised mostly strains of *Arthrobacter*, *Pseudomonas* and *Psychrobacter* strain 94-6PB (**Figure 1B**). Among the strains with lipolytic activity, 46% (13/28) produced only lipases (degradation of C18 fatty acids), 14% (4/28) only esterases (degradation of C4 fatty acids), 21% (6/28) produced both and for 18% (5/28) it was not possible to discriminate (**Supplementary Table S1** and **Supplementary Figure S1**).

In silico Analysis and Gene Identification of P94-6PB Protease

Representing a model organism for cold-adaptation, *Psychrobacter* strain 94-6PB was selected for further characterization. Based on the almost full-length 16S rRNA gene sequence, this strain was found closest to *Psychrobacter glacincola* (99%). It was capable to grow in a temperature range from 0 to 20°C (subzero temperatures were not tested) and showed highest growth rates at 15°C (μ_{max} 0.07 h^{-1}). Little growth was observed at 30°C.

To better elucidate the biotechnological potential of its proteolytic enzymes, we set out to identify the gene sequence of an extracellular protease, clone it in heterologous hosts and characterize the physicochemical properties and activities of the purified enzyme *in vitro*. Because strain 94-6PB had no sequenced genome, we PCR-amplified and sequenced the protease-target gene using primers designed on the basis of comparative analysis of homologous genes in other *Psychrobacter* organisms. Search on the NCBI database for all available *Psychrobacter* genomes, both complete and as scaffolds/contigs, revealed that homologous sequences of a gene coding for an extracellular peptidase (i.e., signal site for extracellular translocation predicted to be at the amino acid in position 31/32) were present in 12 genomes (**Supplementary Table S2**). Multiple sequence alignment of both gene and protein sequences showed that most of the peptidases shared similarity >80%. Highly conserved regions were identified within the peptidase-coding gene of all strains as well as in the flanking untranslated regions (5′-UTR and 3′-UTR) and they were used as target for primer design and PCR amplification. Initially, the gene was amplified in 3 overlapping fragments using the following primer combinations: P1-F/P6-R (product size 2080 bp), P10-F/P10-R (product size 1156 bp) and P24-F/P20-R (product size 555 bp). In addition, internal primers P5-F, P4-R and P9-R were used for sequencing. Finally, using primers P26-F and P20-R, a 2471 bp-long amplicon was obtained without any secondary products. Subsequent sequencing confirmed the full peptidase-coding gene including parts of the 5′ UTR (54 bp upstream) and 3′ UTR (203 bp downstream) (**Figure 2A**, **Supplementary Table S3** and **Supplementary Figure S2**).

Protease Sequence Analysis and Prediction

The gene coding sequence resulted to be 2214 bp long and its translated polypeptide was predicted, by the Expasy translate tool, to be 737 aa long, with an estimated molecular size of approximately 80 kDa (**Figure 2B**). Analysis with the NCBI's CDD tool confirmed the identification of a serine peptidase belonging to the S41 superfamily (E.C. 3.4.21). The protein

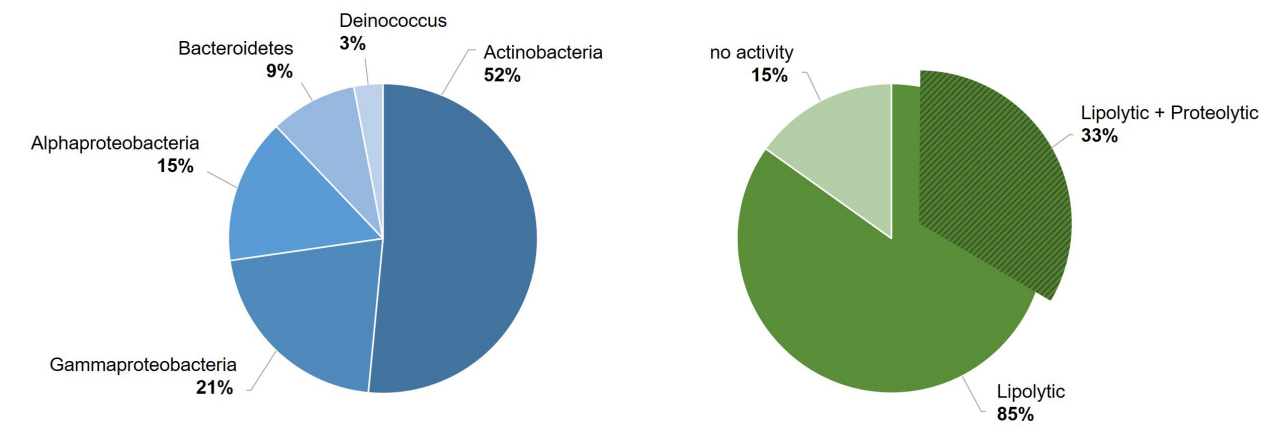


FIGURE 1 | Overview of the bacterial strain collection from the Larsemann Hills (East Antarctica) used in this study for the search of cold-active enzymes. Bacterial isolates were assessed for both biodiversity (i.e., taxonomic affiliation) **(A)** and extracellular enzymatic activities at low temperatures (4–18°C) **(B)**.

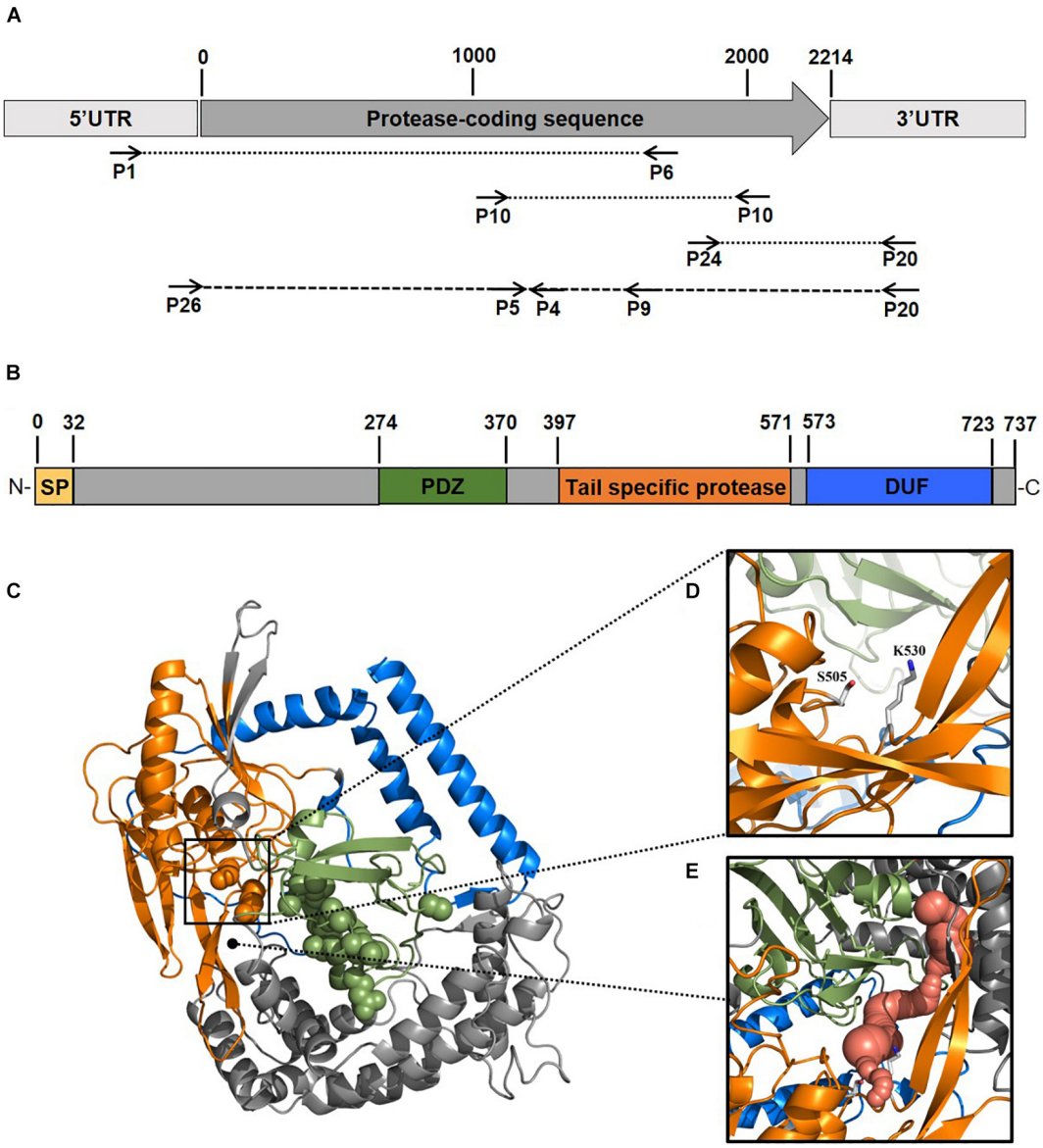


FIGURE 2 | Scheme of the experimental approach used to identify a new protease from the Antarctic bacterium *Psychrobacter* sp. 94-6PB. Target protease-coding gene was PCR-amplified in our strain using a combination of primers designed based on comparative genome analysis across *Psychrobacter* organisms **(A)**. Sequence-based prediction of the domains of the target enzyme (SP, signal peptide; PDZ domain for substrate recognition; tail-specific protease domain with catalytic function; DUF, domain of unknown function) **(B)**. Predicted 3D-macromolecular structure of the target protease **(C)**, with catalytic dyad at position S505 and K530 **(D)** and a connecting tunnel to the PDZ domain for substrate recognition **(E)**.

was predicted to consist of: (i) transmembrane region, (ii) PDZ domain containing a protein-binding site involved in substrate recognition; (iii) tail-specific protease domain with catalytic activity; and (iv) a conserved domain of unknown function (DUF) at the C-terminal region of the protease (**Figure 2B**). Modeling of the 3-dimensional structure (**Figure 2C**) resulted in a globular protein consisting of 18 α-helices and 19 β-strands. The substrate-binding site showed to be well exposed on the surface, while the catalytic dyad was found to be in the inner part of the enzyme. Superimposition of the modeled structure of P94-6PB_SP and the template 6IQR showed the catalytic dyad at position S505 and K530 (**Figure 2D**). Besides, a tunnel connecting the PDZ domain for substrate recognition and the catalytic dyad could be also detected (**Figure 2E**).

In vitro Expression and Purification

To obtain insights into the physicochemical properties and activity of the 94-6PB protease, we first worked to express the enzyme in heterologous hosts. The PCR product containing the protease sequence (with and without signal peptide) was cloned into pET30b expression vector and the recombinant plasmid was transformed into both *E. coli* BL21 (DE3) and *E. coli*

ArcticExpress (DE3). The latter strain, co-expressing cold-active chaperonins, Cpn10 and Cpn60 from *Oleispira antarctica*, that assist in the folding of proteins at low temperatures, was expected to be particularly suited for molecules from psychrophilic organisms. However, expression tests using the two strains, different kanamycin concentrations as well as different media and cell disruption methods did not exhibit higher expression levels of the protease using the *E. coli* ArcticExpress strain. Highest protein yields (33 mg/l cultivation) could be obtained by cultivating BL21 (DE3) in ZYP medium supplemented with 50 μg/ml kanamycin. Finally, the highest protein yields after his-tag purification, 1.65 mg/50 ml cultivation, were obtained using *E. coli* BL21(DE3) as expression strain with no addition of kanamycin and combined with the microfluidizer as extraction method. The his-tagged protease could be successfully expressed and purified via affinity chromatography (**Figure 3A**). To verify the purified protease, we performed mass analysis and detected a peak corresponding to mass 80613 Da for the protease without signal peptide, which is matching with the calculated mass (**Figure 3B**). A summary of the purification steps of the protease is reported in **Table 1**.

Temperature- and pH-Range Activity and Stability

Protease activity assays using azocasein as substrate demonstrated highest activity of P94-6PB at pH 9.0. At higher pH values (pH 10.0 and 11.0), relative protease activity was still above 50%, whereas under acidic conditions (pH 5.0), P94-6PB exhibited no protease activity (**Figure 4A**). Thus, investigation of temperature-dependent protease activity was performed at pH 9.0. Here, P94-6PB exhibited highest activity at 30°C. At 20°C, we still observed a relative protease activity of 71%. Reducing the temperature to 10°C yielded a relative activity of 24%. At 5 and 50°C, the protease exhibited relative protease activities less than 10% (**Figure 4A**). With regard to the effect of temperature to the enzyme stability, P94-6PB was stable at temperatures ≤20°C. When incubated at higher temperatures, large precipitation of the protease was observed, and the residual activity dropped to less than 10% (**Figure 4B**). Overall, our data indicated that P94-6PB is an alkaline cold-active protease.

Effect of Inhibitors and Detergents on Enzyme Activity

Protease activity was measured after pre-incubation with a range of different inhibitors, detergents and surfactants. Incubation with EDTA and Urea had no significant change in protease activity. In contrast, ß-mercaptoethanol (β-ME), DTT and PMSF exhibited an inhibiting effect on the protease. The highest decrease in activity was detected in presence of 2 mM of DTT (relative activity of 58.78%), however, increasing DTT concentration had no further decrease effect, possibly because

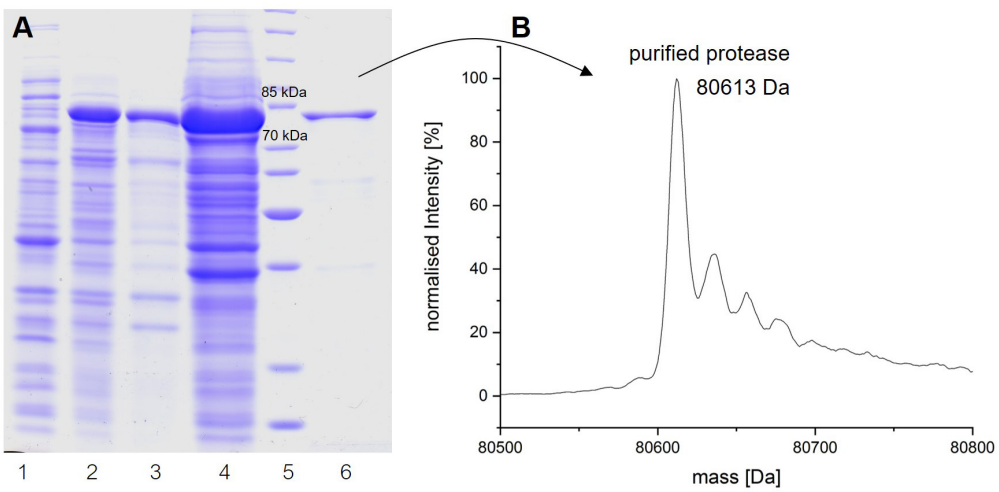


FIGURE 3 | Analysis of the expression and purification of P94-6PB protease. Analysis on 12% SDS-PAGE of enzyme induction, release and purification **(A)**: non-induced (1), induced (2), lysate (3), crude extract (4), protein ladder (5) and purified enzyme (6). Electrospray ionization mass spectrometry (LC-ESI-TOF-MS) analysis and deconvoluted mass spectrum of the purified eluate **(B)**.

TABLE 1 | Purification of the extracellular protease produced by *Psychrobacter* sp. P94-6PB.

Purification step	Fraction volume (ml)	Protein concentration (mg/ml)	Inserted protein amount in activity assay (mg)	Total activity (U)	Specific activity (U/mg)	Purification fold	Yield (%)
Crude extract	80	70	3.5	191	54.6	1	100
Ni-NTA	37.5	2	0.1	56	560.0	10	29

One unit of activity was defined as an increase of 0.01 in absorbance units at 440 nm in the azocasein assay.

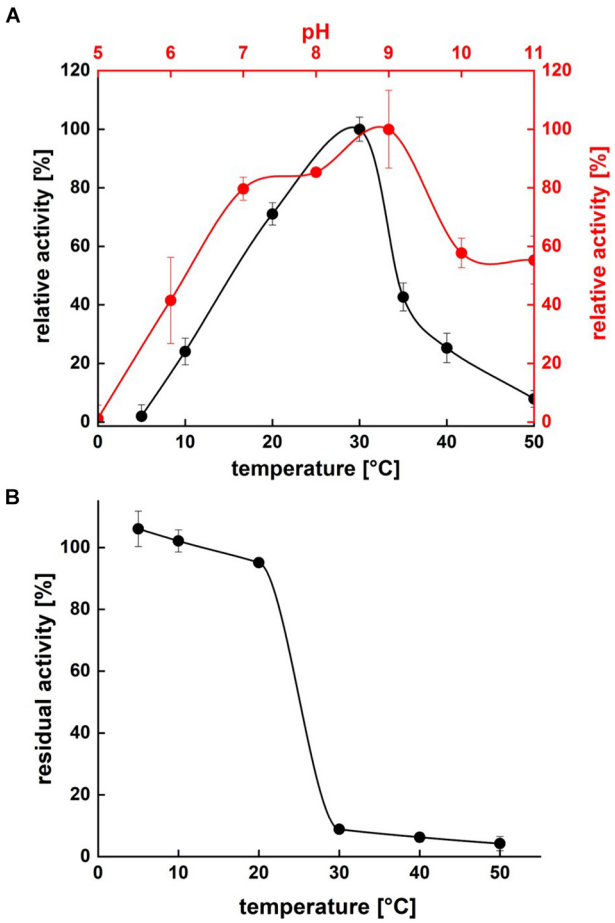


FIGURE 4 | Effect of temperature and pH on P94-6PB protease activity **(A)** and stability **(B)**. Overview of temperature and pH range of the activity and stability of P94-6PB protease. Relative activity was quantified using the azocasein assay.

maximal inhibition was already achieved with 2 mM DTT (**Figure 5A**). We noticed that, under the conditions (time, temperature and pH) we used, PMSF exerted only a partial inhibition of P94-6PB.

Incubation with the detergent Tween 80 had no significant impact on the protease activity, whereas Triton X-100 treatment induced a slight decrease of the relative activity. Incubation with 1% Tween 20 caused a reduction of activity by approximately 24%; however, increasing Tween 20 concentration to 5%, had a slightly positive effect. Apparently, P94-6PB activity was highly dependent on the detergent concentration, as similarly observed by others (Abdel-Hamed et al., 2016). Incubation with 1% SDS caused the complete loss of protease activity. Finally, we could observe that pre-incubation with 1 and 5% of the oxidizing agent H_2O_2 resulted in an increase of the relative activity by 12 and 23%, respectively (**Figure 5B**). Here, it is likely that H_2O_2 contributed to make azocasein more susceptible to proteolytic degradation (Fligiel et al., 1984).

Enzyme Microdiversity in *Psychrobacter* Organisms From Different Cold-Habitats

Finally, we sought to investigate how genetically different is the protease of our strain 94-6PB compared with homologous sequences of other 21 *Psychrobacter* strains from different ecological niches subdivided in three main groups: terrestrial habitats (e.g., soil and permafrost, Arctic and Antarctic), aquatic habitats (e.g., marine water, polluted and pristine, cold and temperate) and host-associated (e.g., skin and intestinal, animals, fish, tunicates). The phylogenetic reconstruction, inferred by using the Maximum Likelihood method based on the Tamura-Nei model, showed a general tendency of the enzyme sequences to cluster on the basis of strain-specific habitat, and this was also reflected in the percentage identity matrix (**Figure 6** and **Supplementary Figure S3**). Strains of aquatic origin such as *Psychrobacter* sp. choline-3u-12, *P. piscatorii* LQ58, *P. pacificensis* DSM 23406 and *Psychrobacter* sp. AntiMin-1 clustered together, and particularly the three latter ones, all isolated from deep seawater in the Pacific ocean, shared high sequence similarity (97–98%). Similarly, also terrestrial strains showed a high degree of relatedness, particularly *P. arcticus* 273-4 and *P. cryohalolentis* K5 (91%), both from permafrost of the Kolyma region in Siberia. However, also *Psychrobacter* sp. G, a seawater isolate, showed high similarity (91 and 98%, respectively) with both permafrost strains. Host-associated *Psychrobacter* sp. JCM 18902, *Psychrobacter* sp. JCM 18903 and *Psychrobacter* sp. P11F6, all isolated from cold/frozen marine organisms shared also high sequence similarity (98%). The overall similarity among all sequences of the serine peptidase enzyme was approximately 80 ± 5%, with some strains such as *Psychrobacter* sp. PRwf-1 and *P. lutiphocae* DSM 21542, both host-associated, being rather dissimilar from the others and having sequence similarity as low as 66 ± 3%. The enzyme of our isolate *Psychrobacter* sp. 94-6PB (from Antarctic soil) was closest (80% sequence identity) to *P. urativorans* R310.10 (from Antarctic soil) and *Psychrobacter* sp. P2G3 (host-associated in the Arctic) (**Figure 6**). While more data also on other enzyme families need to be considered, the role of environmental conditions in shaping the molecular diversity, hence the functionality, may have important implications for the search and/or selection of novel biocatalysts from natural sources.

DISCUSSION

Fast-growing industrial demand for sustainable bioproducts with unique functionalities (e.g., performing under harsh conditions) drives research in the field of bioprospecting toward the continuous search for novel organisms and biomolecules, and also for new methodologies enabling their discovery. Here, we report on the characterization of a cold-active alkaline protease from an Antarctic bacterium, *Psychrobacter* sp. 94-6PB. This enzyme is active over a wide range of temperatures (10–50°C) and pH values (6–11) but performs best at 20–30°C and pH 7-9. It is largely unaffected by the major protease inhibitors and is compatible with most of common detergents and surfactants; only SDS exerted a substantial negative effect. However, the

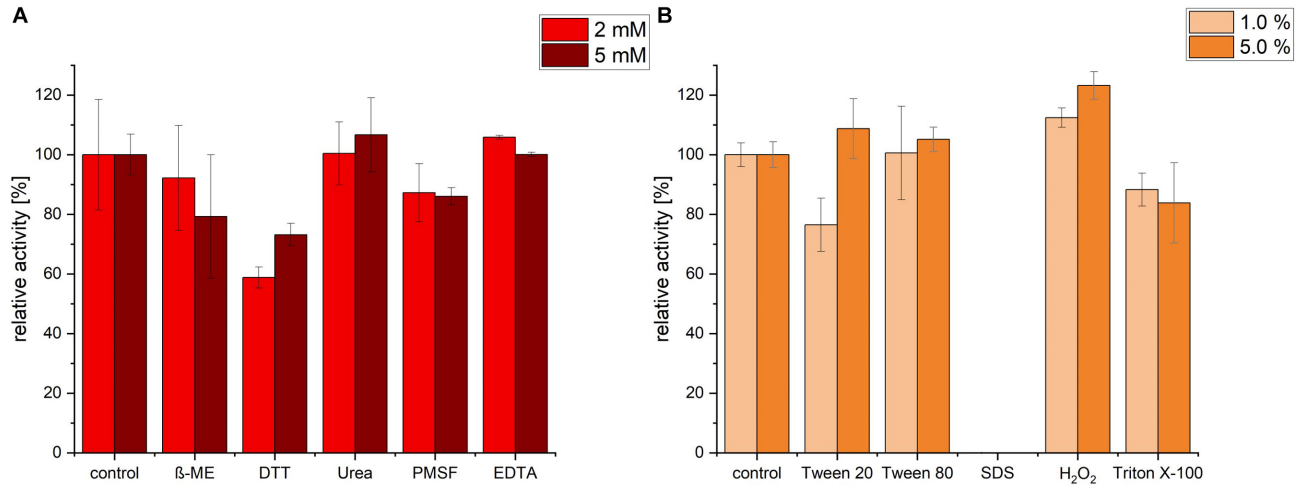


FIGURE 5 | Effect of inhibitors, detergents and surfactants on P94-6PB protease activity. Relative activity measured via the azocasein assay after pre-incubation (30 min) with inhibitors (β-mercaptoethanol (β-ME), Dithiothreitol (DTT), Urea, Phenylmethylsulfonyl fluoride (PMSF) and Ethylenediaminetetraacetic acid (EDTA) at 2 and 5 mM concentration) **(A)** and detergents and surfactants (Tween 20, Tween 80, sodium dodecylsulfate (SDS), Hydrogen peroxide (H_2O_2) and Triton X-100, at 1.0 and 5.0% v/v concentration) **(B)**.

partial tolerance of P94-6PB against the well-known serine protease inhibitor PMSF needs further investigation.

Collectively, these physicochemical properties make it highly suitable for applications especially as detergent additive. Today, companies worldwide are strongly committed to policies for sustainable laundry, which include lowering the washing temperature while maintaining washing performance unaltered. As an example, in 2014 DuPont and P&G won the Sustainable Bio Award following a collaboration to develop energy-saving laundry detergents. Despite the fact that new enzymes of this type are highly attractive for applications (Al-Ghanayem and Joseph, 2020), only 7% of the studies so far reported on proteases produced by psychrophilic bacteria, while the large majority (62%) still focused on proteases from mesophiles (Salwan and Sharma, 2019). Known cold-active proteases of bacterial origin, e.g., *Bacillus subtilis* WLCP1, *Pseudoalteromonas arctica* PAMC 21717, *Chryseobacterium* sp., and *Stenotrophomonas maltophilia* MTCC 7528 all share similar properties with the 94-6PB protease from *Psychrobacter* sp., having optimum pH at 9–10 and optimum temperature at 10–30°C (reviewed by Joshi and Satyanarayana, 2013; Al-Ghanayem and Joseph, 2020). Amongst these, there is only one produced by another *Psychrobacter* organism, *P. proteolyticus* DSM 13887^T (associated with Antarctic krill); similarly to ours, this enzyme showed highest activity in the temperature range 20–30°C, though at a more neutral pH 6.5-7.0 (Denner et al., 2001).

The genus *Psychrobacter*, having colonized a variety of cold habitats (terrestrial, marine and also host-associated) over a wide geographic distribution (Arctic, Antarctic, high altitude), is ideal to examine the effect of environmental adaptation on functional biodiversity while reducing the phylogenetic effect. Our comparative analysis of 22 homologous protease-coding genes from different *Psychrobacter* strains shows several clusters correlating with the habitat, for example the deep-sea marine cluster (*Psychrobacter* sp. choline-3u-12, *P. piscatorii* LQ58, *P. pacificensis* DSM 23406 and *Psychrobacter* sp. AntiMin-1, 97–98% sequence similarity), the marine organism-associated cluster (*Psychrobacter* sp. JCM 18902, *Psychrobacter* sp. JCM 18903 and *Psychrobacter* sp. P11F6) and the terrestrial-permafrost cluster (*P. arcticus* 273-4 and *P. cryohalolentis* K5). The protease of our strain *Psychrobacter* sp. 94-6PB was most similar to *P. urativorans* R310.10 and both organisms shared the same environmental source, being isolated from Antarctic soil. Thus, our data on the correlation between protease diversity and habitat further supports recent findings pointing toward the ecological specialization of *Psychrobacter* organisms (Zhang et al., 2017; Bakermans, 2018). For example, based on whole genome comparison, aquatic and terrestrial strains seem to be better adapted to low temperatures and high salinity than host-associated ones (Bakermans, 2018). It is conceivable that such adaptive traits are then expressed also in proteins and enzymes. While a clearer understanding can be gained as more genomes and comparative analyses become available, the aspect of functional microdiversity holds a great potential to refine the discovery of biomolecules of biotechnological interest.

Experimentally, we applied a recombinant DNA technique, in which we first identified the biosynthetic gene of interest via *in silico* comparative analysis of *Psychrobacter* genomes deposited in public databases, PCR-amplified the target gene in our strain 94-6PB and then cloned, expressed and purified it in heterologous *E. coli* host cells. Expression systems for heterologous production of cold-active enzymes have long suffered of various technical constraints (e.g., intrinsic low thermostability of the recombinant enzyme, protein misfolding), but new tools are becoming accessible to tackle these issues (Bjerga et al., 2016). In our specific case, the expression of the protease P94-6PB did not really benefit from the use of *E. coli* ArcticExpress

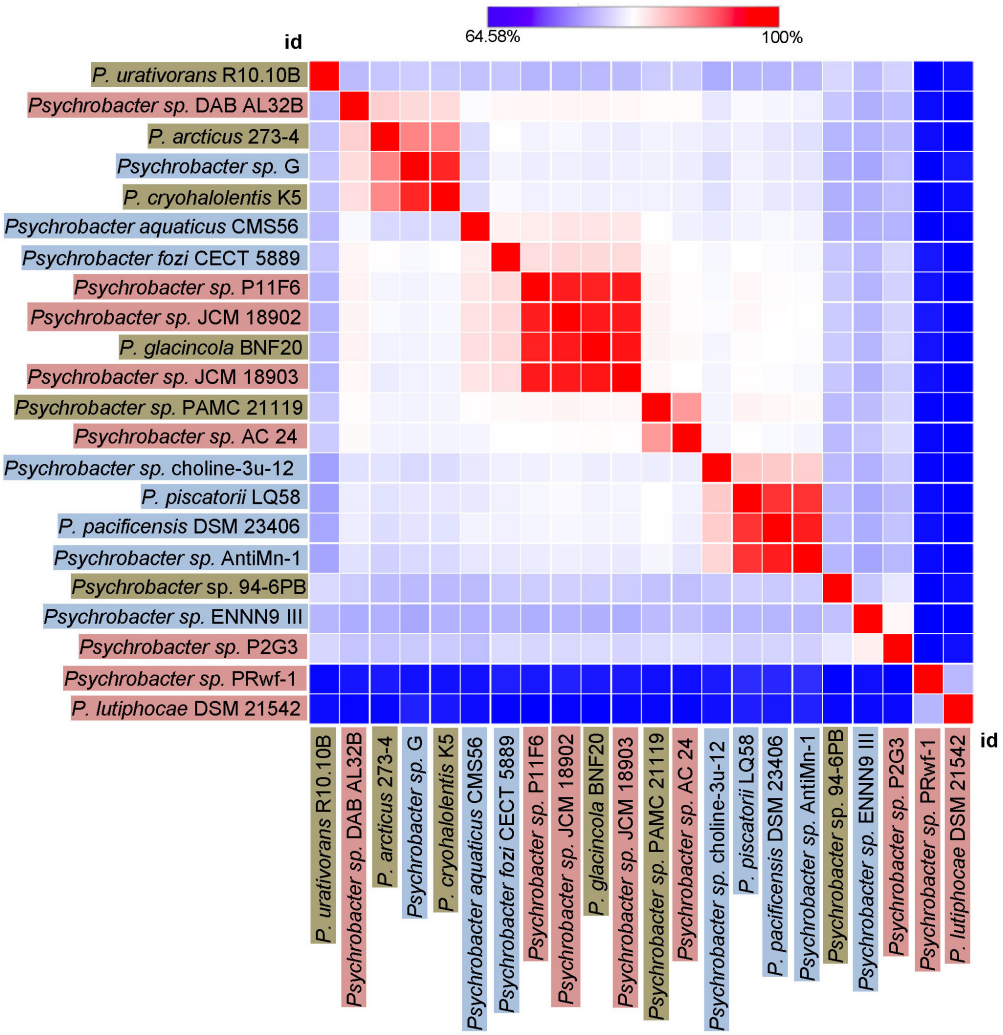


FIGURE 6 | Heat-Map showing percentage identity of the studied protease across the *Psychrobacter* genus. Homologous gene sequences coding for the target protease in 22 strains of *Psychrobacter* sp. (including our strain 94-6PB) are used as entry points of the Heat-Map. Each strain identifier (id) is reported as row and column, and color-coded according to the ecological source (terrestrial-brown, aquatic-blue and host associated-pink). A color bar indicates the correspondence between pairwise identities and the colors displayed in the matrix.

(engineered to co-express Cpn60 and Cpn10 chaperons from *Oleispira antarctica*) compared to a standard *E. coli* BL21(DE3). Likewise, other enzymes from *Psychrobacter* isolates, including nitroreductase (Wang et al., 2019), lipase (Xuezheng et al., 2010), esterase (Novototskaya-Vlasova et al., 2012), have been successfully produced in heterologous *E. coli* cells.

Although not yet fully established, the genome mining/heterologous expression-based approach is becoming more and more popular, as it offers distinct advantages. First, it bypasses the cultivation step, which is often challenging for extremophilic microorganisms. Second, virtually any gene of interest available in public repositories can be simply synthesized and directly used for heterologous expression, with no need for the original organism. Third, having the enzyme-coding sequence at hand opens to further additional study possibilities; for example, it allows running sequence-based modeling to predict *in silico* enzyme activity and characteristics prior of testing, or conduct genetic engineering experiments to improve or customize the enzyme. Finally, it enables to investigate the enzyme microdiversity, which may arise within phylogenetically related groups of microorganisms as a consequence of the adaptation to different environmental niches (Zimmerman et al., 2013; Nguyen et al., 2019). Today, it is estimated that approximately 90% of enzymes used in industrial processes are recombinant forms (Adrio and Demain, 2014).

Thus, in this work we presented three main novelty aspects for the bioprospecting field, namely the use of extremophilic organisms, of genome repositories as a direct source of new functional biomolecules and the habitat-driven microdiversity to further expand the range of molecular variants. Further research in all these research areas is necessary to boost the utilization of microbial biotechnologies at industrial scale.

DATA AVAILABILITY STATEMENT

The data generated for this study can be found in the GenBank MN606318.

AUTHOR CONTRIBUTIONS

AP, DW, and NB conceived the study. E-LN, GF, and AP conducted experimental work and data analysis. AP and GF wrote the manuscript. All authors read and approved the manuscript and contribute to interpret the results.

FUNDING

This project was supported by European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Grant Agreement No. 657473 (BioFrost) to AP. The isolation of the new strains was done in a previous project with support of the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program "Antarctic Research with Comparative Investigations in Arctic Ice Areas" by a grant to DW (WA 1554/18).

ACKNOWLEDGMENTS

We thank Hamid Reza Karbalaei-Heidari (Shiraz University, Shiraz, Iran) for critical reading this manuscript and providing useful suggestions for its improvement. NB thanks Prof. Christian Thomsen, President of the Technical University of Berlin and the European Union SynCrop ETN Grant No. 764591 for support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00881/full#supplementary-material

REFERENCES

Abdel-Hamed, A. R., Abo-Elmatty, D. M., Wiegel, J., and Mesbah, N. M. (2016). Biochemical characterization of a halophilic, alkalithermophilic protease from *Alkalibacillus* sp. *NM-Da*2. *Extremophiles* 20, 885–894. doi: 10.1007/s00792-016-0879-x

Adrio, J. L., and Demain, A. L. (2014). Microbial enzymes: tools for biotechnological processes. *Biomolecules* 4, 117–139. doi: 10.3390/biom4010117

Agostini, F., Völler, J. S., Koksch, B., Acevedo-Rocha, C. G., Kubyshkin, V., and Budisa, N. (2017). Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew. Chem. Int. Ed. Engl.* 56, 9680–9703. doi: 10.1002/anie.201610129

Al-Ghanayem, A. A., and Joseph, B. (2020). Current prospective in using cold-active enzymes as eco-friendly detergent additive. *Appl. Microbiol. Biotechnol.* 104, 2871–2882. doi: 10.1007/s00253-020-10429-x

Bajerski, F., and Wagner, D. (2013). Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. *FEMS Microbiol. Ecol.* 85, 128–142. doi: 10.1111/1574-6941.12105

Bakermans, C. (2018). Adaptations to marine versus terrestrial low temperature environments as revealed by comparative genomic analyses of the genus *Psychrobacter*. *FEMS Microbiol. Ecol.* 94:fiy102. doi: 10.1093/femsec/fiy102

Beisson, F., Tiss, A., Rivière, C., and Verger, R. (2000). Methods for lipase detection and assay: a critical review. *Eur. J. Lipid Sci. Technol.* 102, 133–153. doi: 10.1016/j.biotechadv.2009.06.001

Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017). Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* 364:fnx211. doi: 10.1093/femsle/fnx211

Bjerga, G. E. K., Lale, R., and Williamson, A. K. (2016). Engineering low-temperature expression systems for heterologous production of cold-adapted enzymes. *Bioengineered* 7, 33–38. doi: 10.1080/21655979.2015.1128589

Borchert, E., Jackson, S. A., O'Gara, F., and Dobson, A. D. (2017). "Psychrophiles as a source of novel antimicrobials," in *Psychrophiles: From Biodiversity to Biotechnology*, eds R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday (Cham: Springer), 527–540.

Cavicchioli, R., Charlton, T., Ertan, H., Omar, S. M., Siddiqui, K. S., and Williams, T. J. (2011). Biotechnological uses of enzymes from psychrophiles. *Microb. Biotechnol.* 4, 449–460. doi: 10.1111/j.1751-7915.2011.00258.x

Chovancová, E., Pavelka, A., Beneš, P., Strnad, O., Brezovský, J., Kozlíková, B., et al. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comp. Biol.* 8:e1002708. doi: 10.1371/journal.pcbi.1002708

Dang, H., Zhu, H., Wang, J., and Li, T. (2009). Extracellular hydrolytic enzyme screening of culturable heterotrophic bacteria from deep-sea sediments of the Southern Okinawa Trough. *World J. Microbiol. Biotechnol.* 25, 71–79. doi: 10.1007/s11274-008-9865-5

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. Available online at: http://www.pymol.org (accessed March 2, 2020).

Denner, E. B., Mark, B., Busse, H. J., Turkiewicz, M., and Lubitz, W. (2001). *Psychrobacter proteolyticus* sp. nov., a psychrotrophic, halotolerant bacterium isolated from the Antarctic krill Euphausia superba Dana, excreting a cold-adapted metalloprotease. *Syst. Appl. Microbiol.* 24, 44–53. doi: 10.1078/0723-2020-00006

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol.* 9, 22–34. doi: 10.1111/1751-7915.12309

Fligiel, S. E., Lee, E. C., McCoy, J. P., Johnson, K. J., and Varani, J. (1984). Protein degradation following treatment with hydrogen peroxide. *Am. J. Pathol.* 115, 418–425.

Frazier, W. C., and Rupp, P. (1928). Studies on the proteolytic bacteria of milk I. A medium for the direct isolation of caseolytic milk bacteria. *J. Bacteriol.* 16, 57–63.

Gadberry, M. D., Malcomber, S. T., Doust, A. N., and Kellogg, E. A. (2005). Primaclade - a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* 21, 1263–1264. doi: 10.1093/bioinformatics/bti134

Gerlt, J. A. (2017). Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* 56, 4293–4308. doi: 10.1021/acs.biochem.7b00614

Hoesl, M. G., Acevedo-Rocha, C. G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., et al. (2011). Lipase congeners designed by genetic code engineering. *ChemCatChem* 3, 213–221. doi: 10.1002/cctc.201000253

Isaksen, G. V., Åqvist, J., and Brandsdal, B. O. (2016). Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7822–7827. doi: 10.1073/pnas.1605237113

Joshi, S., and Satyanarayana, T. (2013). Biotechnology of cold-active proteases. *Biology* 2, 755–783. doi: 10.3390/biology2020755

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Nguyen, T. T. H., Myrold, D. D., and Mueller, R. S. (2019). Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. *Front. Microbiol.* 10:413. doi: 10.3389/fmicb.2019.00413

Novototskaya-Vlasova, K., Petrovskaya, L., Yakimov, S., and Gilichinsky, D. (2012). Cloning, purification, and characterization of a cold-adapted esterase produced by *Psychrobacter cryohalolentis* K5T from Siberian cryopeg. *FEMS Microbiol. Ecol.* 82, 367–375. doi: 10.1111/j.1574-6941.2012.01385.x

Perfumo, A., Banat, I. M., and Marchant, R. (2018). Going green and cold: biosurfactants from low-temperature environments to biotechnology applications. *Trends Biotechnol.* 36, 277–289. doi: 10.1016/j.tibtech.2017.10.016

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Salwan, R., and Sharma, V. (2019). Trends in extracellular serine proteases of bacteria as detergent bioadditive: alternate and environmental friendly tool for detergent industry. *Arch. Microbiol.* 201, 863–877. doi: 10.1007/s00203-019-01662-8

Santiago, M., Ramírez-Sarmiento, C. A., Zamora, R. A., and Parra, L. P. (2016). Discovery, molecular mechanisms, and industrial applications of cold-active enzymes. *Front. Microbiol.* 7:1408. doi: 10.3389/fmicb.2016.01408

Sarmiento, F., Peralta, R., and Blamey, J. M. (2015). Cold and hot extremozymes: industrial relevance and current trends. *Front. Bioeng. Biotechnol.* 3:148. doi: 10.3389/fbioe.2015.00148

Struvay, C., and Feller, G. (2012). Optimization to low temperature activity in psychrophilic enzymes. *Int. J. Mol. Sci.* 13, 11643–11665. doi: 10.3390/ijms130911643

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41, 207–234. doi: 10.1016/j.pep.2005.01.016

Tomarelli, R. M., Charney, J., and Harding, M. L. (1949). The use of azoalbumin as a substrate in the colorimetric determination of peptic and tryptic activity. *J. Lab. Clin. Med.* 34, 428–433.

Wang, Y., Hou, Y., Wang, Y., Zheng, L., and Wang, Q. (2019). A novel cold-adapted nitroreductase from *Psychrobacter* sp. *ANT*206: heterologous expression, characterization and nitrobenzene reduction capacity. *Enzyme Microb. Technol.* 131:109434. doi: 10.1016/j.enzmictec.2019.109434

Xuezheng, L., Shuoshuo, C., Guoying, X., Shuai, W., Ning, D., and Jihong, S. (2010). Cloning and heterologous expression of two cold-active lipases from the Antarctic bacterium *Psychrobacter* sp. G. *Polar Res.* 29, 421–429. doi: 10.3402/polar.v29i3.6087

Zhang, S., Song, W., Yu, M., and Lin, X. (2017). Comparative genomics analysis of five *Psychrobacter* strains isolated from world-wide habitats reveal high intra-genus variations. *Extremophiles* 21, 581–589. doi: 10.1007/s00792-017-0927-1

Zhang, Y., Ji, F., Wang, J., Pu, Z., Jiang, B., and Bao, Y. (2018). Purification and characterization of a novel organic solvent-tolerant and cold-adapted lipase from *Psychrobacter* sp. *ZY*124. *Extremophiles* 22, 287–300. doi: 10.1007/s00792-018-0997-8

Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in microbes-a review. *Nat. Prod. Rep.* 33, 988–1005. doi: 10.1039/c6np00025h

Zimmerman, A. E., Martiny, A. C., and Allison, S. D. (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J.* 7, 1187–1199. doi: 10.1038/ismej.2012.176

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 May 2020
doi: 10.3389/fmicb.2020.01050

Co-crystal Structure of *Thermosynechococcus elongatus* Sucrose Phosphate Synthase With UDP and Sucrose-6-Phosphate Provides Insight Into Its Mechanism of Action Involving an Oxocarbenium Ion and the Glycosidic Bond

*Yuying Li[1†], Yuan Yao[2†], Guosong Yang[3], Jun Tang[1], Gabriela Jaramillo Ayala[1], Xumin Li[1], Wenlu Zhang[1], Qiuyu Han[1], Tong Yang[1], Hao Wang[1], Kevin H. Mayo[4] and Jiyong Su[1]**

[1] Engineering Research Center of Glycoconjugates Ministry of Education, Jilin Provincial Key Laboratory of Chemistry and Biology of Changbai Mountain Natural Drugs, School of Life Sciences, Northeast Normal University, Changchun, China, [2] Media Academy, Jilin Engineering Normal University, Changchun, China, [3] Zhongke Biopharm Co., Ltd, Beijing, China, [4] Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, United States

In green species, sucrose can help antagonize abiotic stress. Sucrose phosphate synthase (SPS) is a well-known rate-limiting enzyme in the synthesis of sucrose. To date, however, there is no known crystal structure of SPS from plant or cyanobacteria. In this study, we report the first co-crystal structure of SPS from *Thermosynechococcus elongatus* with UDP and sucrose-6-phosphate (S6P). Within the catalytic site, the side chains of His158 and Glu331, along with two phosphate groups from UDP, form hydrogen bonds with the four hydroxyl groups of the glucose moiety in S6P. This association causes these four hydroxyl groups to become partially negatively charged, thus promoting formation of the C1 oxocarbenium ion. Breakage of the hydrogen bond between His158 and one of the hydroxyl groups may trigger covalent bond formation between the C1 oxocarbenium ion and the C2 hydroxyl of fructose-6-phosphate. Consistent with our structural model, we observed that two SPS mutants, H158A and E331A, lost all catalytic activity. Moreover, electron density of residues from two loops (loop1 and loop2) in the SPS A-domain was not observed, suggest their dynamic nature. B-factor analysis and molecular dynamics stimulations of the full-length enzyme and A-domain indicate that both loops are crucial for binding and release of substrate and product. In addition, temperature gradient analysis shows that SPS exhibits its highest activity at 70°C, suggesting that this enzyme has the potential of being used in industrial production of S6P.

Keywords: sucrose phosphate synthase, *Thermosynechococcus elongatus*, UDP, sucrose-6-phosphate, oxocarbenium ion, glycosidic bond, catalysis mechanism

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Alberto A. Iglesias,
National University of the Littoral, Argentina
Tom Desmet,
Ghent University, Belgium
Quan Luo,
Hubei University, China

***Correspondence:**
Jiyong Su
sujy100@nenu.edu.cn

[†]*These authors share first authorship*

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *02 January 2020*
Accepted: *28 April 2020*
Published: *26 May 2020*

Citation:
Li Y, Yao Y, Yang G, Tang J, Ayala GJ, Li X, Zhang W, Han Q, Yang T, Wang H, Mayo KH and Su J (2020) Co-crystal Structure of Thermosynechococcus elongatus Sucrose Phosphate Synthase With UDP and Sucrose-6-Phosphate Provides Insight Into Its Mechanism of Action Involving an Oxocarbenium Ion and the Glycosidic Bond. Front. Microbiol. 11:1050. doi: 10.3389/fmicb.2020.01050

INTRODUCTION

Sucrose is primarily synthesized in photosynthetic organisms, including cyanobacteria, plants, and some algae (Lunn, 2002; Salerno and Curatti, 2003; Wind et al., 2010), although it is also metabolized in non-photosynthetic chemolitho-autotrophic organisms (Chain et al., 2003), e.g., *Nitrosomonas europaea* (Wu et al., 2015). Following photosynthesis, chlorophyll-containing organisms store carbon (CO_2) and reducing energy (i.e., NADPH) in sucrose via the Calvin cycle (Angermayr et al., 2015). The main enzymes used for sucrose synthesis in cyanobacteria and plants are sucrose phosphate synthase (SPS) and sucrose phosphate phosphatase (SPP) (Winter and Huber, 2000; Maloney et al., 2015). SPS catalyzes sucrose-6-phosphate (S6P) synthesis by using UDP-glucose and fructose-6-phosphate (Chua et al., 2008). SPP removes the phosphate group from sucrose-6-phosphate (Chua et al., 2008), whereas SPS catalysis is the rate limiting step for sucrose synthesis (Rufty and Huber, 1983). The catalytic efficiency of SPS and the amount of this enzyme determine the abundance of sucrose in these organisms (Rufty and Huber, 1983; Weiner et al., 1992).

Phylogenetic analysis indicates that the evolution of sucrose biosynthesis-related enzymes in modern cyanobacteria and plants arises from a common ancestral SPS-like gene (Cumino et al., 2002). In plants and several cyanobacteria (e.g., *Synechocystis* sp. PCC 6803), SPS and SPP are fused together containing regulatory domains within their N- and C-termini (Curatti et al., 1998). However, a bioinformatics study shows that SPS and SPP of *Anabaena* sp. PCC 7120 are not fused and define minimal catalytic units in almost all cyanobacteria (Cumino et al., 2002). All SPSs identified so far belong to the GT-B type glucosyltransferase family and contain two Rossmann-type folds (Lairson et al., 2008). The N-terminal fold is called the A-domain and the C-terminal fold is called the B-domain (Chua et al., 2008).

Sucrose phosphate synthase (SPSs) are highly expressed in plants and can easily be purified from spinach, soybean, and tobacco (Amir and Preiss, 1982; Huber et al., 1984). Early studies of plant SPSs showed that plants can regulate SPSs activity related to diurnal rhythmic changes (Huber et al., 1989; Huber and Huber, 1992). Further studies demonstrated that diurnal regulation is correlated to the phosphorylation state of SPSs (Huber et al., 1989), which can be mediated by various kinases (Huber and Huber, 1991; McMichael et al., 1995) at many sites (Huber et al., 1989; Huber and Huber, 1990), in particular at Ser158 (Toroser et al., 1999). These sites can also be dephosphorylated by protein phosphatase 2A (Siegl et al., 1990). The activity of phosphorylated and dephosphorylated SPS is thereby either inhibited or activated, respectively. Aside from the regulatory effects of phosphorylation (pi), inorganic phosphate can also inhibit SPS activity (Amir and Preiss, 1982; Doehlert and Huber, 1983). Mechanistically, Pi-mediated inhibition is proposed to occur via direct binding to the SPS catalytic site. The regulatory role of phosphorylation and the inhibitory effect of Pi have been proposed as "fine" and "coarse" control of SPS light activation (Weiner et al., 1992). The rapid activation of SPS by light involves cytosolic Pi being transferred to chloroplasts and activation of protein phosphatase 2A by a novel mechanism that may involve (directly or indirectly) a step in protein synthesis.

When cyanobacteria and plants undergo abiotic stress, such as due to the presence of salt (Hershkovitz et al., 1991) and low temperature (Guy et al., 1992), SPS expression is usually upregulated to increase sucrose production. This indicates that these species require sucrose to stabilize proteins and/or membrane structure and function (Lunn, 2002). The high production of sucrose from chlorophyll-containing species has considerable economic value. Genetic engineering of SPS has already been performed in plants (Haigler et al., 2007; Chua et al., 2008; Seger et al., 2015) and cyanobacteria (Wind et al., 2010; Du et al., 2013). Overexpression of SPS in these species could significantly increase the production of sucrose that then could be directly fermented to biofuel (Wind et al., 2010). In this regard, the study of SPS is warranted. However, until now, only the crystal structure of SPS from *Halothermothrix orenii* (HoSPS) has been solved (Chua et al., 2008), showing that this enzyme adopts a typical GT-B fold (Lairson et al., 2008). Comparision of this SPS to those of *Agrobacterium tumefaciens* glycogen synthase-ADP and *E. coli* trehalose-6-phosphate synthase (OtsA)-glucose-6-phosphate (G6P)-UDP complexes indicates that the HoSPS structure adopts a catalytically open form (Gibson et al., 2002; Buschiazzo et al., 2004; Chua et al., 2008). However, the actual catalytic mechanism of SPS remains unclear. On the other hand, the catalytic mechanism of OtsA has been experimentally revealed (Lee et al., 2011). In that model, the OtsA catalytic reaction occurs via an S_Ni mechanism in which a covalent bond between UDP and glucose is broken and one between glucose and G6P is formed (Lee et al., 2011). During this process, a oxocarbenium ion in the glucose residue exists in a transient state. This model nicely explains the catalytic process of glucosyltransferase. However, various details in this model remain unknown, for example how the oxocarbenium ion and new glycosidic bond are formed.

Thermosynechococcus elongatus is a genetically transformable rod-shaped cyanobacterium (Iwai et al., 2004). The most suitable growth temperature for this cyanobacterium is 57°C, which is suitable for industrial use (Yamaoka et al., 1978). In the present study, we solved the co-crystal structure of *Thermosynechococcus elongatus* SPS (TeSPS) (Uniprot code: tll1590) with S6P and UDP. The structure of the A-domain was also solved. Mass spectrometry indicates that TeSPS is an active enzyme that can synthesize S6P from fructose-6-phospahate (F6P) and uridine diphosphate glucose (UDPG). SPP from *Synechocystis* sp. PCC 6803 (Fieulaine et al., 2005), included in the TLC study, can specifically hydrolyze the phosphate group from S6P and produce sucrose. We also generated mutants within the highly conserved catalytic center of TeSPS, and determined their activities. Based on our findings, we propose a catalytic mechanism for TeSPS. Our model provides

clues for utilizing TeSPS in the over-production of sucrose in various species.

MATERIALS AND METHODS

Cloning, Protein Expression, and Purification

The TeSPS gene (Uniprot code: tll1590) was synthesized by SynBio Technologies (Monmouth Junction, United States), and amplified using primers (forward: 5'-CATATGCAAGCACTGAGTACC-3', reverse: 5'-CTCGAGTTAACTTGCTAATGCTGCTTT-3') that contain *Nde*I and *Xho*I restriction sites. PCR products were digested and cloned into a pET28a vector (Novagen, Gibbstown, NJ, United States). The procedure used for site-directed mutagenesis of TeSPS was performed by using the manual of the QuickChange XL site-directed mutagenesis kit (Stratagene, La Jolla, Canada). PCR products of two A-domains (residues 27 to 220, and residues 27 to 220 plus 406 to 426) and B-domain (residues 221 to 405) were also digested and cloned into the pET28a vector (Novagen, Gibbstown, NJ, United States). All constructs were checked by DNA sequencing. The TeSPS construct and the mutants were transformed into *E. coli* BL21 (DE3) cells and plated on LB agar plates supplemented with 100 μg/ml kanamycin. After overnight culture, several *E.coli* colonies were scraped from the LB agar plates and transferred into a 10 ml LB medium containing 100 μg/ml kanamycin. The culture was shaken at 37°C for 16 h. During the following day, LB medium-containing *E.coli* cells were transferred to 1 L of LB medium and shaken at 37°C. When the optical density (OD600) of these cultures reached 1.2–1.5, IPTG was added to a final concentration of 0.5 mM to induce protein over-expression. After induction at 37°C overnight, cells were harvested by centrifugation (6000 *g* for 15 min) and lysed by sonification in a lysis buffer consisting of 10 mM Tris/HCl, pH 8.0, 150 mM NaCl, 20 mM imidazole. The clarified cell extract was used for protein purification on a Ni-NTA Agarose column (Qiagen, Hilden, Germany). After purification, the His-tagged protein was dialyzed in 10 mM Tris/HCl, pH 7.5, 150 mM NaCl, and thrombin (20 units/mg protein; units defined by the National Institutes of Health) was added to remove the His tag. SDS-PAGE showed that all proteins were >90% pure. Proteins were concentrated to approximately 10 mg/ml and stored at −80°C.

The SPP gene (Uniprot code: Q7BII3) was synthesized by SynBio Technologies (Monmouth Junction, United States), and amplified using primers (forward: 5'-CATATGCGTCAGCTGCTGCTG-3', reverse: 5'-CTCGAGTTAGCTCAGAAAATCAAAATG-3') that contain *Nde*I and *Xho*I restriction sites. PCR products were digested and cloned into a pET28a vector (Novagen, Gibbstown, United States). Expression and purification of SPP was the same as that for TeSPS. After purification, the His-tagged protein was dialyzed in 10 mM Tris/HCl, pH 7.5, 150 mM NaCl. As determined by SDS-PAGE, all protein purities were > 90%. Proteins were concentrated to approximately 10 mg/ml and stored at −80°C.

Crystallization, Data Collection, and Structure Determination

Thermosynechococcus elongatus SPS crystals were obtained between 7 and 14 days from hanging drops that contained 1 μl protein and 1 μl solution from the well containing 0.1M sodium citrate, pH 7.0, 10% isopropanol, 0.01M UDP (Sangon, Shanghai, China), 10% PEG 10000 at room temperature. Crystals of the TeSPS A-domain (27-220_406-426) were obtained between 1 and 4 days from hanging drops that contained 1 μl protein and 1 μl solution from the well containing 0.1M sodium acetate, pH 4.6, 0.5 M potassium thiocyanate at room temperature. Prior to X-ray data collection, TeSPS crystals were soaked for 5 min in the reservoir solution supplemented with 10 mM S6P (Sigma, Shanghai, China). 20% (v/v) glycerol was used as the cryoprotectant. Crystals were flash cooled in liquid nitrogen. Data sets were collected at 100 K at the Shanghai Synchrotron Radiation Facility (Shanghai, China).

Data sets were indexed and integrated using XDS (Kabsch, 2010) and scaled using Aimless (Evans and Murshudov, 2013) from the CCP4 software package (Potterton et al., 2003). Structures were determined by using the program Phaser (McCoy et al., 2007) and molecular replacement with the structure of glycosyltransferase MshA (PDB: 3C4Q) (Vetting et al., 2008) as the search model. Structure refinement and water updating were performed using Phenix (Adams et al., 2010) refine and manual adjustment. Final structure validations were performed using MolProbity (Davis et al., 2007; Chen et al., 2010). Figures for all structures were generated using Pymol[1].

Thin Layer Chromatography

The reaction solution for TeSPS contains 10 mM Tris–HCl, pH7.5, 10 mM F6P, 10 mM UDPG (Sigma, Shanghai, China), 2 μg TeSPS. The reaction was performed at 40°C and stopped by low temperature (−20°C). 2 μg SPP (Fieulaine et al., 2005) was added to the reaction solution to hydrolyze the phosphate group of S6P. After the reaction, 10 μl of the above solution was placed on a thin layer chromatography (TLC) plate. The developing agent for TLC contains n-butanol, acetone and water at the volume ratio of 4:3:1, respectively. Following TLC, plates were incubated with 2% aniline acetone solution, 2% diphenylamine acetone solution and 85% phosphoric acid at the volume ratio of 5:5:1, as the color developing agent. The plate was then heated at 85°C until the bands became clear.

Mass Spectrometry

After enzyme reactions were run, resulting compounds were confirmed by high resolution mass spectrometry (MS). The final solution was directly injected into an Q-Exactive MS instrument equipped with an electrospray ion source (Thermo Fisher Scientific, United States). MS data were acquired over the range of m/z 100–800 at a resolution of 70,000. MS was performed in

[1]https://pymol.org/2/

the negative ion mode and operated with following optimized parameters: spray voltage, 3.5 kV; capillary temperature, 320°C; sheath gas flow rate, 20 arbitrary units; aux gas flow rate, 2 arbitrary units; S-Lens RF level, 80%. MS analysis for each enzyme reaction was performed in triplicate.

Molecular Dynamics Stimulation

After removing water molecules, coot was used to add residues that were absent in TeSPS and A-domain structures. Both proteins were placed in the centers of cubic boxes. The distance between the box edges was 5 nm. Molecular dynamics (MD) simulations were performed using GROMACS (Berendsen et al., 1995) v.2019.3. The protein topology was defined with CHARMM36 parameters (Huang et al., 2017). TIP3P water molecules were added using *gmx solvate*. Cl^- or Na^+ were used to neutralize the system. CHARMM36 compatible parameters for the UDP and S6P were obtained using the CGenFF server[2] (Vanommeslaeghe et al., 2010; Yu et al., 2012). The structure was relaxed during energy minimization. After NVT and NPT equilibration, a 20 ns MD stimulation was performed using a time step of 2 fs with LINCS holonomic constraints on all bonds. The particle mesh Ewald (PME) algorithm was used for electrostatic interactions, with a cutoff of 1.2 nm. A single cutoff of 1.2 nm was used for van der Waals interactions. Temperature coupling was performed with the *v-rescale* algorithm. Following MD stimulations, RMSD values were analyzed by xmgrace[3].

RESULTS

Co-crystal Structure of TeSPS With UDP and S6P

Although crystallization of TeSPS alone was unsuccessful, addition of UDP did induce crystal growth, indicating that this ligand stabilized the protein for crystallization. Nevertheless, resolution of the co-crystal structure of TeSPS and UDP was only about 5 Å. Therefore, we decided to soak the TeSPS:UDP co-crystals with S6P in an attempt to stabilze the structure further and increase resolution. In doing so, we were able to obtain a dataset at 3 Å resolution. Apparently, the presence of S6P could stabilize various segments of TeSPS, thereby increasing structural uniformity within the crystals and improving resolution. This dataset allowed us to solve the co-crystal structure of TeSPS with these two ligands.

The crystal structure of TeSPS (residues 27–426) showed that it is a GT-B type glycotransferase as a monomer (**Figure 1**). The electron density of residues at the N- and C- termini was absent, indicating that these segments were either very flexible or were hydrolyzed by some proteases. Overall, our resolved structure showed that TeSPS has 16 α helices and 14 β sheets, with UDP and S6P being bound at the interface of A- and B-domains. Structural statistics are provided in **Table 1**.

The N-terminus (residues 27–220) and the α6 helix (residues 406–426) formed the A-domain, and the C-terminus (residues 221–405) formed the B-domain, with two loops (connecting α6 and α6' and α7' and β8, respectively) linking these two domains. Other GT-B type glycotransfereases also adopt similar dumbbell-shaped structures, with some of them adopting an open conformation and others having a closed conformation. In the open conformation, the catalytic interface is not formed (Buschiazzo et al., 2004), and represents the structural state pre- or post-catalysis. On the other hand, the closed conformation could represent the actual catalytic state of the enzyme (Gibson et al., 2002). In our structure, TeSPS is in the closed conformational state.

The primary structure of TeSPS is conserved compared to other cyanobacteria and plant SPSs (**Figure 2**), including An-SPS-A, An-SPS-B (Cumino et al., 2002), *H. orenil*-SPS (Chua et al., 2008), Spinach-SPS (Amir and Preiss, 1982), *Synechocystis*-SPS (Curatti et al., 1998). The number of amino acids in spinach- and *Synechocystis*-SPS are larger than that in TeSPS. Other results show that these two enzymes have functional SPP domains or

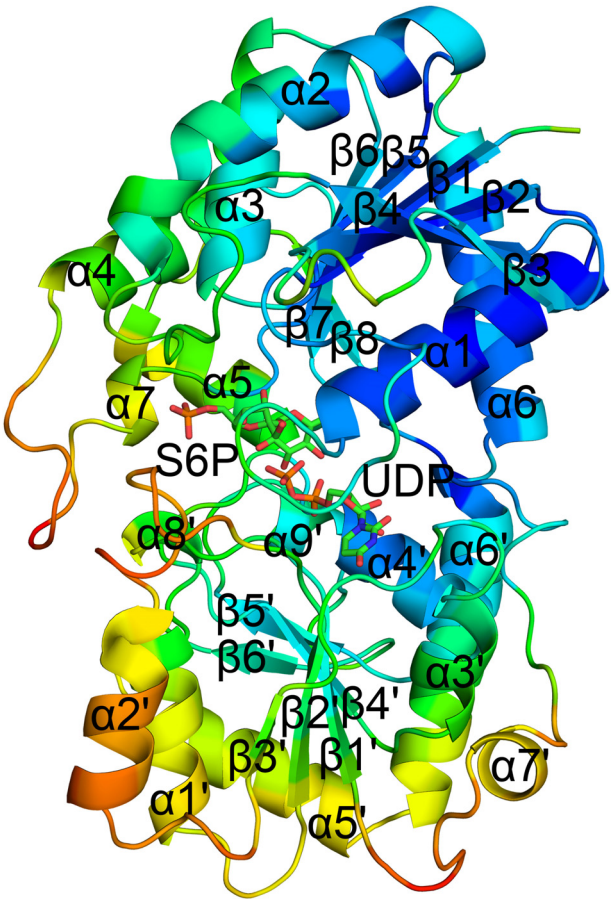


FIGURE 1 | Co-crystal structure of TeSPS with UDP and S6P. Here, we show that TeSPS adopts the closed conformation (Gibson et al., 2002) in its structure. TeSPS contains two Rossmann-type domains (A-domain and B-domain). Seven α helices (α1-7) and eight β strands (β1-8) constitute the A-domain. Nine α helices (α'1-9) and six β strands (β'1-6) constitute the B-domain.

[2]https://cgenff.umaryland.edu/

[3]http://plasma-gate.weizmann.ac.il/Grace/

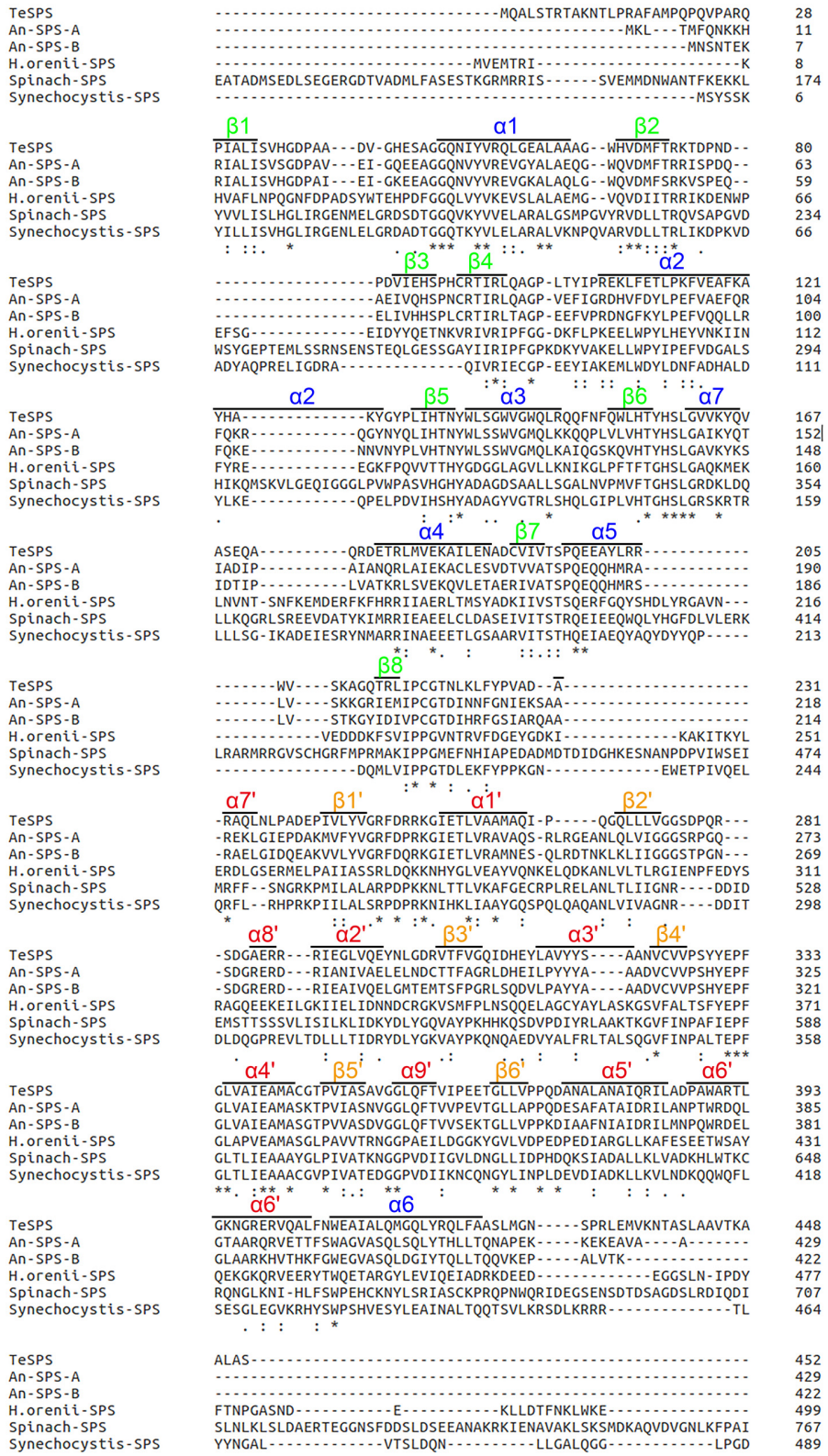


FIGURE 2 | Amino acid sequence comparisons. The amino acid sequence of TeSPS is aligned with those of An-SPS-A, An-SPS-B (Cumino et al., 2002), *H. orenil*-SPS (Chua et al., 2008), spinach-SPS (Amir and Preiss, 1982), and *Synechocystis*-SPS (Curatti et al., 1998). TeSPS is highly similar to An-SPS-A and An-SPS-B. Residues within quotes "." indicates highly conserved residues, ":" indicates those that are more conserved than ".", and residues with "*" indicate those that are most conserved.

domains crucial for binding other proteins (Toroser et al., 1998). The alignment also indicates that these SPSs contain many highly conserved residues that are important for folding and/or activity of these enzymes. To assess this, we mutated several conserved residues within the catalytic center of TeSPS to identify the roles that these residues play during catalysis.

Structure of the A-Domain

The spinach SPS A-domain was previously overexpressed in *E.coli* and purified (Salvucci and Klein, 1993). Mass spectrometry and HPLC showed that this domain is functional and could bind UDP and UDPG. In order to study the function of TeSPS A- and B- domains, we generated three truncated proteins. The first one was composed of residues 27 to 220 without the α6 helix. In the second one, the last α6 helix was switched to the C-terminus of the first one, with this truncated protein having the entire A-domain. The third protein was composed of the entire B-domain with residues 221 to 405. All three truncated proteins were easily purified from *E.coli*. However, only the second protein (residues 27–220 and 406–426) could be crystallized (**Figure 3A**) and resolved to 1.92 Å with structural statistics provided in **Table 1**. The other two truncated proteins could not be crystallized.

The A-domain could be merged perfectly with the A-domain of the full-length enzyme (**Figure 3B**) with an RMSD of 1.3 Å. This indicates that its folded structure is similar to the A-domain from the full-length enzyme. Unlike the first truncated protein, this one has the α6 helix at the C-terminus, implying that the α6 helix in this position is important for domain folding. Without this construction, the global fold of the first truncated protein could not be stabilized, which likely impeded crystallization. However, in the second truncated protein, electron densities of two loops (loop 1, residues 37-52, and loop 2, residues 160–173) were relatively poor and thus their structures could not be solved. In the full-length enzyme, loop 1 is key to binding UDP and loop 2 is close to S6P. Because of these interactions, the conformations of these two loops in TeSPS are stabilized, allowing their structures to be solved. In apo-TeSPS, these loops are likely flexible and appear to be important for enzyme-substrate binding.

B factor analysis of the full-length enzyme showed that the B-domain is more dynamic than the A-domain when UDP and S6P are bound to TeSPS (**Figure 4**). The flexibility of this domain likely impedes its crystallization. In contrast, even though the A-domain is relatively inflexible, loops 1 and 2 are flexible, and thus could not be resolved in the second truncated protein.

We used molecular dynamics (MD) stimulations for insight into this flexibility (**Figure 5**). The backbone RMSD of the full-length enzyme is about 2 Å, whereas the 3 Å RMSD of loop 2 is much higher, indicating that it is highly dynamic in solution. When loop 1 interacts with UDP, the RMSD value of this loop is only about 2 Å. MD analysis of the A-domain shows that RMSD values of loop 1 and loop 2 are higher than those in the full-length enzyme. The differential flexibility of loop 1 before and after substrate binding indicates that this loop is indeed crucial to substrate binding. Loop 2 (that is close to the S6P binding site) is always flexible, whether or not TeSPS is bound to substrate, suggesting that the flexibility of this loop in TeSPS may play a dual role in catalysis. One role is to bind substrate, and the other role is to release product. Aside from loop 2, the α3' helix in the B-domain is also flexible. Overall, loop 2 and this helix form a gate at the top of the catalytic site (**Figure 5A**). Fluctuations of these two segments, therefore, might play a role in opening the closed catalytic site, as well as release of products.

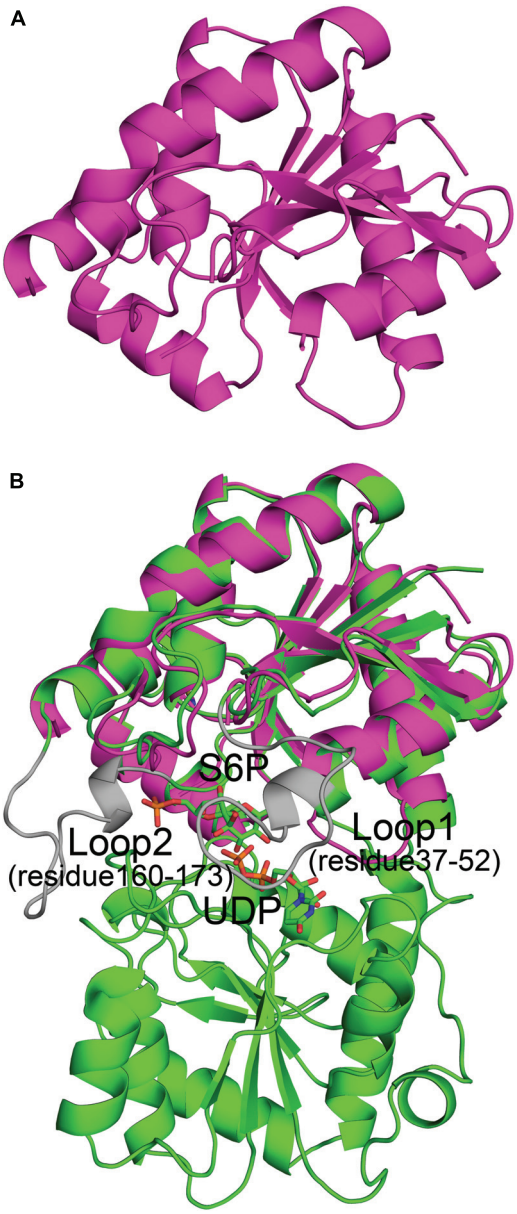


FIGURE 3 | Crystal structure of the A-domain (27-220_406-426). **(A)** The A-domain (27-220_406-426) adopts a Rossmann fold conformation. **(B)** The A-domain can be merged perfectly with the structure of the full-length enzyme. However, residues from loop 1 (residues 37–52) and loop 2 (residues 160–173) in the A-domain gave no electron density, and thus their structures could not be solved. This implies that these two loops are important for substrate binding or product release.

TABLE 1 | Data collection and refinement statistics.

PDB code	6KIH	6LDQ
Resolution (Å)	19.90–3.00 (3.06–3.00)	19.40–1.92 (1.97–1.92)
Space group	P1211	P1211
Unit cell parameters (a, b, c) (Å), (α, β, γ) (°)	(116.66, 170.84, and 160.55), (90.0, 96.43, and 90.0)	(50.02, 134.06, 60. and 79), (90.0, 90.8, and 90.0)
No. of measured reflections	411622 (19701)	202422 (13779)
No. of unique reflections	122717 (5683)	60298 (4059)
Completeness (%)	98.7 (92.2)	99.0 (99.2)
Multiplicity	3.4 (3.5)	3.4 (3.4)
R_{merge} (%)	11.9 (85.5)	11.8 (62.5)
$<I/\delta(I)>$	7.6 (1.2)	6.7 (1.8)
R_{model} (%)	27.1	24.2
R_{free} (%)	27.7	29.3
Rmsd bond lengths (Å)	0.01	0.01
Rmsd bond angles (°)	1.23	0.89
Ramachandran plot[f] residues in favored regions (%)	94	96
Substrate/Ligand	UDP and S6P	–

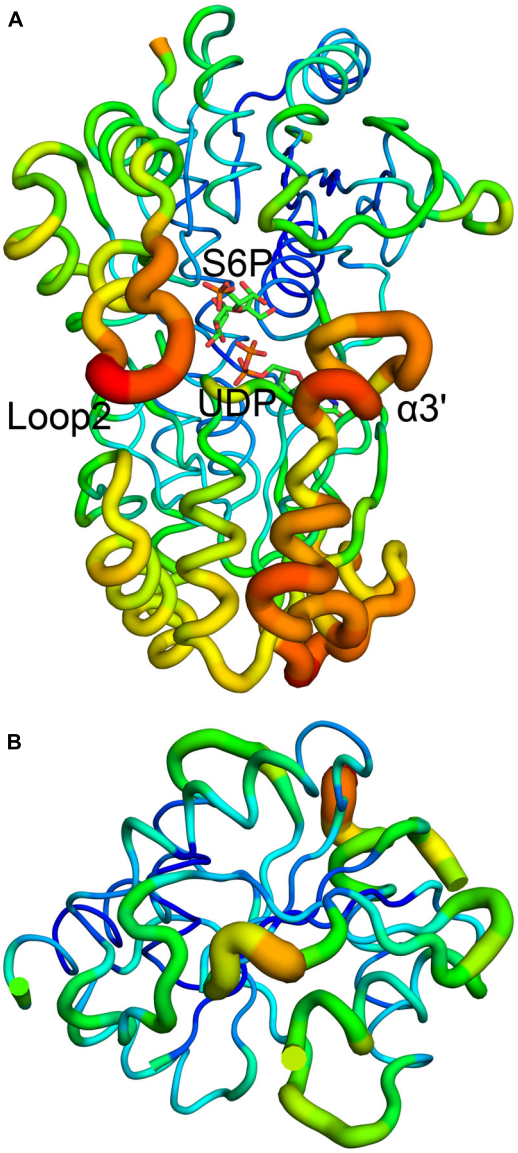


FIGURE 4 | B factor analysis of the full-length enzyme and A-domain. **(A)** Loop 2 and the α3' helix that form a gate at the top of the catalytic center, have relative high B factors. **(B)** B factor analysis of the A-domain indicates that with the exception of loops 1 and 2, the domain is relatively stable. Blue indicates low B factors, whereas green and yellow indicate mid-range B factors, and red indicates high B factors.

Catalytic Center

Although atomic resolution of the co-crystal structure of TeSPS with UDP and S6P is only 3 Å, the electron density map clearly shows that the two substrates are bound at the catalytic site (**Figure 6A**). The clear profiles of these two molecules allowed us to refine their structures within the catalytic center (**Figure 6B**). As already mentioned, addition of both substrates make the enzyme more compact and increases structural resolution. In particular, UDP and S6P stabilize conformations of the loops in the catalytic center in which the uracil moiety of UDP is inserted into a cavity formed by loops 1, 3, 4, and 5. Two hydroxyl groups of the UDP ribose moiety are stabilized by Glu339 (**Figure 7**) that is a highly conserved among glycosyltransferases. Hydrophobic residue Leu335 that is proximal to two phosphate groups from UDP is forced to reorient, and the terminal phosphate (P1) of UDP is stabilized by interactions with basic amino acids, including Arg249 and Arg253.

Previous studies demonstrated that inorganic phosphate inhibits SPS activity (Amir and Preiss, 1982; Doehlert and Huber, 1983). In the present study, our structure shows that the phosphate group of S6P is stabilized by numerous basic amino acid residues, including Arg105, Arg178, Arg249, and Arg253 (**Figure 7**). This indicates that inorganic phosphate may influence SPS binding to and/or release of F6P and S6P, and thus inhibits SPS catalytic activity.

In the catalytic center, His158 and Glu331 form hydrogen bonds with the 6-OH and 3-OH groups of glucose, respectively. In the following section, we mutated these residues for insight into their roles at the catalytic center. In addition, our co-crystal structure showed that P1 of UDP is close to the glycosidic bond between glucose and fructose rings of S6P. The distance between the oxygen atom of this phosphate group and the oxygen atom at the glycosidic bond is only 3.3 Å. In addition, the distances between oxygen atoms of two phosphate groups and the O2 and O4 groups from glucose are 2.4 and 2.6 Å, respectively, indicating that the hydrogen bonds formed by these groups are relatively strong.

Determination of SPS Activity

We used thin layer chromatography (TLC) to assess the activity of TeSPS (**Figure 8**). Because the phosphate groups of substrates and products (e.g., UDP, F6P, and S6P) are highly polar and negatively charged, these molecules remain mostly stationary

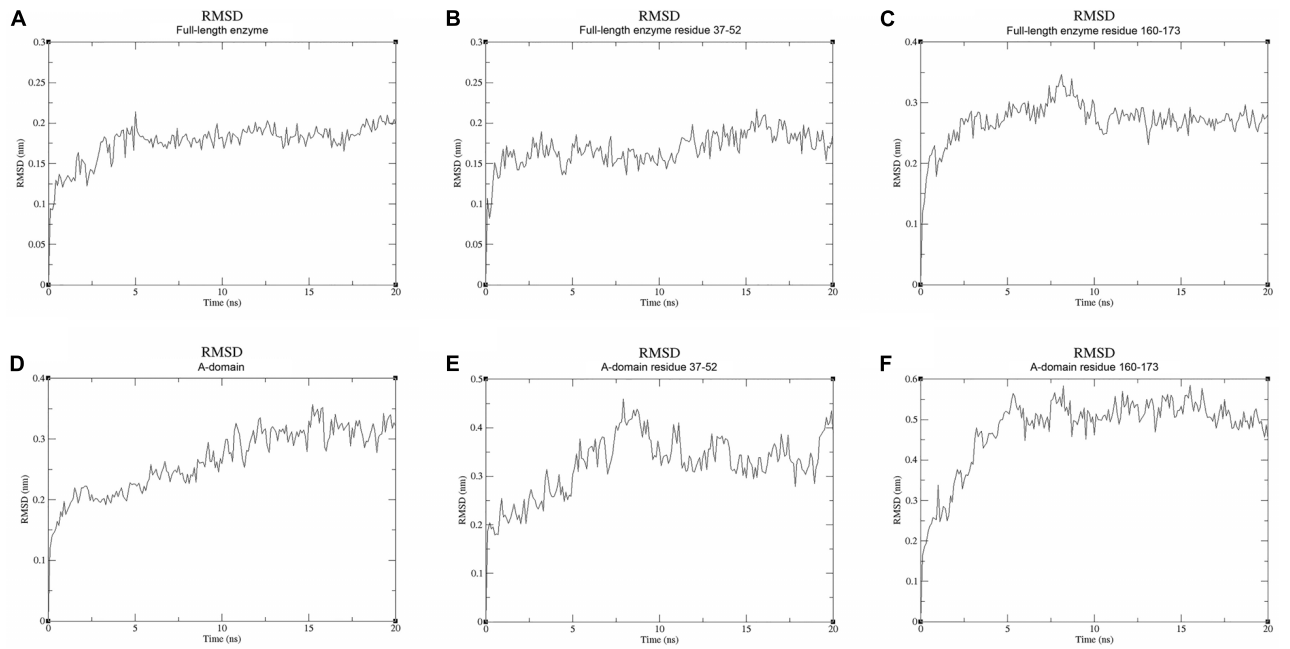


FIGURE 5 | Molecular dynamics stimulations of the full-length enzyme and A-domain. **(A)** The backbone RMSD value of the full-length enzyme is ∼2 Å. **(B)** The RMSD value for loop 1 of the full-length enzyme is also ∼2 Å. **(C)** The RMSD value of loop 2 is ∼3 Å. Because loop 1 is stabilized by UDP, the RMSD value of loop 1 is lower than that of loop 2. **(D)** The backbone RMSD value of the A-domain is ∼3 Å, a value that is greater than that in the full-length enzyme. Without stabilization of the B-domain and substrates, the A-domain is more flexible than the full-length enzyme. **(E,F)** High RMSD values for loops 1 and 2 indicate these two parts are highly flexible in the A-domain. This suggests that these both loops play a role in binding substrate and releasing product.

on the TLC plates (**Figures 8A,B**). In this regard, only sucrose could migrate on the TLC plate (lane1 in **Figure 8B**). TeSPS can catalyze the conversion from UDPG and F6P to UDP and S6P, respectively. Therefore, we used a specific SPP enzyme from *Synechocystis* sp. PCC6803 (Fieulaine et al., 2005) to specifically dephosphorylate S6P in order to observe sucrose on the TLC plate (**Figure 8C**). The amount of sucrose produced reflects the activity of TeSPS.

Mass spectrometry was used to identify S6P and sucrose following the reaction (**Figure 9**). These data demonstrated that S6P is present in solution following the reaction catalyzed by TeSPS. MS also showed that SPP hydrolyzed the phosphate group of S6P to produce sucrose. S6P itself was barely detected by MS after SPP hydrolysis. Therefore, the sucrose observed by TLC could be directly used to assess TeSPS activity. The negative control showed that with the inactivation of TeSPS, UDPG, and F6P could not be converted to UDP and S6P. Overall, our MS results demonstrate that TeSPS is a SPS enzyme, and SPP could specifically hydrolyze S6P to sucrose.

We also followed the time and temperature dependence of catalysis (**Figure 10**). The time dependence showed that a large amount of sucrose can be synthesized in a very short time, i.e., 30 s time scale. This indicates that the enzyme very quickly converts UDPG and F6P to UDP and S6P, respectively. The temperature dependence showed that the enzyme exhibits its greatest activity at 70°C, suggesting that it may be used for industrial production of S6P. Interestingly, TeSPS can also synthesize sucrose at low temperature (10°C), thus being able to protect *Thermosynechococcus elongates* from abiotic stress.

As mentioned above, we generated three truncated proteins, and mixed the B-domain and two A-domain proteins to test whether they had enzymatic activity (**Figures 11A,B**). However, the mixture of those proteins could not recover enzymatic activity, indicating that the two loops connecting the A- and B-domains in the full-length enzyme are crucial for maintaining TeSPS activity. In the absence of these two loops, the A- and B-domains are too free to form the catalytic interface.

Within the TeSPS catalytic center, Arg105 and Arg178 coordinate with the phosphate group of S6P via ionic bonds. Arg249 and Arg253 also play roles in stabilizing the terminal phosphate group of UDP. In order to study the functions of these basic residues during catalysis, we produced the mutants R105A, R178A, R249A, and R253A. In the enzymatic assay, all four variants exhibited lower activity compared to the wild type enzyme (**Figure 11C**). Because activity was not fully absent, it appears that these basic residues are not directly involved in catalysis, but may be involved in binding substrate or releasing product. In the native enzyme, Arg249 forms two ionic bonds with the terminal phosphate group of UDP, thus likely explaining why the R249A mutant displays the lowest activity among these variants. Without this residue, TeSPS cannot properly interact with UDP, thus leading to reduced activity.

We also mutated His158 and Glu331 to alanine to study their roles in catalysis. Both residues are highly conserved among SPSs (**Figure 2**). His158 forms a hydrogen bond with the 6-OH from

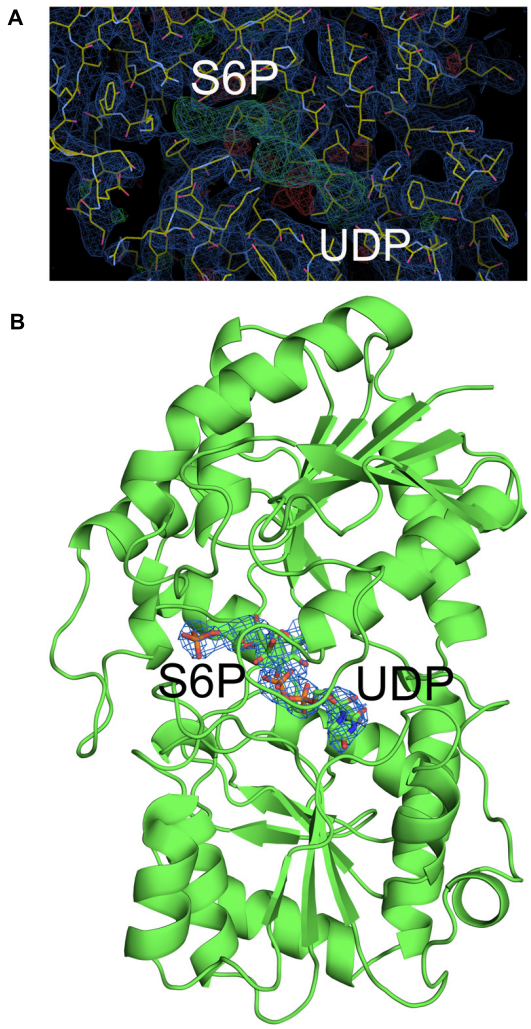


FIGURE 6 | Electron density map of the catalytic site of TeSPS. **(A)** The 2 | F_o|-|F_c|, αc map contoured at 1.0δ is shown in blue, and the | F_o| -| F_c|, αc map contoured at 3.0δ is shown in green. **(B)** Based on the | F_o| -| F_c| map, UDP and S6P were resolved within the catalytic center of TeSPS.

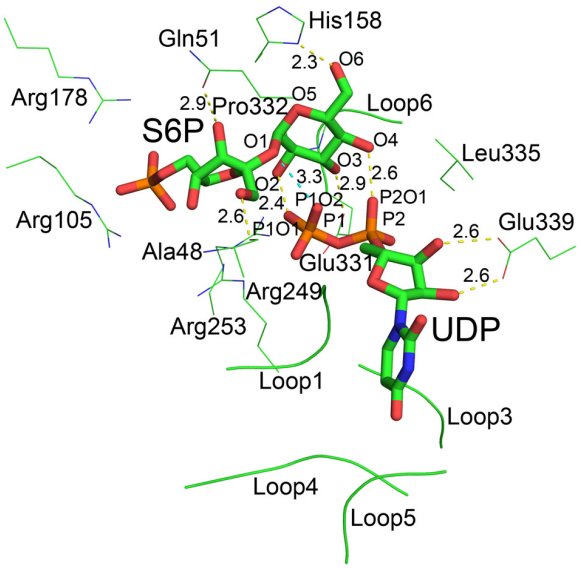


FIGURE 7 | The catalytic center of TeSPS. Loops 1, 3, 4, and 5 form a cave that binds to the uracil moiety of UDP. Glu339 stabilizes the ribose ring via formation of two hydrogen bonds. Leu335 forces two phosphate groups in UDP to reorient. Several basic amino acids, including Arg105, Arg178, Arg249, and Arg253, interact with the phosphate groups of UDP and S6P via ionic bonds. Pro332 at the turn of loop 6 interacts with the pyranose ring via CH/π bonds. All hydroxyl groups (O2, O3, O4, and O6) of the glucose moiety of S6P form hydrogen bonds with phosphate groups or the side chains of various amino acids. "O2" forms a hydrogen bond with P1O1 of the P1 phosphate group of UDP. "O3" forms a hydrogen bond with carboxyl group of Glu331. "O4" forms a hydrogen bond with P2O1 of P2 phosphate group of UDP. "O6" forms a hydrogen bond with the imidazole side chain of His158. The distances between groups are indicated in the figure.

the glucose residue of S6P, and Glu331 is hydrogen bonded to the 3-OH group of this glucose. Chemical modification of plant SPS already showed that a histidine residue is crucial to SPS catalytic activity (Sinha et al., 1998; Chua et al., 2008). Here, our enzymatic assay showed that H158A and E331A either had no activity in converting UDPG and F6P to UDP and S6P, or activity was too small to be determined in the TLC assay. Overall, His158 and Glu331 are critical for TeSPS activity.

DISCUSSION

In the present study, we identified TeSPS (Uniprot code: tll1590) from *T. elongatus* and demonstrated that the enzyme is biologically active. TeSPS is functional at higher temperatures, exhibiting its greatest activity at 70°C. The co-crystal structure of TeSPS complexed with UDP and S6P shows that the enzyme is very compact, likely explaining its resistance to unfolding/degradation at higher temperatures (Berezovsky and Shakhnovich, 2005). TeSPS has 8 Trp, 18 Tyr and 14 Phe residues, in contrast to An-SPS-A from *Anabaena* sp. PCC 7120 which only functions at normal, physiological temperatures. An-SPS-A has 54% identity to TeSPS, but only has 5 Trp, 14 Tyr and 15 Phe residues, with a generally lower number of hydrophobic core residues. In TeSPS, the higher number of hydrophobic residues likely contributes to stabilizing its structure and makes it more resistant to thermal denaturation (Taylor and Vaisman, 2010; Tsukamoto et al., 2016).

In plants, the sucrose synthesis pathway has been known for many years (Winter and Huber, 2000). However, only the structure of SPS from *Halothermothrix orenii* had been reported. The lack of other SPS structures may result from its relatively flexible dumbbell shape that may inhibit its crystallization. In order to crystallize TeSPS, we employed its ligands UDP and S6P to stabilize the protein structure and solve its crystal structure to a resolution at 3 Å. Because TeSPS is from a cyanobacteria, it is closely related to plant SPSs and can be used as a model to understand the function of plant SPSs.

Thermosynechococcus elongatus SPS has 452 residues, making it shorter than plant SPSs that have an addition N-terminal

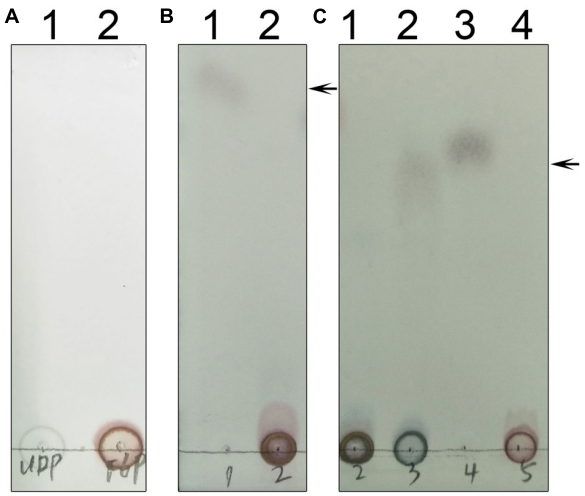


FIGURE 8 | TLC of UDP, F6P, S6P and sucrose. **(A)** Lane 1 and 2 indicate standard UDP and F6P that could not migrate on the TLC plate. **(B)** Lane 1 shows that sucrose could migrate on the TLC plate. An arrow indicates the sucrose band on the plate. Lane 2 shows standard S6P that could not migrate on the plate. **(C)** Lane 1 shows that S6P produced from the TeSPS reaction could not migrate on the plate. Lane 2 shows that SPP hydrolyzes S6P (same as lane 1) to sucrose that migrates on the plate. Lane 3 is the sucrose standard. Lane 4 shows that standard F6P could not migrate on the plate. An arrow indicates the sucrose band on the plate.

domain. Therefore, TeSPS may not be phosphorylated like plant SPSs which allows them to be regulated by diurnal rhythms. Moreover, the expression levels of plant SPSs can also regulate their activities. Similarly, it is likely that *T. elongatus* may also control TeSPS activity by regulating its expression and/or degradation. However, the exact mechanism required clarification.

Thermosynechococcus elongatus SPS is a kind of sucrose-phosphate synthase (EC 2.4.1.14). Based on The Carbohydrate-Active EnZymes database (CAZy) classification[4], TeSPS belongs to the glycosyltransferase family 4 and has a conserved glycogen phosphorylase GT (GPGTP) motif (Wrabl and Grishin, 2001). Almost all known glycosyltransferases have this motif that is formed primarily by helix 4 and the loop connecting helix 4 and strand 4, being referred to as positions 1 and 2, respectively (Wrabl and Grishin, 2001). This GPGTP has been proposed to be crucial for maintaining glycosyltransferase activity. Therefore, TeSPS is basically a glycosyltransferase that does not change the configuration of the anomeric carbon of glucose upon catalysis.

Initial reaction velocity studies of MshA indicate a sequential mechanism, with UDP-GlcNAc almost certainly binding first followed by the binding of 1-$_L$-*myo*-inositol-1-phosphate (Vetting et al., 2008). Our crystal structures and MD stimulations indicate that TeSPS also follows a sequential mechanism. In the

[4]http://www.cazy.org/

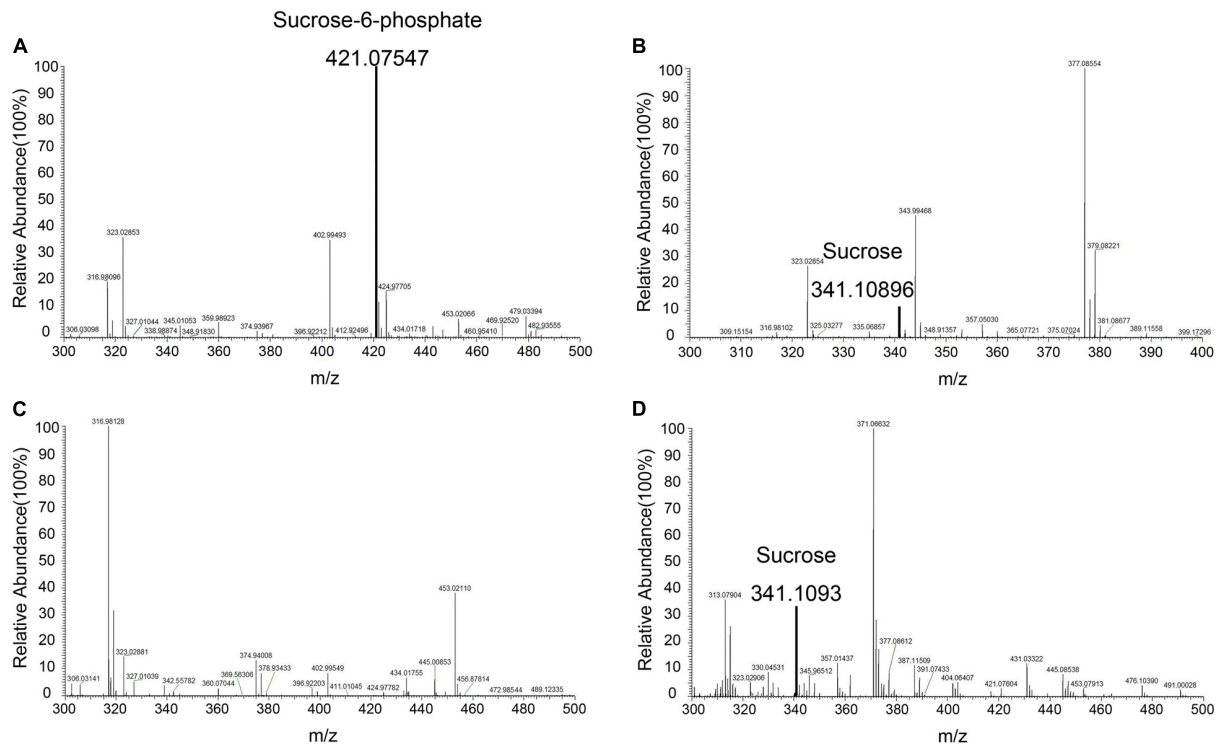


FIGURE 9 | Mass spectroscopy. **(A)** Mass spectroscopy of S6P produced by the enzyme reaction catalyzed by TeSPS. **(B)** Mass spectroscopy of sucrose that was produced from the coupled enzyme reaction catalyzed by TeSPS and SPP. This indicates that following TeSPS catalysis of UDPG and F6P to S6P, SPP could fully hydrolysis S6P to sucrose. **(C)** Prior to the enzyme reaction, TeSPS was inactivated. Mass spectroscopy indicates that S6P could not be produced. **(D)** sucrose standard.

absence of UDP binding, loop 1 was crystallographically invisible in the A-domain structure. However, upon UDP binding to the full-length enzyme, the B factors of loop 1 were reduced and RMSD values of that loop in MD stimulations were also lower than those of the full-length enzyme. Therefore, UDPG (or UDP) first binds to the interface of the A- and B-domains and promotes partial formation of the catalytic center via an induced fit mechanism. Consequently, basic residues at the gate of the catalytic center (**Figure 7**) captures the phosphate group of F6P, and Ala48 and Gln51 form hydrogen bonds with the hydroxyl groups of the fructose residue of F6P to stabilize binding.

Pro332 is located at "position 2" of the TeSPS GPGTP motif. In the TeSPS co-crystal structure, "position 2" forms a loop (loop 5). This proline residue interacts with the pyranose ring of glucose via CH/π bonds (Zondlo, 2013), which stabilizes binding of S6P to the enzyme. When UDPG binds to the catalytic site, Pro332 or the loop force the glucose residue of UDPG into that conformation and promote formation of the transition state to product. Although Pro332 is not conserved among known glycosyltransferases (Wrabl and Grishin, 2001), other residues within that loop may also play the same role as this proline.

Acidic residues asparate and glutamate (Glu331 in TeSPS) are highly conserved within the "position 2" GPGTP motif of glycosyltransferases (Wrabl and Grishin, 2001). In this study, therefore, we mutated Glu331 to alanine and obtained an E331A variant. This mutation totally abolished TeSPS catalytic activity, a result that is consistent with studies on other glycosyltransferases. In *Acetobacter xylinium* mannosyltransferase AceA, mutation of Glu287 (conserved like Glu331 in TeSPS) at "position 2" in the GPGTP motif also causes the enzyme to lose activity (Abdian et al., 2000). In addition, mutation of Glu510 at "position 2" (E510A) in human muscle glycogen synthase also inactivates the enzyme (Cid et al., 2000). In our co-crystal structure, Glu331 directly coordinates with the 3-OH group of the S6P glucose moiety, suggesting that this coordination is important for catalysis.

The crystal structures of known glycosyltransferases indicate that there is always a histidine residue coordinating the 6-OH of the monomeric sugar residues (Wrabl and Grishin, 2001; Gibson et al., 2002; Buschiazzo et al., 2004; Chua et al., 2008). A previous report showed that chemical modification of histidine residues of plant SPSs abolishes activity (Sinha et al., 1998), indicating that this conserved histidine is directly involved in catalysis. Here, we mutated the conserved histidine (His158) to alanine, and showed that H158A has no catalytic activity (**Figure 11C**), indicating that the hydrogen bond formed between His158 and the 6-OH group of the S6P glucose is important for catalysis. Aside from the 3-OH and 6-OH groups forming hydrogen bonds with Glu331 and His158, the 2-OH and 4-OH groups of glucose form strong hydrogen bonds with oxygen atoms of the two phosphate groups (P1 and P2) from UDP, respectively, indicating that the two phosphate groups are also important for catalysis. This in turn implies that UDPG could assist the glucose residue in entering the catalytic transition state upon binding.

An S_Ni (substitution nucleophilic internal)-like catalytic mechanism for *Neisseria menignitidis* glycosyltransferase has been proposed based on its co-crystal structure with acceptor and donor substrate analogs (Lee et al., 2011). The co-crystal structures of two known glycosyltransferases (OtsA and MshA) with substrates support this mechanism (Gibson et al., 2002; Buschiazzo et al., 2004). In addition, free energy relationships confirm that the inhibitors of OtsA are synergistic transition state mimics that support front-to-face nucleophilic attack involving hydrogen bonds between the leaving group (donor or UDPG) and nucleophile (acceptor or G6P). Kinetic isotope effects of donor and acceptor substrates of OtsA indicate a highly dissociative oxocarbenium ion-like transition state (Lee et al., 2011). Our co-crystal structure of TeSPS with UDP and

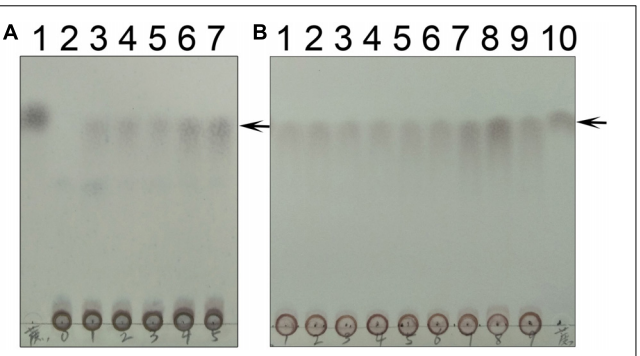


FIGURE 10 | Time and temperature dependence. **(A)** The time dependence is shown. Lane 1 is the sucrose standard. Lanes 2–7 show reaction times of 0, 10 s, 30 s, 1 min, 10 min, and 30 min. **(B)** The temperature dependence is shown. Lanes 1–9 show results of reaction temperatures of 0, 10, 20, 30, 40, 50, 60, 70, and 80°C. TeSPS has the highest activity at 70°C. Lane 10 shows the sucrose standard. The arrows indicate the sucrose bands on the plate.

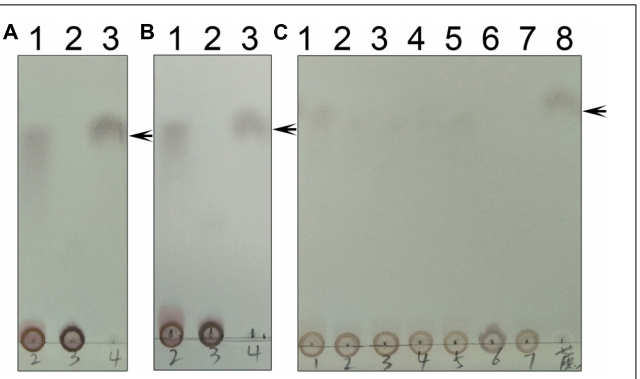


FIGURE 11 | Enzyme assays of the mutants. **(A)** Lane 1 shows the catalytic reaction of wild-type TeSPS where sucrose could be produced. Lane 2 shows results for the mixture of the A-domain (residues 27–220) and B-domain (residues 221-405) that could not catalyze the reaction. Lane 3 is for the sucrose standard. **(B)** Lane 1 shows the catalytic reaction of wild-type TeSPS where sucrose could be produced. Lane 2 shows results of the mixture of the A-domain (residues 27-220_4-6-426) and B-domain (residues 221–405) that could not catalyze the reaction. Lane 3 shows the sucrose standard. **(C)** Lane 1 shows the catalytic reaction of wild-type TeSPS. Lanes 2–7 show results of the reaction catalyzed by mutants R105A, R178A, R249A, R253A, H158A, and E331A, respectively. R105A, R178A, R249A, and R253A show some activity, whereas H158A and E331A could not catalyze the reaction. Lane 8 is for the sucrose standard. The arrows indicate the sucrose bands on the plate.

S6P is consistent with this S_Ni catalytic mechanism. However, how the oxocarbenium ion is formed remains unclear. Based on the hydrogen bonding network with the S6P glucose residue in the catalytic center of TeSPS, we proposed a model to explain the generation of the oxocarbenium ion and formation of the covalent bond between F6P and this glucose residue (**Figure 12**).

As mentioned previously, the glucose hydroxyl groups are fully coordinated, forming hydrogen bonds with His158, Glu331, and the oxygen atoms of the phosphates. Formation of these strong hydrogen bonds induce the oxygen atoms of the hydroxyl groups to become partially negatively charged. The glucose pyranose ring might share those partially negative charges just like a peptide bond. Because of this effect, the covalent bond formed between UDP and the glucose residue is likely broken, thus allowing the oxocarbenium ion to form. The positively charged oxocarbenium ion could neutralize these partially negative charges via resonance effects within the pyranose ring. In the following steps, the oxocarbenium ion, the oxygen atom of UDP, and the hydrogen of F6P would form a catalytic triad, as proposed by Seung et al. (Lee et al., 2011). We hypothesize that dissociation of the hydrogen bond between His158 and the glucose 6-OH group triggers formation of the glycosidic bond between F6P and glucose. The hydrogen bond formed by His158 and this 6-OH group is weaker than the hydrogen bonds formed between Glu331 and the phosphate and hydroxyl groups. Glu331 and these phosphate groups are fully negatively

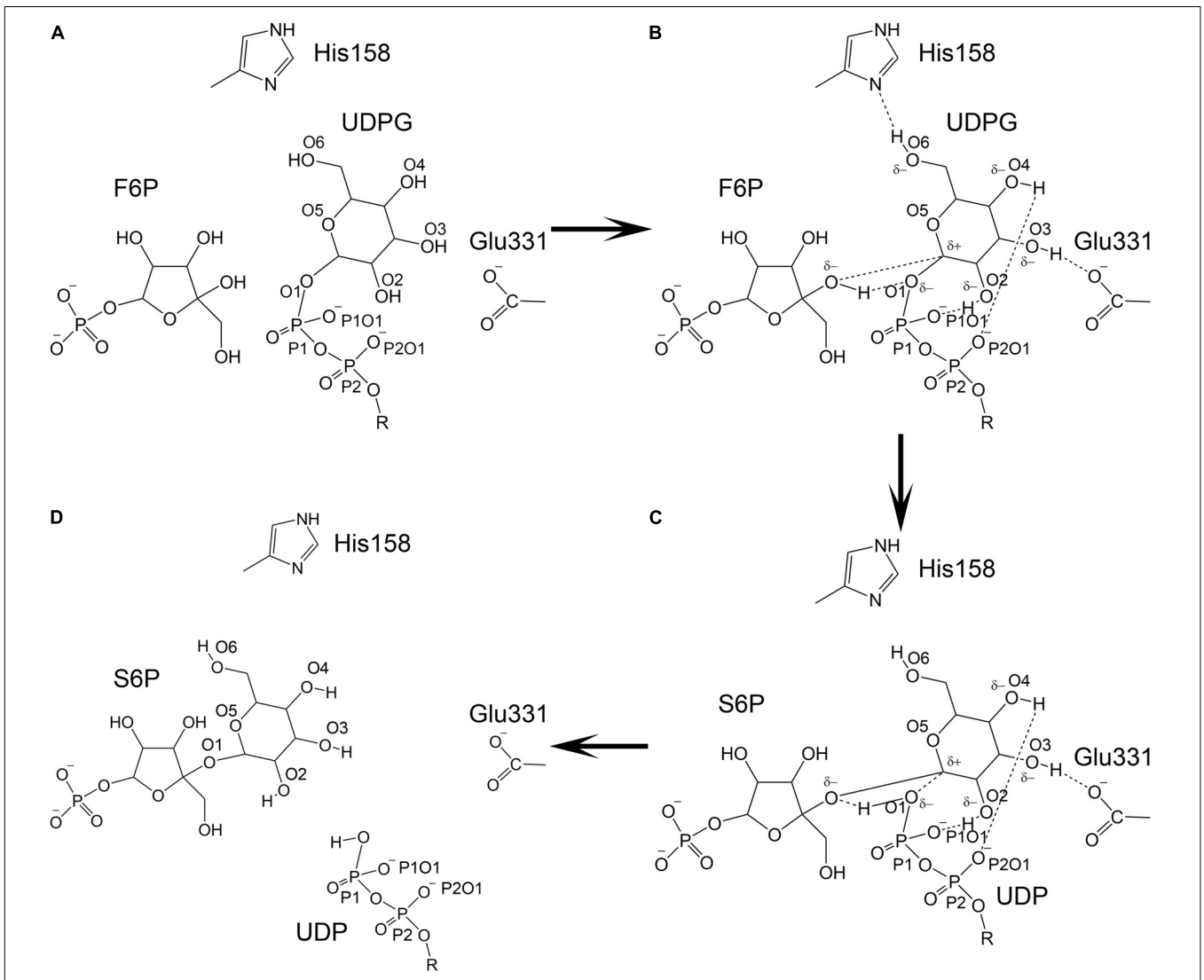


FIGURE 12 | Catalytic model of TeSPS. **(A)** The state prior to the reaction is shown. **(B)** The glucose residue of UDPG forms hydrogen bonds between/among the phosphate groups, His158, Glu331, and F6P. Due to formation of these hydrogen bonds, the pyranose ring of the glucose becomes negatively charged to promote C1 to form an oxocarbenium ion. **(C)** The relatively weak hydrogen bond formed by His158 and O6 is broken, which causes the pyranose ring to loose some negative charge character and force the C1 oxocarbenium ion to form a covalent bond with the F6P oxygen atom. **(D)** UDP and S6P are released from the catalytic center.

charged, whereas His158 can easily acquire or loose an electron. Therefore, the hydrogen bond formed by this residue would not be stable. Moreover, the 6-OH group is within the flexible part of the hexose ring. In many co-crystal structures of hexose-bound proteins, the 6-OH group is often not observed (Su et al., 2015; Si et al., 2016). Therefore, the hydrogen bond between His158 and the 6-OH group could be broken. If this occurs, then the pyranose ring would have less negative charge and could not neutralize the positive charge on the oxocarbenium ion. This in turn would force the oxocarbenium ion to find another negatively charged atom in order to form a covalent bond. At that point, the hydrogen atom of F6P would be essentially captured by the oxygen atom of the UDP phosphate group, and the oxocarbenium ion could quickly form a new covalent bond with the F6P oxygen atom. Overall, it is the fluctuation of the charge on His158 and the flexibility of the 6-OH group of glucose that triggers formation of S6P.

CONCLUSION

In conclusion, we structurally characterized SPS from *T. elongatus*. Furthermore, because this SPS retains activity at 70°C, it may be useful for the industrial production of S6P, as well as for possibly increasing crop production for farmers. Based on our co-crystal structure of ligand-bound TeSPS, we proposed a model for the catalytic mechanism of action. In the *T. elongatus* genome, another protein (Uniprot code: Q8DLB4) has also been predicted to be a SPS (Nakamura et al., 2002). This protein contains 716 amino acids and exhibits high sequence identity to a functionally characterized SPS of *Synechococcus elongatus* PCC 7942 (Martinez-Noel et al., 2013). If this protein were biological active, then the question as to which enzyme is the main SPS of *T. elongatus* needs to be addressed.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Protein Data Bank with accession numbers 6KIH and 6LDQ.

AUTHOR CONTRIBUTIONS

YL, YY, and GY participated in most experiments, including protein overexpression, TLC, protein crystallization, and solving the structures. JT and WZ ran the SDS-PAGE and performed the enzyme assay. TY, XL, GA, and QH participated in some experiments. HW performed the mass spectroscopy. JS conceived of the study and participated in its design and coordination. JS and KM analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by Science and Technology Project of Jilin Provincial Department of Education during the Thirteenth Five-Year Plan Period, China, grant number JJKH20190287KJ.

ACKNOWLEDGMENTS

We thank the staff from BL17B/BL18U/BL19U1/BL19U2/BL01B beamline of National Facility for Protein Science Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility, for assistance during data collection.

REFERENCES

Abdian, P. L., Lellouch, A. C., Gautier, C., Ielpi, L., and Geremia, R. A. (2000). Identification of essential amino acids in the bacterial alpha -mannosyltransferase aceA. *J. Biol. Chem.* 275, 40568–40575. doi: 10.1074/jbc.M007496200

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 2), 213–221. doi: 10.1107/S0907444909052925

Amir, J., and Preiss, J. (1982). Kinetic characterization of spinach leaf sucrose-phosphate synthase. *Plant Physiol.* 69, 1027–1030. doi: 10.1104/pp.69.5.1027

Angermayr, S. A., Gorchs Rovira, A., and Hellingwerf, K. J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol.* 33, 352–361. doi: 10.1016/j.tibtech.2015.03.009

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). Gromacs-a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* 91, 43–56. doi: 1 0.1016/0010-4655(95)00042-E

Berezovsky, I. N., and Shakhnovich, E. I. (2005). Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12742–12747. doi: 10.1073/pnas.0503890102

Buschiazzo, A., Ugalde, J. E., Guerin, M. E., Shepard, W., Ugalde, R. A., and Alzari, P. M. (2004). Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *EMBO J.* 23, 3196–3205. doi: 10.1038/sj.emboj.7600324

Chain, P., Kurtz, S., Ohlebusch, E., and Slezak, T. (2003). An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. *Brief. Bioinform.* 4, 105–123. doi: 10.1093/bib/4.2.105

Chen, V. B., Arendall, W. B. lll, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 1), 12–21. doi: 10.1107/S0907444909042073

Chua, T. K., Bujnicki, J. M., Tan, T. C., Huynh, F., Patel, B. K., and Sivaraman, J. (2008). The structure of sucrose phosphate synthase from Halothermothrix orenii reveals its mechanism of action and binding mode. *Plant Cell* 20, 1059–1072. doi: 10.1105/tpc.107.051193

Cid, E., Gomis, R. R., Geremia, R. A., Guinovart, J. J., and Ferrer, J. C. (2000). Identification of two essential glutamic acid residues in glycogen synthase. *J. Biol. Chem.* 275, 33614–33621. doi: 10.1074/jbc.M005358200

Cumino, A., Curatti, L., Giarrocco, L., and Salerno, G. L. (2002). Sucrose metabolism: anabaena sucrose-phosphate synthase and sucrose-phosphate phosphatase define minimal functional domains shuffled during evolution. *FEBS Lett.* 517, 19–23. doi: 10.1016/s0014-5793(02)02516-4

Curatti, L., Folco, E., Desplats, P., Abratti, G., Limones, V., Herrera-Estrella, L., et al. (1998). Sucrose-phosphate synthase from *Synechocystis* sp. strain PCC 6803: identification of the spsA gene and characterization of the enzyme expressed in *Escherichia coli*. *J. Bacteriol.* 180, 6776–6779.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383. doi: 10.1093/nar/gkm216

Doehlert, D. C., and Huber, S. C. (1983). Regulation of Spinach Leaf Sucrose Phosphate Synthase by Glucose-6-Phosphate, Inorganic Phosphate, and pH. *Plant Physiol.* 73, 989–994. doi: 10.1104/pp.73.4.989

Du, W., Liang, F., Duan, Y., Tan, X., and Lu, X. (2013). Exploring the photosynthetic production capacity of sucrose by cyanobacteria. *Metab. Eng.* 19, 17–25. doi: 10.1016/j.ymben.2013.05.001

Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* 69(Pt 7), 1204–1214. doi: 10.1107/S0907444913000061

Fieulaine, S., Lunn, J. E., Borel, F., and Ferrer, J. L. (2005). The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. *Plant Cell* 17, 2049–2058. doi: 10.1105/tpc.105.031229

Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R., and Davies, G. J. (2002). Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.* 9, 1337–1346. doi: 10.1016/s1074-5521(02)00292-2

Guy, C. L., Huber, J. L., and Huber, S. C. (1992). Sucrose phosphate synthase and sucrose accumulation at low temperature. *Plant Physiol.* 100, 502–508. doi: 10.1104/pp.100.1.502

Haigler, C. H., Singh, B., Zhang, D., Hwang, S., Wu, C., Cai, W. X., et al. (2007). Transgenic cotton over-producing spinach sucrose phosphate synthase showed enhanced leaf sucrose synthesis and improved fiber quality under controlled environmental conditions. *Plant Mol. Biol.* 63, 815–832. doi: 10.1007/s11103-006-9127-6

Hershkovitz, N., Oren, A., and Cohen, Y. (1991). Accumulation of trehalose and sucrose in cyanobacteria exposed to matric water stress. *Appl. Environ. Microbiol.* 57, 645–648.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., et al. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14, 71–73. doi: 10.1038/nmeth.4067

Huber, J. L., and Huber, S. C. (1992). Site-specific serine phosphorylation of spinach leaf sucrose-phosphate synthase. *Biochem. J.* 283(Pt 3), 877–882. doi: 10.1042/bj2830877

Huber, J. L., Huber, S. C., and Nielsen, T. H. (1989). Protein phosphorylation as a mechanism for regulation of spinach leaf sucrose-phosphate synthase activity. *Arch Biochem. Biophys.* 270, 681–690. doi: 10.1016/0003-9861(89)90551-1

Huber, S. C., and Huber, J. L. (1990). Activation of sucrose-phosphate synthase from darkened spinach leaves by an endogenous protein phosphatase. *Arch. Biochem. Biophys.* 282, 421–426. doi: 10.1016/0003-9861(90)90138-o

Huber, S. C., and Huber, J. L. (1991). In vitro phosphorylation and inactivation of spinach leaf sucrose-phosphate synthase by an endogenous protein kinase. *Biochim. Biophys. Acta* 1091, 393–400. doi: 10.1016/0167-4889(91)90205-c

Huber, S. C., Rufty, T. W., and Kerr, P. S. (1984). Effect of photoperiod on photosynthate partitioning and diurnal rhythms in sucrose phosphate synthase activity in leaves of soybean (Glycine max L. [Merr.]) and Tobacco (Nicotiana tabacum L.). *Plant Physiol.* 75, 1080–1084. doi: 10.1104/pp.75.4.1080

Iwai, M., Katoh, H., Katayama, M., and Ikeuchi, M. (2004). Improved genetic transformation of the thermophilic cyanobacterium, Thermosynechococcus elongatus BP-1. *Plant Cell Physiol.* 45, 171–175. doi: 10.1093/pcp/pch015

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 2), 125–132. doi: 10.1107/S0907444909047337

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi: 10.1146/annurev.biochem.76.061005.092322

Lee, S. S., Hong, S. Y., Errey, J. C., Izumi, A., Davies, G. J., and Davis, B. G. (2011). Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nat. Chem. Biol.* 7, 631–638. doi: 10.1038/nchembio.628

Lunn, J. E. (2002). Evolution of sucrose synthesis. *Plant Physiol.* 128, 1490–1500. doi: 10.1104/pp.010898

Maloney, V. J., Park, J. Y., Unda, F., and Mansfield, S. D. (2015). Sucrose phosphate synthase and sucrose phosphate phosphatase interact in planta and promote plant growth and biomass accumulation. *J. Exp. Bot.* 66, 4383–4394. doi: 10.1093/jxb/erv101

Martinez-Noel, G. M., Cumino, A. C., Kolman Mde, L., and Salerno, G. L. (2013). First evidence of sucrose biosynthesis by single cyanobacterial bimodular proteins. *FEBS Lett.* 587, 1669–1674. doi: 10.1016/j.febslet.2013.04.012

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40(Pt 4), 658–674. doi: 10.1107/S0021889807021206

McMichael, R. W. Jr., Bachmann, M., and Huber, S. C. (1995). Spinach leaf sucrose-phosphate synthase and nitrate reductase are phosphorylated/inactivated by multiple protein kinases in vitro. *Plant Physiol.* 108, 1077–1082. doi: 10.1104/pp.108.3.1077

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., et al. (2002). Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. *DNA Res.* 9, 123–130. doi: 10.1093/dnares/9.4.135

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* 59(Pt 7), 1131–1137. doi: 10.1107/S2059798317016035

Rufty, T. W., and Huber, S. C. (1983). Changes in starch formation and activities of sucrose phosphate synthase and cytoplasmic fructose-1,6-bisphosphatase in response to source-sink alterations. *Plant Physiol.* 72, 474–480. doi: 10.1104/pp.72.2.474

Salerno, G. L., and Curatti, L. (2003). Origin of sucrose metabolism in higher plants: when, how and why? *Trends Plant Sci.* 8, 63–69. doi: 10.1016/S1360-1385(02)00029-8

Salvucci, M. E., and Klein, R. R. (1993). Identification of the uridine-binding domain of sucrose-phosphate synthase. Expression of a region of the protein that photoaffinity labels with 5-azidouridine diphosphate-glucose. *Plant Physiol.* 102, 529–536. doi: 10.1104/pp.102.2.529

Seger, M., Gebril, S., Tabilona, J., Peel, A., and Sengupta-Gopalan, C. (2015). Impact of concurrent overexpression of cytosolic glutamine synthetase (GS1) and sucrose phosphate synthase (SPS) on growth and development in transgenic tobacco. *Planta* 241, 69–81. doi: 10.1007/s00425-014-2165-4

Si, Y., Wang, Y., Gao, J., Song, C., Feng, S., Zhou, Y., et al. (2016). Crystallization of galectin-8 linker reveals intricate relationship between the n-terminal tail and the linker. *Int. J. Mol. Sci.* 17:2088. doi: 10.3390/ijms17122088

Siegl, G., MacKintosh, C., and Stitt, M. (1990). Sucrose-phosphate synthase is dephosphorylated by protein phosphatase 2A in spinach leaves. Evidence from the effects of okadaic acid and microcystin. *FEBS Lett.* 270, 198–202. doi: 10.1016/0014-5793(90)81267-r

Sinha, A. K., Pathre, U. V., and Sane, P. V. (1998). Essential histidyl residues at the active site(s) of sucrose-phosphate synthase from *Prosopis juliflora*. *Biochim. Biophys. Acta* 1388, 397–404. doi: 10.1016/s0167-4838(98)00199-x

Su, J., Zhang, T., Wang, P., Liu, F., Tai, G., and Zhou, Y. (2015). The water network in galectin-3 ligand binding site guides inhibitor design. *Acta Biochim. Biophys. Sin. (Shanghai)* 47, 192–198. doi: 10.1093/abbs/gmu132

Taylor, T. J., and Vaisman, I. I. (2010). Discrimination of thermophilic and mesophilic proteins. *BMC Struct. Biol.* 10(Suppl. 1):S5. doi: 10.1186/1472-6807-10-S1-S5

Toroser, D., Athwal, G. S., and Huber, S. C. (1998). Site-specific regulatory interaction between spinach leaf sucrose-phosphate synthase and 14-3-3 proteins. *FEBS Lett.* 435, 110–114. doi: 10.1016/s0014-5793(98)01048-5

Toroser, D., McMichael, R. Jr., Krause, K. P., Kurreck, J., Sonnewald, U., Stitt, M., et al. (1999). Site-directed mutagenesis of serine 158 demonstrates its role in spinach leaf sucrose-phosphate synthase modulation. *Plant J.* 17, 407–413. doi: 10.1046/j.1365-313x.1999.00389.x

Tsukamoto, T., Mizutani, K., Hasegawa, T., Takahashi, M., Honda, N., Hashimoto, N., et al. (2016). X-ray crystallographic structure of thermophilic rhodopsin: implications for high thermal stability and optogenetic function. *J. Biol. Chem.* 291, 12223–12232. doi: 10.1074/jbc.M116.719815

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690. doi: 10.1002/jcc.21367

Vetting, M. W., Frantom, P. A., and Blanchard, J. S. (2008). Structural and enzymatic analysis of MshA from Corynebacterium glutamicum: substrate-assisted catalysis. *J. Biol. Chem.* 283, 15834–15844. doi: 10.1074/jbc.M801017200

Weiner, H., McMichael, R. W., and Huber, S. C. (1992). Identification of factors regulating the phosphorylation status of sucrose-phosphate synthase in vivo. *Plant Physiol.* 99, 1435–1442. doi: 10.1104/pp.99.4.1435

Wind, J., Smeekens, S., and Hanson, J. (2010). Sucrose: metabolite and signaling molecule. *Phytochemistry* 71, 1610–1614. doi: 10.1016/j.phytochem.2010.07.007

Winter, H., and Huber, S. C. (2000). Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit. Rev. Biochem. Mol. Biol.* 35, 253–289. doi: 10.1080/10409230008984165

Wrabl, J. O., and Grishin, N. V. (2001). Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J. Mol. Biol.* 314, 365–374. doi: 10.1006/jmbi.2001.5151

Wu, R., Asencion Diez, M. D., Figueroa, C. M., Machtey, M., Iglesias, A. A., Ballicora, M. A., et al. (2015). The crystal structure of nitrosomonas europaea sucrose synthase reveals critical conformational changes and insights into sucrose metabolism in prokaryotes. *J. Bacteriol.* 197, 2734–2746. doi: 10.1128/JB.00110-15

Yamaoka, T., Satoh, K., and Katoh, S. (1978). Photosynthetic activities of a thermophilic blue-green alga. *Plant Cell Physiol.* 19, 943–954. doi: 10.1093/oxfordjournals.pcp.a075684

Yu, W., He, X., Vanommeslaeghe, K., and MacKerell, A. D. Jr. (2012). Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* 33, 2451–2468. doi: 10.1002/jcc.23067

Zondlo, N. J. (2013). Aromatic-proline interactions: electronically tunable CH/pi interactions. *Acc. Chem. Res.* 46, 1039–1049. doi: 10.1021/ar300087y

Conflict of Interest: GY was employed by Zhongke Biopharm Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 June 2020
doi: 10.3389/fmicb.2020.00941

Cloning, Expression, and Structural Elucidation of a Biotechnologically Potential Alkaline Serine Protease From a Newly Isolated Haloalkaliphilic *Bacillus lehensis* JO-26

*Hitarth B. Bhatt and Satya P. Singh**

UGC-CAS Department of Biosciences, Saurashtra University, Rajkot, India

An alkaline protease gene of *Bacillus lehensis* JO-26 from saline desert, Little Rann of Kutch, was cloned and expressed in *Escherichia coli* BL21 (DE3). A 1,014-bp ORF encoded 337 amino acids. The recombinant protease (APrBL) with Asp 97, His 127, and Ser 280 forming catalytic triad belongs to the subtilase S8 protease family. The gene was optimally expressed in soluble fraction with 0.2 mM isopropyl β-D-thiogalactopyranoside (IPTG), 2% (w/v) NaCl at 28°C. APrBL, a monomer with a molecular mass of 34.6 kDa was active over pH 8–11 and 30°C–70°C, optimally at pH 10 and 50°C. The enzyme was highly thermostable and retained 73% of the residual activity at 80°C up to 3 h. It was significantly stimulated by sodium dodecyl sulfate (SDS), Ca^{2+}, chloroform, toluene, n-butanol, and benzene while completely inhibited by phenylmethylsulfonyl fluoride (PMSF) and Hg^{2+}. The serine nature of the protease was confirmed by its strong inhibition by PMSF. The APrBL gene was phylogenetically close to alkaline elastase YaB (P20724) and was distinct from the well-known commercial proteases subtilisin Carlsberg (CAB56500) and subtilisin BPN′ (P00782). The structural elucidation revealed 31.75% α-helices, 22.55% β-strands, and 45.70% coils. Although high glycine and fewer proline residues are a characteristic feature of the cold-adapted enzymes, the similar observation in thermally active APrBL suggests that this feature cannot be solely responsible for thermo/cold adaptation. The APrBL protease was highly effective as a detergent additive and in whey protein hydrolysis.

Keywords: recombinant alkaline protease, gene expression, Little Rann of Kutch, structure–function relationship, detergent additive, whey protein hydrolysis

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Ram Karan,
King Abdullah University of Science and Technology, Saudi Arabia
Hui Ni,
Jimei University, China

***Correspondence:**
Satya P. Singh
satyapsingh@yahoo.com

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *31 January 2020*
Accepted: *20 April 2020*
Published: *03 June 2020*

Citation:
Bhatt HB and Singh SP (2020) Cloning, Expression, and Structural Elucidation of a Biotechnologically Potential Alkaline Serine Protease From a Newly Isolated Haloalkaliphilic Bacillus lehensis JO-26. Front. Microbiol. 11:941. doi: 10.3389/fmicb.2020.00941

INTRODUCTION

Proteases, being among the most important groups of the industrial enzymes, represent 60% of the total enzyme market (Jaouadi et al., 2008). The demand has increased due to its applications in various sectors such as detergent, leather, food, dairy, pharmaceutical, textile, and peptide synthesis (Haki and Rakshit, 2003; Jain et al., 2012; Purohit and Singh, 2013; Sinha and Khare, 2013; Raval et al., 2014).

Majority of the industrial processes are carried out under the extreme conditions of temperature, pH, and high concentrations of organic solvents, where majority of the enzymes fail to function. Therefore, enzymes of haloalkaliphile origin will serve the purpose of stability and activity at multiple of extremities (Raval et al., 2018). Proteases from halophilic microorganisms are in demand for various applications due to their ability to function at alkaline pH, high temperatures, and high salt concentrations (Sinha and Khare, 2013; Raval et al., 2014). Thus, they are also suitable for food processing under saline conditions. Proteases are widely used in detergents due to inefficiency of non-enzymatic detergents for the removal of protein-containing materials from the textile fibers. The enzymes not only enhance the washing efficiency but also shorten the washing duration at moderate temperatures.

Whey contains approximately 20% of the total soluble milk proteins (Walstra and Jenness, 1984). In protein hydrolysis, peptide bonds are cleaved, resulting in peptides of different sizes and free amino acids. Protein hydrolysis is carried out either chemically or enzymatically. Chemical methods usually yield products with modified amino acids and reduced nutritional values. Therefore, the enzymatic hydrolysis is performed under moderate conditions without side reactions and loss of nutritional values (Tavano, 2013).

Bacillus strains are suitable sources of the commercial enzymes due to their ability to secrete enzymes in a short time (Maurer, 2004). As a matter of fact, different species of *Bacillus* genus are among the most common sources of commercial proteases (Contesini et al., 2018). Although many *Bacillus* strains produce alkaline proteases (Patel et al., 2006; Dodia et al., 2008; Jain et al., 2012; Raval et al., 2014), we need enzymes with better efficiency and features to perform under stressful conditions. Therefore, the search for suitable enzymes continues for various applications. Alkaline proteases are reported from bacteria and actinomycetes of different marine habitats (Thumar and Singh, 2009; Gohel and Singh, 2012, 2018; Annamalai et al., 2014; Raval et al., 2014, 2015; Thakrar and Singh, 2019; Sharma et al., 2020). However, only limited studies are available on the microorganisms of desert origin. The Little Rann of Kutch is a saline desert largely unexplored for its microbial diversity and biotechnological potential (Bhatt and Singh, 2016; Bhatt et al., 2017, 2018).

Further, a high level of production is desirable for the commercialization of any enzyme from microbial sources. In view of the investigations on the gene expression and cost-effectiveness of the enzymes, the protease gene was cloned and expressed in mesophilic host, *Escherichia coli*. Only few reports exist on the gene cloning and characterization of the recombinant proteases of haloalkaliphilic microorganisms. Besides, sequence information of the protease gene will help in elucidating the structure–function relationship and identifying various functional regions of the encoded enzymes.

In consideration of the above scenario, we cloned an alkaline protease (APrBL) gene from a newly isolated haloalkaliphilic *Bacillus lehensis* and expressed into *E. coli* BL21 (DE3), a mesophilic host. The gene expression was optimized with respect to induction, growth temperature, and NaCl concentrations. We purified and characterized the recombinant protease and assessed its potential as detergent additive and in whey protein hydrolysis. The study established the potential of this serine alkaline protease in detergent and food industries. This study represents the first account on the gene cloning, analysis of the expression, and characterization of a recombinant protease from a haloalkaliphilic bacterium of the saline desert, Little Rann of Kutch.

MATERIALS AND METHODS

Bacterial Strains and Vectors

A haloalkaliphilic bacterium *B. lehensis* JO-26 was isolated from the saline desert of Little Rann of Kutch (India). *E. coli* BL21 (DE3) (Merck Millipore, Germany) was used for the expression studies and was grown in LB supplemented with kanamycin (50 μg/ml) for the recombinant protein expression. pET28a (+) (Novagen, CA, United States) was used as an expression vector. Plasmid preparation and DNA purification were carried out using commercial kits (Qiagen, Germany). Restriction enzymes and other molecular biology reagents were from commercial sources (Roche, Germany).

Genomic DNA Extraction, Primer Designing, and Gene Amplification

Genomic DNA was extracted from 2 ml of activated growth cultures (OD600 ∼0.8) using the DNeasy Blood & Tissue Kit for DNA (Qiagen, Germany). Isolated DNA was used as a template for the amplification of the alkaline protease (APrBL) gene. ORF finder tool of NCBI was used to identify the open reading frame of the APrBL gene. Based on the whole genome sequence of *B. lehensis* available in the NCBI database, a set of primer pair (APrBL *Nde*I F′ CCG**CATATG**GCGCAGGTTGGAACATTTTG and APrBL *Xho*I R′ CCG**CTCGAG**CGAGTAGGTCTCTTTTGCAG) was designed to obtain the APrBL gene without secretion signal. Restriction sites are underlined and highlighted in bold. Using APrBL F & R primers, the APrBL gene was amplified using Emerald green master mix (Takara, Japan) under the optimized PCR program (denaturation step at 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 50 s, annealing at 56°C for 45 s and extension at 72°C for 1 min 30 s with final extension at 72°C for 7 min) in 25-μl reaction mixture using a thermocycler (Eppendorf, Germany).

Construction of the Recombinant APrBL-pET28a

A set of primers APrBL Forward and Reverse with *Nde*I and *Xho*I restriction sites was used to amplify the APrBL gene. The amplified product APrBL (200 ng) and expression vector pET28a (2 μg) were digested with *Nde*I and *Xho*I restriction enzymes as per the manufacturer's instructions. The digested amplified product and vector were purified using the HiPurA PCR product purification kit (Himedia, India). Double-digested vector pET28a (+) (133 ng) and APrBL (88 ng) were ligated using T4 DNA ligase

(Roche, Germany) at 16°C overnight. The ligated product was transformed into competent *E. coli* BL21 (DE3) cells. Colony PCR and double digestion of the plasmids with *Nde*I and *Xho*I were carried out to confirm the positive APrBL-pET28a clones.

Gene Expression: Effect of Isopropyl β-D-Thiogalactopyranoside, Temperature, and Salt

Seed culture of *E. coli* BL21 harboring APrBL-pET28a vector was prepared by growing cells on a rotary shaker (200 rpm) for overnight at 37°C. LB medium (50 ml) containing kanamycin (50 μg/ml) was inoculated with 1% (v/v) inoculum and incubated until optical density reaches to 0.5–0.6 at 600 nm. It was then followed by the addition of isopropyl β-D-thiogalactopyranoside (inducer, IPTG) and incubation for 20 h on a rotary shaker (200 rpm). Protease expression was examined at the IPTG concentrations of 0.2 and 1 mM and at 28 and 37°C to optimize the distribution of the expressed protease in soluble fraction. The cells were harvested at 14,031 × g (Beckman Coulter Allegra 64R Centrifuge, United States) for 10 min at 4°C, and the cell pellet was washed with buffer containing 50 mM NaH_2PO_4 and 100 mM NaCl (pH 8). The cells were disrupted by sonication for 30 s in six cycles with Sartorius Labsonic M at 30 Hz. Between each cycle, the samples were chilled in ice for 30 s. The resulting supernatant was considered as soluble fraction. The pellets were treated with 8 M urea for 30 min at 30°C, followed by centrifugation at 9,744 × *g* for 5 min at 4°C to obtain supernatant considered as the insoluble fraction (Purohit and Singh, 2013). Insoluble and soluble fractions of the induced and uninduced cells were analyzed by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) on a 12% polyacrylamide gel. Similarly, the effect of NaCl concentrations (0–4%, w/v) was examined on the expression of protease and its distribution in soluble fraction with other optimum conditions of 0.2 mM IPTG and 28°C growth temperature. The expression was analyzed by plate diffusion assay and enzyme assay as described below.

In silico Analysis of APrBL

Nucleotide sequence of APrBL gene was used for phylogenetic analysis and was compared against NCBI database using BLASTn tool. Similarly, the protein sequence of APrBL was compared against the protein database using BLASTp tool. Multiple sequence alignment (MSA) was achieved using Clustal W software. The phylogenetic tree was constructed using the Neighbor-Joining method of MEGA 6 software (Tamura et al., 2013). Distances were calculated using the Kimura correction in a pairwise deletion manner (Kimura, 1980). The N-terminal signal peptide analysis was carried out by the SignalP version 5.0 program (Armenteros et al., 2019). The hydropathicity of the protein was determined using the Kyte and Doolittle scale (Kyte and Doolittle, 1982). Primary and secondary structure properties of the protease and amino acids were studied by Expasy and NCBI tools. Expasy's Protparam server[1] was used for detection of the physicochemical properties of these proteins. The number of amino acids, theoretical isoelectric point (p*I*), molecular weight, aliphatic index, instability index, grand average hydropathy (GRAVY), and total number of positive and negative residues were computed for each protein. Expasy's PROSITE tool[2] was used to determine the catalytic triad of the enzyme. The secondary structure configuration of APrBL was predicted by Endscript 2.0 (Gouet et al., 2003) and SOPMA tool (Geourjon and Deleage, 1995). Three-dimensional structure of the serine protease was modeled using the I-TASSER server for three-dimensional (3D) structure prediction and validation (Yang et al., 2015). The model was also validated by structure assessment tool of SWISS-MODEL (Kiefer et al., 2009). Further, the hydropathy profile was plotted according to the method of Kyte and Doolittle using pscale tool (Kyte and Doolittle, 1982).

Homology Modeling

APrBL amino acid sequence without signal peptide was used for homology modeling. SWISS-MODEL Workspace server was used to identify the template and building protein model[3] (Kiefer et al., 2009).

Purification of the Recombinant Alkaline Protease

The recombinant protease (APrBL) was purified under non-denaturing conditions using Novagen Ni-NTA His·Bind resin. The IPTG-induced recombinant *E. coli* BL21 (DE3) cells were sonicated in binding buffer containing 50 mM phosphate buffer (pH 8), 100 mM NaCl, 10 mM $MgCl_2$, 1 mg ml^1 lysozyme, 5 mM β-mercaptoethanol (β-ME), and 5% (v/v) glycerol. After sonication, debris was spun down and the supernatant containing APrBL was collected for purification. The chromatographic column was packed with Ni-NTA His·Bind Resin (Merck) and allowed to settle to generate 2 ml of bed volume. The packed Ni-NTA resin was washed with a 5-column volume of the deionized water to remove the ethanol. The column was then equilibrated with a 5-column volume of the binding buffer. Thereafter, the supernatant was passed through the column several times for binding of the 6 × histidine-containing recombinant protein to the Ni-NTA resin. After binding, 2-column volume of wash buffer (binding buffer + 20 mM imidazole) was passed through the column in order to wash out non-specific proteins and other impurities. The recombinant protein was eluted with one column volume of the elution buffer (binding buffer + imidazole) with 250 mM and 500 mM of imidazole in increasing gradient, and fractions were collected. The fractions were examined for purity of the enzyme on 12% SDS–PAGE. The protease activity of the purified APrBL was confirmed by enzyme assay.

Protease Assay and Protein Estimation

The protease activity was measured by Anson–Hagihara's method using tyrosine as standard (0–100 g/ml) (Hagihara, 1958). The enzyme (0.5 ml appropriately diluted enzyme) was added to 3.0 ml casein [0.6% (w/v) in 20 mM NaOH–Borax buffer, pH 10],

[1]http://web.expasy.org/protparam/

[2]https://prosite.expasy.org/

[3]http://swissmodel.expasy.org/workspace

and the reaction mixture was incubated at 50°C for 10 min. The reaction was terminated by the addition of 3.2 ml of TCA mixture (0.11 M trichloro acetic acid, 0.22 M sodium acetate, and 0.33 M acetic acid) and incubated at room temperature for 30 min. The precipitates were removed by filtration through Whatman-1 filter paper, and the absorbance of the filtrate was measured at 280 nm. One unit of alkaline protease activity was defined as the amount of enzyme liberating 1 μg of tyrosine per minute under the assay conditions. Protein concentrations were determined by the method of Bradford (1976) using bovine serum albumin as a standard.

Kinetic Parameters of APrBL

V_{max} and K_m were determined using Lineweaver–Burk double reciprocal (1/V versus 1/S) plot by measuring the activity at various concentrations of casein substrate (0.25–10 mg/ml) under the standard conditions. The value of the turnover number (K_{cat}) was calculated using the following equation:

$$K_{cat} = \frac{V_{max}}{[Et]}$$

where [E] refers to the active enzyme concentration, V_{max} refers to the maximum reaction rate, and K_{cat} is defined as the maximum number of chemical conversions of substrate molecules per second that a single catalytic site will execute for a given enzyme concentration (Mechri et al., 2017). For the determination of K_{cat} and K_{cat}/K_m, the value of V_{max} was expressed in terms of μmoles/ml/min.

Enzyme Secretion by Plate Diffusion Assay

The soluble fractions of the expressed enzyme were analyzed for the enzyme activity using plate diffusion assay. The enzyme preparations were dispensed into wells created in the gelatin agar plate (3% Agar + 1% gelatin in 50 mM glycine-NaOH buffer, pH 10). The plates were then incubated for 24 h at 50°C followed by the observation of the zone of gelatin hydrolysis around the wells.

Biochemical Characterization of APrBL

Effect of Temperature and pH on the Activity and Stability of the Enzyme

The temperature profile of the enzyme was determined using standard assay at varying temperatures in the range of 30–80°C. For thermal stability, the enzyme was preincubated for 30 min at the temperatures in the range of 40–80°C. Aliquots were withdrawn at 1, 3, and 24 h, and the protease activity was determined. The residual activities were determined by taking the activity prior to incubation as 100%. The assay was performed in 20 mM borax-NaOH buffer (pH 10) using 0.6% (w/v) casein as a substrate.

The effect of pH on the enzyme activity was determined in buffers of different pH. To determine the optimum pH, sodium phosphate buffer (pH 7.0), Tris–HCl buffer (pH 8–9), borax-NaOH buffer (pH 10), and glycine-NaOH buffer (pH 11–12) were used. The pH stability was determined by preincubating the enzyme at pH 8–12. Aliquots were withdrawn at 1, 3, and 24 h, and the protease activities were determined.

Effect of Metal Ions on the Enzyme Activity

The effect of various metal ions on APrBL activity was assessed by preincubating the enzyme with 5 mM concentration of different metal ions: $CaCl_2$, $MgSO_4$, $HgCl_2$, $CoCl_2$, $MnCl_2$, $FeSO_4$, and $ZnCl_2$ for 30 min. The residual activities were measured using casein as substrate at pH 10 and 50°C. The activity measured without metal ions was considered as 100%.

Effect of Solvents on the Enzyme Activity

The effect of organic solvents on APrBL activity was examined by preincubating the enzyme with 10% (v/v) of n-hexane, n-butanol, iso-propanol, ethanol, methanol, chloroform, glycerol, benzene, and toluene for 1 h. The residual activities were measured using casein as substrate at pH 10 and 50°C. For the control, the organic solvent was replaced with an equivalent amount of borax-NaOH buffer (20 mM, pH 10). The relative activities were calculated taking the activity of the control as 100%.

Effect of Inhibitors and Surfactants on Enzyme Activity

To evaluate the effects of inhibitors and surfactants on the enzyme activity, APrBL was preincubated with 5 mM of different inhibitors such as ethylenediaminetetraacetic acid (EDTA), β-ME, phenylmethylsulfonyl fluoride (PMSF), dithiothreitol (DTT), and 1% (w/v) of surfactants such as SDS, Triton X-100, and Tween 80 for 30 min at room temperature before the residual activities were measured. The activity without any additive was considered as 100%.

Applications of the Recombinant Alkaline Protease APrBL

Detergent Additive: Wash Performance Analysis

Application of the recombinant alkaline protease APrBL was assessed as detergent additive by wash performance analysis. New cotton cloth pieces (6 cm × 6 cm) were stained with human blood (0.5 ml). Five different commercial detergents Ariel, Tide, Surf, Nirma, and Wheel were used at the concentration of 7 mg/ml to simulate washing conditions. Endogenous protease present in these detergents was inactivated by incubating the solid detergents at 70°C for 1 h prior to the addition of APrBL enzyme. The following sets were prepared:

1. Flask with tap water (25 ml).
2. Five flask with tap water (25 ml) + Each different detergent.
3. Five flask with tap water (25 ml) + Each different boiled detergent + APrBL (200 U).
4. Flask with tap water (25 ml) + APrBL (200 U).

Then, the stained cloth pieces were shake-incubated (100 rpm) in abovementioned flasks with different wash conditions at 40°C for 10 min containing a total volume of 25 ml of tap water. After incubation, cloth pieces were taken out, rinsed with water, and dried. Untreated stained cloth pieces and treated with only tap water were taken as control. The performance

of the recombinant alkaline protease in bloodstain removal was evaluated by visual examination.

Whey Protein Hydrolysis

Whey solution was adjusted to pH 10 with 3 M NaOH solution. APrBL at 300 U was added into the whey solutions. Two negative controls were included to compare the effect of APrBL for whey protein hydrolysis. In one control, enzyme was replaced with equal amounts of borax-NaOH buffer (20 mM, pH 10), while in another control, denatured alkaline protease APrBL was added. The reaction mixture was incubated for 8 h at 45°C with continuous shaking. The aliquots were withdrawn at regular time intervals, and hydrolysis reactions were terminated by heating at 90°C for 15 min in water bath (Shu et al., 2018). Total protein content of these sets was determined before and after hydrolysis using Bradford's method (Bradford, 1976).

Degree of hydrolysis (DH) was calculated using soluble protein content method (Morais et al., 2013; Sinha and Khare, 2015). One milliliter of 10% (w/v) trichloroacetic acid was added to 1 ml of the aliquot collected after enzymatic hydrolysis and incubated at 37°C for 30 min. The mixture was then centrifuged at 10,000 rpm for 10 min at room temperature, and the total content of the soluble protein was determined by Bradford's method (Bradford, 1976). The DH was calculated as follows:

$$\text{Degree of Hydrolysis (\%)} = \frac{\text{Soluble protein content in 10\% TCA (mg)} \times 100}{\text{Total protein content (mg)}}$$

RESULTS AND DISCUSSION

Gene Amplification, Sequencing, and Phylogenetic Analysis

The APrBL gene without the signal sequence was amplified using primer pair (APrBL *Nde*I F/*Xho*I R), designed based on whole genome sequence of *B. lehensis* G1 (CP003923) available in NCBI database. Signal peptides usually have no role except the export of protein in the periplasmic space. Therefore, production of recombinant protein without signal peptide in *E. coli* would not alter its biochemical characteristics and function. Moreover, removal of signal peptide has been reported to increase the expression of recombinant protein (Gopal and Kumar, 2013). Some reports describe the expression of proteases without signal peptide (Joshi and Satyanarayana, 2013; Cheng et al., 2015; Devi et al., 2016).

The APrBL gene coding full-length ORF of 1,014 bp with a start and stop codon consisted of 337 amino acid residues: a pro-peptide of 71 and a mature protein of 266 amino acids. The molecular mass was 34.5 kDa, and isoelectric point (pI) was 4.66. The corresponding molecular weights with pro-peptide regions are reported earlier (Jaouadi et al., 2008; Deng et al., 2011; Cheng et al., 2015; Zhang et al., 2016). PROSITE analysis of the ORF revealed the catalytic triads of APrBL to be Asp 97, His 127, and Ser 280, a common feature of the serine proteases of subtilase superfamily (Siezen and Leunissen, 1997; Purohit and Singh, 2014).

BLASTn analysis of the cloned gene revealed 98.2% similarity with the alkaline protease of *B. lehensis* G1 (CP003923), 97.9% with YaB alkaline elastase subtilisin gene of *Bacillus subtilis* (M28537), and 95.4% with AprN alkaline protease of *Bacillus* sp. (AB005792). Whereas the BLASTp analysis revealed highest sequence similarity of 99.3% with both alkaline protease of *B. lehensis* MLB2 (AFK08970) and alkaline elastase YaB of *Bacillus* sp. YaB (P20724) followed by 97.8% with AprN alkaline protease of *Bacillus* sp. (BAA25184). Further, APrBL has shown 82.3 and 80.4% similarity with the subtilisin savinase of *Bacillus* sp. (P29600) and alkaline protease of *Bacillus clausii* KSM-K16 (Q99405), respectively. The APrBL protease was further compared with other proteases with respect to amino acid composition, structural features, and phylogenetic analysis. APrBL closely related to alkaline elastase YaB of *Bacillus* sp. YaB and alkaline protease BLAP of *B. lehensis* MLB2 (AFK08970). On the other hand, it distantly related to subtilisin Carlsberg (CAB56500) of *Bacillus licheniformis* and subtilisin BPN' (P00782) from *Bacillus amyloliquefaciens* (**Figure 1** and **Table 1**). The alkaline protease gene sequence of *B. lehensis* JO-26 (APrBL) has been submitted to GenBank (Accession No. MN104891).

Expression of the Alkaline Protease Gene in *Escherichia coli*

Effect of Isopropyl β-D-Thiogalactopyranoside and Temperature on Gene Expression

It is highly desirable to get the expressed protein in soluble fraction due to its easy recovery, cost-effective purification, and minimum activity loss and time consumption (Sørensen and Mortensen, 2005). Among the strategies to limit trapping of the expressed proteins in insoluble fraction and thus reducing the formation of the inclusion bodies, growth of the host organism at low temperatures with lower induction level are important (Sørensen and Mortensen, 2005; Rosano and Ceccarelli, 2014). In the present report, the growth of the host was carried out at various temperatures and different levels of induction. At 37°C, while growth of the organisms was highest, the expressed protease was largely trapped into the inclusion bodies, i.e., in insoluble fraction. On the contrary, growth at 28°C enhanced the solubility of the expressed protein despite lower growth of the host. There are several reports on the expression of alkaline protease in soluble fraction at lower temperatures of 27–30°C (Gohel and Singh, 2012; Joshi and Satyanarayana, 2013; Purohit and Singh, 2014; Devi et al., 2016).

The synergistic effect of IPTG induction and growth temperatures on APrBL expression was reflected by the fact that despite reduced growth of the recombinant *E. coli* at 1 mM IPTG, the expression level was quite comparable at both 0.2 mM and 1 mM IPTG concentrations at both 28°C and 37°C temperatures (**Figure 2A**).

Although expression was higher at 1 mM IPTG induction, considering the correlation of the growth and expression level, IPTG induction at 0.2 mM can be considered optimum for the APrBL expression. Similarly, a study on the expression of cellobiose phosphorylase from *cellvibrio gilvus* revealed that

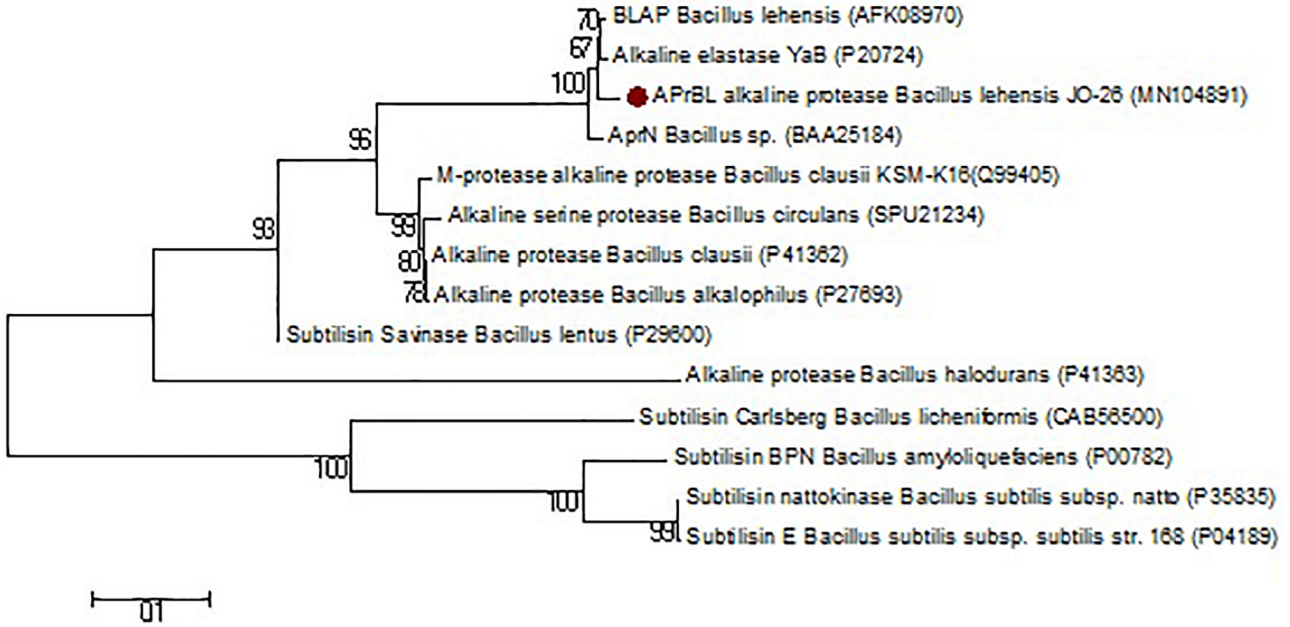


FIGURE 1 | Phylogenetic tree based on a comparison of the APrBL amino acid sequence and some of their closest phylogenetic relatives retrieved from the NCBI database. The tree was reconstructed by the Neighbor-Joining method using MEGA 6 software. The numbers on the tree indicate the percentages of bootstrap support derived from 1,000 replications. The scale bar corresponds to 0.1 substitutions per nucleotide position.

TABLE 1 | Comparison of primary structure profile of APrBL with other reported proteases including phylogenetic neighbors.

Enzymes	Hydrophobic aa (%)	Aromatic aa (%)	Asp + Glu/Arg + Lys	Theoritical PI	Gravy index	Instability index	Aliphatic index
AprBL	55.3	7.8	2.14	4.66	−0.036	26.23	85.44
YaB (P20724)	56	7.2	1.84	4.66	0.014	26.01	87.54
Subtilisin Carlsberg (X03341)	54.3	8.9	0.9	8.73	0.036	13.96	83.69
AprN (BAA25184)	55.5	7.3	1.78	4.77	−0.002	26.28	87.28
BLAP (AFK08970)	56	7.5	1.84	4.72	0.025	26.08	87.54
Subtilisin BPN′ (NC_014551)	52.6	8.6	0.89	8.73	−0.105	26.29	78.51

enzyme activity was higher in soluble fraction when cells were grown at 25°C as compared with 37°C (Singh et al., 2010). Furthermore, the enzyme activity was higher in soluble fraction at 0.1 mM IPTG induction as compared to 1 mM IPTG (Singh et al., 2010).

Effect of Salt Concentration

There is scarce information on the effect of salt on the gene expression in *E. coli*. The main reason to study the effect of salt on the expression of APrBL protease gene was the haloalkaliphilic nature of the native strain *B. lehensis* JO-26. The maximum expression was achieved with 2% NaCl (w/v). Growth of *E. coli* along with protease expression increased up to 1% NaCl (w/v) (**Figure 3**). Despite reduced growth, the enzyme production was higher at 2% NaCl (w/v) concentration. At 4% NaCl (w/v) concentration, both growth of the host and protease production decreased. The data clearly suggested that with increasing salt concentration, expression of functional recombinant APrBL increased up to 2% NaCl (w/v) (**Figure 3** and **Supplementary Figure S1**).

Purification of the Recombinant Protease

The recombinant protease appeared to be located in the cytoplasmic space as extracellular fraction did not contain any protease activity. The maximum expressed enzyme in soluble fraction was observed at 0.2 mM IPTG induction after 20 h of growth at 28°C and 200 rpm agitation. The soluble fraction of APrBL was purified onto Ni-NTA matrix using 250 mM imidazole for elution. The purified APrBL resolved as a single band of 34.6 ± 1.0 kDa on SDS–PAGE (**Figure 2B**). The purified recombinant enzyme exhibited a 16.93-fold increase in the specific activity with a yield of 84.6% (**Table 2**). The molecular masses of microbial alkaline proteases with some exceptions range between 15 and 40 kDa (Dodia et al., 2008; Haddar et al., 2009; Sinha and Khare, 2013; Raval et al., 2014). Alkaline proteases of similar molecular mass to APrBL have been cloned into mesophilic hosts, such as 34 kDa from *Bacillus pumilus* CBS (Jaouadi et al., 2008), 34.4 kDa from *Bacillus vallismortis* (Cheng et al., 2015), and 35.6 kDa from *Planococcus* sp. (Zhang et al., 2016). Zone of the utilization of the substrate was observed in the plate diffusion assay of

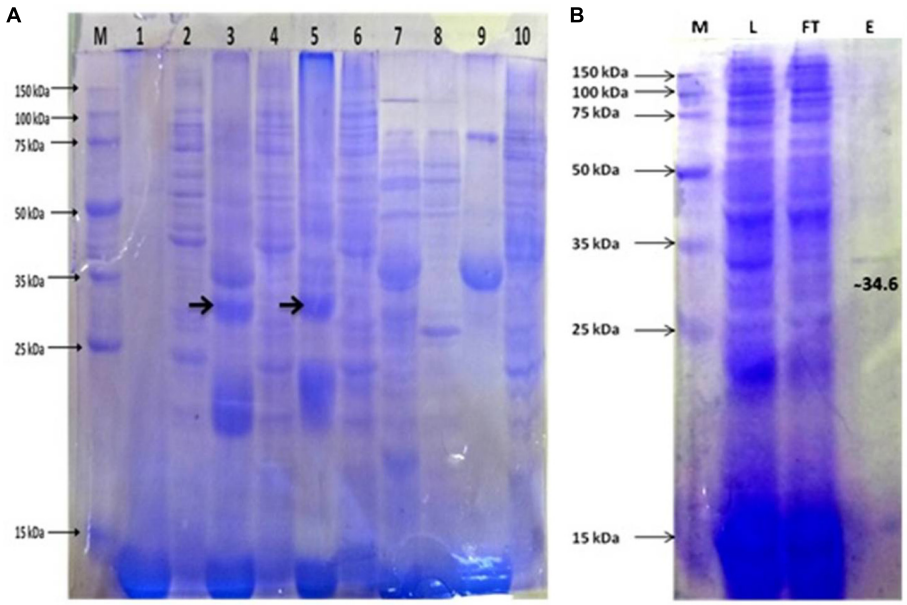


FIGURE 2 | (A) Protein expression profile of APrBL. M: 15–150 KD, Lane 1: Preinduced soluble fraction of *Escherichia coli* BL21 (DE3) harboring recombinant APrBL-pET28a, Lane 2: Preinduced insoluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a, Lane 3: Soluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a induced with 0.2 mM IPTG, Lane 4: Insoluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a induced with 0.2 mM IPTG, Lane 5: Soluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a induced with 1 mM IPTG, Lane 6: Insoluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a induced with 1 mM IPTG, Lane 7: Soluble fraction of *E. coli* BL21 (DE3), Lane 8: Insoluble fraction of *E. coli* BL21 (DE3), Lane 9: Uninduced soluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a, Lane 10: Uninduced insoluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a. **(B)** Protein purification using Ni-NTA column after expression of the protein in pET28a, 28°C, 0.2 mM IPTG. M: 15–150 KD, L: Soluble fraction of *E. coli* BL21 (DE3) harboring recombinant APrBL-pET28a induced with 0.2 mM IPTG, FT: Flow through from the column, E: Column elution in 250 mM imidazole.

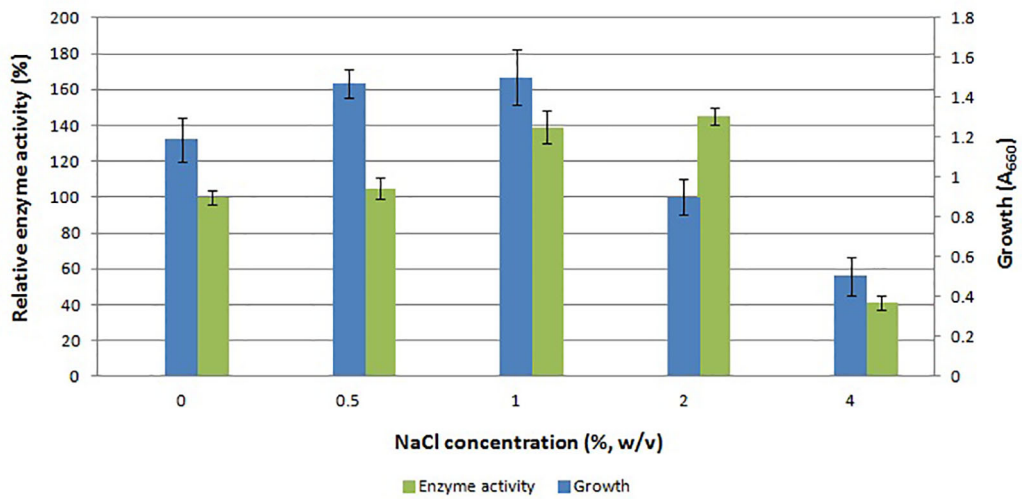


FIGURE 3 | Effect of NaCl concentration on growth and APrBL activity.

the purified APrBL, confirming the activity of the enzyme (**Supplementary Figure S2**).

Kinetic Parameters: K_m, V_{max}, and K_{cat}

K_m and V_{max} were computed as 1.38 mg/ml and 27.14 μmol mg^{-1} min^{-1} (212.76 U/ml), respectively. The K_m is inversely proportional to the enzyme's affinity with the substrate. The APrBL protease has low K_m value compared to other reported proteases from *Bacillus* sp. (1.53–2.5 mg/ml), metagenomic alkaline protease (1.70 mg/ml), *Bacillus halotolerance* (10 mg/ml), *Bacillus marismortui* (2.5 mg/ml), *Shewanella arctica* (1.75 mg/ml), *Idiomarina* sp. (3.76 mg/ml) (Patel et al., 2006;

TABLE 2 | Purification of the recombinant alkaline protease APrBL by Ni-NTA affinity chromatography.

Enzyme preparation	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Purification fold	Yield (%)
Recombinant crude fraction	1,392	4.8	290	1	100
Ni-NTA affinity Chromatography Purified fraction	1,179	0.24	4,912	16.93	84.6

Dodia et al., 2008; Sinha and Khare, 2013; Qoura et al., 2015; Devi et al., 2016; Dorra et al., 2018; Zhou et al., 2018). V_{max} is the catalytic activity of an enzyme usually desired as high as possible. The studied enzyme had higher V_{max} than other reported proteases from *B. vallismortis* (49.8 μg ml^{-1} min^{-1}), *Bacillus circulans* (1.8 μmol/min), *B. lehensis* (25 nmol mg^{-1} s^{-1}), and metagenomic protease (278.2 U/mg/min) (Joshi and Satyanarayana, 2013; Cheng et al., 2015; Devi et al., 2016; Patil et al., 2016). The results signify high affinity and catalytic efficiency of APrBL compared to other proteases.

K_{cat} is the maximum number of conversions of the substrate molecules per second by a single catalytic site at a given enzyme concentration (Mechri et al., 2017). K_{cat} is a constant independent of the amount of the enzyme. With respect to the present enzyme, the catalytic constant K_{cat} was deduced as 0.549 S^{-1}. In comparison, Cheng et al. (2015) reported a K_{cat} value of 4.35 min^{-1} for an alkaline protease from *B. vallismortis*. K_{cat} values of 0.13 S^{-1} and 3.99 × 10^{-2} S^{-1} were reported for the alkaline proteases of *Bacillus* sp. and *Bacillus pseudofirmus*, respectively (Sinha and Khare, 2013; Raval et al., 2014). However, a higher K_{cat} value of 289 S^{-1} has also been earlier reported with low V_{max} (Joshi and Satyanarayana, 2013). Overall, the kinetic parameters; K_m, V_{max}, and K_{cat} of the APrBL suggest its potential and suitability as a biocatalyst in comparison to the previously reported alkaline proteases.

Characterization of Recombinant Protease

Effect of pH on the Enzyme Activity and Stability

The APrBL was active in a broad range of pH 8–12 with an optimum at pH 10, thus confirming the alkaline nature of the enzyme (**Figure 4B**). The trend of the pH optima is consistent with the proteases of *Bacillus* sp. (Patel et al., 2006), *B. pumilus* (Jaouadi et al., 2008), haloalkaliphilic bacterium strain AH-6 (Dodia et al., 2008), *Nocardiopsis alba* (Gohel and Singh, 2012), *Bacillus* sp. (Jain et al., 2012), and *B. pseudofirmus* (Raval et al., 2014). The trends, however, vary from the pH optima of the alkaline proteases of *B. vallismortis* (optimum pH, 6.5) (Cheng et al., 2015), *Bacillus mojavensis* (optimum pH, 8.5) (Haddar et al., 2009), *Bacillus halotolerans* (optimum pH, 9) (Dorra et al., 2018), *Bacillus* sp. (optimum pH, 9) (Sinha and Khare, 2013), and *Bacillus* sp. (optimum pH, 9) (Briki et al., 2016).

The recombinant protease APrBL was highly stable in the range of pH 8–12 (**Figure 4D**). The enzyme was stable at pH 8–12, for 3 h, with significant stability displayed even after prolonged incubation up to 24 h at pH 9–10. The enzyme retained more than 80% of the residual activity at pH 9 and 10, while it retained 50–60% activity at pH 8. The trends, however, differ from the alkaline protease of *B. vallismortis* which was stable at pH 5.6–9.6 (Cheng et al., 2015). Effective functioning of the protease over a broad pH range is an important criterion for the application as detergent additives. The APrBL protease fulfills this condition as it was active and stable in the pH range of 8–12. The APrBL appears more efficient at alkaline pH as compared to the commercial detergent enzyme Alcalase (Novozymes A/S) produced by *B. licheniformis*, having a maximal activity at pH 8–9 (Van Kampen and Merget, 2002). The enzymatic features are quite comparable with Purafect (Genencor International Inc., United States), a genetically engineered donor *B. lentus* expressed in *Bacillus* sp. having the optimum activity at pH 10 (Van Kampen and Merget, 2002) and Maxatase (Gist-Brocades), produced by *B. licheniformis* with a maximum activity at pH 9–10 (Beg and Gupta, 2003).

Effect of Temperature on the Activity and Stability

The enzyme was active at temperatures 30–70°C with optimum activity at 50°C (**Figure 4A**). The enzyme retained 76% of the residual activity at 60°C, whereas at 40°C, 37% of the activity was evident. A sharp increase in the activity was observed from 40 to 50°C, while the activity decreased from 50 to 60°C, clearly suggesting the thermally active nature of the enzyme. These results, however, deviate from the alkaline proteases of *Planococcus* sp. (optima at 35°C) and *Bacillus* sp. (optima at 37°C) (Briki et al., 2016; Zhang et al., 2016). Interestingly, the enzyme also differs from its phylogenetic relative, BLAP alkaline protease, of the same species *B. lehensis* (Joshi and Satyanarayana, 2013). Despite sharing a common temperature optima of 50°C, while BLAP rapidly loses activity at 60°C, the APrBL retains 76% of the residual activity at this temperature.

The APrBL is stable in the range of 40°C–80°C (**Figure 4C**). The protease was highly stable at 40°C even after 20 h. While at 50°C, it was highly stable up to 1 h, the enzyme retained 72 and 62% of the residual activities after 3 and 20 h of incubation, respectively. At 60 and 80°C, 85 and 73% of the residual activities were evident up to 3 h, respectively, whereas 40% of the activity was retained at 60°C, and total loss of the activity was observed at 80°C after 3 h of incubation (**Figure 4C**). The stability profile of APrBL suggests its superiority when compared to other alkaline proteases (Haddar et al., 2009; Jellouli et al., 2011; Cheng et al., 2015).

Effect of Metal Ions, Surfactants, Inhibitors, and Solvents

Among the cations, while Ca^{2+} slightly stimulated the APrBL activity, Hg^{2+} completely inhibited the activity (**Table 3**). The trends corresponded with some earlier reports on alkaline proteases (Patel et al., 2006; Cheng et al., 2015). Hg^{2+} reacts with the aromatic ring of tryptophan oxidizing its indole ring (Liu et al., 2010). APrBL activity was marginally inhibited by Mg^{2+}, Mn^{2+}, Zn^{2+}, and Co^{2+}, while Fe^{2+} significantly reduced the activity by 60% (**Table 3**). In a similar manner, an alkaline protease from *B. halotolerans*

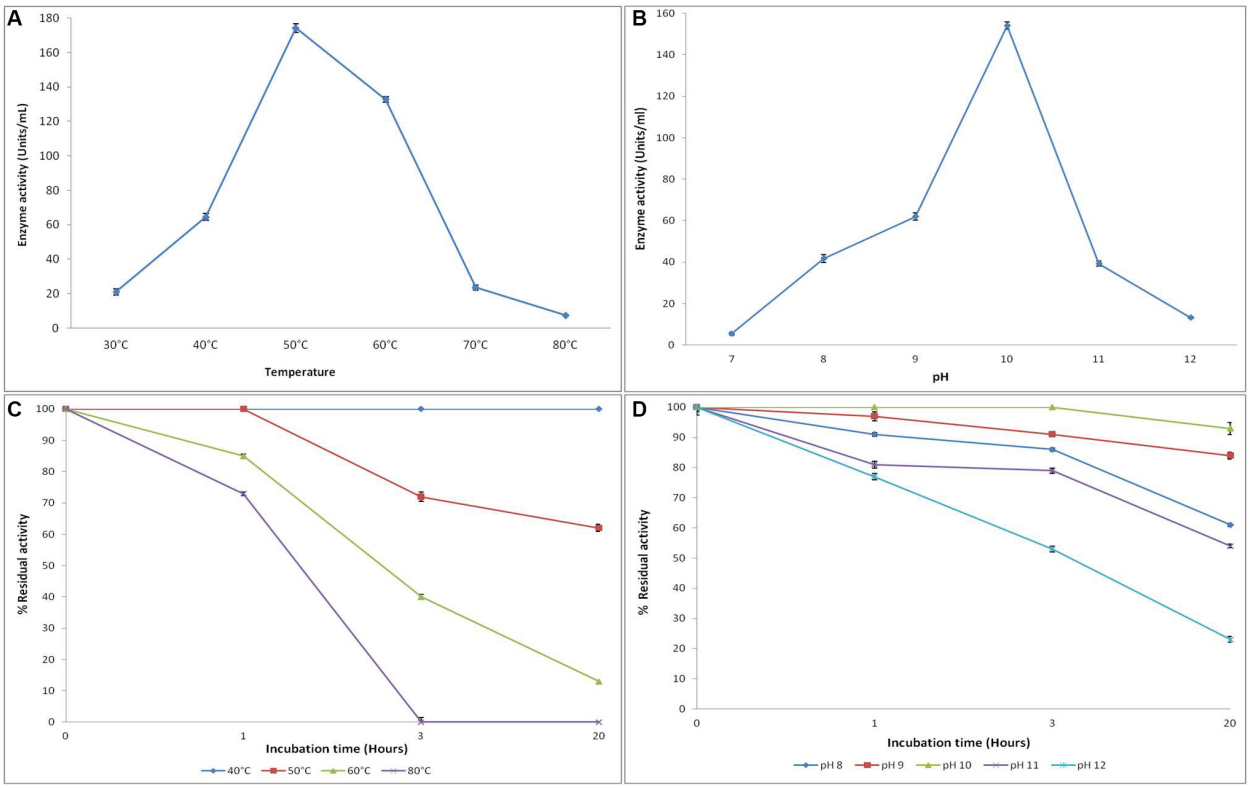


FIGURE 4 | Effect of temperature and pH on the activity **(A,B)** and stability **(C,D)** of recombinant alkaline protease APrBL, respectively.

was marginally inhibited by Mg^{2+}, Zn^{2+}, and Co^{2+} while strongly inhibited by Fe^{2+} (Dorra et al., 2018). In comparison, Zn^{2+} inhibited the activity by 50%, while Fe^{2+} marginally inhibited the activity (Devi et al., 2016). Majority of the proteases of *Bacillus* sp. are induced by Ca^{2+} (Dodia et al., 2008; Haddar et al., 2009). Ca^{2+} marginally stimulated the activity of alkaline protease as also reported earlier for serine proteases of *Bacillus* sp. (Jain et al., 2012). Despite some benefits, calcium dependency raises limitations of using enzymes in detergents containing chelating agents (Sato et al., 1990). Calcium-independent protease variants can be obtained by removing the calcium-binding loop using directed mutagenesis (Strausberg et al., 1995). Therefore, the calcium-independent character of the APrBL adds to its suitability as a detergent additive.

Various inhibitors were assessed for their effect on the protease activity. The activity was completely inhibited by PMSF, suggesting its serine nature (**Table 3**). The essential serine residue in the active site is sulfonated by PMSF, leading to the total loss of the activity (Mechri et al., 2017). Further, APrBL was inhibited by DTT, β-ME, and EDTA, retaining 37, 31, and 66% activities, respectively (**Table 3**).

Among the surfactants, SDS stimulated APrBL activity by 2.35-fold (**Table 3**). Previously, 1.55- and 3.35-fold enhancement in the activity are reported by SDS in *Nocardiopsis alba* and *Bacillus* sp. (Gohel and Singh, 2012; Jain et al., 2012). APrBL enzyme was relatively less stable in non-ionic surfactants such as Triton X-100 and Tween-80 as compared to some earlier reported proteases (Annamalai et al., 2014). Recently, a 50% induction of an alkaline protease activity by SDS was reported (Dorra et al., 2018). However, majority of the serine alkaline protease are reported as SDS labile (Haddar et al., 2009; Jellouli et al., 2011; Raval et al., 2014; Zhang et al., 2016). Due to the amphiphilic nature, SDS reacts with the amino acids, leading to protein unfolding and loss of activity. On the contrary, SDS favorably changes the confirmation of the APrBL protease and stimulates the activity. The recombinant protease might be slightly misfolded yet retains full activity after the interaction with SDS.

The alkaline proteases have a potential role in ester and peptide synthesis under non-aqueous conditions (Jellouli et al., 2011). The advantages of non-aqueous catalysis include the use of water-insoluble substrates, reduced side reactions, reduced microbial contamination, easy product separation and recovery, and thermodynamic equilibrium favoring synthesis (Klibanov, 2001). However, enzymes lose their conformation in organic solvents, leading to the loss of structural flexibility due to the stripping off of the essential water layer (Doukyu and Ogino, 2010). Hence, enzyme activity and stability in organic solvents become major challenge. APrBL activity was significantly enhanced by chloroform and benzene and moderately enhanced by n-butanol and toluene (**Table 3**). On the contrary, adverse

TABLE 3 | Effect of various additives on the purified APrBL enzyme activity.

Additives	Residual activity (%)
Solvents [10% (v/v)]	
Control	100 ± 1
n-Hexane	85.39 ± 3.8
DMF	35.35 ± 3.07
Propanol	75.36 ± 1.55
Chloroform	378.51 ± 1.55
Glycerol	22.27 ± 0.69
Toluene	113.07 ± 0.78
n-Butanol	118.2 ± 2.84
Benzene	146.95 ± 3.89
Methanol	76.15 ± 0.79
Ethanol	53.83 ± 1.52
Metal ions (5 mM)	
Control	100 ± 1
Ca^{2+}	107.8 ± 4.26
Mg^{2+}	77.7 ± 2.55
Zn^{2+}	63.2 ± 2.55
Mn^{2+}	76.4 ± 4.24
Hg^{2+}	0
Co^{2+}	84.3 ± 1.70
Fe^{2+}	38.5 ± 3.40
Surfactants/inhibitors	
Control	100 ± 1
EDTA (5 mM)	46.15 ± 2.17
PMSF (5 mM)	0
DTT (5 mM)	36.92 ± 4.35
β-ME (5 mM)	30.765 ± 2.17
SDS [1% (w/v)]	275.38 ± 4.35
Triton X-100 (1%)	36.92 ± 2.17
Tween 80 [1% (v/v)]	54.61 ± 5.44

effect on the activity of alkaline protease by benzene and toluene has been reported earlier (Joshi and Satyanarayana, 2013). Further, methanol, ethanol, and isopropanol positively affected the APrBL activity, as against the toluene and hexane, which considerably acted against the enzyme (Devi et al., 2016). In the present study, dimethylformamide (DMF) and glycerol significantly inhibited the APrBL activity, while the activity was moderately inhibited by n-hexane, propanol, methanol, and ethanol (**Table 3**). There are several studies on the stimulation of protease activity by methanol and ethanol (Thumar and Singh, 2009; Sinha and Khare, 2014; Mechri et al., 2017). APrBL enzyme was studied in a broad range of organic solvents of variable log *P*-value (−2.32 to 3.9). Solvents with logP < 4.0 are considered as extremely toxic due to their greater partitioning into the aqueous and hydrophobic layers (Mechri et al., 2017). Thus, due to the disruption of the hydrogen bonds and hydrophobic interactions depriving the water hydration shell of the protein, the structural flexibility and functionality are lost (Barberis et al., 2006; Mechri et al., 2017). The haloalkaliphilic bacteria decrease the water activity in the cytoplasm by accumulating salt and thus mimic a non-aqueous system. The enzymes by adapting to high salt are able to function in non-aqueous media. The APrBL was stable in all solvents except those of log $P < 0.24$, which might be due to a larger number of acidic amino acids on its surface as evident from its structure (**Table 1**). The negative charges enhance protein solubility through hydrated ion network with cations or by preventing the protein aggregation *via* electrostatic repulsion on the protein surface (Jain et al., 2012). The experiments were carried out in triplicates, and the results are presented as the mean values.

Structure–Function Analysis of APrBL

The physicochemical properties of APrBL elucidated by ProtParam, including the amino acid composition, instability index, theoretical pI, molecular weight, grand average of hydropathicity (GRAVY), aliphatic index, and total number of negatively and positively charged residues, are shown in **Table 3**. Proteins with an instability index of below 40 are predicted stable, while those with values above 40 are predicted unstable (Guruprasad et al., 1990). Instability index was as low as 26.23, clearly suggesting the stability of the protein. Further, aliphatic index and grand average of hydropathicity (GRAVY) index were 85.44 and −0.036, respectively. The GRAVY index is the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence. A low GRAVY index of APrBL indicates low hydrophobicity and high hydrophilicity of the protein, suggesting a better interaction with water. The aliphatic index is measured by the relative volume occupied by the aliphatic side chains. It may enhance the thermostability of the globular proteins (Ikai, 1980). A high aliphatic index of the protein indicates its enhanced thermostability over a broad range of temperatures, as reflected by the APrBL and supported by the experimental data. A high number of negatively charged amino acids (Asp + Glu: 30) as compared to positively charged amino acids (Arg + Lys: 14) is believed to confer alkali-halo stability of the protease. The ratio of Asp + Glu/Arg + Lys was found to be 2.14, which is highest for APrBL in comparison with other phylogenetic relatives (**Table 1**). Normally, in the presence of salt ions, the enzyme gets precipitated. However, with respect to the halophilic enzymes, negative charges stabilize water and/or ion-binding essential for the tertiary or quaternary structure (Takenaka et al., 2011). The negative charges promote protein refolding and prevent aggregation. The APrBL exhibited 8.90% negatively charged residues, which is higher than other alkaline proteases, M-protease (3.36%), subtilisin BPN' (5.08%), and subtilisin Carlsberg (7.38%) (Jacobs et al., 1985; Gallagher et al., 1996; Shirai et al., 1997). The features of the primary structure of the APrBL align with some earlier reported proteases: BLAP protease (Joshi and Satyanarayana, 2013), YaB protease (Kaneko et al., 1989), and AprN (Masui et al., 1998) (**Table 1**). On the contrary, non-halophilic subtilisin Carlsberg and subtilisin BPN' have a reduced number of negatively charged and hydrophobic residues, high number of aromatic amino acids, high GRAVY index, and low aliphatic index. Moreover, APrBL exhibited as high as 55.3% hydrophobic amino acids which might be responsible for solvent stability (Karan et al., 2011). Solvent stability of the enzyme was also supported by the hydropathy profile of the amino acid sequence of APrBL, which reveals

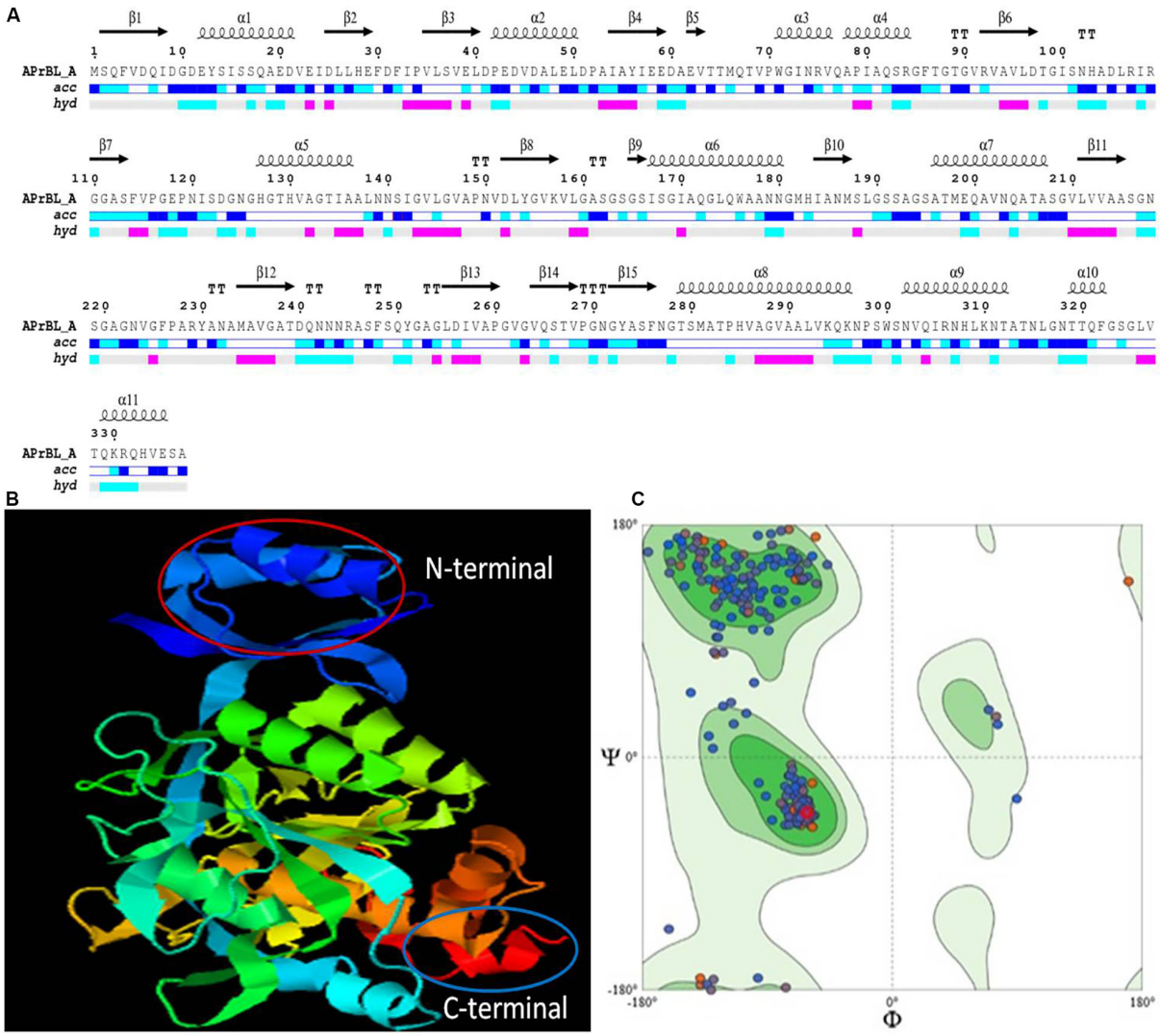


FIGURE 5 | (A) Secondary structure prediction of APrBL by ENDscript 2.0. On the top are secondary structures (squiggles: helices, arrows: beta strands, TT letters: beta turns). On the bottom, ***acc*** denotes bar of accessibility (dark blue: accessible, light blue: intermediate, white: buried) and ***hyd*** denotes a bar of hydropathy (pink: hydrophobic, light blue: hydrophilic, white neutral). **(B)** Three-dimensional (3D) structure showing N-terminal and C-terminal of APrBL generated using I-TASSER server. **(C)** Ramachandran plot for the predicted 3D structure of APrBL by SWISS-MODEL. A total of 95.22% of the residues are distributed in the most favored region which predicts its stability.

hydrophobic nature of the protein (**Supplementary Figure S3**). Interestingly, although high glycine and less proline content is a characteristic feature of the cold-adapted enzymes, it was observed in thermo active APrBL, which suggests that this feature may not be solely responsible for thermal/cold adaptation.

Secondary and Three-Dimensional Structure Modeling of the APrBL

Secondary structure of APrBL as analyzed by END script 2.0 indicated 10 α-helices and 16 β-sheets (**Figure 5A**). The SOPMA tool revealed that the APrBL contained 31.75% α-helices, 22.55% β-strands, and 45.70% coils, suggesting the dominance of the coils. The halophilic and alkaliphilic enzymes need to maintain a balance between enough rigidity to prevent the unfolding and flexibility to allow the motions necessary for the catalysis (Zorgani et al., 2014). Therefore, the coil regions in APrBL provide a flexible conformation enhancing its activity and stability under the saline and alkaline conditions. On the contrary, non-halophilic enzymes contain more α-helix and β-strand-forming regions (Zorgani et al., 2014). A similar trend was observed in an alkaline protease from *B. pumilus* (Baweja et al., 2016). Further, low K_m and high K_{cat} values of APrBL could be attributed to the dominance of coils which confers structural flexibility to the enzyme. Structural flexibility in the active site of the enzyme might facilitate efficient substrate binding and catalysis. Nevertheless, stability of APrBL at high temperatures and alkaline pH, solvents, and detergent is presumably due to the

α-helices and β-strands that provide rigidity to the structure (Fields et al., 2015).

It is hypothesized that a high number of hydrophobic amino acids facilitated APrBL folding in a way to generate more grooves to avoid water molecule. High hydrophobicity in protein promotes stability in stressful conditions. However, the balance of the hydrophobic and hydrophilic regions is important for the interaction of the protein with its surrounding medium, at the same time to increase the substrate binding capacity in the core region required for the efficient catalysis (Panja et al., 2020).

The tertiary structure model was generated by I-TASSER which combines the methods of threading, *ab initio* modeling, and structural refinement (**Figure 5B**). The 3D model explained that amino acids forming catalytic triad are in close proximity in tertiary structure but located far apart in primary structure. Therefore, appropriate folding of the enzyme brings amino acids together which are crucial to the catalytic activity in the tertiary structure of the enzyme. The structure had a good C score of 1.56 (range −5 to 2) and TM score of 0.93. A TM score of > 0.5 indicates a model of correct topology. The quality of 3D structural model was validated by structure analysis tool of SWISS-MODEL. High score value for local quality plot (> 0.6) supports the structure of the APrBL as a good feature. In addition, Ramachandran plot indicated that 95.22% of the amino acid residues are located in the most favored region, which further validates as a good quality model (**Figure 5C**).

Homology Modeling Analysis of APrBL

SWISS-MODEL workspace of expasy platform was used to search template for homology modeling. Homology search of APrBL has shown highest structural similarity of 80.97% with the crystal structure of alkaline M-protease (1wsd.1.A protein) from *Bacillus clausii* strain KSM-K16 (PubMed ID: 9278275). An alkaline M-protease optimally acts at pH 12.3 (Shirai et al., 1997). In homology modeling, QMEAN, a composite estimator, provides both global (i.e., for the entire structure) and local (i.e., per residue) quality estimates based on the single model (swissmodel.expasy.org). The QMEAN Z-score provides an estimate of the "degree of nativeness" of the structure which indicates whether this model would fit into the experimental structures of similar size (swissmodel.expasy.org). QMEAN *Z*-score near zero suggests correspondence between the model and the experimental structure. Lower scores are indicative of low quality of the model. A QMEAN *Z*-score of −0.12 for the APrBL supports the homology model.

Applications of the Recombinant Alkaline Protease APrBL

Wash Performance Analysis of APrBL

Stain removal ability of the APrBL was analyzed on cotton cloths. The washing test of the APrBL revealed significant removal of bloodstain from the cotton cloth when compared to controls (**Figure 6**). The washing efficiency of the detergents alone revealed that a complete stain removal with Ariel and Tide, while the detergents Surf, Nirma, and Wheel could not completely remove the stain. However, the stain was completely removed with the supplementation of the APrBL in the detergents (**Figure 6**). During the removal of the proteinaceous substrates in the cloths, the proteins are degraded into smaller peptides due to the action of the detergent and/or supplemented proteases. In case of the smaller peptides, they are either solubilized into the washing solution or get deposited back on the fabric. Therefore, the suitable enzymes with improved hydrolysis would aid the stain removal and prevent redeposition. In this investigation, the treatment with APrBL plus detergent not only removed the stain by hydrolyzing the protein but also prevented protein redeposition on the fabric. The application of alkaline proteases in bloodstain removal is described in the literature (Beg and Gupta, 2003; Jaouadi et al., 2008; Patil et al., 2016; Mechri et al., 2017; Rekik et al., 2019). Noticeably, APrBL efficiently removed the bloodstains within 10 min compared to 30–60 min by the previously reported proteases (Beg and Gupta, 2003; Jaouadi et al., 2008; Patil et al., 2016; Mechri et al., 2017; Rekik et al., 2019). These findings further add to the potential of APrBL in detergent industry.

Whey Protein Hydrolysis

Whey, a by-product of the dairy industry, provides important nutritional ingredients (Sinha et al., 2007). The most important proteins in whey are α-lactalbumin and β-lactoglobulin, representing 70–80% of its total protein content (Smithers et al., 1996). Besides, whey possesses antioxidant properties attributed to high presence of sulfur-containing amino acids. However, due to its high organic content, it is also considered as one of the most polluting agents of water bodies (Moreno-Indias et al., 2009).

Effect of APrBL on whey protein led to a decreased total protein content (**Figure 7A**). Hydrolysis of the whey protein results into small peptides of varied size and free amino acids. SDS–PAGE profile reveals action of APrBL in the gradual conversion of high-molecular-weight whey proteins into low-molecular-weight smaller peptides (**Figure 7C**). The protein hydrolysates exhibit antimicrobial, antihypertensive, antioxidant, and lipid-lowering properties (Silvestre et al., 2013). Hydrolyzed whey protein-based formulas are favored for infants intolerant to cow's milk protein (Sinha et al., 2007). In our study, 19% of the hydrolysis was achieved in 8 h by APrBL recombinant protease (**Figure 7B**). In majority of the cases, the DH for whey proteins varies in the range of 5–23% (Sinha and Khare, 2014). A 28 and 35% hydrolysis of whey protein was obtained by free and immobilized protease, respectively (Sinha and Khare, 2014). The protein hydrolysates provide growth-supportive peptides, eliminating the need of expensive nutrients in the growth medium (Pintado et al., 1999). The enzymatic processes provide environment-friendly processing conditions with fewer undesirable effects as compared to other conventional methods.

CONCLUSION

We successfully cloned an alkaline protease gene (APrBL) from *B. lehensis* JO-26 and optimally expressed in *E. coli* BL21

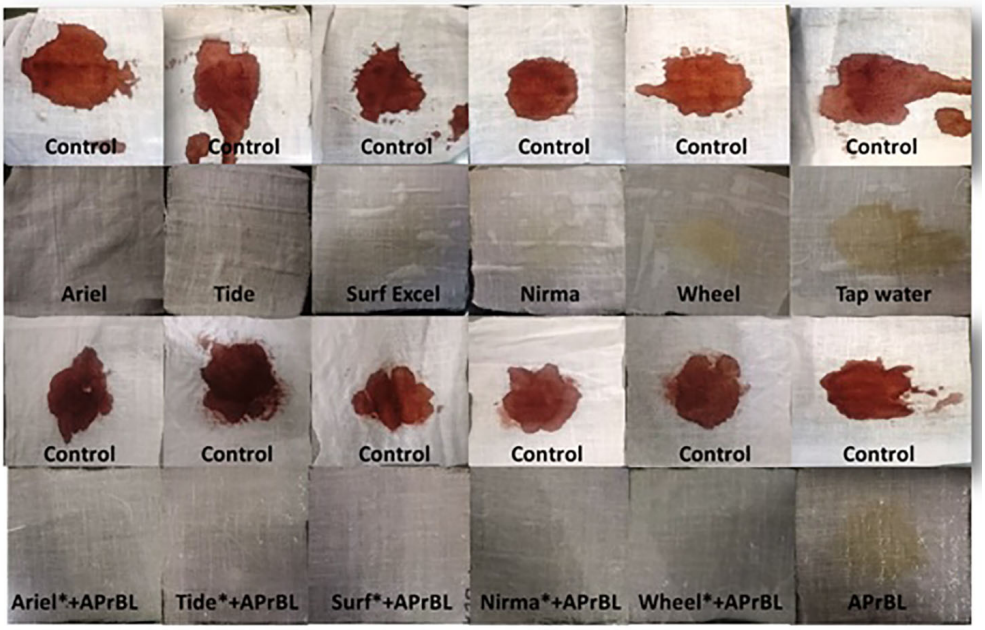


FIGURE 6 | Wash performance analysis showing different washing conditions for bloodstain removal.

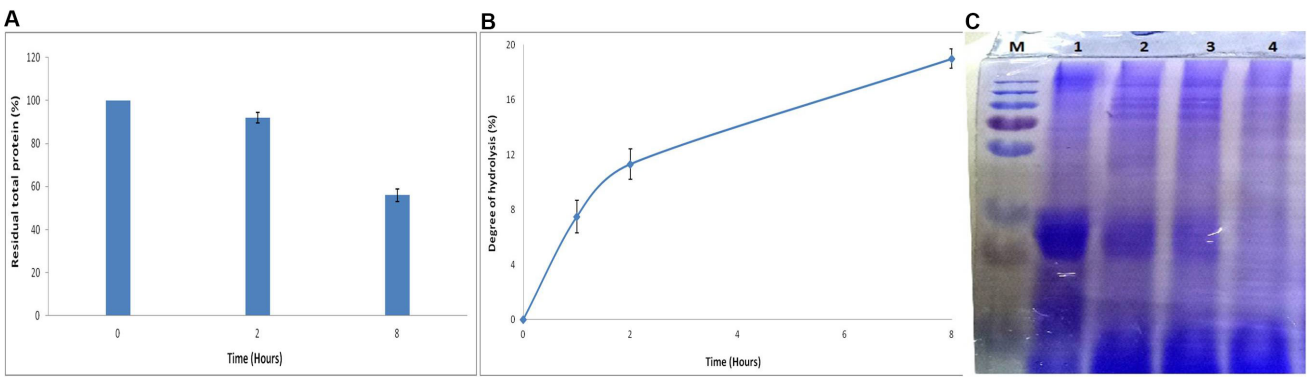


FIGURE 7 | (A) Effect of APrBL enzyme on total protein in milk whey. **(B)** Effect of APrBL enzyme on degree of hydrolysis of milk whey. **(C)** Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) profile of untreated and enzymatically hydrolyzed milk whey proteins by APrBL enzyme. M: Marker; Lane 1: untreated whey; Lane 2: treated, 1 h; Lane 3: treated, 2 h; Lane 4: treated, 8 h.

(DE3). The effect of NaCl on the expression of the alkaline protease was demonstrated. Maximum enzyme expression was achieved with 0.2 mM IPTG, 2% NaCl (w/v), and 28°C growth temperature. The APrBL has shown high activity and excellent stability over a wide range of pH and temperatures, with the tolerance against surfactants and reducing agents. Kinetic parameters K_m, V_{max}, and K_{cat} of the APrBL protease suggest its potential as a biocatalyst. Structural features of APrBL corroborate the experimental data. It was also evident that high glycine and reduced proline content are not solely responsible for the thermal/cold adaptation. The enzyme can also be used in peptide synthesis under non-aqueous environment due to its stability and activity in organic solvents. Applications of this enzyme as a detergent supplement and in whey protein hydrolysis clearly suggest its suitability in detergent and food industries. We understand this is the first report on the cloning, expression, characterization, and structural elucidation of an alkaline serine protease from a saline desert.

DATA AVAILABILITY STATEMENT

The alkaline protease gene sequence of Bacillus lehensis JO-26 (APrBL) has been submitted to GenBank (Accession No. MN104891).

AUTHOR CONTRIBUTIONS

HB and SS conceived and designed the experiments. HB performed the experiments, analyzed data and wrote the manuscript. SS helped in analysis and interpretation of the data. SS supervised the work.

ACKNOWLEDGMENTS

HB is thankful to the Council of Scientific and Industrial Research (CSIR), New Delhi, India, for the award of Senior Research Fellowship (CSIR–SRF) and UGC, New Delhi, for the award of BSR-Meritorious Fellowship. The authors are grateful to the University Grants Commission (UGC), India, and Saurashtra University for the infrastructural and financial support. Support under the UGC-Centre of Advanced Study, DST-FIST program and DBT-Multi Institutional Project is also acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00941/full#supplementary-material

REFERENCES

Annamalai, N., Rajeswari, M. V., and Balasubramanian, T. (2014). Extraction, purification and application of thermostable and halostable alkaline protease from *Bacillus alveayuensis* CAS 5 using marine wastes. *Food Bioprod Process* 92, 335–342. doi: 10.1016/j.fbp.2013.08.009

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Barberis, S., Quiroga, E., Morcelle, S., Priolo, N., and Luco, J. M. (2006). Study of phytoproteases stability in aqueous-organic biphasic systems using linear free energy relationships. *J. Mol. Catal.* 38, 95–103. doi: 10.1016/j.molcatb.2005.11.011

Baweja, M., Tiwari, R., Singh, P. K., Nain, L., and Shukla, P. (2016). An alkaline protease from *Bacillus pumilus* MP 27: functional analysis of its binding model toward its applications as detergent additive. *Front. Microbiol.* 7:1195. doi: 10.3389/fmicb.2016.01195

Beg, Q. K., and Gupta, R. (2003). Purification and characterization of an oxidation-stable, thiol-dependent serine alkaline protease from *Bacillus mojavensis*. *Enzyme Microb. Technol.* 32, 294–304. doi: 10.1016/S0141-0229(02)00293-4

Bhatt, H. B., Begum, M. A., Chintalapati, S., Chintalapati, V. R., and Singh, S. P. (2017). *Desertibacillus haloalkaliphilus* gen. nov., sp. nov., isolated from a saline desert. *Int. J. Syst. Evol. Microbiol.* 67, 4435–4442. doi: 10.1099/ijsem.0.002310

Bhatt, H. B., Gohel, S. D., and Singh, S. P. (2018). Phylogeny, novel bacterial lineage and enzymatic potential of haloalkaliphilic bacteria from the saline coastal desert of Little Rann of Kutch, Gujarat, India. *3 Biotech* 8:53. doi: 10.1007/s13205-017-1075-0

Bhatt, H. B., and Singh, S. P. (2016). "Phylogenetic and phenogram based diversity of haloalkaliphilic bacteria from the saline desert," in *Microbial Biotechnology*, eds B. Bhukya and A. Tangutur (New Jersey: Apple Academic Press), 373–386.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Briki, S., Hamdi, O., and Landoulsi, A. (2016). Enzymatic dehairing of goat skins using alkaline protease from *Bacillus* sp. *SB*12. *Protein Exp. Purif.* 121, 9–16. doi: 10.1016/j.pep.2015.12.021

Cheng, Q., Xu, F., Hu, N., Liu, X., and Liu, Z. (2015). A novel Ca2+-dependent alkaline serine-protease (Bvsp) from *Bacillus* sp. with high fibrinolytic activity. *J. Mol. Catal.* 117, 69–74. doi: 10.1016/j.molcatb.2015.04.006

Contesini, F. J., Melo, R. R. D., and Sato, H. H. (2018). An overview of *Bacillus* proteases: from production to application. *Crit Rev Biotechnol.* 38, 321–334. doi: 10.1080/07388551.2017.1354354

Deng, A., Wu, J., Zhang, G., and Wen, T. (2011). Molecular and structural characterization of a surfactant-stable high-alkaline protease AprB with a novel structural feature unique to subtilisin family. *Biochimie* 93, 783–791. doi: 10.1016/j.biochi.2011.01.011

Devi, S. G., Fathima, A. A., Sanitha, M., Iyappan, S., Curtis, W. R., and Ramya, M. (2016). Expression and characterization of alkaline protease from the metagenomic library of tannery activated sludge. *J. Biosci. Bioeng.* 122, 694–700. doi: 10.1016/j.jbiosc.2016.05.012

Dodia, M. S., Bhimani, H. G., Rawal, C. M., Joshi, R. H., and Singh, S. P. (2008). Salt dependent resistance against chemical denaturation of alkaline protease from a newly isolated haloalkaliphilic *Bacillus* sp. *Bioresour. Technol.* 99, 6223–6227. doi: 10.1016/j.biortech.2007.12.020

Dorra, G., Ines, K., Imen, B. S., Laurent, C., Sana, A., Olfa, T., et al. (2018). Purification and characterization of a novel high molecular weight alkaline protease produced by an endophytic *Bacillus halotolerans* strain CT2. *Int. J. Biol. Macromol.* 111, 342–351. doi: 10.1016/j.ijbiomac.2018.01.024

Doukyu, N., and Ogino, H. (2010). Organic solvent-tolerant enzymes. *Biochem. Eng.* 48, 270–282. doi: 10.1016/j.bej.2009.09.009

Fields, P. A., Dong, Y., Meng, X., and Somero, G. N. (2015). Adaptations of protein structure and function to temperature: there is more than one way to 'skin a cat'. *J. Exp. Biol.* 218, 1801–1811. doi: 10.1242/jeb.114298

Gallagher, T., Oliver, J., Bott, R., Betzel, C., and Gilliland, G. L. (1996). Subtilisin BPN'at 1.6 Å resolution: analysis for discrete disorder and comparison of crystal forms. *Acta Cryst. D* 52, 1125–1135.

Geourjon, C., and Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11, 681–684. doi: 10.1093/bioinformatics/11.6.681

Gohel, S. D., and Singh, S. P. (2012). Purification strategies, characteristics and thermodynamic analysis of a highly thermostable alkaline protease from a salt-tolerant alkaliphilic actinomycete. *Nocardiopsis alba* OK-5. J. Chromatogr. B 889, 61–68. doi: 10.1016/j.jchromb.2012.01.031

Gohel, S. D., and Singh, S. P. (2018). Thermodynamics of a Ca2+ dependent, highly thermostable and detergent compatible purified alkaline serine protease from *Nocardiopsis xinjiangensis* strain OM-6. *Int. J. Biol. Macromol.* 113, 565–574. doi: 10.1016/j.ijbiomac.2018.02.157

Gopal, G. J., and Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *Protein J.* 32, 419–425. doi: 10.1007/s10930-013-9502-5

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* 31, 3320–3323. doi: 10.1093/nar/gkg556

Guruprasad, K., Reddy, B. B., and Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng Des Sel.* 4, 155–161.

Haddar, A., Agrebi, R., Bougatef, A., Hmidet, N., Sellami-Kamoun, A., and Nasri, M. (2009). Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. *Bioresour. Technol.* 100, 3366–3373. doi: 10.1016/j.biortech.2009.01.061

Hagihara, B. (1958). *The Enzymes*, Vol. 4. New York, NY: Academic press Inc.

Haki, G. D., and Rakshit, S. K. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresour. Technol.* 89, 17–34. doi: 10.1016/S0960-8524(03)00033-6

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898. doi: 10.1093/oxfordjournals.jbchem.a133168

Jacobs, M., Eliasson, M., Uhlén, M., and Flock, J. I. (1985). Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res.* 13, 8913–8926. doi: 10.1093/nar/13.24.8913

Jain, D., Pancha, I., Mishra, S. K., Shrivastav, A., and Mishra, S. (2012). Purification and characterization of haloalkaline thermoactive, solvent stable and SDS-induced protease from *Bacillus* sp.: a potential additive for laundry detergents. *Bioresour. Technol.* 115, 228–236. doi: 10.1016/j.biortech.2011.10.081

Jaouadi, B., Ellouz-Chaabouni, S., Rhimi, M., and Bejar, S. (2008). Biochemical and molecular characterization of a detergent-stable serine alkaline protease from *Bacillus pumilus* CBS with high catalytic efficiency. *Biochimie* 90, 1291–1305. doi: 10.1016/j.biochi.2008.03.004

Jellouli, K., Ghorbel-Bellaaj, O., Ayed, H. B., Manni, L., Agrebi, R., and Nasri, M. (2011). Alkaline-protease from *Bacillus licheniformis* MP1: purification, characterization and potential application as a detergent additive and for shrimp waste deproteinization. *Process Biochem.* 46, 1248–1256. doi: 10.1016/j.procbio.2011.02.012

Joshi, S., and Satyanarayana, T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis*. *Bioresour. Technol.* 131, 76–85. doi: 10.1016/j.biortech.2012.12.124

Kaneko, R., Koyama, N., Tsai, Y. C., Juang, R. Y., Yoda, K., and Yamasaki, M. (1989). Molecular cloning of the structural gene for alkaline elastase YaB, a new subtilisin produced by an alkalophilic *Bacillus* strain. *J. Bacteriol.* 171, 5232–5236. doi: 10.1128/jb.171.9.5232-5236.1989

Karan, R., Singh, R. K., Kapoor, S., and Khare, S. K. (2011). Gene identification and molecular characterization of solvent stable protease from a moderately haloalkaliphilic bacterium. *Geomicrobium* sp. *EMB2*. J. Microbiol. Biotechnol. 21, 129–135. doi: 10.4014/jmb.1007.07064

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 37(Suppl._1), D387–D392. doi: 10.1093/nar/gkn750

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Klibanov, A. M. (2001). Improving enzymes by using them in organic solvents. *Nature* 409, 241–246. doi: 10.1038/35051719

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. doi: 10.1016/0022-2836(82)90515-0

Liu, W., Shi, P., Chen, Q., Yang, P., Wang, G., Wang, Y., et al. (2010). Gene cloning, overexpression, and characterization of a xylanase from *Penicillium* sp. *CGMCC* 1669. *Appl. Biochem. Biotechnol.* 162, 1–12. doi: 10.1007/s12010-009-8719-4

Masui, A., Fujiwara, N., Yamamoto, K., Takagi, M., and Imanaka, T. (1998). Rational design for stabilization and optimum pH shift of serine protease aprN. *J. Ferment. Bioengineer.* 85, 30–36. doi: 10.1016/S0922-338X(97)80349-2

Maurer, K. H. (2004). Detergent proteases. *Curr. Opin. Biotechnol.* 15, 330–334. doi: 10.1016/j.copbio.2004.06.005

Mechri, S., Berrouina, M. B. E., Benmrad, M. O., Jaouadi, N. Z., Rekik, H., Moujehed, E., et al. (2017). Characterization of a novel protease from *Aeribacillus pallidus* strain VP3 with potential biotechnological interest. *Int. J. Biol. Macromol.* 94, 221–232. doi: 10.1016/j.ijbiomac.2016.09.112

Morais, H. A., Silve, M. P. C., Silva, V. D. M., Silva, M. R., Cristina, A. S., and Silveira, J. N. (2013). Correlation between the degree of hydrolysis and the peptide profile of whey protein concentrate hydrolysates: effect of the enzyme type and reaction time. *Am. J. Food Technol.* 8, 1–16. doi: 10.3923/ajft.2013.1.16

Moreno-Indias, I., Castro, N., Morales-delaNuez, A., Sánchez-Macías, D., Assunção, P., Capote, J., et al. (2009). Farm and factory production of goat cheese whey results in distinct chemical composition. *J. Dairy Sci.* 92, 4792–4796. doi: 10.3168/jds.2009-2215

Panja, A. S., Maiti, S., and Bandyopadhyay, B. (2020). protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Sci. Rep.* 10, 1–9. doi: 10.1038/s41598-020-58825-7

Patel, R. K., Dodia, M. S., Joshi, R. H., and Singh, S. P. (2006). Purification and characterization of alkaline protease from a newly isolated haloalkaliphilic *Bacillus* sp. *Process. Biochem.* 41, 2002–2009. doi: 10.1007/s12257-010-0410-7

Patil, U., Mokashe, N., and Chaudhari, A. (2016). Detergent-compatible, organic solvent-tolerant alkaline protease from *Bacillus circulans* MTCC 7942: purification and characterization. *Prep Biochem Biotech.* 46, 56–64. doi: 10.1080/10826068.2014.979205

Pintado, M. E., Pintado, A. E., and Malcata, F. X. (1999). Controlled whey protein hydrolysis using two alternative proteases. *J. Food Eng.* 42, 1–13.

Purohit, M. K., and Singh, S. P. (2013). A metagenomic alkaline protease from saline habitat: cloning, over-expression and functional attributes. *Int. J. Biol. Macromol.* 53, 138–143. doi: 10.1016/j.ijbiomac.2012.10.032

Purohit, M. K., and Singh, S. P. (2014). Cloning, over expression and functional attributes of serine proteases from *Oceanobacillus iheyensis* OM A18 and Haloalkaliphilic bacterium OM E12. *Process Biochem.* 49, 61–68. doi: 10.1016/j.procbio.2013.07.009

Qoura, F., Kassab, E., Reisse, S., Antranikian, G., and Brueck, T. (2015). Characterization of a new, recombinant thermo-active subtilisin-like serine protease derived from *Shewanella arctica*. *J. Mol. Catal.* 116, 16–23. doi: 10.1016/j.molcatb.2015.02.015

Raval, V. H., Bhatt, H. B., and Singh, S. P. (2018). "Adaptation Strategies in Halophilic Bacteria," in *Extremophiles*, eds C. Gerday and N. Glansdorff (Boca Raton, FL: CRC Press), 137–164.

Raval, V. H., Pillai, S., Rawal, C. M., and Singh, S. P. (2014). Biochemical and structural characterization of a detergent-stable serine alkaline protease from seawater haloalkaliphilic bacteria. *Process. Biochem.* 49, 955–962. doi: 10.1016/j.procbio.2014.03.014

Raval, V. H., Rawal, C. M., Pandey, S., Bhatt, H. B., Dahima, B. R., and Singh, S. P. (2015). Cloning, heterologous expression and structural characterization of an alkaline serine protease from sea water haloalkaliphilic bacterium. *Ann Microbiol.* 65, 371–381. doi: 10.1007/s13213-014-0869-0

Rekik, H., Frikha, F., Jaouadi, N. Z., Gargouri, F., Jmal, N., Bejar, S., et al. (2019). Gene cloning, expression, molecular modeling and docking study of the protease SAPRH from *Bacillus safensis* strain RH12. *Int. J. Biol. Macromol.* 125, 876–891. doi: 10.1016/j.ijbiomac.2018.12.103

Rosano, G. L., and Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5:172. doi: 10.3389/fmicb.2014.00172

Sato, M., Yoshikawa, K., and Minagawa, M. (1990). The effect of builders on the activity of protease enzymes. *J. Am. Oil Chem. Soc.* 67, 711–716. doi: 10.1007/BF02540477

Sharma, A. K., Kikani, B. A., and Singh, S. P. (2020). Biochemical, thermodynamic and structural characteristics of a biotechnologically compatible alkaline protease from a haloalkaliphilic. *Nocardiopsis dassonvillei* OK-18. *Int. J. Biol. Macromol.* 153, 680–696. doi: 10.1016/j.ijbiomac.2020.03.006

Shirai, T., Suzuki, A., Yamane, T., Ashida, T., Kobayashi, T., Hitomi, J., et al. (1997). High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Protein Eng.* 10, 627–634. doi: 10.1093/protein/10.6.627

Shu, G., Huang, J., Bao, C., Meng, J., Chen, H., and Cao, J. (2018). Effect of different proteases on the degree of hydrolysis and angiotensin I-Converting enzyme-inhibitory activity in goat and cow milk. *Biomolecules* 8:101. doi: 10.3390/biom8040101

Siezen, R. J., and Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* 6, 501–523. doi: 10.1002/pro.5560060301

Silvestre, M. P. C., Morais, H. A., Silva, V. D. M., and Silva, M. R. (2013). Whey as source of peptides with high antioxidant activity: use of a pancreatin and an *Aspergillus sojae* protease. *Publicatio UEPGCiênc. Biol. Saúde* 19, 143–147.

Singh, S. P., Purohit, M. K., Aoyagi, C., Kitaoka, M., and Hayashi, K. (2010). Effect of growth temperature, induction, and molecular chaperones on the solubilization of over-expressed cellobiose phosphorylase from *Cellvibrio gilvus* under in vivo conditions. *Biotechnol. Bioproc. E* 15, 273–276. doi: 10.1007/s12257-009-0023-1

Sinha, R., and Khare, S. K. (2013). Characterization of detergent compatible protease of a halophilic *Bacillus* sp. *EMB*9: differential role of metal ions in stability and activity. *Bioresour. Technol.* 145, 357–361. doi: 10.1016/j.biortech.2012.11.024

Sinha, R., and Khare, S. K. (2014). Effect of organic solvents on the structure and activity of moderately halophilic *Bacillus* sp. *EMB*9 protease. *Extremophiles* 18, 1057–1066. doi: 10.1007/s00792-014-0683-4

Sinha, R., and Khare, S. K. (2015).). Immobilization of halophilic *Bacillus* sp. *EMB*9 protease on functionalized silica nanoparticles and application in whey protein hydrolysis. *Bioproc. Biosyst. Eng.* 38, 739–748. doi: 10.1007/s00449-014-1314-2

Sinha, R., Radha, C., Prakash, J., and Kaul, P. (2007). Whey protein hydrolysate: functional properties, nutritional quality and utilization in beverage formulation. *Food Chemistry* 101, 1484–1491.

Smithers, G. W., Ballard, F. J., Copeland, A. D., de Silva, K. J., Dionysius, D. A., Francis, G. L., et al. (1996). New opportunities from the isolation and utilization of whey proteins. *J. Dairy Sci.* 79, 1454–1459. doi: 10.3168/jds.S0022-0302(96)76504-9

Sørensen, H. P., and Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fac.* 4:1. doi: 10.1186/1475-2859-4-1

Strausberg, S. L., Alexander, P. A., Gallagher, D. T., Gilliland, G. L., Barnett, B. L., and Bryan, P. N. (1995). Directed evolution of a subtilisin with calcium-independent stability. *Bio Technol.* 13, 669–673. doi: 10.1038/nbt0795-669

Takenaka, S., Yoshida, N., Yoshida, K. I., Murakami, S., and Aoki, K. (2011). Molecular cloning and sequence analysis of two distinct halotolerant extracellular proteases from *Bacillus subtilis* FP-133. *Biosci. Biotechnol. Biochem.* 75, 148–151. doi: 10.1271/bbb.100588

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tavano, O. L. (2013). Protein hydrolysis using proteases: an important tool for food biotechnology. *J. Mol. Catal. B Enzymatic* 90, 1–11. doi: 10.1016/j.molcatb.2013.01.011

Thakrar, F. J., and Singh, S. P. (2019). Catalytic, thermodynamic and structural properties of an immobilized and highly thermostable alkaline protease from a haloalkaliphilic actinobacteria. *Nocardiopsis alba TATA*-5. *Bioresour. Technol.* 278, 150–158. doi: 10.1016/j.biortech.2019.01.058

Thumar, J. T., and Singh, S. P. (2009). Organic solvent tolerance of an alkaline protease from salt-tolerant alkaliphilic *Streptomyces clavuligerus* strain Mit-1. *J. Ind. Microbiol. Biot.* 36:211. doi: 10.1007/s10295-008-0487-6

Van Kampen, V., and Merget, R. (2002). Berufliche Atemwegssensibilisierungen durch Subtilisine. *Pneumologie* 56, 182–186. doi: 10.1055/s-2002-20552

Walstra, P., and Jenness, R. (1984). *Dairy Chemistry & Physics*. Hoboken, NJ: John Wiley & Sons.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12, 7–8. doi: 10.1038/nmeth.3213

Zhang, H., Mu, H., Mo, Q., Sun, T., Liu, Y., Xu, M., et al. (2016). Gene cloning, expression and characterization of a novel cold-adapted protease from *Planococcus* sp. *J. Mol. Catal.* 130, 1–8. doi: 10.1016/j.molcatb.2016.04.002

Zhou, C., Qin, H., Chen, X., Zhang, Y., Xue, Y., and Ma, Y. (2018). A novel alkaline protease from alkaliphilic *Idiomarina* sp. *C*9-1 with potential application for eco-friendly enzymatic dehairing in the leather industry. *Sci. Rep.* 8, 1–18. doi: 10.1038/s41598-018-34416-5

Zorgani, M. A., Patron, K., and Desvaux, M. (2014). New insight in the structural features of haloadaptation in α-amylases from halophilic Archaea following homology modeling strategy: folded and stable conformation maintained through low hydrophobicity and highly negative charged surface. *J. Comput. Aided Mol. Des.* 28, 721–734. doi: 10.1007/s10822-014-9754-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

MINI REVIEW
published: 02 July 2020
doi: 10.3389/fmicb.2020.01263

New Insights Into DNA Repair Revealed by NucS Endonucleases From Hyperthermophilic Archaea

Likui Zhang[1,4], Donghao Jiang[1], Mai Wu[1], Zhihui Yang[2]* and Philippe M. Oger[3]**

[1]College of Environmental Science and Engineering, Marine Science and Technology Institute, Yangzhou University, Yangzhou, China, [2]College of Plant Protection, Agricultural University of Hebei, Baoding, China, [3]Univ Lyon, INSA de Lyon, CNRS UMR 5240, Villeurbanne, France, [4]Guangling College, Yangzhou University, Yangzhou, China

Hyperthermophilic Archaea (HA) thrive in high temperature environments and their genome is facing severe stability challenge due to the increased DNA damage levels caused by high temperature. Surprisingly, HA display spontaneous mutation frequencies similar to mesophilic microorganisms, thereby indicating that the former must possess more efficient DNA repair systems than the latter to counteract the potentially enhanced mutation rates under the harsher environment. Although a few repair proteins or enzymes from HA have been biochemically and structurally characterized, the molecular mechanisms of DNA repair of HA remain largely unknown. Genomic analyses of HA revealed that they lack MutS/MutL homologues of the mismatch repair (MMR) pathway and the recognition proteins of the nucleotide excision repair (NER) pathway. Endonucleases play an essential role in DNA repair. NucS endonuclease, a novel endonuclease recently identified in some HA and bacteria, has been shown to act on branched, mismatched, and deaminated DNA, suggesting that this endonuclease is a multifunctional enzyme involved in NER, MMR, and deaminated base repair in a non-canonical manner. However, the catalytic mechanism and the physiological function of NucS endonucleases from HA need to be further clarified to determine how they participate in the different DNA repair pathways in cells from HA. In this review, we focus on recent advances in our understanding of the function of NucS endonucleases from HA in NER, MMR, and deaminated DNA repair, and propose directions for future studies of the NucS family of endonucleases.

Keywords: hyperthermophilic Archaea, NucS endonuclease, mismatch repair, nucleotide excision repair, deaminated DNA repair

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Xi-Peng Liu,
Shanghai Jiao Tong University, China
Peter Friedhoff,
Justus-Liebig
University of Giessen, Germany

****Correspondence:***
Likui Zhang
lkzhang@yzu.edu.cn
Zhihui Yang
bdyzh@hebau.edu.cn
Philippe M. Oger
philippe.oger@insa-lyon.fr

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *18 February 2020*
Accepted: *18 May 2020*
Published: *02 July 2020*

Citation:
Zhang L, Jiang D, Wu M, Yang Z and Oger PM (2020) New Insights Into DNA Repair Revealed by NucS Endonucleases From Hyperthermophilic Archaea. Front. Microbiol. 11:1263. doi: 10.3389/fmicb.2020.01263

INTRODUCTION

Archaea compose the third domain of life, and the second domain of the prokaryotes. As such, they have a single-cell ultrastructure with no cell nucleus, similar to bacteria. However, Archaea share high similarities to eukaryotic cells for many central cellular processes, such as DNA replication and RNA polymerization, suggesting that they are a simplified version of eukaryotes for several functions, which is coherent with the currently proposed origins of eukaryotes from an archaeal ancestor (Eme et al., 2017). Archaeal DNA replication, DNA repair, and DNA recombination (White and Allers, 2018) are also share extensive similarities

to that of eukaryotes. Hyperthermophilic Archaea (HA) are the Archaea with an optimal growth temperature above 80°C (Stetter, 2013), where they live in high temperature environments such as high-pressure hydrothermal vents, volcanic vents, and hot springs. Since the isolation of the first hyperthermophilic crenarchaeon *Sulfolobus acidocaldarius* from Yellowstone Park, more than 90 species of HA have been identified, comprising most members of the Euryarchaea and the Crenarchaea (Stetter, 2013; Eme et al., 2017). Studying HA helps us to understand adaptation mechanisms to life in high temperature environments and develop thermostable enzyme resources, and also provides new insights into the origin and evolution of life.

Endogenous and exogenous factors are capable of causing DNA damage, following which the replication of the damaged DNA would lead to serious consequences for the cells, such as the accumulation of mutations and eventual cell death, if it is not repaired. Thus, efficient DNA repair is an essential cellular function to maintain genomic integrity. The high temperature environments in which HA thrive can further increase their genomic DNA damages, and thus increase drastically the challenge their genomes is facing (Grogan, 1998, 2015). Firstly, high temperature accelerates the deamination rates of bases in DNA (Lindahl and Nyberg, 1974), forming deaminated bases. Hypoxanthine, xanthine, and uracil are common deaminated bases, which are created by deamination of adenine, guanine, and cytosine, respectively. Secondly, high temperature also enhances the hydrolysis rates of DNA bases, potentially leading to elevated apurinic/apyrimidinic (AP) site levels in the cells of HA (Lindahl and Nyberg, 1972). For instance, the hyperthermophilic euryarchaeon *Pyrococcus abyssi* harbors a level of AP sites 10-fold higher than that of *Escherichia coli* (Palud et al., 2008). However, how HA maintain their genome integrity remains elusive (Grogan, 2000).

The molecular mechanisms of DNA repair, including nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), homologous recombination (HR) repair, and non-homologous end joining (NHEJ), have been extensively studied. These pathways are essentially conserved from prokaryotes to eukaryotes (Zatopek et al., 2018). As mentioned above, although they live in high temperature environments and experience higher mutagenic potential, HA harbor spontaneous mutation rates similar to those of *E. coli* (Jacobs and Grogan, 1997; Grogan et al., 2001). Thus, HA are expected to have developed more efficient repair systems than other organisms.

To date, the DNA repair pathways of HA have been receiving much attention because they remain to be fully understood. Genomic analyses of HA showed that they could encode a variety of proteins involved in the NER, BER, alkyl transfer, damage reversion, and TLS pathways (Stetter, 2013), but they lack MutS/MutL homologues of the conserved MMR pathway and the recognition proteins of the conserved NER pathway (Grogan, 2004). Although biochemical and structural data of several DNA repair proteins from HA have provided some insights into the catalytic mechanisms and repair pathways (Stetter, 2013), most of the DNA repair mechanisms of HA remain unclear. Endonucleases play an important role in DNA repair pathway since they can cleave the damaged DNA (Li et al., 2019).

TABLE 1 | Substrate specificity of NucS endoncleases from hyperthermophilic Archaea (HA).

Organism	Substrate specificity	References
Pyrococcus abyssi	Branched and splayed DNA	(Ren et al., 2009)
Thermococcus kodakarensis	Mismatched DNA; I-containing dsDNA	(Ishino et al., 2016)
Thermococcus gammatolerans	U- and I-containing dsDNA; branched and splayed DNA; mismatched DNA	(Zhang et al., 2020)

Recently, a novel endonuclease, NucS [nuclease specific for single-stranded DNA (ssDNA)], was identified and characterized in several HA and some bacteria. The biochemical, structural, and genetic analyses showed that NucS endonucleases from HA are multifunctional endonucleases that can cleave branched DNA, mismatched, and deaminated, potentially providing an alternative pathway for the removal of mismatches, deaminated bases, and helix-distorting DNA lesions from DNA in HA. The substrate specificity of NucS endonucleases from HA is summarized in **Table 1**.

ARCHAEAL NucS ENDONUCLEASES IN NUCLEOTIDE EXCISION REPAIR

It is now well-known that NER is employed to repair helix-distorting DNA lesions, including UV-induced DNA damage (TT dimer and 6,4-photoproduct), intra-strand crosslinks, and bulky adducts (Grogan, 2015; White and Allers, 2018). NER proceeds in three steps: damage recognition, dsDNA opening by a helicase, and DNA cutting on both sides of the lesion by an endonuclease. Genomic analysis of HA have demonstrated that they possess homologues of only a fraction of the eukaryotic NER proteins, which comprise the XPF and XPG endonucleases, and the XPB and XPD helicases (Rouillon and White, 2011; Grogan, 2015; White and Allers, 2018) while they lack those for damage recognition proteins XPC and XPA (Grogan, 2015; White and Allers, 2018). Thus, it is widely admitted that HA do not possess a classic NER system as observed in other organisms and that their corresponding repair might involve other mechanisms.

Helix-distorting DNA damage can lead to stalled-fork DNA. To repair such stalled fork, HA could use a HR-mediated stalled-fork DNA repair pathway (Fujikane et al., 2010; Grogan, 2015). The stalled replication fork can be cleaved by a flap endonuclease that can act on branched DNA, and the corresponding lesion is then removed by double-stranded end processing. Genetic and biochemical data suggest that the 3'-flap endonuclease XPF/Hef (Hef represents euryarchaeal XPF) participates in an initial step of HR-mediated stalled-fork DNA repair (Nishino et al., 2005; Roberts and White, 2005). However, the 5'-flap endonuclease XPG/FEN-1 and other flap endonucleases do not seem to be involved in the helix-distorting DNA repair pathways of HA.

HA genome sequences showed that they possess structure-specific endonucleases (Grogan, 2004), which might play a

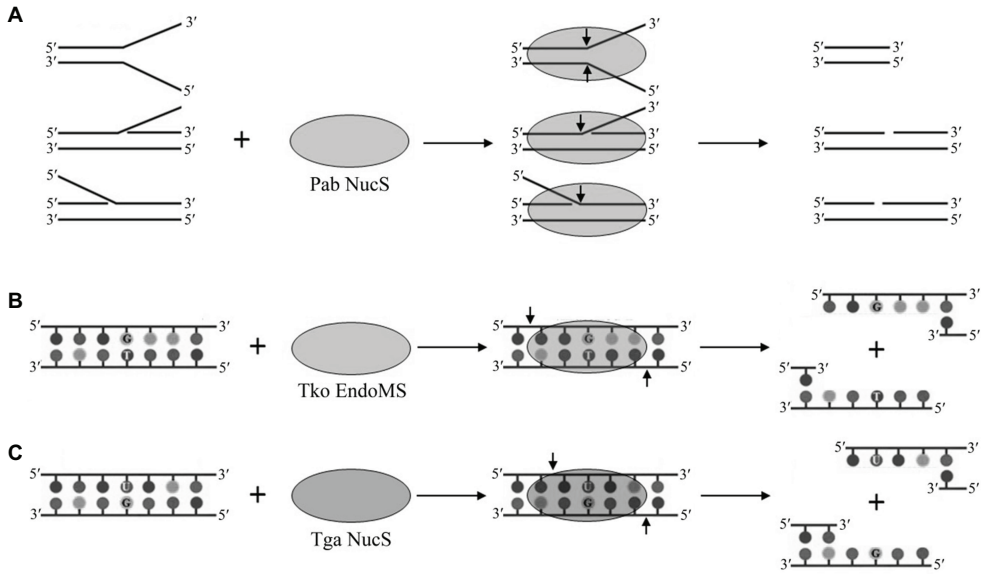


FIGURE 1 | Cleavage of branched, mismatched, and deaminated DNA by archaeal NucS endonucleases. **(A)** Cleavage of branched DNA by Pab NucS. **(B)** Cleavage of mismatched DNA by Tko EndoMS. **(C)** Cleavage of deaminated DNA by Tga NucS.

potential role in the recognition of helix-distorting DNA damage in their putative NER pathway, including the NucS endonuclease. NucS endonuclease from the hyperthermophilic euryarchaeon *P. abyssi* was first characterized *in vitro*, demonstrating that it can cleave the branched-structure DNA substrates that contain flapped and splayed DNA by removing the large ssDNA segments at the branch points (**Figure 1A**; Ren et al., 2009; Creze et al., 2011, 2012). The cleavage of flapped and splayed DNA substrates by Pab NucS suggests that this endonuclease is a potential archaeal NER endonuclease that can remove helix-distorting DNA damages (Rouillon and White, 2011). Furthermore, Rezgui et al. (2014) revealed that NucS endonuclease is a bipolar structure-specific endonuclease. *In vivo*, the engineered NucS-knockout strains of *S. acidocaldarius* are sensitive to DNA adducts, however, they did not show higher mutation rates (Suzuki and Kurosawa, 2019). Sac NucS does not seem to be involved in the DNA repair of UV-induced DNA damage, double strand break (DSB), or oxidative DNA damage, but this endonuclease seems essential for DNA repair of intra-strand crosslinks and works with XPF *via* the homologous recombination-mediated stalled-fork DNA repair for the removal of helix-distorting DNA lesions (Suzuki and Kurosawa, 2019). Combined with the cleavage of branched DNA substrates by Pab NucS, the role of Sac NucS *in vivo* in *Sulfolobus* cells suggests that the archaeal NucS endonuclease may be important in the NER pathway as a flap endonuclease.

ARCHAEAL NucS ENDONUCLEASES IN MISMATCH REPAIR

Most HA lack a complete mismatch repair machinery as the one conserved in bacteria and eukaryotes since they have no homologs of MutS and MutL (Grogan, 2004), the genes which are essential to recognize and remove mismatched bases. The search for an alternative MMR pathway in the hyperthermophilic euryarchaeon *Pyrococcus furiosus* leads to the isolation of a novel endonuclease, which was capable of cleaving mismatch-specific DNA in a non-canonical manner. This enzyme was named EndoMS, suggesting that this endonuclease is involved in DNA mismatch repair (Ishino et al., 2016). EndoMS proved to be identical to the NucS endonuclease previously characterized in *P. abyssi* (Ren et al., 2009). *Thermococcus kodakarensis* EndoMS (Tko EndoMs) is capable of cleaving a DNA double strand with a mismatched base, leaving a 5 base 5' overhang (**Figure 1B**; Ishino et al., 2016). Tko EndoMS displays a higher affinity for mismatched DNA than for branched or ssDNA. Furthermore, Tko EndoMS exhibits mismatched cleavage activity against G:T, G:G, T:T, T:C, and A:G mismatches, but cannot cleave C:C, A:C, or A:A mismatches *in vitro* (Ishino et al., 2016), which is consistent with the involvement of this enzyme in the excision of mismatched G or T. Structural data of Tko EndoMS suggest that the mismatched DNA substrate is wrapped by this endonuclease (Nakae et al., 2016), and two mismatched bases are flipped from the DNA substrate and cleaved by this endonuclease in a manner similar to type II restriction endonucleases (Mariko and Kosuke, 2016).

Besides HA and halophilic Archaea, a few bacteria that lack the MutS/MutL homologues also encode a NucS endonuclease (**Figure 2A**). Recent studies revealed that the deletion of the NucS endonuclease gene in *Mycobacterium smegmatis* leads to about 100-fold increased mutation rates, thereby creating a hypermutational phenotype (Castaneda-Garcia et al., 2017). Further mutational spectrum data demonstrated that the increased

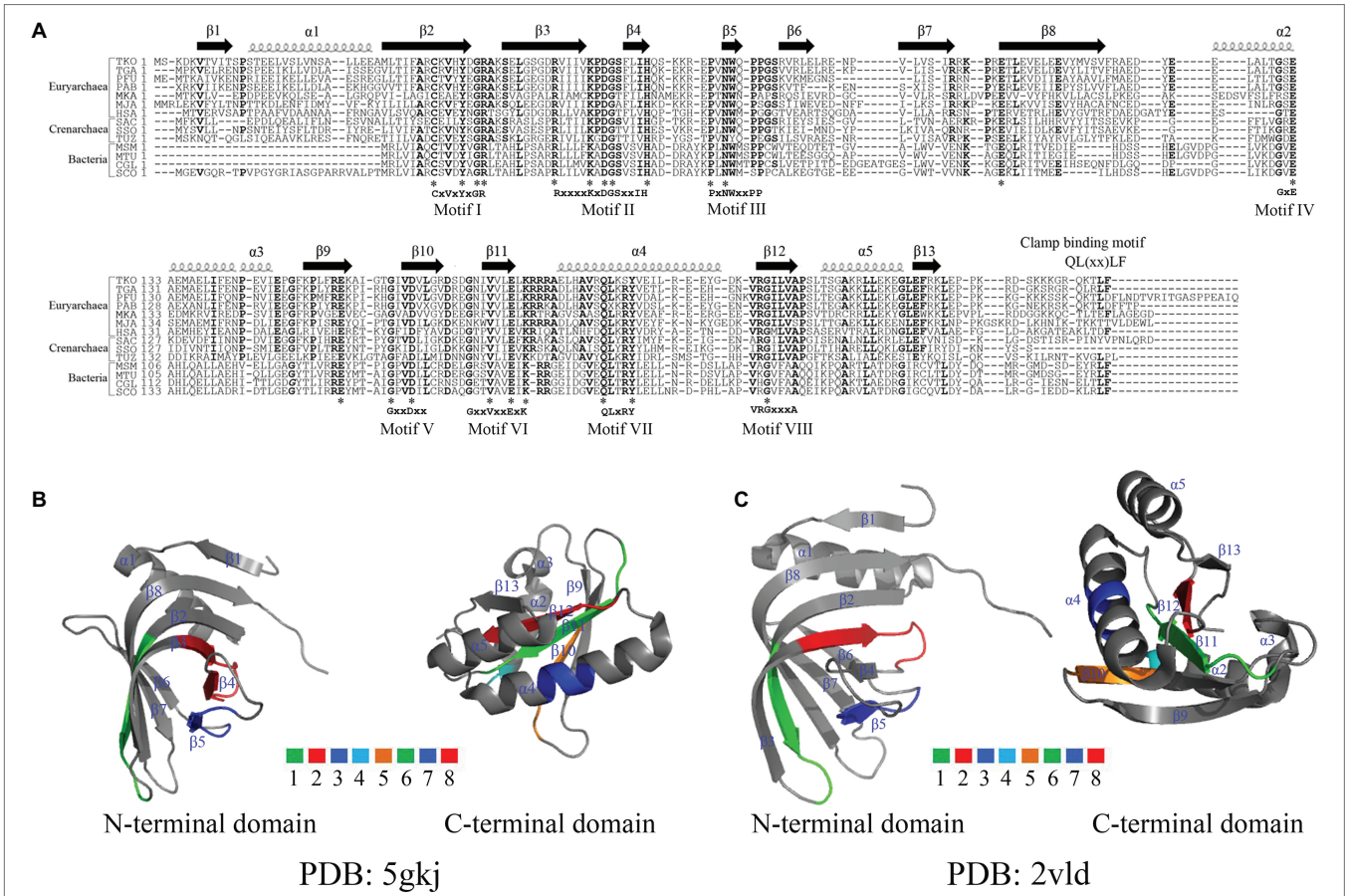


FIGURE 2 | Alignment of amino acid residues and crystal structures of NucS endonucleases. **(A)** Alignment of amino acid residues of NucS endonucleases. Abbreviations used are as follows: TKO, *T. kodakarensis* (NCBI reference sequence: WP_011250849.1); TGA, *T. gammatolerans* (NCBI reference sequence: WP_015857754.1); PFU, *Pyrococcus furiosus* (NCBI reference sequence: WP_014835498.1); PAB, *Pyrococcus abyssi* (PDB: 2VLD_B); MKA, *Methanopyrus kandleri* (UniProtKB/Swiss-Prot: Q8TY00.1); MJA, *Methanocaldococcus jannaschii* (UniProtKB/Swiss-Prot: Q57678.1); HSA, *Halobacterium salinarum* (GenBank: QCC45644.1); SAC, *Sulfolobus acidocaldarius* (UniProtKB/Swiss-Prot: Q4JC60.1); SSO, *Sulfolobus solfataricus* (UniProtKB/Swiss-Prot: Q97WK8.1); TUZ, *Thermofilum uzonense* (NCBI reference sequence: WP_052884192.1); MSM, *Mycolicibacterium smegmatis* (NCBI reference sequence: WP_003896320.1); MTU, *Mycobacterium tuberculosis* (GenBank: SIP67590.1); CGL, *Corynebacterium glutamicum* (GenBank: QDQ24257.1); SCO, *Streptomyces coelicolor* (NCBI reference sequence: WP_003973608.1). The conserved amino acid residues are labeled in bold. *Represents same amino acid residues in the selected organisms. **(B,C)** The crystal structures and conserved motifs of Tko EndoMS **(B)** and Pab NucS **(C)**. Tko EndoMS and Pab NucS are represented with α-helix and β-sheet, which are adapted from the references (Ren et al., 2009 and Nakae et al., 2016) by PyMOL. The conserved motifs in **(B)** and **(C)** are labeled with colors.

mutation rates are due to elevated levels of base conversion (A:T to G:C conversion or G:C to A:T conversion), which are the mutations created by a typical MMR defect (Castaneda-Garcia et al., 2017). Similar effects were observed in *Corynebacterium glutamicum* and *Streptomyces coelicolor* (Ishino et al., 2018; Takemoto et al., 2018). The *C. glutamicum* EndoMS was further confirmed as the mismatch-specific endonuclease and its function was fully dependent on its physical interaction with the sliding clamp. *In vivo*, a combination of the *C. glutamicum* EndoMS gene disruption and a mutation in the *dnaE* gene can lead to the increased mutation rates, thereby confirming the role of EndoMS in replication error correction (Ishino et al., 2018). *In vitro*, the *C. glutamicum* EndoMS can specifically cleave G/T, G/G, and T/T mismatches (Ishino et al., 2018), consistent with the mutation spectrum observed by genome-wide analyses. Thus, the bacterial NucS endonucleases resemble their archaeal homologues, and are capable of acting on mismatched bases in DNA to potentially complete the MMR DNA repair pathway.

ARCHAEAL NucS ENDONUCLEASES IN DEAMINATED BASE REPAIR

As mentioned above, high temperature leads to the accumulation of increased levels of uracil, hypoxanthine, or AP sites in genomic DNA. BER present in all organisms is a classical pathway for repairing the above-mentioned damaged bases (Grasso and Tell, 2014). Besides BER, the alternative excision repair (AER) pathway was identified as an alternative route to repair uracil, hypoxanthine, or AP sites. Generally, the AER pathway is initiated by an endonuclease that can cleave a

phosphodiester bond of the DNA in the vicinity of the damaged base, thus creating a nick (Yasui, 2013). The endonuclease V (EndoV) was the first endonuclease described to mediate the AER pathway, capable of recognizing and cleaving hypoxanthine-containing DNA with a cleavage site downstream of the damaged base of the second phosphodiester bond. The genomes of most HA encode a conserved EndoV suggesting that they can use the AER pathway for the repair of uracil, hypoxanthine, or AP sites (Liu et al., 2000; Kanugula et al., 2005; Kiyonari et al., 2014; Wang et al., 2018). The HA EndoVs can cleave hypoxanthine-containing DNA in a similar manner as other EndoVs.

Recently, another endonuclease, EndoQ was characterized from the hyperthermophilic euryarchaeon *P. furiosus* (Ishino et al., 2015). Similar to EndoVs, Pfu EndoQ is capable of cleaving the 5′-end DNA phosphodiester bond of uracil, hypoxanthine, or abasic sites forming a nick (Shiraishi et al., 2015). Further studies suggested that EndoQ and EndoV from *P. furiosus* act alone on the repair of damaged bases (Ishino et al., 2015). Interestingly, two incisions on either side of the damaged base by EndoQ and EndoV would generate a short hypoxanthine-containing ssDNA fragment, which could subsequently be repaired by a DNA helicase, a DNA polymerase, and a DNA ligase (Shiraishi et al., 2015). However, Ishino et al. (2015) suggested that EndoV and EndoQ are not involved in the same DNA repair pathway, which leaves open the question of how the AER pathway might work in HA.

Recently, our lab characterized the NucS endonuclease from the radioresistant hyperthermophilic euryarchaeon *Thermococcus gammatolerans* (Tga NucS). In addition to previously described activities of G and T mismatched DNA repair, we found that the Tga NucS can also recognize and cleave DNA with deaminated bases (uracil and hypoxanthine; **Figure 1C**; Zhang et al., 2020). Despite a very close proximity with Tko EndoMS (86% similarity), Tga NucS displays very divergent capabilities. Firstly, Tga NucS can cleave U- and I-containing dsDNA on both strands to produce a double strand break with a 4-nt 5'-overhang (**Figure 1C**), rather than a 5-nt 5'-overhang as observed for Tko EndoMS (Ishino et al., 2016). Both Tko EndoMS and Tga NucS can cleave hypoxanthine-containing DNA with the same efficiency as G/T mismatch (Ishino et al., 2016; Zhang et al., 2020). Tga NucS is more thermoactive than Tka EndoMS with a higher activity at physiological temperature, e.g., 80°C (Zhang et al., 2020), when Tko EndoMS works optimally at 55°C (Ishino et al., 2016). Taken together, it is clear that the Tga NucS exhibits several distinct biochemical properties from canonical NucS/EndoMS, which lead us to propose that Tga NucS might be involved in a novel alternative pathway for the repair of deaminated bases (uracil and hypoxanthine), in addition to BER and AER.

STRUCTURE AND FUNCTION OF ARCHAEAL NucS ENDONUCLEASES

Genomic analyses revealed that NucS endonucleases are widely distributed in Euryarchaea, Crenarchaea, and bacteria. Alignments of the amino acid sequences of NucS endonucleases from Archaea and bacteria showed that they possess eight highly conserved motifs, comprising mostly negatively and positively charged amino acid residues (**Figure 2A**). In addition, NucS endonucleases from HA harbor have a conserved QLxxLF motif, which is known to be necessary for the interaction with proliferating cell nuclear antigen (PCNA), a clamp protein as a DNA polymerase processivity (Creze et al., 2012; Ishino et al., 2018). Pab NucS interacts with Pab PCNA and the latter can modulate the activity of the former (Creze et al., 2012). Similar effects were observed in *T. kodakarensis* (Grasso and Tell, 2014; Ishino et al., 2018).

The structure of the EndoMS/NucS proteins from *T. kodakarensis* and *P. abyssi* have been solved recently (**Figures 2B,C**; Ren et al., 2009; Nakae et al., 2016). The data show that Tko EndoMS and Pab NucS share similar overall structures, possessing 13 β-sheet and 5 α-helix structures (**Figures 2B,C**; Ren et al., 2009; Nakae et al., 2016), however, the location and length of these β-sheet and α-helix structures differ between the two endonucleases. The ssDNA-binding sites for the flapped-DNA substrate (Model I) for the Pab NucS are the conserved residues W75 and Y39, which might forge stacking interactions with DNA. The ssDNA-binding sites of the splayed-DNA substrate (Model II) are residues D160, E174, K176, K44, E127, Q187, and Y191, which is a conserved motif also present in recB (Aravind et al., 2000). Mutational analysis showed that residue Y191 is a key active site residue, and residues R42, R70, and W75 are DNA binding residues of Pab NucS (Ren et al., 2009). In contrast, the key residues for mismatch base binding of the Tko EndoMS in complex with mismatched dsDNA are Y41, N76, and W77 and the catalytic activity sites are D165, E179, and K181 (Nakae et al., 2016; Ishino et al., 2018). Mutation at position D163 in the Tga NucS, which is analogous to the D165 residue of the Tko EndoMS, leads to a mutant partially lacking its uracil-containing dsDNA cleaving activity, suggesting that these residues, is also one of the catalytic activity sites in Tga NucS (Zhang et al., 2020).

CONCLUSIONS AND FUTURE DIRECTIONS

NucS endonucleases from HA are multifunctional enzymes that can cleave fork-shaped DNA, mismatched DNA, and deaminated DNA, suggesting that they can participate in several DNA repair pathways. Thus, although HA lack MutS/MutL homologues, a complete MMR system or the recognition proteins of the NER system, the NucS endonucleases, which can cleave mismatched and branched DNA, could confer HA alternatives to complete their MMR and NER pathways but also their repair pathway of deaminated bases in DNA.

However, these alternative DNA repair pathways mediated NucS endonucleases in HA still remain to be fully elucidated. Firstly, DSBs created by cleaving mismatched and deaminated DNA by the Tko EndoMS and Tga NucS appear to be a dangerous strategy for a cell, more so if under harsh environmental conditions, unless cellular HR is very effective.

This is possibly the case in *Thermococcales* to which *P. abyssi*, *T. kodakarensis*, and *T. gammatolerans* belong since they are highly meroploid (Spaans et al., 2015) and have therefore multiple copies of their genomes available for DNA repair by HR (Hildenbrand et al., 2011).

Furthermore, genomic analyses have showed that in HA the *radA* gene, which codes for RadA that mediate HR and the *nucS* gene share the same promoter, suggesting that a DSB produced by the archaeal NucS endonuclease might be taken over immediately by the RadA protein. However, whether the archaeal RadA protein interacts with the archaeal NucS endonuclease to complete the endonuclease-generated DSB repair remains to be tested.

The role of the archaeal NucS endonucleases *in vivo* in cleavage of mismatched, deaminated DNA, and branched DNA remains to be clarified. Pull-down experiments will be performed by using the free archaeal NucS endonucleases or the NucS endonuclease-DNA (mismatched, deaminated, and branched DNA) as the baits to target proteins/enzymes that can interact with NucS endonucleases, which would allow to describe the complete DNA repair triggered by these archaeal NucS endonucleases.

Although several mutants of archaeal NucS endonucleases have been characterized and the crystal structures of two NucS proteins (Tko EndoMS and Pab NucS) have been solved, the detailed catalytic mechanism of the archaeal NucS endonucleases remains elusive. Furthermore, the fine characterization of additional mutants is required to provide insights into the molecular mechanism of NucS endonucleases activity since the key residues of the two enzymes differ so strongly. Finally, solving the structures of the complexes formed by archaeal NucS endonucleases with their deaminated and branched DNA substrates will be helpful to reveal structural and functional relationships of these endonucleases.

Last, both Tko EndoMS and Tga NucS, which are thermostable endonucleases, can cleave dsDNA in a manner similar to type II restriction endonucleases, suggesting that these endonucleases might be potentially used in the future for genetic engineering.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Research in Prof. Zhihui Yang's laboratory was supported by the Earmarked Fund for Modern Agro-industry Technology Research System in Hebei Province, China (HBCT2018080205). Research in Dr. Likui Zhang's laboratory was supported by the Provincial Natural Science Foundation of Jiangsu (No. BK20191219), the Academic Leader of Middle and Young People of Yangzhou University Grant and State Key Laboratory of Microbial Metabolism, Shanghai JiaoTong University (No. MMLKF18-05).

REFERENCES

Aravind, L., Makarova, K. S., and Koonin, E. V. (2000). Survey and summary: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* 28, 3417–3432. doi: 10.1093/nar/28.18.3417

Castaneda-Garcia, A., Prieto, A. I., Rodriguez-Beltran, J., Alonso, N., Cantillon, D., Costas, C., et al. (2017). A non-canonical mismatch repair pathway in prokaryotes. *Nat. Commun.* 8:14246. doi: 10.1038/ncomms14246

Creze, C., Lestini, R., Kuhn, J., Ligabue, A., Becker, H. F., Czjzek, M., et al. (2011). Structure and function of a novel endonuclease acting on branched DNA substrates. *Biochem. Soc. Trans.* 39, 145–149. doi: 10.1042/BST0390145

Creze, C., Ligabue, A., Laurent, S., Lestini, R., Laptenok, S. P., Khun, J., et al. (2012). Modulation of the *Pyrococcus abyssi* NucS endonuclease activity by replication clamp at functional and structural levels. *J. Biol. Chem.* 287, 15648–15660. doi: 10.1074/jbc.M112.346361

Eme, L., Spang, A., Lombard, J., Stairs, C. W., and Ettema, T. J. G. (2017). Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15, 711–723. doi: 10.1038/nrmicro.2017.133

Fujikane, R., Ishino, S., Ishino, Y., and Forterre, P. (2010). Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Genes Genet. Syst.* 85, 243–257. doi: 10.1266/ggs.85.243

Grasso, S., and Tell, G. (2014). Base excision repair in Archaea: back to the future in DNA repair. *DNA Repair* 21, 148–157. doi: 10.1016/j.dnarep.2014.05.006

Grogan, D. W. (1998). Hyperthermophiles and the problem of DNA instability. *Mol. Microbiol.* 28, 1043–1049. doi: 10.1046/j.1365-2958.1998.00853.x

Grogan, D. W. (2000). The question of DNA repair in hyperthermophilic Archaea. *Trends Microbiol.* 8, 180–185. doi: 10.1016/S0966-842X(00)01729-7

Grogan, D. W. (2004). Stability and repair of DNA in hyperthermophilic Archaea. *Curr. Issues Mol. Biol.* 6, 137–144. doi: 10.21775/cimb.006.137

Grogan, D. W. (2015). Understanding DNA repair in hyperthermophilic Archaea: persistent gaps and other reasons to focus on the fork. *Archaea* 2015:942605. doi: 10.1155/2015/942605

Grogan, D. W., Carver, G. T., and Drake, J. W. (2001). Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7928–7933. doi: 10.1073/pnas.141113098

Hildenbrand, C., Stock, T., Lange, C., Rother, M., and Soppa, J. (2011). Genome copy numbers and gene conversion in methanogenic Archaea. *J. Bacteriol.* 193, 734–743. doi: 10.1128/JB.01016-10

Ishino, S., Makita, N., Shiraishi, M., Yamagami, T., and Ishino, Y. (2015). EndoQ and EndoV work individually for damaged DNA base repair in *Pyrococcus furiosus*. *Biochimie* 118, 264–269. doi: 10.1016/j.biochi.2015.06.015

Ishino, S., Nishi, Y., Oda, S., Uemori, T., Sagara, T., Takatsu, N., et al. (2016). Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res.* 44, 2977–2986. doi: 10.1093/nar/gkw153

Ishino, S., Skouloubris, S., Kudo, H., l'Hermitte-Stead, C., Es-Sadik, A., Lambry, J. C., et al. (2018). Activation of the mismatch-specific endonuclease EndoMS/NucS by the replication clamp is required for high fidelity DNA replication. *Nucleic Acids Res.* 46, 6206–6217. doi: 10.1093/nar/gky460

Jacobs, K. L., and Grogan, D. W. (1997). Rates of spontaneous mutation in an archaeon from geothermal environments. *J. Bacteriol.* 179, 3298–3303. doi: 10.1128/JB.179.10.3298-3303.1997

Kanugula, S., Pauly, G. T., Moschel, R. C., and Pegg, A. E. (2005). A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits O-6-alkylguanine-DNA alkyltransferase and endonuclease V activities. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3617–3622. doi: 10.1073/pnas.0408719102

Kiyonari, S., Egashira, Y., Ishino, S., and Ishino, Y. (2014). Biochemical characterization of endonuclease V from the hyperthermophilic archaeon, *Pyrococcus furiosus*. *J. Biochem.* 155, 325–333. doi: 10.1093/jb/mvu010

Li, Y., Shi, H., and Zhang, L. (2019). Research progress of hyperthermophilic archaeal DNA repair endonucleases. *Acta Microbiol. Sin.* 59, 1889–1896. doi: 10.13343/j.cnki.wsxb.20180576

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11, 3610–3618. doi: 10.1021/bi00769a018

Lindahl, T., and Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 13, 3405–3410. doi: 10.1021/bi00713a035

Liu, J., He, B., Qing, H., and Kow, Y. W. (2000). A deoxyinosine specific endonuclease from hyperthermophile, *Archaeoglobus fulgidus*: a homolog of *Escherichia coli* endonuclease V. *Mutat. Res.* 461, 169–177. doi: 10.1016/S0921-8777(00)00054-9

Mariko, A., and Kosuke, M. (2016). A dual base flipping mechanism for archaeal mismatch repair. *Structure* 24, 1859–1861. doi: 10.1016/j.str.2016.10.004

Nakae, S., Hijikata, A., Tsuji, T., Yonezawa, K., Kouyama, K., Mayanagi, K., et al. (2016). Structure of the EndoMS-DNA complex as mismatch restriction endonuclease. *Structure* 24, 1960–1971. doi: 10.1016/j.str.2016.09.005

Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2005). Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure* 13, 1183–1192. doi: 10.1016/j.str.2005.04.024

Palud, A., Villani, G., L'Haridon, S., Querellou, J., Raffin, J. P., and Henneke, G. (2008). Intrinsic properties of the two replicative DNA polymerases of *Pyrococcus abyssi* in replicating abasic sites: possible role in DNA damage tolerance? *Mol. Microbiol.* 70, 746–761. doi: 10.1111/j.1365-2958.2008.06446.x

Ren, B., Kuhn, J., Meslet-Cladiere, L., Briffotaux, J., Norais, C., Lavigne, R., et al. (2009). Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J.* 28, 2479–2489. doi: 10.1038/emboj.2009.192

Rezgui, R., Lestini, R., Kuhn, J., Fave, X., McLeod, L., Myllykallio, H., et al. (2014). Differential interaction kinetics of a bipolar structure-specific endonuclease with DNA flaps revealed by single-molecule imaging. *PLoS One* 9:e113493. doi: 10.1371/journal.pone.0113493

Roberts, J. A., and White, M. F. (2005). An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. *J. Biol. Chem.* 280, 5924–5928. doi: 10.1074/jbc.M412766200

Rouillon, C., and White, M. F. (2011). The evolution and mechanisms of nucleotide excision repair proteins. *Res. Microbiol.* 162, 19–26. doi: 10.1016/j.resmic.2010.09.003

Shiraishi, M., Ishino, S., Yamagami, T., Egashira, Y., Kiyonari, S., and Ishino, Y. (2015). A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*. *Nucleic Acids Res.* 43, 2853–2863. doi: 10.1093/nar/gkv121

Spaans, S. K., van der Oost, J., and Kengen, S. W. (2015). The chromosome copy number of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *Extremophiles* 19, 741–750. doi: 10.1007/s00792-015-0750-5

Stetter, K. O. (2013). A brief history of the discovery of hyperthermophilic life. *Biochem. Soc. Trans.* 41, 416–420. doi: 10.1042/BST20120284

Suzuki, S., and Kurosawa, N. (2019). Endonucleases responsible for DNA repair of helix-distorting DNA lesions in the thermophilic crenarchaeon *Sulfolobus acidocaldarius* in vivo. *Extremophiles* 23, 613–624. doi: 10.1007/s00792-019-01120-9

Takemoto, N., Numata, I., Su'etsugu, M., and Miyoshi-Akiyama, T. (2018). Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. *Nucleic Acids Res.* 46, 6152–6165. doi: 10.1093/nar/gky481

Wang, Y., Zhang, L., Zhu, X., Li, Y., Shi, H., and Oger, P. (2018). Biochemical characterization of a thermostable endonuclease V from the hyperthermophilic euryarchaeon *Thermococcus barophilus* Ch5. *Int. J. Biol. Macromol.* 117, 17–24. doi: 10.1016/j.ijbiomac.2018.05.155

White, M. F., and Allers, T. (2018). DNA repair in the Archaea—an emerging picture. *FEMS Microbiol. Rev.* 42, 514–526. doi: 10.1093/femsre/fuy020

Yasui, A. (2013). Alternative excision repair pathways. *Cold Spring Harb. Perspect. Biol.* 5:a012617. doi: 10.1101/cshperspect.a012617

Zatopek, K. M., Gardner, A. F., and Kelman, Z. (2018). Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms. *FEMS Microbiol. Rev.* 42, 477–488. doi: 10.1093/femsre/fuy017

Zhang, L., Shi, H., Gan, Q., Wang, Y., Wu, M., Yang, Z., et al. (2020). An alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease. *DNA Repair* 85:102734. doi: 10.1016/j.dnarep.2019.102734

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 July 2020
doi: 10.3389/fmicb.2020.01466

Molecular and Biochemical Characterization of Salt-Tolerant Trehalose-6-Phosphate Hydrolases Identified by Screening and Sequencing Salt-Tolerant Clones From the Metagenomic Library of the Gastrointestinal Tract

*Yanxia Yang[1], Yunjuan Yang[1,2,3], Qin Fan[1], Zunxi Huang[1,2,3], Junjun Li[1,2,3], Qian Wu[1,2,3], Xianghua Tang[1,2,3], Junmei Ding[1,2,3], Nanyu Han[1,2,3] and Bo Xu[1,2,3]**

[1] School of Life Sciences, Yunnan Normal University, Kunming, China, [2] Engineering Research Center of Sustainable Development and Utilization of Biomass Energy, Ministry of Education, Kunming, China, [3] Key Laboratory of Yunnan for Biomass Energy and Biotechnology of Environment, Kunming, China

OPEN ACCESS

Edited by:
Sunil Khare, Indian Institute of Technology Delhi, India

Reviewed by:
Guozeng Wang, Fuzhou University, China
María-Eugenia Guazzaroni, University of São Paulo, Brazil

***Correspondence:**
Bo Xu
xubo128028@163.com

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *18 December 2019*
Accepted: *04 June 2020*
Published: *07 July 2020*

Citation:
Yang Y, Yang Y, Fan Q, Huang Z, Li J, Wu Q, Tang X, Ding J, Han N and Xu B (2020) Molecular and Biochemical Characterization of Salt-Tolerant Trehalose-6-Phosphate Hydrolases Identified by Screening and Sequencing Salt-Tolerant Clones From the Metagenomic Library of the Gastrointestinal Tract. Front. Microbiol. 11:1466. doi: 10.3389/fmicb.2020.01466

The exploration and utilization of microbial salt-tolerant enzymatic and genetic resources are of great significance in the field of biotechnology and for the research of the adaptation of microorganisms to extreme environments. The presence of new salt-tolerant genes and enzymes in the microbial metagenomic library of the gastrointestinal tract has been confirmed through metagenomic technology. This paper aimed to identify and characterize enzymes that confer salt tolerance in the gastrointestinal tract microbe. By screening the fecal metagenomic library, 48 salt-tolerant clones were detected, of which 10 salt-tolerant clones exhibited stronger tolerance to 7% (wt/vol) NaCl and stability in different concentrations of NaCl [5%–9% (wt/vol)]. High-throughput sequencing and biological information analysis showed that 91 potential genes encoded proteins and enzymes that were widely involved in salt tolerance. Furthermore, two trehalose-6-phosphate hydrolase genes, namely, *tre_P2* and *tre_P3*, were successfully cloned and expressed in *Escherichia coli* BL21 (DE3). By virtue of the substrate of *p*-nitrophenyl-α-D-glucopyranoside (*p*NPG) which can be specifically hydrolyzed by trehalose-6-phosphate hydrolase to produce glucose and *p*-nitrophenol, the two enzymes can act optimally at pH 7.5 and 30°C. Steady-state kinetics with *p*NPG showed that the K_M and K_{cat} values were 15.63 mM and 10.04 s^{-1} for rTRE_P2 and 12.51 mM and 10.71 s^{-1} for rTRE_P3, respectively. Characterization of enzymatic properties demonstrated that rTRE_P2 and rTRE_P3 were salt-tolerant. The enzymatic activity increased with increasing NaCl concentration, and the maximum activities of rTRE_P2 and rTRE_P3 were obtained at 4 and 3 M NaCl, respectively. The activities of rTRE_P2 increased by approximately 43-fold even after 24 h of incubation with 5 M NaCl. This study is the first to report the identification as well as molecular and

biochemical characterization of salt-tolerant trehalose-6-phosphate hydrolase from the metagenomic library of the gastrointestinal tract. Results indicate the existence of numerous salt-tolerant genes and enzymes in gastrointestinal microbes and provide new insights into the salt-tolerant mechanisms in the gastrointestinal environment.

Keywords: trehalose-6-phosphate hydrolase, salt tolerance, gastrointestinal tract microbe, metagenomic library, high-throughput sequencing

INTRODUCTION

Salt-tolerant bacteria mainly live in extreme salt-rich environments. These bacteria adapt to the special external environmental pressure by producing numerous salt-tolerant genes and enzymes. Exploring and utilizing microbial salt-tolerant enzymatic and genetic resources are of great significance in the study of biological adaptation to extreme environments and for broadening the application of salt-tolerant enzymes and genes in the field of biotechnology.

A large number of salt environments exist in natural habitats, where halophilic or salt-tolerant microorganisms are widely distributed. Researchers use pure culture techniques to obtain various salt-tolerant genes and enzymes from halophilic or salt-tolerant microorganisms such as *IoXyl2p* gene (Han et al., 2020), *pprI* gene (Hossein Helalat et al., 2019), α-amylase (Deutch and Yang, 2020), and β-xylosidase (Xu et al., 2019). However, the number of salt-tolerant genes and enzymes is still lower than that of salt-tolerant bacteria isolated. With the advent of the post-genome era, the emergence of metagenomic technology addresses the problem of microbial separation and culture, expands the utilization of microbial resources, and provides a new research strategy for discovering new salt-tolerant functional genes and enzymes.

The gastrointestinal tract is considered the densest microbial ecosystem on the planet (Whitman et al., 1998; Egert et al., 2006) and is a huge resource for exploring biocatalysts. Microbes inhabiting in the gastrointestinal tract are exposed to stresses, such as acids, bile salts, and osmosis pressure (Louis and O'Byrne, 2010). Hyperosmosis gastrointestinal cavity (equivalent to 0.3 M) (Alvarez-Ordonez et al., 2011) and dietary changes (De Dea Lindner et al., 2007) can lead to the production of osmotic pressure in the gastrointestinal tract. In this external environmental pressure, some special genes and enzymes of microorganisms may be activated to adapt to pressure.

Few studies have investigated on salt-tolerant genes and enzymes from the metagenome of the gastrointestinal tract. Since 2012, Culligan et al. (2012) obtained salt-tolerant clones from the metagenomic library of human gut microbiomes as well as salt-tolerant genes and enzymes, such as *galE*, *murB*, *mazG* (Culligan et al., 2012), *stlA* (Culligan et al., 2013), *sdtR* (Culligan et al., 2014a), and *brpA* (Culligan et al., 2014b). *galE* encodes UDP-glucose 4-epimerase. *murB* encodes UDP-*N*-acetylenolpyruvoyl glucosamine reductase. *mazG* encodes the MazG family of proteins (nucleoside triphosphate pyrophosphohydrolase), which can delay the procedural death of cells. The *stlA* gene is a salt-tolerant gene that is unique to the human gut; this gene can gain a competitive advantage in a high-salt environment by obtaining adjacent genes through lateral gene transfer (LGT) events. *sdtR* encodes putative transcriptional regulators. The *brpA* gene belongs to the brp/blh family of β-carotene 15,15′-monooxygenase, which encodes a membrane protein. Verma et al. (2018) screened salt-tolerant clones, namely, SR6 and SR7, from the human fecal metagenomic library, performed subclone analysis, and found that *TMSRP1*, *ABCTPP*, and *TLSRP1* are associated with salt tolerance. Ilmberger et al. (2012) obtained salt-tolerant hydrolase from the gastrointestinal microbial metagenome (celA84 in the GH5 family can retain 50% relative activity after 34 days in 4 M NaCl). Hence, gastrointestinal microorganisms contain abundant salt-tolerant genes and enzymes that can be obtained through metagenomic technology.

Nycticebus pygmaeus is a low-class primate and omnivorous animal that inhabits tropical rainforests, seasonal rainforests, and subtropical evergreen broad-leaved forests. *Bos frontalis* is a rare bovine species living in free-range conditions within tropical rainforest ecosystems which lie merely in some regions in India, Bangladesh, Bhutan, China, and Myanmar (Mondal et al., 2004). The animal is believed to be domesticated for as long as 8000 years, during which common salt was exclusively provided to the animal by local farmers as an additional feed (Das et al., 2011). This study screened salt-tolerant clones from the constructed fecal metagenomic library of *N. pygmaeus* and *B. frontalis* (Xu et al., 2014; Dong et al., 2016). We analyzed potential salt-tolerant genes and enzymes through high-throughput sequencing and bioinformatics technology. We then selected trehalose-6-phosphate hydrolase for cloning, heterologous expression, purification, and characterization. Recombinant rTRE_P2 and rTRE_P3 were found to be involved in salt tolerance. This study lays a theoretical basis for exploring the salt-tolerance mechanism and for the development and utilization of salt-tolerant genes and enzymes in gastrointestinal microorganisms.

MATERIALS AND METHODS

Materials, Bacterial Strain, and Culture Conditions

Taq DNA polymerase, restriction endonucleases, and PCR reagents were obtained from Takara (Beijing, China). In-Fusion HD Cloning Kit (Takara Beijing, China), plasmid pEASY-E2, and *E. coli* BL21(DE3) competent cells (TransGen, Beijing, China) were adopted in the cloning process. The purification was performed with Ni-NTA Agarose resin (Qiagen, Hilden, Germany). *p*-Nitrophenyl-α-D-glucopyranoside (*p*NPG),

p-nitrophenyl-β-D-galactopyranoside, *p*-nitrophenyl-β-D-glucopyranoside, *p*-nitrophenyl-α-D-mannopyranoside, and 2-nitrophenyl-β-D-glucopyranoside were supplied by Shanghai Yuanye Bio-Technology (China). The vector pCC1FOS and *E. coli* EPI300 were purchased from Epicentre. The remaining chemicals reagents were of analytical grade except for those specified. The cultivation environment of the strains *E. coli* is at 37°C in the Luria–Bertani (LB) liquid medium, in which the salt concentrations can reach their final stages, encompassing the salt showed in the normal LB medium. In order to select certain plasmids, 12.5 μg/ml chloramphenicol (Cm) or ampicillin (100 μg/ml) was used as an alternative of the medium. Previously constructed fecal microbial metagenomic libraries of *N. pygmaeus* and *B. frontalis* were used (Xu et al., 2014; Dong et al., 2016).

Screening of Salt-Tolerant Clones

Augmented fecal microbial metagenomic library culture was inoculated into an LB broth complemented with 12.5 μg/ml chloramphenicol (Cm) and 7% (wt/vol) NaCl (1 ml of the former into 20 ml of the later). The flakes were nurtured at 37°C for 24 consecutive hours and unceasing shaking condition at 180 rpm. On top of that, new flasks with 20 ml of the identical components received 1 ml of the culture. The enrichment process for salt-tolerant clones was done for three times. The agar with the same medium was covered by the cultures. Noticeable colonies then emerged on the plates following the incubation for 72 h at 37°C. The isolated colonies were analyzed by an enzyme labeling apparatus. The optical density value (OD_{600}) of the culture was determined after activation of salt-tolerant clones. The activated culture (20 μl) was shifted to 20 ml of the LB liquid medium [including 12.5 μg/ml Cm and 7% (wt/vol) NaCl] and nurtured for 24 h, 37°C, 180 rpm. The optical density value (OD_{600}) was recorded using 200 μl of the culture. *E. coli* EPI300-C culture containing the empty pCC1FOS was set as the control group. Clones with high optical density value were selected for study of tolerance to different salt levels [5%–9% (wt/vol) NaCl].

Fosmid DNA Extraction

Bacterial culture (5 ml) was brought into existence after the cultivation of 12.5 μg/ml Cm for a night. Near 1 ml of the culture was added to the broth with 4 ml fresh LB. Moreover, 5 μl of the 1000 copy control induction solution (Epicentre Biotechnologies), together with 12.5 μg/ml Cm, was put in the culture. The mixture was cultured at 37°C for 5 h with vigorous shaking (200–250 rpm) to ensure maximum aeration. All 5-ml stimulated cultures generated cells after the 1-min centrifugation at 12000 rpm. ZR BAC DNA Miniprep Kit (Zymo Research) was used to extract fosmid based on the manufacturer's directions.

High-Throughput Sequencing and Bioinformatics Analysis

The fosmid DNA of salt-tolerant clones was sequenced with the Illumina Solexa Genome Analyzer platform. Clean data were obtained by removing the low-quality base at both ends of the raw data and removing the sequences containing the adapter and consisting of a low average base mass, multiple N, and very short length. The sequences were assembled using SPAdes (v3.11.1) to obtain high-quality scaffold fragments. Scaffolds with length > 300 bp were retained for bioinformatics analyses. Megablast against the NCBI plasmid database[1] was adopted to compare and filter the results in line with identity ≥ 30% and math length ≥ 200. cBar (v1.2) and PlasmidFinder (v1.3) were used to identify plasmids and select scaffold fragments as plasmid for genetic prediction. Prodigal (v2.6.3) was used as prediction software. Gene function annotation and secretory protein prediction analysis were carried out using Blast (v2.7.1 plus) and Diamond (v0.9.10) against Gene Ontology (GO), Cluster of Orthologous Groups of Protein (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and nonredundant protein databases (NR). Low-reliability comparison results were filtered out according to identity, e-value, and score values. The optimal results were filtered as functional annotation results for the gene.

Cloning of Trehalose-6-Phosphate Hydrolase

Tre_P2 and *tre_P3* were identified in salt-tolerant clones 1A and 5_1_1 in *N. pygmaeus*, and *tre_P4* was identified in salt-tolerant clone 21_9A in *B. frontalis* by high-throughput sequencing. The plasmid DNA of clones 1A, 5_1_1, and 21_9A were used as template for PCR amplification. The primer pairs are as follows: *tre_P2*-F: 5′-TAA GAA GGA GAT ATA CAT ATG GAA TTG ATG AGT GAA CAA GAC TGG-3′ and *tre_P2*-R: 5′-GTG GTG GTG GTG GTG CTC GAG CTT CGC CGT TTC AAC CAC AAT TGC-3′. *tre_P3*-F: 5′-TAA GAA GGA GAT ATA CAT ATG GAA TTG ATG GAA TTG ATG AGT GAA CAA GAC TGG-3′, and *tre_P3*-R: 5′-GTG GTG GTG GTG GTG CTC GAG GTG GTG GTG GTG GTG GTG-3′, *tre_P4*-F: TAA GAA GGA GAT ATA CAT ATG GAA TTG ATG GAA TGG CAA GTA GGT ATC, and *tre_P4*-R: GTG GTG GTG GTG GTG CTC GAG ATA TGA CTT ACT AAT AAC TAT CGC (primers highlighted with underlines stand for *Nde* I and *Xho* I sites, specifically). The cloning process of amplified fragments into the pEASY-E2 vector was done after PCR. The verification of gene sequence was achieved through DNA sequencing.

Expression and Purification

E. coli BL21 (DE3) transformed with pEASY-E2 containing *tre_P2* or *tre_P3* fragments was incubated overnight on the LB broth supplemented with 100 μg/ml ampicillin in a shaking incubator at 180 rpm and 37°C. What followed was the shifting process of the culture (1 ml) to new flasks with the same portion of 200 ml. 0.7 mM isopropyl-D-thiogalactopyranoside (IPTG) was used to induce thriving cells in the optimum density of 0.6 at 600 nm, which further grew for 20 h at 20°C, 160 rpm. Cells were gathered after centrifugation (6000 × *g* for 8 min) at 4°C, suspended in binding buffer (0.5 M NaCl, 20 mM imidazole, 20 mM Tris–HCl, pH 8.0), and the disrupted with a high-pressure cell disruptor. The centrifugal process was then done to the cell lysate (12000 × *g* and 4°C for 20 min). For the analysis of the subsequent supernatant, sodium dodecyl–sulfate

[1] ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plasmid/

polyacrylamide gel electrophoresis (SDS-PAGE) and enzyme assay were adopted. By virtue of the Ni-NTA His·Bind column and immobilized metal-affinity chromatography, the enzyme was purified. A Ni-NTA His·Bind column, which was equilibrated earlier using binding buffer, was adopted to carry the sample. The column used was washed with binding buffer and washing buffer (concentration: 0.5 M NaCl, 80 or 100 mM imidazole, 20 mM Tris–HCl, pH 8.0). Eluting buffer (500 mM imidazole, 0.5 M NaCl, 20 mM Tris–HCl, pH 8.0) was used to elute the bound protein. Those fractions carrying the recombinant protein were gathered and preserved at −20°C.

Enzyme Assay and Characterization

The following experiments were repeated three times.

Trehalose-6-phosphate hydrolase hydrolyzes *p*NPG to produce glucose and *p*-nitrophenol, which is deemed as a specific process as *p*NPG does not hydrolyze itself in the absence of trehalose-6-phosphate hydrolases, nor will it be hydrolyzed by the periplasmic trehalase or maltodextrin-degrading enzymes (Rimmele and Boos, 1994). *p*NPG was employed, on which rTRE_P2 and rTRE_P3 activities were testified, adopting the method conveyed by Chuang et al. (2012). The combination for further reaction was with 5 mM *p*NPG, 100 mM NaCl, and 50 mM HEPES–NaOH buffer (pH 7.5) and a certain amount of the enzyme reagent. First cultivated for 10 min at 30°C, the reaction was then boiled for 5 min, which in the end caused it to cease. Centrifuge was used to separate the sample, and A410 of the supernatant was registered. Delineation of a single unit of rTRE_P2 and rTRE_P3 activities was considered as related to the quantity of the enzyme capable of discharging 1 μmol *p*-nitroaniline per minute at 30°C.

Temperature influence on the activities of rTRE_P2 and rTRE_P3 was reviewed at the range of 0°C–70°C. On the substrate of *p*NPG operated the reactions in standard conditions. In a bid to study the influence of temperature on enzyme stability, a control group containing different temperatures (37°C, 45°C, and 50°C) was conducted on rTRE_P2 and rTRE_P3 for an hour. Regardless of the designated temperature, the activity of the residual enzyme was circumscribed in a 10-min span during the given time.

The pH with which the activities of rTRE_P2 and rTRE_P3 were examined was from 3 to 10. The activities of rTRE_P2 and rTRE_P3 were assayed at 30°C in citric acid–Na_2HPO_4 buffer (pH 3–5) and HEPES-NaOH buffer (6–12). The pH stability changed with different status in incubating the purified enzyme in different buffers during the 1-h period at 30°C, for which the untreated enzyme solution was adopted as means of controlling.

Determination of the effect of different NaCl concentrations on rTRE_P2 and rTRE_P3: different concentrations of NaCl solution (0.5–5.0 M) were prepared with HEPES–NaOH (pH 7.5) containing 5 mM *p*NPG, and residual enzyme activity was determined at the optimum condition, with no NaCl enzymatic reaction added as the control group. Determination of NaCl stability of rTRE_P2 and rTRE_P3: the purified recombinant rTRE P2 and rTRE P3 were added into 0.5–5 M NaCl solution and treated for 1 or 24 h at 30°C, respectively. The residual enzyme activity was determined at the optimum pH and optimum temperature, and the enzymatic reaction without NaCl treatment under the same conditions was used as the control group.

In evaluating the effects of metal ions and chemicals on the activities of rTRE_P2 and rTRE_P3 (which include GuHCl, KCl, NaCl, $FeSO_4$, $CoCl_2$, iodoacetic acid, pig bile salt, Tween 80, Triton X-100, $FeCl_3$, $ZnSO_4$, CTAB, EDTA, $NiSO_4$, $MnSO_4$, $AgNO_3$, β-mercaptoethanol, DTT, EGTA, $CuSO_4$, PEG20000, SDS, NBS, $MgCl_2$, $HgCl_2$, $AlCl_3$, LiCl, $PbCl_2$, and $CaCl_2$), a 10-min incubation process was carried out on purified enzyme aliquots at 30°C in 50 mM HEPES–NaOH buffer (pH 7.5). Residual activity was assessed according to standard assay situations. Enzymatic activity in exclusion of the metal ions and chemicals beyond the experiment was supposed as 100%.

In environments with a temperature of 30°C and a pH of 7.5, multiform *p*-nitrophenyl derivatives were utilized to delve the specificities brought from the substrate of rTRE_P2 and rTRE_P3. The substrate composites applied in the experiment involved *p*NPG, *p*-nitrophenyl-β-D-galactopyranoside, *p*-nitrophenyl-β-D-glucopyranoside, *p*-nitrophenyl-α-D-mannopyranoside, and 2-nitrophenyl-β-D-glucopyranoside. The substrates used to incubate the purified enzyme also include glucose, glucose-6-phosphate, trehalose, and trehalose-6-phosphate [0.5% (wt/vol)], and the environment in which it grew for 2 h was in 50 mM HEPES–NaOH buffer (200 μl, pH 7.5) at 30°C. The enzymatic fruits (2 μl) here were located on a silica plate. Water, acetic acid, and 1-butanol were mixed (1:1:2, vol/vol/vol) for the operation of thin-layer chromatography (TLC). The specimen was dipped into 50 ml of acetone containing 1 g of diphenylamine, added with 1 ml of phenylid and 5 ml of 85% phosphorous acid for 30 s, and heated at 120°C for 15 min.

To study the steady-state kinetics, the samples were monitored at 30°C for 12 min (first-order reaction time) in a reaction buffer with *p*NPG concentrations of 0.5–10 mM. Moreover, based on the least-square fits of initial rates, kinetic parameters were evaluated, making their functions of substrate concentration stand out. The figuration of K_M and K_{cat} values was achieved through gauging the gradient of the linear part of the Michaelis–Menten plot.

RESULTS

Screening of Salt-Tolerant Clones

Seventeen and 31 salt-tolerant clones were detected from the metagenomic fosmid libraries of *N. pygmaeus* and *B. frontalis*, respectively. The growth of partial salt-tolerant clones is shown in **Figure 1A**. After 24 h of culture, the OD_{600} of 48 salt-tolerant clones was 0.236–1.471, indicating that these strains were salt-tolerant. Clone 5_1_1 had a high ability to resist 7% (wt/vol) NaCl, and clone 1_2B has the weakest tolerance to 7% NaCl (wt/vol). The control *E. coli* EPI300-C strain was cultured in 7% (wt/vol) NaCl medium, and no significant change was noted. Hence, microbial salt-tolerant gene fragments may have been inserted into the pCC1FOS vector. Ten clones (1A, 3-1, 5_1_1, and 5_1_5 from the *N. pygmaeus* gastrointestinal

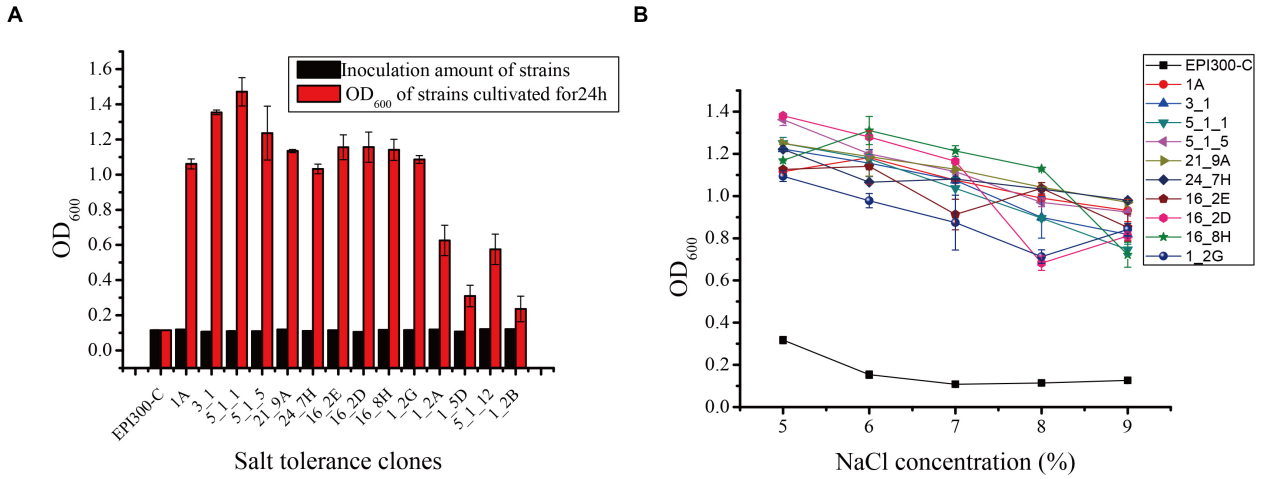


FIGURE 1 | (A) Screening of some salt-tolerant clones of the fosmid library. **(B)** Salt tolerance of some salt-tolerant clones against different salt stress levels. Error bars represent the mean ± SD (n = 3).

tract and 21_9A, 24_7H, 16_2E, 16_2D, 16_8H, and 1_2G from the *B. frontalis* gastrointestinal tract) with high salt tolerance were selected for the study of tolerance to different salt levels (**Figure 1B**). With increasing NaCl concentration (5% to 9%), the OD_{600} of the strains remained basically stable, showing significant salt tolerance compared with the controls.

High-Throughput Sequencing and Functional Classification of Unigene

To obtain salt-tolerant genes and enzymes from the gastrointestinal tract of microbes, we sequenced the fosmid DNA of 10 salt-tolerant clones by high-throughput sequencing. We obtained 100,909,958 raw reads with a total of 15,136,493,700 bps. After trimming, 94,533,758 clean reads remained. Low-quality reads and adapter sequence were filtered. There were 323,181 bp effective sequences (average length: 5,572.09 bp) produced from the gastrointestinal tract of *N. pygmaeus* and 898,646 bp effective sequences (average length: 3,907.16 bp) from the gastrointestinal tract of *B. frontalis* (**Supplementary Table S1**).

Gene Ontology analysis showed that most genes are associated with biological processes including metabolic and cellular processes (**Supplementary Figure S1**). Clones 3_1 and 1A, but not clone 5_1_5, are related to metabolic processes alone. Only clone 21_9A is associated with cellular components, which include membrane parts. Most genes, including clones 3_1 and 16_2D, are involved in molecular functions, such as binding but not catalytic activity.

We also classified the predicted genes by aligning them to COG and KEGG databases. Based on the annotation against the COG database, these genes are mainly associated with carbohydrate transport and metabolism, transcription, replication, recombination and repair, inorganic ion transport and metabolism, and defense mechanism: mobilomes, prophages, and transposons (**Figure 2**). We found that the predicted functional genes in the gastrointestinal tract of *N. pygmaeus* were not involved in secondary metabolite biosynthesis, transport, and catabolism (**Figures 2A–D**), whereas the predicted functional genes in the gastrointestinal tract of *B. frontalis* were involved in such processes (**Figures 2E–J**). The contrasting finding may be due to dietary differences. The KEGG analysis of the salt-tolerant clones showed presumed components of carbohydrate metabolism, membrane transport, drug resistance, and nucleotide metabolism, which perhaps implicated diverse metabolic pathways related to salt tolerance (**Supplementary Figure S2**). Hence, salt-tolerant strains may have a specific membrane structure, transcriptional regulation process, high transport, and metabolic capacities.

Screening and Analysis of Potential Salt-Tolerant Genes and Enzymes

We annotated genes to the Nr database. A total of 1,285 genes and enzymes of salt-tolerant clones were significantly annotated (data not shown). After comparative analysis, 91 genes and enzymes were assumed to be widely involved in salt resistance. These genes and enzymes include UDP-glucose 4-epimerase, general stress protein, transcriptional regulator, alcohol dehydrogenase, glucose-6-phosphate 1-dehydrogenase, trehalose-6-phosphate hydrolase, transporters, etc. (**Supplementary Table S2**). Fifty-two potential salt-tolerant genes and enzymes were identified from the gastrointestinal tract of *B. frontalis*, which are more than the genes and enzymes from the gastrointestinal tract of *N. pygmaeus* (39 genes). UDP-glucose 4-epimerase, catalase, and cadmium transporter were only detected in the gastrointestinal tract of *N. pygmaeus*. Trehalose-6-phosphate hydrolase is considered irreducible in the assimilation of trehalose (Chuang et al., 2012), with which the inactivating or denaturizing of proteins and cellular membranes can be prevented even in conditions such as desiccation, hypohydration, torridness, coldness, and oxidation (Elbein et al., 2003). Thus, we chose trehalose-6-phosphate hydrolase for further study. In addition, we found multiple functional genes that are resistant to other stresses, such as glutaredoxin and cold-shock proteins.

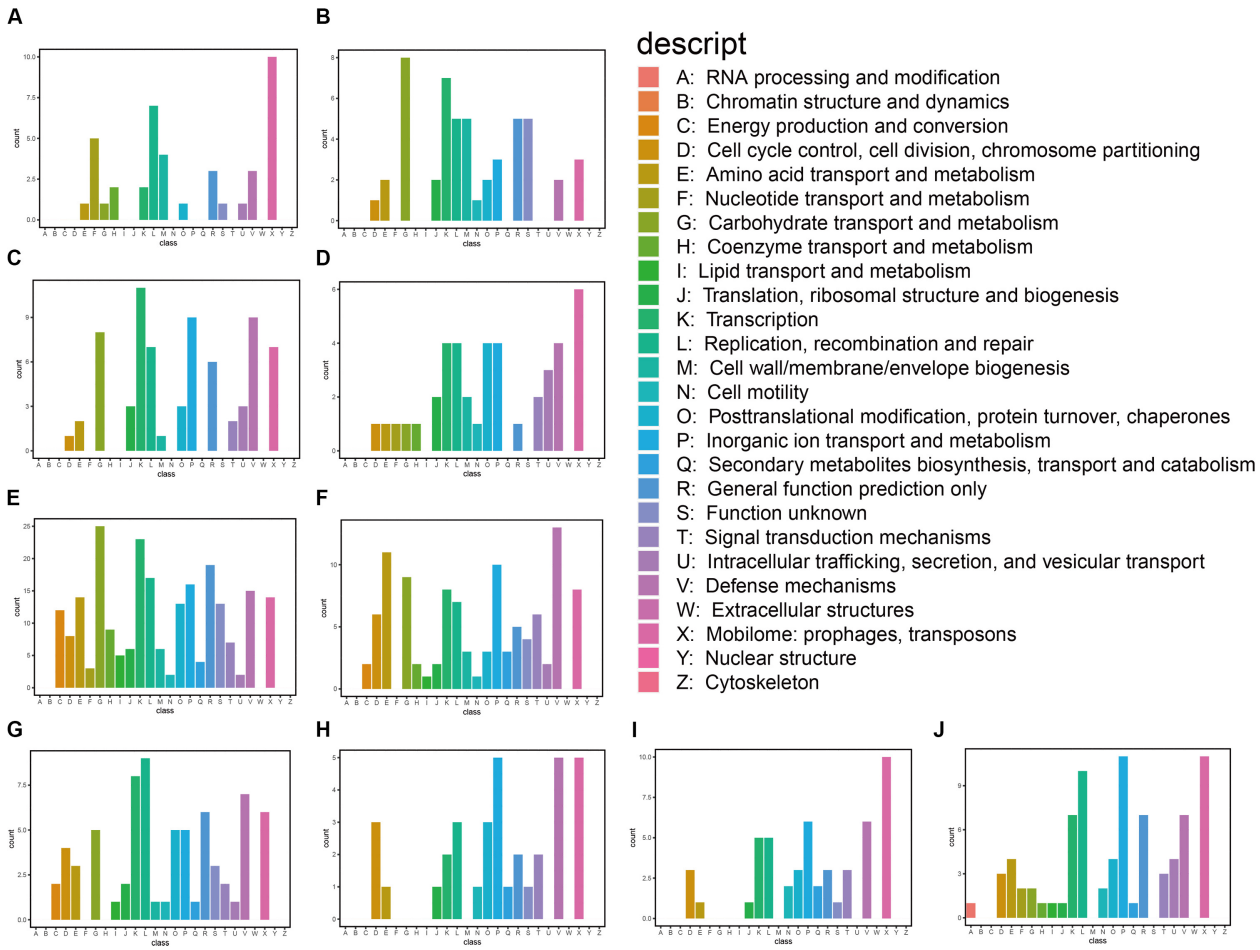


FIGURE 2 | Functional classification of annotated transcripts by COG. **(A)** 3_1; **(B)** 1A; **(C)** 5_1_1; **(D)** 5_1_5; **(E)** 21_9A; **(F)** 24_7H; **(G)** 16_2E; **(H)** 16_2D; **(I)** 16_8H; and **(J)** 1_2G.

Cloning and Sequence Analysis of Trehalose-6-Phosphate Hydrolase

PCR amplification proved unsuccessful for *tre_P4* from clone 21_9A in the gastrointestinal tract of *B. frontalis*, possibly owing to special structures of this gene. As such, *tre_P2* and *tre_P3* from salt-tolerant clones 1A and 5_1_1 in the gastrointestinal tract of *N. pygmaeus* were chosen for further investigation. *tre_P2* (1638 bp) and *tre_P3* (1644 bp) were obtained through PCR amplification, and the nucleotide sequence was preserved in GenBank (MN830270, MN830271). SignalP server[2] was adopted, as a means of testing signal peptide, to analyze the translated amino acid sequence in the gene. The results indicate that TRE_P2 and TRE_P3 have no signal peptides. The homology search conducted on NCBI through the BLASTP tool suggested that TRE_P2 had the highest sequence identity (100%) and TRE_P3 had the highest sequence identity (99.82%) to trehalose-6-phosphate hydrolase from *Staphylococcus warneri* (WP_124228261) and *Staphylococcus warneri* (WP_058709819), respectively. However, these proteins were not characterized. The deduced amino acid sequences of TRE_P2 and TRE_P3 were compared with those of the available trehalose-6-phosphate hydrolase from different microorganisms by using CLUSTALX (2.1) program and mapped by ESPript.[3] From the results, we found a site Asn327 marked by a pentangle (**Figure 3**). The site is considered essential for binding to chloride ions according to Ong et al. (2014). Moreover, the findings may help explain why TRE_P2 and TRE_P3 confer salt tolerance.

A dendrogram was constructed in accordance with amino acid sequences to attest the evolution correlation of TRE_P2 and TRE_P3 with other disclosed oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase, trehalose-6-phosphate hydrolase, amylosucrase, sucrose phosphorylase, isomaltulose synthase, trehalose synthase, cyclomaltodextrinase, maltogenic amylase, and neopullulanase that originated from various microorganisms. As shown in **Figure 4**, TRE_P2 and TRE_P3 were divided into one group and had a closer relationship to trehalose-6-phosphate hydrolase.

[2]http://www.cbs.dtu.dk/services/SignalP-4.0/

[3]http://espript.ibcp.fr/ESPript/ESPript/

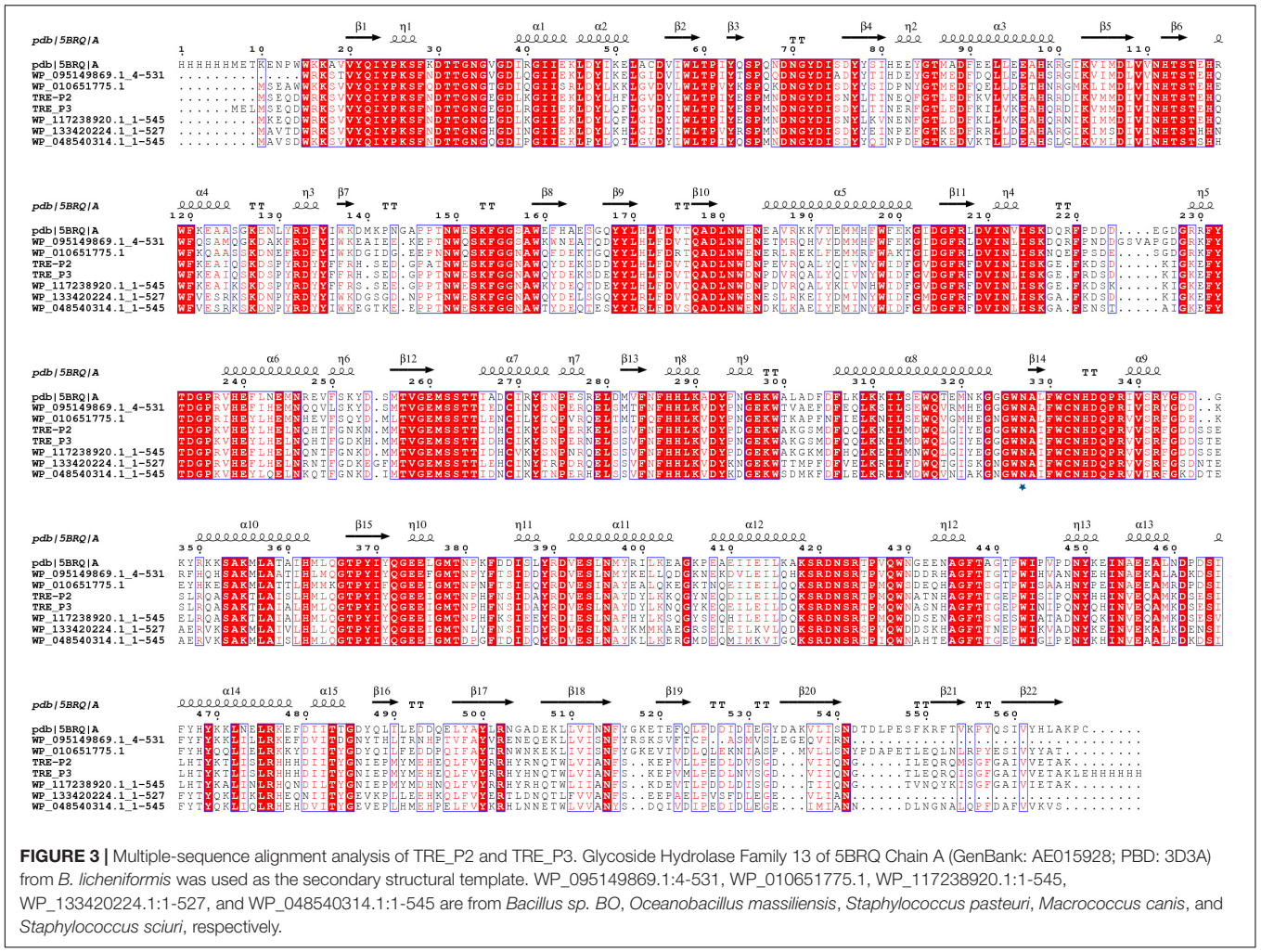


FIGURE 3 | Multiple-sequence alignment analysis of TRE_P2 and TRE_P3. Glycoside Hydrolase Family 13 of 5BRQ Chain A (GenBank: AE015928; PBD: 3D3A) from *B. licheniformis* was used as the secondary structural template. WP_095149869.1:4-531, WP_010651775.1, WP_117238920.1:1-545, WP_133420224.1:1-527, and WP_048540314.1:1-545 are from *Bacillus sp. BO*, *Oceanobacillus massiliensis*, *Staphylococcus pasteuri*, *Macrococcus canis*, and *Staphylococcus sciuri*, respectively.

Expression, Purification, and Molecular Size Determination of rTRE_P2 and rTRE_P3

The expression of rTRE_P2 and rTRE_P3 as an N-terminal His-tag fusion protein to characterize biochemical properties was conducted through the pEASY-E2 expression system, in which the T7 lac promoter in *E. coli* BL21 (DE3) was used as a controller. A high-pressure cell disruptor was adopted to break apart the cells collected. Following purification with the Ni-NTA column and assay using SDS-PAGE analysis, roughly 65 kDa molecular mass of the purified enzymes was gathered (**Supplementary Figure S3**). The comparative molecular masses of rTRE_P2 and rTRE_P3 were estimated to be about 65.21 and 65.67 kDa, respectively, proving them to be monomeric proteins.

Biochemical Properties of the Recombinant Enzyme

The purified enzyme was tested at a temperature range of 0°C–70°C for 10 min to get its apparent temperature dependence. As indicated in **Figures 5A,B**, the enzymes reach their most dynamic status at 30°C and their most stable status at 37°C.

The optimum pH activity assay of rTRE_P2 and rTRE_P3 was performed under standard conditions but with different buffer components. In 50 mM HEPES–NaOH buffer, the optimum pH for rTRE_P2 and rTRE_P3, 7.5, exists (**Figure 5C**). At pH 10, the rTRE_P3 activity was reduced to 4.95% and rTRE_P2 completely lost its activity. Recombinant rTRE_P2 and rTRE_P3 were stable within the pH between 7 and 8 (**Figure 5D**).

Enzymatic activities of rTRE_P2 and rTRE_P3 were stimulated by NaCl, implying certain effect (**Figure 5E**). The enzymatic activity increased with increasing NaCl concentration, and the maximum activities of rTRE_P2 and rTRE_P3 were obtained at 4 and 3 M NaCl, respectively (approximately 35-fold). The activity also increased after 1 h of incubation with the involvement of 0.5–5 M NaCl, and this trend remained after 24 h. The activity of rTRE_P2 even increased by approximately 43-fold after incubation with 5 M NaCl after 24 h (**Figure 5F**).

Immersed in HEPES–NaOH buffer (50 mM, pH 7.5, 30°C) for 10 min, various chemicals were pre-incubated to examine their possible effects on the activities of rTRE_P2 and rTRE_P3. The activities were increased by GuHCl, KCl, NaCl, EDTA, β-mercaptoethanol, DTT, and EGTA but completely inhibited by $PbCl_2$, CTAB, $AgNO_3$, $CuSO_4$, $HgCl_2$, and $ZnSO_4$ (**Table 1**).

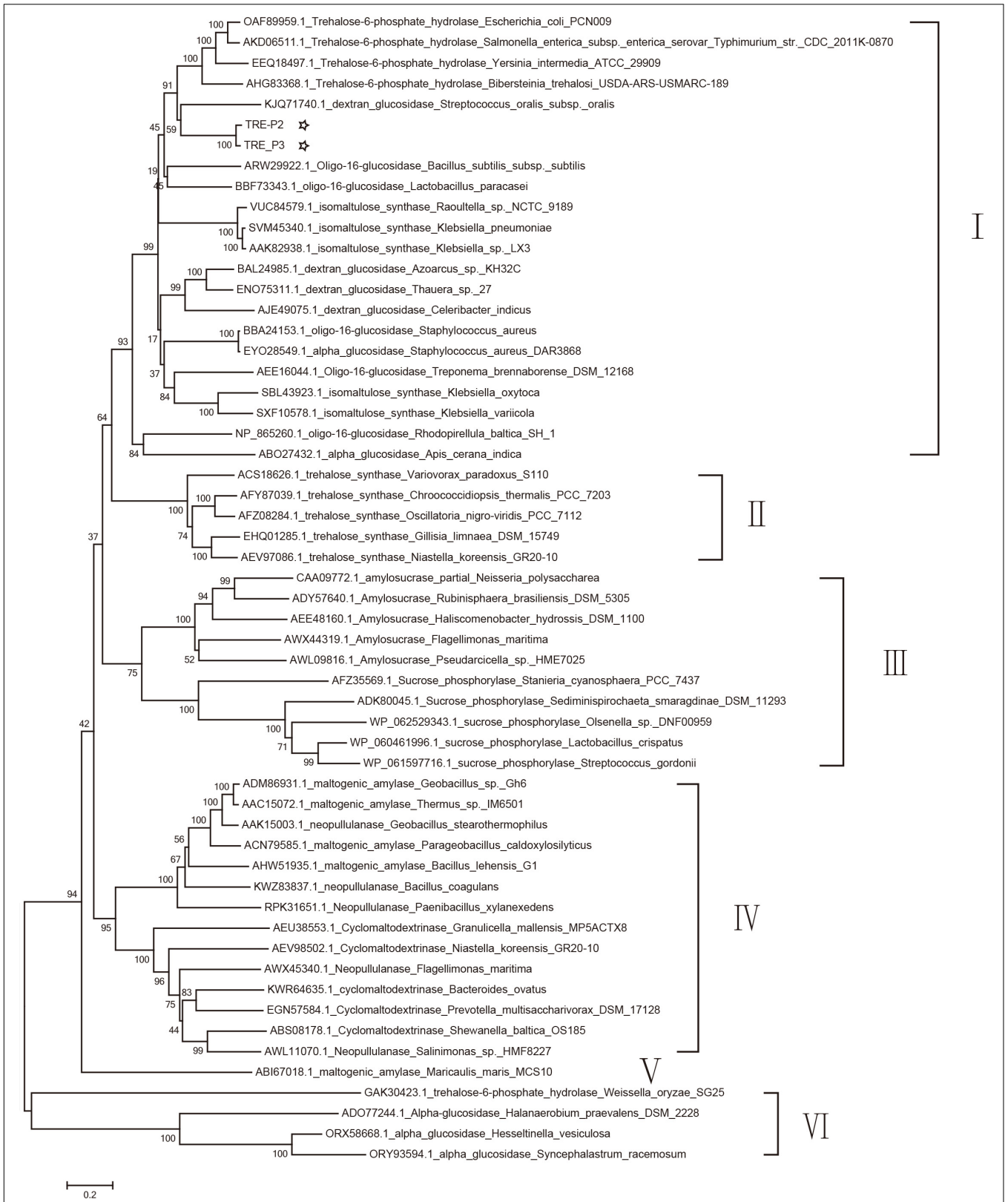


FIGURE 4 | Phylogenetic tree analysis of oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase, trehalose-6-phosphate hydrolase, amylosucrase, sucrose phosphorylase, isomaltulose synthase, trehalose synthase, cyclomaltodextrinase, maltogenic amylase, and neopullulanase which originated from various microorganisms homologous to TRE_P2 and TRE_P3 by the neighbor-joining method.

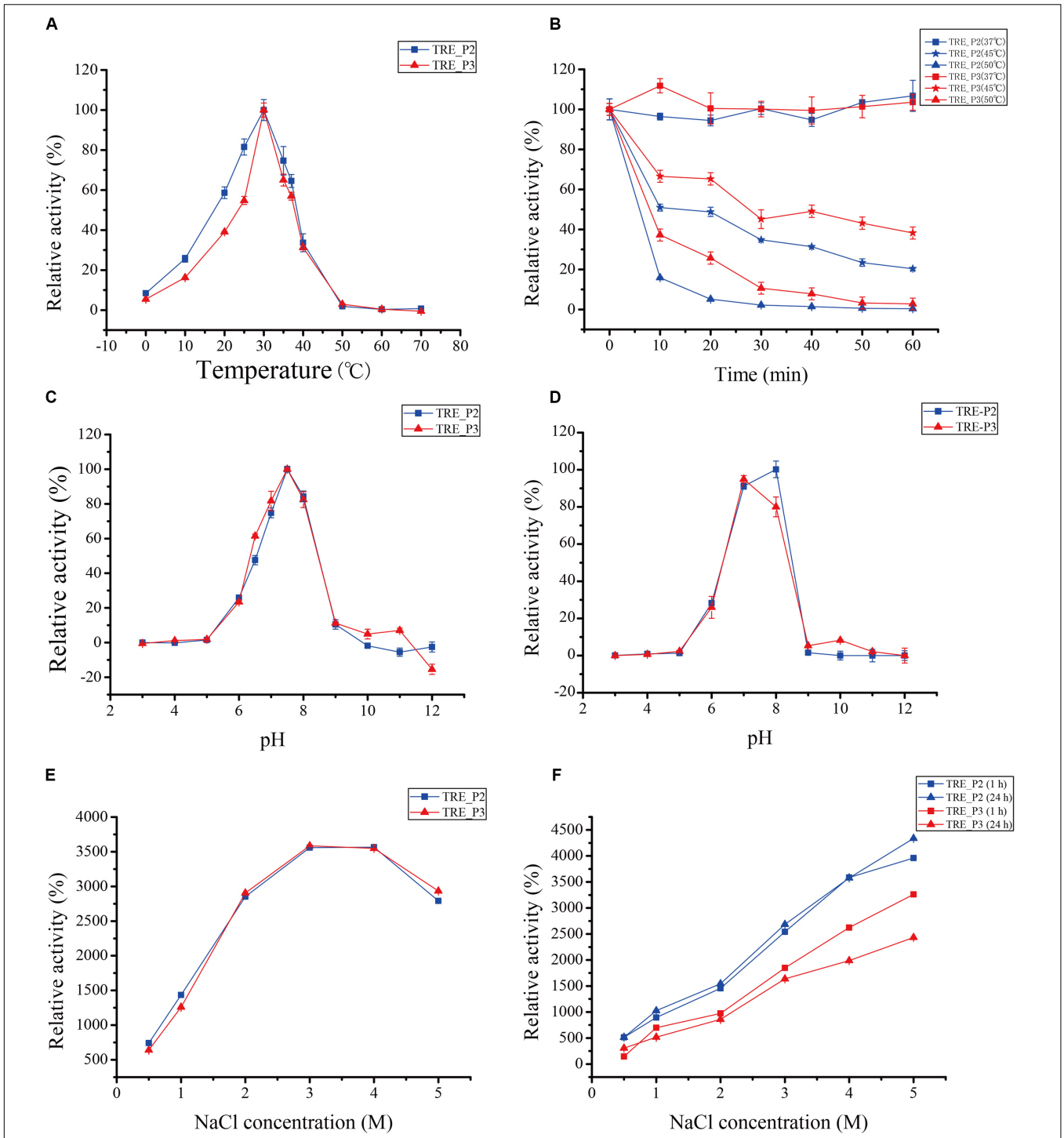


FIGURE 5 | Effects of temperature, pH, and salts on rTRE_P2 and rTRE_P3 activities. **(A)** optimum temperature; **(B)** temperature stability; **(C)** optimum pH; **(D)** pH stability; **(E)** effect of NaCl; and **(F)** NaCl stability. Error bars represent the mean ± SD (n = 3).

Recombinant rTRE_P2 and rTRE_P3 activities were slightly affected by iodoacetic acid, NBS, Triton X-100, PEG20000, and Tween 80. $MgCl_2$ and $MnSO_4$ increased the rTRE_P2 activity but inhibited the rTRE_P3 activity. The remaining ions significantly inhibited the activities of rTRE_P2 and rTRE_P3.

Four various *p*-nitrophenyl derivatives were adopted in the examining of purified rTRE_P2 and rTRE_P3 substrate specificity. These enzymes hydrolyzed *p*NPG efficiently but not *p*-nitrophenyl-β-D-galactopyranoside, *p*-nitrophenyl-β-D-glucopyranoside, *p*-nitrophenyl-α-D-mannopyranoside, and

TABLE 1 | Effects of various chemicals on relative activity of recombinant rTRE_P2 and rTRE_P3.

Chemicals (1 mM)	Relative activity of rTRE_P2	Relative activity of rTRE_P3
Metal ions		
KCl	110.54 ± 9.22%	122.77 ± 2.64%
NaCl	104.48 ± 4.21%	112.28 ± 2.08%
$FeSO_4$	26.50 ± 5.57%	44.65 ± 0.84%
$CoCl_2$	72.50 ± 5.43%	68.19 ± 0.66%
$FeCl_3$	29.23 ± 1.07%	7.70 ± 0.66%
$ZnSO_4$	0.65 ± 0.24%	1.46 ± 0.22%
$NiSO_4$	72.58 ± 8.76%	42.25 ± 3.38%
$MnSO_4$	111.08 ± 6.49%	70.31 ± 4.14%
$AgNO_3$	0.32 ± 0.39%	−7.60 ± 5.72%
$CuSO_4$	−0.42 ± 0.90%	−0.02 ± 0.42%
$CaCl_2$	22.50 ± 7.83%	53.86 ± 3.85%
$MgCl_2$	103.90 ± 9.23%	76.04 ± 2.77%
$HgCl_2$	−0.45 ± 0.37%	−0.16 ± 0.06%
$AlCl_3$	36.58 ± 16.09%	38.8 ± 4.47%
LiCl	53.28 ± 1.85%	58.4 ± 1.33%
$PbCl_2$	−2.66 ± 4.32%	1.87 ± 0.95%
Nonmetal ions		
NBS	116.02 ± 3.34%	96.4 ± 2.01%
GuHCl	110.35 ± 13.92%	125.32 ± 3.11%
Iodoacetic acid	99.20 ± 10.91%	103.26 ± 1.53%
Pig bile salt	72.73 ± 6.74%	74.44 ± 4.96%
Tween 80	90.85 ± 7.39%	96.64 ± 6.03%
Triton X-100	85.34 ± 5.05%	99.26 ± 1.12%
CTAB	−0.06 ± 0.02%	0.26 ± 0.13%
EDTA	131.17 ± 6.51%	142.71 ± 2.13%
β-Mercaptoethanol	128.38 ± 0.70%	119.66 ± 1.67%
DTT	127.52 ± 2.35%	136.50 ± 2.84%
EGTA	126.21 ± 1.77%	120.66 ± 4.84%
PEG20000	108.99 ± 3.37%	94.89 ± 1.27%
SDS	11.17 ± 3.91%	4.64 ± 0.20%

2-nitrophenyl-β-D-glucopyranoside. With *p*NPG as the substrate, the K_M and K_{cat} values of rTRE_P2 were 15.63 mM and 10.04 s^{-1}, and those of rTRE_P3 were 12.51 mM and 10.71 s^{-1}, respectively. TLC was then adopted to further assay rTRE_P2 and rTRE_P3 their specific substrate quality. As demonstrated in **Supplementary Figure S4**, no hydrolysis of trehalose was observed in the environment provided. It is believed that rTRE_P2 and rTRE_P3 have cleaved the trehalose-6-phosphate effectively and glucose and glucose-6-phosphate was thus generated.

DISCUSSION

Exploring salt-tolerant genes and enzymes in different environments has profound significance to reveal the salt-tolerant mechanism of microorganisms and direct their application in the high-salinity environment. At present, salt-tolerant enzymes and genes from marine bacteria (Zhang et al., 2017), soil (Petrovskaya et al., 2016), and gastrointestinal tract (Culligan et al., 2014b) as well as pond (Kapardar et al., 2010a), lake (Martinez-Martinez et al., 2013), biogas (Ilmberger et al., 2012), and wastewater treatment plants (Yang et al., 2016) have been reported. Salt-tolerant genes and enzymes are widely applied in high-salt food processing, seafood processing, washing, and other biotechnology fields with high salt concentrations (Margesin and Schinner, 2001). It is even more interesting that the gastrointestinal tract has its own unique salt-tolerant genes, which, through more sensitive and innovative screening assays, may stand out from salt-tolerant genes found in other environments when it comes to human-related biotech, medicine, or health researches (Culligan et al., 2013). The metagenomic technology explores the diversity and functions of microorganisms avoiding pure cultures and exhibits potential for discovering novel genes, developing new microbial active substances, and studying microbial community structure and function. In this study, we obtained 17 and 31 salt-tolerant clones from the metagenomic libraries of *N. pygmaeus* and *B. frontalis*, respectively. We selected ten clones with high salt tolerance for sequencing using high-throughput sequencing and function annotation.

Gene Ontology analysis showed that most genes in salt-tolerant clones were associated with molecular functions including catalytic activity. The KEGG analysis of the salt-tolerant clones showed presumed components of membrane transport, which perhaps implicated diverse metabolic pathways related to salt tolerance. Of course, whether these genes get involved in salt tolerance needs to be further experimentally demonstrated by individual cloning of each gene or random mutagenesis of these genes using a transposon assay. To maintain osmotic balance, bacteria have to alter their physiology for activating or deactivating specific enzymes or transporters to adapt to osmotic changes (Kempf and Bremer, 1998). COG functional analysis showed that the most abundant genes in salt-tolerant clones are mobilome: prophages and transposons, which can gain a nearby gene through LGT events, giving them a dominant position in a stress environment. Culligan et al. (2013) found that the salt-tolerant gene *stlA* in human gut microbiomes may belong to this family, which is only found in the human gut. Hence, gut microbes may have their own unique salt-tolerance mechanism. The COG database analysis also indicated the association of the genes with carbohydrate transport and metabolism. The results are consistent with the report of Harding et al. (2017), which stated that carbohydrate metabolism is induced by high salt conditions. The pathway enrichment analysis indicated signs that the salt-tolerant clones are involved in carbohydrate metabolism, illustrating that genes with these functions confer salt tolerance. Many genes were classified to be involved in transcription, replication, recombination, and repair. Wecker et al. (2009) studied the response of *Rhodopirellula baltica* to salt stress and found that it is associated with transcription in COG. Ivanova et al. (2011) sequenced the complete genome of the extremely halophilic *Halanaerobium praevalens* and discovered 115 genes that are in relation to replication, recombination, and repair and 150 genes connected with carbohydrate transport and metabolism.

A total of 1285 genes were detected from the salt-tolerant clones by high-throughput sequencing. Ninety-one potential genes coded proteins, such as general stress proteins, transcription regulators, trehalose-6-phosphate hydrolase, and zinc-dependent alcohol dehydrogenases, whose homologous proteins are closely related to salt tolerance. Transcription regulators are usually associated with bacterial response to various stresses (Hengge-Aronis et al., 1991; Cheville et al., 1996; Battesti et al., 2011; Hoffmann et al., 2013). Culligan et al. (2014a) subcloned and expressed the transcription regulator gene *sdtR* and found that it confers salt tolerance. Zinc-dependent alcohol dehydrogenase (Musa et al., 2008) is highly active and selective in non-water media. Trehalose-6-phosphate hydrolase may accumulate compatible solutes, which are widely used to cope with changing salinity concentrations (Roesser and Muller, 2001). Chuang et al. (2012) found that trehalose-6-phosphate hydrolase is important in the assimilation of trehalose. Moreover, the inactivating or denaturizing of proteins and cellular membranes can be prevented even in conditions such as desiccation, hypohydration, torridness, coldness, and oxidation by trehalose (Elbein et al., 2003). As we know, bacteria normally react to osmotic pressure changes in a so-called staged response, to be specific, (1) fast aggregation of K^+; (2) succeeding synthesis or accumulating of osmoprotectant compounds (Kempf and Bremer, 1998; Sleator and Hill, 2002; Epstein, 2003; Kunte, 2006); and (3) an assisting mechanism that may include a broad scope of genes such as *htrA* gene (Sleator and Hill, 2005), GspM and EchM proteins (Kapardar et al., 2010a), ClpS protein (Kapardar et al., 2010b), *galE*, *murB*, and *mazG* genes (Culligan et al., 2012), *stlA* gene (Culligan et al., 2013), *sdtR* gene (Culligan et al., 2014a), *brpA* gene (Culligan et al., 2014b), and *TMSRP1*, *ABCTPP*, and *TLSRP1* genes (Verma et al., 2018). As such, further studies on the relations between these genes and enzymes and the microbial salt-tolerance mechanisms might prove much need.

Trehalose-6-phosphate hydrolase TRE_P2 and TRE_P3 are from *Staphylococcus warneri* according to sequence analysis. According to current findings, trehalose-6-phosphate hydrolases are identified in *E. coli* (Rimmele and Boos, 1994), *Bacillus subtilis* (Sniezko et al., 1998), *Bacillus* sp. GP16 (Karelov et al., 2003), *Bacillus licheniformis* (Chuang et al., 2012; Ong et al., 2014; Chen et al., 2015; Lin et al., 2016), *Lactobacillus acidophilus* (Duong et al., 2006), or *Klebsiella pneumoniae* (Wu et al., 2011), among which *Lactobacillus acidophilus* is a probiotic microorganism existing in the human gastrointestinal environment. At present, the largest amount of reported trehalose-6-phosphate hydrolases found were from *Bacillus*.

The biochemical characterization research of trehalose-6-phosphate hydrolase mainly focused on the substrate specificity (Rimmele and Boos, 1994), kinetic parameters (Karelov et al., 2003), structure (Ong et al., 2014), and the effect of sugar osmolytes on the refolding of trehalose-6-phosphate hydrolase (Chen et al., 2015). For example, Karelov et al. (2003) used T6P and *p*NPG as substrates to determine the K_M and K_{cat} values of trehalose-6-phosphate hydrolase at different pHs, temperatures, and NaCl conditions, respectively. Rimmele and Boos (1994) mainly determined the substrate specificity of trehalose-6-phosphate hydrolase. Ong et al. (2014) explored the effect of Arg201, Asn327, and Tyr365 on chloride ion binding of trehalose-6-phosphate hydrolase through mutation and biophysical analysis. Only trehalose-6-phosphate hydrolase BlTreA (Chuang et al., 2012) from *Bacillus licheniformis* have been studied in detail. However, there is no report of trehalose-6-phosphate hydrolase from the gastrointestinal tract microbial metagenome.

The cloning and expressing of *tre_P2* and *tre_P3* was successfully carried out in *E. coli*, and their products were purified and characterized. The predominant molecular mass of native rTRE_P2 and rTRE_P3 is approximately 65 kDa, indicating that the enzymes mainly exist as monomer, similar to that of most trehalose-6-phosphate hydrolases (Rimmele and Boos, 1994; Karelov et al., 2003; Chuang et al., 2012). Compared with trehalose-6-phosphate hydrolase (BlTreA) from *B. licheniformis* (Chuang et al., 2012), the optimum temperature of rTRE_P2 and rTRE_P3 is similar to that of BlTreA (30°C), and the optimum pH (7.5) is lower than that of BlTreA (pH 8.0). The activity of enzymes rose with the increasing NaCl concentration. The activity of BlTreA was elevated for roughly 3.1 times when being exposed in 100 mM NaCl, whereas the activities of rTRE_P2 and rTRE_P3 were activated by 7.4- to 35.7- and 6.4- to 35.9-fold, respectively, under the condition of 0.5–5 M NaCl. This finding may stem from an increase in solubility caused by an salt-in effect. A certain concentration of salt solution can affect the structure of proteins by salting out (decreasing protein solubility) and salting in (increasing protein solubility) (Gotsche and Dahl, 1995). Another reason may be the presence of NaCl, which causes increases in the affinity of the enzyme to the substrate (Karelov et al., 2003). Besides, the considerable amount of negatively charged acid residues dispersed on the outer layer of the protein may also play a part in it as it constituted a solvation case, preventing the protein layer from dehydrating and helping the protein adapt salinity (De Santi et al., 2016; Wang et al., 2016). Ong et al. (2014) also found out that the efficiency of BlTreA and Y365A enzymes toward *p*NPG was limited when they were desalted, maintaining less than 15.8% of the TreA activity. Given this salt-tolerance property, rTRE_P2 and rTRE_P3 are useful for reporter genes (Sniezko et al., 1998) and other salt-tolerant environments.

The experimental results show that the inhibition of rTRE_P2 and rTRE_P3 activities caused by $PbCl_2$, $AgNO_3$, $CuSO_4$, $HgCl_2$, and $ZnSO_4$ probably lay in the incompatibility effect between the exogenous ions and the protein-associated cation. Chuang et al. (2012) suggested that BlTreA activity was activated by Ca^{2+} ions but was not significantly inhibited or stimulated by Co^{2+}, Mn^{2+}, Mg^{2+}, Ni^{2+}, and Fe^{2+}; meanwhile, the rTRE_P2 and rTRE_P3 activities decreased to varying degrees, except for the rTRE_P2 activity that increased to 111.08% ± 6.49% and 103.90% ± 9.23% after adding Mg^{2+} and Mn^{2+}, respectively. Recombinant rTRE_P2 and rTRE_P3 have dissevered trehalose-6-phosphate effectively, and glucose and glucose-6-phosphate were thus generated but do not hydrolyze trehalose. This finding is similar to that reported from *B. licheniformis* (Chuang et al., 2012) and *E. coli* (Rimmele and Boos, 1994). Hence, trehalose-6-phosphate hydrolase is probably essential for trehalose assimilation (Chuang et al., 2012) and is probably involved in the secondary response, synthesis, or accumulation

of osmoprotectant compounds (Kempf and Bremer, 1998; Sleator and Hill, 2002; Epstein, 2003; Kunte, 2006).

CONCLUSION

We studied the salt-tolerant genes and enzymes from the constructed gut microbial metagenomic libraries of *N. pygmaeus* and *B. frontalis* for the first time. Based on the sequencing and analysis, two salt-tolerant trehalose-6-phosphate hydrolases were identified, recombinant expressed, and characterized. TRE_P2 and TRE_P3 are the first reported salt-tolerant trehalose-6-phosphate hydrolases from the fecal microbial metagenome.

This study contributes to the limited studies of microbial salt-tolerant genes and enzymes in the gastrointestinal tract and to provide a unique genetic resource of salt-tolerant genes and enzymes for biotechnology application. Further studies of the connection between the genes and their host by transposon mutations or gene knockouts will be conducted, including identification of other enzymes involved in the trehalose metabolism. To that end, more other salt-tolerant genes and enzymes are required to fully explore the salt-tolerance mechanisms and the development and utilization of salt-tolerant genes and enzymes in gastrointestinal microorganisms.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank accession numbers MN830270 and MN830271.

AUTHOR CONTRIBUTIONS

BX designed the study. YaY, QF, and QW screened the salt-tolerant clones. BX, YuY, JD, NH, and ZH analyzed the sequencing data and provided the technical and scientific discussion. YaY, JL, and XT performed the cloning, expression, and characterization of enzymes. The manuscript was written by YaY and revised by BX. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financed by the National Natural Science Foundation of China (Grant Nos. 31860299 and 31560305) and the National Key Research and Development Program of China (Grant No. 2017YFB0308401).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01466/full#supplementary-material

FIGURE S1 | Functional classification of annotated transcripts by GO.

FIGURE S2 | Functional classification of annotated transcripts by KEGG. **(A)** 3_1; **(B)** 1A; **(C)** 5_1_1; **(D)** 5_1_5; **(E)** 21_9A; **(F)** 24_7H; **(G)** 16_2E; **(H)** 16_2D; **(I)** 16_8H; and **(J)**: 1_2G.

FIGURE S3 | SDS-PAGE analysis of the purified recombinant rTRE_P2 and rTRE_P3. M, marker proteins; 1, purified rTRE_P3; 2, unpurified rTRE_P3; 3, extracts of IPTG-induced *E. coli* BL21 (DE3) containing the empty pEASY-E2; 4, purified rTRE_P2; 5, unpurified rTRE_P2.

FIGURE S4 | Hydrolysis of trehalose and trehalose-6-phosphate by rTRE_P2 and rTRE_P3. The reaction combinations were smeared on the TLC plate using a capillary tube, followed the 2 h incubation at 30°C. Buffer system mentioned above was employed to develop the combinations. **(A)** 1, glucose; 2, glucose + rTRE_P2; 3, glucose-6-phosphate; 4, glucose-6-phosphate + rTRE_P2; 5, trehalose; 6, trehalose + rTRE_P2; 7, trehalose-6-phosphate; 8, trehalose-6-phosphate + rTRE_P2. **(B)** 1, glucose; 2, glucose + rTRE_P3; 3, glucose-6-phosphate; 4, glucose-6-phosphate + rTRE_P3; 5, trehalose; 6, trehalose + rTRE_P3; 7, trehalose-6-phosphate; and 8, trehalose-6-phosphate + rTRE_P3.

TABLE S1 | Statistics of the trimmed sequences.

TABLE S2 | Detection of putative salt tolerance enzymes and genes of the salt tolerance clones.

REFERENCES

Alvarez-Ordonez, A., Begley, M., Prieto, M., Messens, W., Lopez, M., Bernardo, A., et al. (2011). *Salmonella* spp. survival strategies within the host gastrointestinal tract. *Microbiology* 157(Pt 12), 3268–3281. doi: 10.1099/mic.0.050351-0

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65, 189–213. doi: 10.1146/annurev-micro-090110-102946

Cairns, R. A., Harris, I. S., and Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nat. Rev. Cancer* 11, 85–95. doi: 10.1038/nrc2981

Chen, J. H., Chi, M. C., Lin, M. G., Lin, L. L., and Wang, T. F. (2015). Beneficial effect of sugar osmolytes on the refolding of guanidine hydrochloride-denatured trehalose-6-phosphate hydrolase from *Bacillus licheniformis*. *Biomed. Res. Int.* 2015:806847. doi: 10.1155/2015/806847

Cheville, A. M., Arnold, K. W., Buchrieser, C., Cheng, C. M., and Kaspar, C. W. (1996). rpoS regulation of acid, heat, and salt tolerance in *Escherichia coli* O157:H7. *Appl. Environ. Microbiol.* 62, 1822–1824. doi: 10.1128/aem.62.5.1822-1824.1996

Chuang, T. T., Ong, P. L., Wang, T. F., Huang, H. B., Chi, M. C., and Lin, L. L. (2012). Molecular characterization of a novel trehalose-6-phosphate hydrolase, TreA, from *Bacillus licheniformis*. *Int. J. Biol. Macromol.* 50, 459–470. doi: 10.1016/j.ijbiomac.2012.01.011

Culligan, E. P., Marchesi, J. R., Hill, C., and Sleator, R. D. (2014a). Combined metagenomic and phenomic approaches identify a novel salt tolerance gene from the human gut microbiome. *Front. Microbiol.* 5:189. doi: 10.3389/fmicb.2014.00189

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2014b). Metagenomic identification of a novel salt tolerance gene from the human gut microbiome which encodes a membrane protein with homology to a brp/blh-family beta-carotene 15,15'-monooxygenase. *PLoS One* 9:e103318. doi: 10.1371/journal.pone.0103318

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2012). Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J.* 6, 1916–1925. doi: 10.1038/ismej.2012.38

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2013). Functional environmental screening of a metagenomic library identifies stlA; a unique salt tolerance locus from the human gut microbiome. *PLoS One* 8:e82985. doi: 10.1371/journal.pone.0082985

Das, K. C., Haque, N., Baruah, K. K., Rajkhowa, C., and Mondal, M. (2011). Comparative nutrient utilization, growth, and rumen enzyme profile of mithun

(*Bos frontalis*) and Tho-tho cattle (*Bos indicus*) fed on tree-leaves-based ration. *Trop. Anim. Health Prod.* 43, 209–214. doi: 10.1007/s11250-010-9676-1

De Dea Lindner, J., Canchaya, C., Zhang, Z., Neviani, E., Fitzgerald, G. F., van Sinderen, D., et al. (2007). Exploiting Bifidobacterium genomes: the molecular basis of stress response. *Int. J. Food Microbiol.* 120, 13–24. doi: 10.1016/j.ijfoodmicro.2007.06.016

De Santi, C., Leiros, H. K., Di Scala, A., de Pascale, D., Altermark, B., and Willassen, N. P. (2016). Biochemical characterization and structural analysis of a new cold-active and salt-tolerant esterase from the marine bacterium *Thalassospira* sp. *Extremophiles* 20, 323–336. doi: 10.1007/s00792-016-0824-z

Deutch, C. E., and Yang, S. (2020). Genomic sequencing of Gracilibacillus dipsosauri reveals key properties of a salt-tolerant alpha-amylase. *Antonie Van Leeuwenhoek* 113, 1049–1059. doi: 10.1007/s10482-020-01417-2

Dong, M., Yang, Y., Tang, X., Shen, J., Xu, B., Li, J., et al. (2016). NaCl-, protease-tolerant and cold-active endoglucanase from *Paenibacillus sp*. YD236 isolated from the feces of *Bos frontalis*. *Springerplus* 5:746. doi: 10.1186/s40064-016-2360-9

Duong, T., Barrangou, R., Russell, W. M., and Klaenhammer, T. R. (2006). Characterization of the tre locus and analysis of trehalose cryoprotection in *Lactobacillus acidophilus* NCFM. *Appl. Environ. Microbiol.* 72, 1218–1225. doi: 10.1128/aem.72.2.1218-1225.2006

Egert, M., de Graaf, A. A., Smidt, H., de Vos, W. M., and Venema, K. (2006). Beyond diversity: functional microbiomics of the human colon. *Trends Microbiol.* 14, 86–91. doi: 10.1016/j.tim.2005.12.007

Elbein, A. D., Pan, Y. T., Pastuszak, I., and Carroll, D. (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology* 13, 17r–27r. doi: 10.1093/glycob/cwg047

Epstein, W. (2003). The roles and regulation of potassium in bacteria. *Prog. Nucleic Acid Res. Mol. Biol.* 75, 293–320. doi: 10.1016/s0079-6603(03)75008-9

Ermolenko, D. N., and Makhatadze, G. I. (2002). Bacterial cold-shock proteins. *Cell Mol. Life Sci.* 59, 1902–1913. doi: 10.1007/pl00012513

Furuya, K., and Hutchinson, C. R. (1998). The DrrC protein of *Streptomyces peucetius*, a UvrA-like protein, is a DNA-binding protein whose gene is induced by daunorubicin. *FEMS Microbiol. Lett.* 168, 243–249. doi: 10.1111/j.1574-6968.1998.tb13280.x

Gebicka, L., and Didik, J. (2009). Catalytic scavenging of peroxynitrite by catalase. *J. Inorg. Biochem.* 103, 1375–1379. doi: 10.1016/j.jinorgbio.2009.07.011

Gotsche, S., and Dahl, M. K. (1995). Purification and characterization of the phospho-alpha(1,1)glucosidase (TreA) of Bacillus subtilis 168. *J. Bacteriol.* 177, 2721–2726. doi: 10.1128/jb.177.10.2721-2726.1995

Guo, M., Wang, R., Wang, J., Hua, K., Wang, Y., Liu, X., et al. (2014). ALT1, a Snf2 family chromatin remodeling ATPase, negatively regulates alkaline tolerance through enhanced defense against oxidative stress in rice. *PLoS One* 9:e112515. doi: 10.1371/journal.pone.0112515

Han, X., Hu, X., Zhou, C., Wang, H., Li, Q., Ouyang, Y., et al. (2020). Cloning and functional characterization of xylitol dehydrogenase genes from *Issatchenkia orientalis* and T*orulaspora delbrueckii*. *J. Biosci. Bioeng.* 130, 29–35. doi: 10.1016/j.jbiosc.2020.02.012

Harding, T., Roger, A. J., and Simpson, A. G. B. (2017). Adaptations to high salt in a halophilic protist: differential expression and gene acquisitions through duplications and gene transfers. *Front. Microbiol.* 8:944. doi: 10.3389/fmicb.2017.00944

Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M., and Boos, W. (1991). Trehalose synthesis genes are controlled by the putative sigma factor encoded by rpoS and are involved in stationary-phase thermotolerance in *Escherichia coli*. *J. Bacteriol.* 173, 7918–7924. doi: 10.1128/jb.173.24.7918-7924.1991

Hoffmann, R. F., McLernon, S., Feeney, A., Hill, C., and Sleator, R. D. (2013). A single point mutation in the listerial betL sigma(A)-dependent promoter leads to improved osmo- and chill-tolerance and a morphological shift at elevated osmolarity. *Bioengineered* 4, 401–407. doi: 10.4161/bioe.24094

Hossein Helalat, S., Bidaj, S., Samani, S., and Moradi, M. (2019). Producing alcohol and salt stress tolerant strain of *Saccharomyces cerevisiae* by heterologous expression of pprI gene. *Enzyme Microb. Technol.* 124, 17–22. doi: 10.1016/j.enzmictec.2019.01.008

Huang, J., Zhang, H., Wang, J., and Yang, J. (2003). Molecular cloning and characterization of rice 6-phosphogluconate dehydrogenase gene that is up-regulated by salt stress. *Mol. Biol. Rep.* 30, 223–227.

Ilmberger, N., Meske, D., Juergensen, J., Schulte, M., Barthen, P., Rabausch, U., et al. (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids–linking thermostability and halotolerance. *Appl. Microbiol. Biotechnol.* 95, 135–146. doi: 10.1007/s00253-011-3732-2

Ivanova, N., Sikorski, J., Chertkov, O., Nolan, M., Lucas, S., Hammon, N., et al. (2011). Complete genome sequence of the extremely halophilic *Halanaerobium praevalens* type strain (GSL). *Stand. Genomic Sci.* 4, 312–321. doi: 10.4056/sigs.1824509

Kapardar, R. K., Ranjan, R., Grover, A., Puri, M., and Sharma, R. (2010a). Identification and characterization of genes conferring salt tolerance to *Escherichia coli* from pond water metagenome. *Bioresour. Technol.* 101, 3917–3924. doi: 10.1016/j.biortech.2010.01.017

Kapardar, R. K., Ranjan, R., Puri, M., and Sharma, R. (2010b). Sequence analysis of a salt tolerant metagenomic clone. *Indian J. Microbiol.* 50, 212–215. doi: 10.1007/s12088-010-0041-x

Karelov, D. V., Krasikov, V. V., Surjik, M. A., and Firsov, L. M. (2003). Expression, isolation, purification, and biochemical properties of trehalose-6-phosphate hydrolase from thermoresistant strain *Bacillus* sp. GP16. *Biochemistry* 68, 1012–1019. doi: 10.1023/a:1026068714197

Kempf, B., and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* 170, 319–330. doi: 10.1007/s002030050649

Kumari, N., and Jagadevan, S. (2016). Genetic identification of arsenate reductase and arsenite oxidase in redox transformations carried out by arsenic metabolising prokaryotes - A comprehensive review. *Chemosphere* 163, 400–412. doi: 10.1016/j.chemosphere.2016.08.044

Kunte, H. J. (2006). Osmoregulation in bacteria: compatible solute accumulation and osmosensing. *Environ. Chem.* 3, 94–99. doi: 10.1071/EN06016

Kvint, K., Nachin, L., Diez, A., and Nystrom, T. (2003). The bacterial universal stress protein: function and regulation. *Curr. Opin. Microbiol.* 6, 140–145. doi: 10.1016/s1369-5274(03)00025-0

Lin, M. G., Chi, M. C., Naveen, V., Li, Y. C., Lin, L. L., and Hsiao, C. D. (2016). *Bacillus licheniformis* trehalose-6-phosphate hydrolase structures suggest keys to substrate specificity. *Acta Crystallogr. D Struct. Biol.* 72(Pt 1), 59–70. doi: 10.1107/s2059798315020756

Liu, J., Qu, W., and Kadiiska, M. B. (2009). Role of oxidative stress in cadmium toxicity and carcinogenesis. *Toxicol. Appl. Pharmacol.* 238, 209–214. doi: 10.1016/j.taap.2009.01.029

Livermore, D. M. (1995). beta-Lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* 8, 557–584. doi: 10.1128/cmr.8.4.557-584.1995

Louis, P., and O'Byrne, C. P. (2010). Life in the gut: microbial responses to stress in the gastrointestinal tract. *Sci. Prog.* 93(Pt 1), 7–36. doi: 10.3184/003685009x12605525292307

Margesin, R., and Schinner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles* 5, 73–83. doi: 10.1007/s007920100184

Martinez-Martinez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., et al. (2013). Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol.* 79, 3553–3562. doi: 10.1128/aem.00240-13

Maul, B., Volker, U., Riethdorf, S., Engelmann, S., and Hecker, M. (1995). sigma B-dependent regulation of gsiB in response to multiple stimuli in *Bacillus subtilis*. *Mol. Gen. Genet.* 248, 114–120. doi: 10.1007/bf02456620

Mondal, M., Dhali, A., Rajkhowa, C., and Prakash, B. S. (2004). Secretion patterns of growth hormone in growing captive mithuns (*Bos frontalis*). *Zoolog. Sci.* 21, 1125–1129. doi: 10.2108/zsj.21.1125

Musa, M. M., Ziegelmann-Fjeld, K. I., Vieille, C., and Phillips, R. S. (2008). Activity and selectivity of W110A secondary alcohol dehydrogenase from *Thermoanaerobacterethanolicus* in organic solvents and ionic liquids: mono- and biphasic media. *Org. Biomol. Chem.* 6, 887–892. doi: 10.1039/b717120j

Nemoto, Y., and Sasakuma, T. (2000). Specific expression of glucose-6-phosphate dehydrogenase (G6PDH) gene by salt stress in wheat (*Triticum aestivum* L.). *Plant Sci.* 158, 53–60. doi: 10.1016/s0168-9452(00)00305-8

Nguyen, T. T., Klueva, N., Chamareck, V., Aarti, A., Magpantay, G., Millena, A. C., et al. (2004). Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Mol. Genet. Genomics* 272, 35–46. doi: 10.1007/s00438-004-1025-5

Oleksiewicz, U., Liloglou, T., Field, J. K., and Xinarianos, G. (2011). Cytoglobin: biochemical, functional and clinical perspective of the newest member of the globin family. *Cell Mol. Life Sci.* 68, 3869–3883. doi: 10.1007/s00018-011-0764-9

Ong, P.-L., Chuang, T.-T., Wang, T.-F., and Lin, L.-L. (2014). Identification of critical amino acid residues for chloride binding of *Bacillus licheniformis* trehalose-6-phosphate hydrolase. *Biologia* 69, 1–9. doi: 10.2478/s11756-013-0290-3

Petrovskaya, L. E., Novototskaya-Vlasova, K. A., Spirina, E. V., Durdenko, E. V., Lomakina, G. Y., Zavialova, M. G., et al. (2016). Expression and characterization of a new esterase with GCSAG motif from a permafrost metagenomic library. *FEMS Microbiol. Ecol.* 92:fiw046. doi: 10.1093/femsec/fiw046

Prija, F., and Prasad, R. (2017). DrrC protein of Streptomyces peucetius removes daunorubicin from intercalated dnrI promoter. *Microbiol. Res.* 202, 30–35. doi: 10.1016/j.micres.2017.05.002

Rani, A., Kumar, A., Lal, A., and Pant, M. (2014). Cellular mechanisms of cadmium-induced toxicity: a review. *Int. J. Environ. Health Res.* 24, 378–399. doi: 10.1080/09603123.2013.835032

Rimmele, M., and Boos, W. (1994). Trehalose-6-phosphate hydrolase of *Escherichia coli*. *J. Bacteriol.* 176, 5654–5664. doi: 10.1128/jb.176.18.5654-5664.1994

Roesser, M., and Muller, V. (2001). Osmoadaptation in bacteria and archaea: common principles and differences. *Environ. Microbiol.* 3, 743–754. doi: 10.1046/j.1462-2920.2001.00252.x

Shen, Z. J., Zhang, S. D., Liu, Y. J., Liu, X. M., Li, Z., Zhang, Q. W., et al. (2018). Functional analysis by RNAi of an glutaredoxin gene in *Helicoverpa armigera*. *J. Insect Physiol.* 106(Pt 2), 98–105. doi: 10.1016/j.jinsphys.2017.10.011

Sleator, R. D., and Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol. Rev.* 26, 49–71. doi: 10.1111/j.1574-6976.2002.tb00598.x

Sleator, R. D., and Hill, C. (2005). A novel role for the LisRK two-component regulatory system in listerial osmotolerance. *Clin. Microbiol. Infect.* 11, 599–601. doi: 10.1111/j.1469-0691.2005.01176.x

Sniezko, I., Dobson-Stone, C., and Klein, A. (1998). The treA gene of *Bacillus subtilis* is a suitable reporter gene for the archaeon *Methanococcus voltae*. *FEMS Microbiol. Lett.* 164, 237–242. doi: 10.1111/j.1574-6968.1998.tb13092.x

Somani, R. R., and Chaskar, P. K. (2018). Arginine deiminase enzyme evolving as a potential antitumor agent. *Mini Rev. Med. Chem.* 18, 363–368. doi: 10.2174/1389557516666160817102701

Thevenod, F., and Lee, W. K. (2013). Cadmium and cellular signaling cascades: interactions between cell death and survival pathways. *Arch. Toxicol.* 87, 1743–1786. doi: 10.1007/s00204-013-1110-9

Van Assche, F., Cardinaels, C., and Clijsters, H. (1988). Induction of enzyme capacity in plants as a result of heavy metal toxicity: dose-response relations in *Phaseolus vulgaris* L., treated with zinc and cadmium. *Environ. Pollut.* 52, 103–115. doi: 10.1016/0269-7491(88)90084-x

Verma, M. K., Ahmed, V., Gupta, S., Kumar, J., Pandey, R., Mandhan, V., et al. (2018). Functional metagenomics identifies novel genes *ABCTPP*, *TMSRP1* and *TLSRP1* among human gut enterotypes. *Sci. Rep.* 8:1397. doi: 10.1038/s41598-018-19862-5

Wang, G., Wang, Q., Lin, X., Ng, T. B., Yan, R., Lin, J., et al. (2016). A novel cold-adapted and highly salt-tolerant esterase from *Alkalibacterium* sp. SL3 from the sediment of a soda lake. *Sci. Rep.* 6:19494. doi: 10.1038/srep19494

Wecker, P., Klockow, C., Ellrott, A., Quast, C., Langhammer, P., Harder, J., et al. (2009). Transcriptional response of the model planctomycete *Rhodopirellula baltica* SH1(T) to changing environmental conditions. *BMC Genomics* 10:410. doi: 10.1186/1471-2164-10-410

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578

Wholey, W. Y., and Jakob, U. (2012). Hsp33 confers bleach resistance by protecting elongation factor Tu against oxidative degradation in *Vibrio cholerae*. *Mol. Microbiol.* 83, 981–991. doi: 10.1111/j.1365-2958.2012.07982.x

Wu, M. C., Lin, T. L., Hsieh, P. F., Yang, H. C., and Wang, J. T. (2011). Isolation of genes involved in biofilm formation of a *Klebsiella pneumoniae* strain causing pyogenic liver abscess. *PLoS One* 6:e23500. doi: 10.1371/journal.pone.0023500

Xu, B., Dai, L., Zhang, W., Yang, Y., Wu, Q., Li, J., et al. (2019). Characterization of a novel salt-, xylose- and alkali-tolerant GH43 bifunctional beta-xylosidase/alpha-l-arabinofuranosidase from the gut bacterial genome. *J. Biosci. Bioeng.* 128, 429–437. doi: 10.1016/j.jbiosc.2019.03.018

Xu, B., Yang, F., Xiong, C., Li, J., Tang, X., Zhou, J., et al. (2014). Cloning and characterization of a novel alpha-amylase from a fecal microbial metagenome. *J. Microbiol. Biotechnol.* 24, 447–452. doi: 10.4014/jmb.1310.10121

Yaish, M. W., Patankar, H. V., Assaha, D. V. M., Zheng, Y., Al-Yahyai, R., and Sunkar, R. (2017). Genome-wide expression profiling in leaves and roots of date palm (*Phoenix dactylifera* L.) exposed to salinity. *BMC Genomics* 18:246. doi: 10.1186/s12864-017-3633-6

Yang, C., Xia, Y., Qu, H., Li, A. D., Liu, R., Wang, Y., et al. (2016). Discovery of new cellulases from the metagenome by a metagenomics-guided strategy. *Biotechnol. Biofuels* 9:138. doi: 10.1186/s13068-016-0557-3

Yang, H., and Shu, Y. (2015). Cadmium transporters in the kidney and cadmium-induced nephrotoxicity. *Int. J. Mol. Sci.* 16, 1484–1494. doi: 10.3390/ijms16011484

Zhang, Y., Hao, J., Zhang, Y. Q., Chen, X. L., Xie, B. B., Shi, M., et al. (2017). Identification and characterization of a novel salt-tolerant esterase from the deep-sea sediment of the South China Sea. *Front. Microbiol.* 8:441. doi: 10.3389/fmicb.2017.00441

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 17 September 2020
doi: 10.3389/fmicb.2020.551109

Xylanolytic Extremozymes Retrieved From Environmental Metagenomes: Characteristics, Genetic Engineering, and Applications

*Digvijay Verma[1] and Tulasi Satyanarayana[2]**

[1] Department of Microbiology, Babasaheb Bhimrao Ambedkar (Central) University, Lucknow, India, [2] Department of Biological Sciences and Engineering, Netaji Subhas University of Technology, Dwarka, New Delhi, India

Xylanolytic enzymes have extensive applications in paper, food, and feed, pharmaceutical, and biofuel industries. These industries demand xylanases that are functional under extreme conditions, such as high temperature, acidic/alkaline pH, and others, which are prevailing in bioprocessing industries. Despite the availability of several xylan-hydrolyzing enzymes from cultured microbes, there is a huge gap between what is available and what industries require. DNA manipulations as well as protein-engineering techniques are also not quite satisfactory in generating xylan-hydrolyzing extremozymes. With a compound annual growth rate of 6.6% of xylan-hydrolyzing enzymes in the global market, there is a need for xylanolytic extremozymes. Therefore, metagenomic approaches have been employed to uncover hidden xylanolytic genes that were earlier inaccessible in culture-dependent approaches. Appreciable success has been achieved in retrieving several unusual xylanolytic enzymes with novel and desirable characteristics from different extreme environments using functional and sequence-based metagenomic approaches. Moreover, the Carbohydrate Active Enzymes database includes approximately 400 GH-10 and GH-11 unclassified xylanases. This review discusses sources, characteristics, and applications of xylanolytic enzymes obtained through metagenomic approaches and their amelioration by genetic engineering techniques.

Keywords: extremozymes, xylanolytic enzymes, metagenomes, metagenomics, GH-10 and GH-11 xylanases, genetic engineering

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Hongge Chen,
Henan Agricultural University, China
Ramón Alberto Batista-García,
Universidad Autónoma del Estado de Morelos, Mexico

***Correspondence:**
Tulasi Satyanarayana
tsnarayana@gmail.com

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *11 April 2020*
Accepted: *21 August 2020*
Published: *17 September 2020*

Citation:
Verma D and Satyanarayana T (2020) Xylanolytic Extremozymes Retrieved From Environmental Metagenomes: Characteristics, Genetic Engineering, and Applications. Front. Microbiol. 11:551109. doi: 10.3389/fmicb.2020.551109

INTRODUCTION

Extensive biotechnological applications of xylanolytic extremozymes have raised interest as well as their demand in several industrial processes. Extremozymes are expected to withstand extreme conditions posed during the downstream processing of lignocellulosic materials. For example, thermo-alkali-stable xylanases are well suited for the paper industry for their use in bio-bleaching of paper pulp (Kumar et al., 2016; **Figure 1**). The bread industry prefers xylanases/β-xylosidases of thermo-acid stable properties (Uhl and Daniel, 1999; Harris and Ramalingam, 2010; Kumar et al., 2016). Salt-tolerant hemicellulases are the choice of seafood processing industries (Al-Darkazali et al., 2017). Similarly, barophilic, psychrophilic, and xylan-degrading enzymes functional at

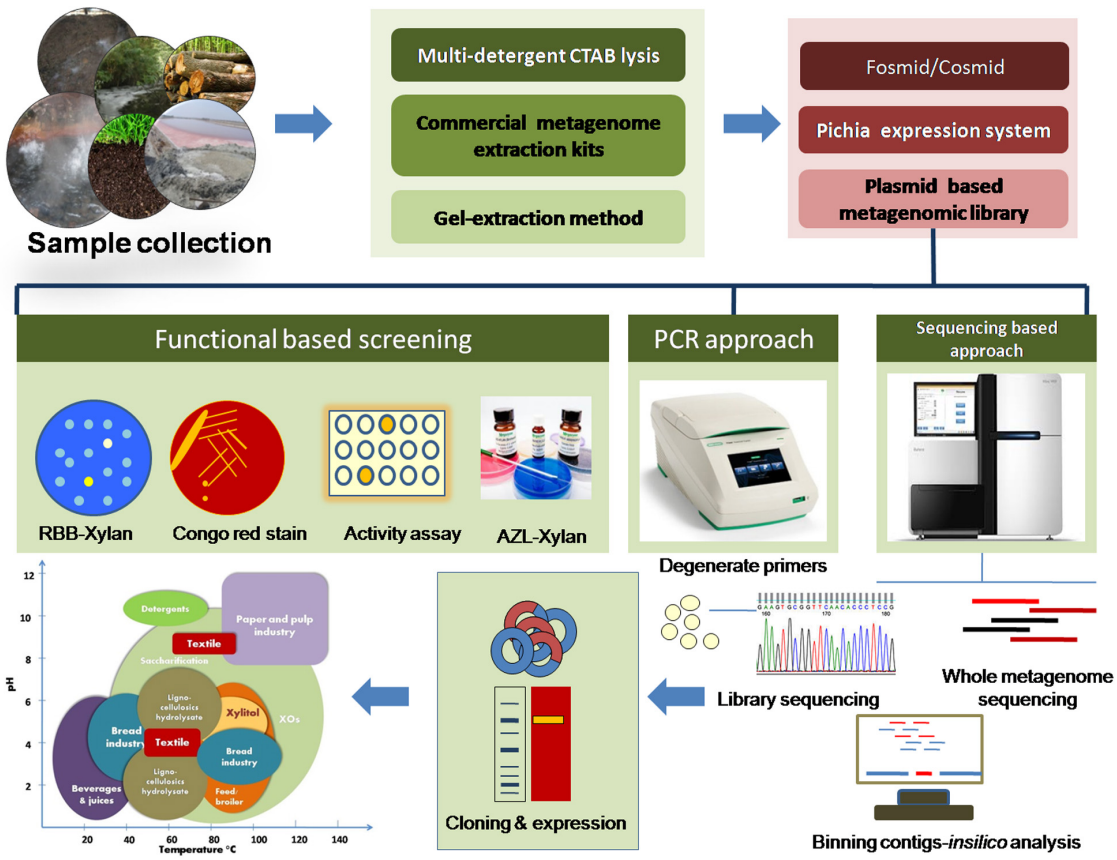


FIGURE 1 | Schematic representation for retrieving xylan-degrading enzymes using metagenomic approaches.

low water activity find multiple applications in food and beverage processing. Microorganisms exhibit a high degree of genomic and metabolic flexibility for adapting to extreme environmental conditions, and therefore, they occur ubiquitously and act as a reservoir for proteins/enzymes capable of withstanding extreme environmental conditions (Badhai et al., 2015). Several extreme environments have been explored for obtaining extremophilic microbes for their exploitation in harnessing extremozymes (Shi H. et al., 2013; Wang et al., 2015). Many xylan-depolymerizing enzymes have also been identified from innumerable prokaryotic and eukaryotic microorganisms from extreme environments (Kumar et al., 2013a; Verma et al., 2019). However, the limitations associated with these microbial enzymes cannot be ignored. Xylanolytic enzymes of bacterial origin are mostly preferred over the fungal xylanases due to the retention of activity by the former under extreme environmental conditions (Kumar et al., 2016). Broad pH and temperature optima and fair stability under extreme conditions in addition to broad substrate specificity make the bacterial xylanases preferable over those from fungi although the production levels are of the former are lower than those of the latter (Verma et al., 2019).

Due to the problems encountered in simulating environmental conditions in the laboratory and limited culturability, we lose a major chunk of microorganisms present in environmental samples. Metagenomics has emerged as an alternative tool for discovering novel bioactive molecules from various environments (Lorenz and Eck, 2005; Berini et al., 2017; Fredriksen et al., 2019). Several novel xylanolytic enzymes with unique properties have also been obtained from various extreme environmental samples using metagenomic approaches. Wang et al. (2016) further emphasize digging out the entire enzyme assemblages from environments that can depolymerize the target complex polysaccharides efficiently. Such a cocktail of enzymes usually share similar biophysical properties due to their common origin, which further enhances their compatibility for complete substrate hydrolysis (Berini et al., 2017). Several such lignocellulases have been successfully derived using the metagenomic approaches (Thidarat et al., 2012; Wang et al., 2015; Liu et al., 2018). Unfortunately, a majority of the discovered xylanolytic enzymes are still uncharacterized, which can be seen in the Carbohydrate Active Enzymes (CAZy) database[1]. In addition, most of those characterized are limited to their biophysical properties only and not well explored for their applicability under industrial process conditions. This review summarizes various facets of metagenomic xylanolytic extremozymes, such as their characteristics, comparison with the available xylan-degrading enzymes, improvement by genetic/protein engineering, and potential applications.

[1]www.cazydatabse.com

TIMELINE OF METAGENOMIC XYLANOLYTIC ENZYME DISCOVERY

It took nearly 10 years to report the first xylan-degrading enzyme through the metagenomic approach after its introduction. Sunna and Bergquist (2003) were the pioneers who reported the first xylanase from the hot pool metagenome using a genome-walking PCR (GW-PCR) approach. The retrieved xylanase shared similarity with GH-10 xylanase with unusual features; thus, this can be considered the first report of the GH-10 family xylanase of metagenomic origin. It was followed by Brennan et al. (2004), who reported the first GH-11 and GH-8 xylanases from the insect gut metagenome using direct cloning of the environmental DNA. Lee et al. (2006) reported the second GH-8 xylanase from the lagoon of a dairy farm. The first multifunctional glycosyl hydrolases of GH-5 and GH-26 were reported by Palackal et al. (2007). Attempts have been made for the first time to enhance the sensitivity of PCR using rolling circle amplification of the horse gut metagenome that revealed five GH-10 xylanases (Yamada et al., 2008). Of various environmental samples, the soil was first explored for cloning community DNA by Hu et al. (2008) to uncover hidden metagenomic xylanases. Similarly, Kwon et al. (2010) investigated the compost soil metagenome and reported five xylanases out of 12,380 fosmid clones. The claim of Jeong et al. (2012) for reporting the first xylanase from the compost metagenomic library, therefore, stands corrected. Verma et al. (2013a) first reported an industrially relevant metagenomic endoxylanase that finds applicability in bio-bleaching of pulp samples.

A novel xylanosome operon having three contiguous genes of family GH-8, GH-10, and GH-11 was identified from a wood-feeding termite gut metagenome in 2012 (Thidarat et al., 2012). Further, Zhang et al. (2014) studied the role of human gut inhabitant Bacteroidetes by reporting two GH-10 family endoxylanases from the human gut metagenome. Several new findings associated with the xylan-depolymerizing enzymes of metagenomic origin have been published during the past 5 years (Wang et al., 2015; Rashamuse et al., 2016; Kim et al., 2018; Ellila et al., 2019; Alves et al., 2020; Victorica et al., 2020; **Tables 1**, **2**). Recently, a novel computational method, thermal activity prediction for xylanase (TAXyl), was developed to predict the thermal activity of GH-10 and GH-11 xylanases (Shahraki et al., 2019). Although several xylan-degrading enzymes are known today, using culture-independent metagenomic approaches, less than 100 of them have been adequately characterized. Therefore, there is an urgent need to characterize the remaining metagenomic xylanases.

STRATEGIES TO BOOST THE RETRIEVAL OF METAGENOMIC XYLANOLYTIC ENZYMES

Metagenomics demands a rational approach to attain the desired xylanolytic enzymes, and it starts from sample collection to screening to final retrieval of the genes (**Figure 1**). The following are the few criteria that should be considered for attaining xylan-degrading enzymes from environmental metagenomes.

Selection of the Xylan Substrates

High-throughput screening strategies enhance the possibility of searching the novel bioactive molecules multifold. For xylanase screening, the traditional Congo red staining method is highly popular; however, it is tedious and labor-intensive, as it demands rigorous replica plating for screening metagenomic libraries. Incorporation of Remazol Brilliant Blue (RBB) cross-linked xylan into the cultivation medium makes the screening efficient, and an appropriately diluted library can be screened for detecting colonies displaying a halo zone of xylan hydrolysis. However, RBB-xylan is quite expensive and can be prepared in the laboratory according to the protocol of Bailey et al. (1992). The fluorescence-based method EnzChek® Ultra Xylanase Assay Kit (Invitrogen, Carlsbad, CA, United States) and Xylazyme tablet (Megazyme, Bray, Ireland) account for the other screening techniques; Azurine-crosslinked xylan (AZCL xylan) is employed as dye conjugated substrates (Selvarajan and Veena, 2017). Several clones have shown a blue halo from fosmid-based metagenomic libraries using AZCL xylan (Ellila et al., 2019; Knapik et al., 2019).

Selection of xylan type is also decisive in obtaining positive clones during functional screening. Birchwood xylan and others are recommended over oat–spelt xylan, Larchwood xylan, and arabinoxylan (Shi H. et al., 2013; Verma et al., 2013a). Birchwood xylan comprises the highest amount of xylose sugar with 89.3% (Kormelink and Voragen, 1993) as compared to beechwood xylan (80.8% xylose; Megazyme) and arabinoxylan (65.8% xylose; Gruppen et al., 1992). Functional screening permits recovery of several full-length metagenomic xylanase-encoding genes by using RBB xylan/AZCL xylan as substrates (Lee et al., 2006; Zhao et al., 2010; Verma and Satynarayana, 2011).

Fosmid Libraries Are the Preferred Choice

The success rate of functional screening of libraries is significantly higher than the sequence-based approaches in obtaining more positive clones as the latter provides usual partial sequences (Lorenz and Eck, 2005; Berini et al., 2017). Several such metagenomic libraries have been constructed in plasmids (Verma et al., 2013b), cosmids (Mo et al., 2010; Chang et al., 2011; Bao et al., 2012), and fosmids (Jeong et al., 2012; Knapik et al., 2019) for retrieving genes that encode xylan-degrading enzymes (Wang et al., 2015; Kim et al., 2018; Thornbury et al., 2019). Among them, a majority of the metagenomic libraries have been constructed in the fosmid vectors (pCC2FOS^{TM} and pCC1FOS^{TM}) with its host strain EPI300-T1R^{TM}, provided by Epicentre, for obtaining genes encoding xylan-degrading enzymes (Jeong et al., 2012; Kim et al., 2018; Liu et al., 2019). Fosmids can harbor large stretches of DNA (40–60 kb) that enhance the number of positive hits. Alves et al. (2020) report

TABLE 1 | An update of metagenomic GH-10 and GH-11 endoxylanases.

S. No.	Source	Name	Vector/host/ approaches	Library type/ approach	Positive hits	Opt_{pH}	Opt_{Temp}.(°C)	Molecular mass (kDa)	Family	References
1.	Camel rumen metagenome	PersiXyn1	NA	Direct sequencing		8.0	40	43	GH-11	Ariaeenejad et al., 2019
2.	Compost metagenome	XYL21, XYL38	CopyControl Fosmid Library Production kit	Fosmid	40	5.5–7.0	80 (XYL38)	41.9 (XYL38)	GH-10	Ellila et al., 2019
3.	Arctic mid-ocean ridge vent	AMOR_GH10A	Metagenomic data set	NA	NA	5.6	80		GH-10	Fredriksen et al., 2019
4.	Camel rumen metagenome	XylCMS	454 Pyrosequencing	NA	NA	6.0	55	46	GH-11	Ghadikolaei et al., 2019
5.	Lobios hot spring sediment	XynA3	Fosmid	Fosmid	1/150,000	6.5	80	41	GH-11	Knapik et al., 2019
6.	Saline-alkaline soil	Xyn22	pMD18T/*E. coli* DH5α	Plasmid	150/–	7.0	60	25	GH-11	Li et al., 2019
7.	Termite gut metagenome		CopyControl fosmid library production kit	Fosmid	101/50,000	ORF7: 7.0 ORF21:7.0	ORF7: 60 ORF21: 60	ORF7: 45 ORF21: 53.1	GH-10	Liu et al., 2019
8.	Soda lake Dabusu metagenome	-NA-	Illumina HiSeq2500	NA	NA	NA	NA	NA	NA	Wang et al., 2019
9.	Black goat rumen	KG42	CopyControl fosmid library production kit	Fosmid	17/115,200	5.0	50	41	GH-10	Kim et al., 2018
10.	Termite gut metagenome	Xyl1	CopyControl fosmid library production kit	Fosmid	14/40,000	6.0	50	55	GH-11	Rashamuse et al., 2016
11.	Hu sheep rumen metagenome	xyn-lxy	EpiFOSTM Fosmid Library Production Kit	Fosmid	18/12704	6.0	50	80	GH-10	Wang et al., 2015
12.	Cattle rumen metagenome	Xyln-SH1	EpiFOSTM Fosmid Library Production Kit	Fosmid	1/–	6.5	40	41.5	GH-10	Cheng et al., 2012
13.	Compost soil metagenome	Mxyl	p18GFP vector	Plasmid	1/36,000	9.0	80	43	GH-11	Verma et al., 2013b
14.	Chicken cecum metagenome	XynAMG1	CopyControl fosmid library production kit	Fosmid	NM	6.0	45	40	GH-10	Al-Darkazali et al., 2017
15.	Cow dung compost	xyn10CD18	PCR approach using degenerate primers	NA	NA	7.0	75	43.3	GH-10	Sun et al., 2015
16.	Soil	XynH	pHBM803 and pHBM625/*E. coli* BL21(DE3) and XL10-Gold	Plasmid	1/24,000	7.8	40	39	GH-10	Hu et al., 2008

(Continued)

TABLE 1 | Continued

S. No.	Source	Name	Vector/host/ approaches	Library type/approach	Positive hits	Opt_{pH}	Opt_{Temp}.(°C)	Molecular mass (kDa)	Family	References
17.	Compost metagenome	Xyn10J	CopyControl fosmid library production kit	Fosmid	5/–	7.0	40	40	GH-10	
18.	Compost metagenome	X1, X2, X3, X4, and X5	CopyControl fosmid library production kit	Fosmid	5/12,380	5.5–6.0	50–55	ND	ND	Kwon et al., 2010
19.	Soil and pulp enrichment culture	X2P1	pUC19/*E. coli* DH5α	Plasmid	1/40,000	ND	ND	ND	ND	Mori et al., 2014
20.	*Globitermesbrachycerastes* gut metagenome	Xyn7	Fosmid library	Fosmid	ND	7	55		GH-11	Qian et al., 2015
21.	Termite gut metagenome	xynGH11-7	PCR approach using degenerate primers	NA	NA	6	30	26.7	GH-11 and many more	Sheng et al., 2015
22.	Alpine Tundra soil	NM	PCR approach using degenerate primers	NA	NA	ND	ND	ND	ND	Wang et al., 2010
23.	POME (Pal oil mill effluent)	-NM-	CopyControl fosmid library production kit	Fosmid	6/–	NM	NM	NM	GH-10 and others	Benbelgacem et al., 2018
24.	Fungus-growing termite, *Pseudacanthotermes militaris*	Xyn3	CopyControl fosmid library production kit	Fosmid	42/16,000-gut 12/24,000-comb	NM	NM	NM	NM	Bastien et al., 2013
25.	Goats ruminal liquid metagenome	Multiple xylanolytic enzymes	CopyControl fosmid library production kit	Fosmid	22/15,0000 endo-xylanase	5–7	50	ND	GH-11	Duque et al., 2018
26.	Thermophilic methanogenic digester community enriched with paper	Multiple xylanolytic enzymes	CopyControl fosmid library production kit	Fosmid	37/9700	5–6	60–75	ND	GH-11	Wang et al., 2015
27.	Kinema soyabean metagenome	NM	Whole metagenome sequencing	NA	NA	ND	ND	ND	GH-11	Kumar et al., 2019

(Continued)

TABLE 1 | Continued

S. No.	Source	Name	Vector/host/ approaches	Library type/approach	Positive hits	Opt_{pH}	Opt_{Temp}.(°C)	Molecular mass (kDa)	Family	References
28.	Grass hay-fed dairy cow rumen metagenome	xyn10N18	BAC library	BAC	ND	6.5	35	54.5	GH-10	Gong et al., 2013
29.	Wood-feeding higher termite	X1098.3	pCC1FOS	Fosmid	12/88,000	8	55	ND	GH-10 and 11	Thidarat et al., 2012
30.	Sugarcane bagasse	Xyn11 One of 5	CopyControl fosmid library production kit	Fomid	05/120,000	6	80	45	GH-11	Kanokratana et al., 2014
31.	Hot springs, Tattapani, Chhattisgarh	Illumina HiSeq1000	NA	MG-RAST and SEED databases	NA	NA	NA	NA	NA	Kaushal et al., 2018
32.	Rice straw metagenome	Umxyn10A	pWEB:TNC Cosmid	Cosmid	1/12,000	6.5	75	44	GH-10	Mo et al., 2010
33.	Cow dung compost metagenome	Xyn10CD18	PCR approach using degenerate primers	NA	NA	7	75	43.3	GH-10	Sun et al., 2015
34.	Termite gut metagenome	XYL6806 XYL6807 XYL6805 XYL6419	pET101/D-TOPO vector	*E. coli* BL21	NM	5–6	50	NM	XYL6806: GH8 Rest were of GH11	Brennan et al., 2004
35.	Cow manure metagenome	manf-x10	Lambda ZAP II vector and Gigapack III Packaging Extract	Bacteriophage	$1/2 \times 10^5$ plaque units	7	40	50.3	GH-10	Li et al., 2009
36.	Pawan hot spring metagenome	*xyn* PW8	PCR approach using GH-11 primers	NA	NA	7.0	50	23.3	GH-11	Helianti, 2007
37.	Vermiform appendixes of horses metagenome	*XynVA1*	Inverse PCR (IPCR)	NA	NA	ND	ND	42–63	GH-10	Yamada et al., 2008
38.	Hot pool metagenome	NM	Genomic walking PCR (GWPCR)	NA	NA	6	100	ND	GH-10	Sunna and Bergquist, 2003
38.	Hot pool metagenome	NM	Genomic walking PCR (GWPCR)	NA	NA	6	100	ND	GH-10	Sunna and Bergquist, 2003

NM, not mentioned; ND, not determined; NA, not applicable.

TABLE 2 | An update of metagenomic xylanolytic enzymes other than GH-10 and GH-11 xylanases.

S. No.	Source	Name	Vector/host/ approaches	Library type/approach	Positive hits	Opt_{pH}	$Opt_{Temp.}$ (°C)	Molecular weight (kDa)	Family	References
1.	Yak rumen metagenome	Rubg3A and Rubg3B	pWEB Cosmid (Epicentre, United States)	Cosmid	2/4000	4.5–5.5	35–40	80	GH-3	Bao et al., 2012
2.	Yak rumen metagenome	RuBGX1	pWEB Cosmid (Epicentre, United States)	Cosmid	1/5000	6	50	80	GH-3	Zhou et al., 2011
3.	Yak rumen metagenome	RuXyn1 and RuXyn2	pWEB Cosmid (Epicentre, United States)	Cosmid	NM	5.5–7.0	40–55	42 and 50	GH-43	Zhou et al., 2012
4.	Cow rumen metagenome	RUM630-BX	Lambda ZAP II Vector and Gigapack III Packaging Ext., Agilent	Lambda phage	NM	7.0	25	ND	GH-43	Jordan et al., 2016
5.	Termite gut metagenome	Xyl1 and xyl8	CopyControl fosmid library production kit	Fosmid	12/8/12,000	5–6	40–50	53–90	GH43 and GH-3	Liu et al., 2018
6.	Wheat straw degrading microbial consortium	xylM1989	Microbial consortium	NA	NA	8	20	37.5	GH-43	Maruthamuthu et al., 2017
7.	Soil	MeXyl31	CopyControl fosmid library production kit	Fosmid	1/50,000	5.5	45	77	GH-31	Matsuzawa et al., 2016
8.	Compost	Biof1_09	CopyControl fosmid library production kit	Fosmid	1/–	4.5	50	12.8	GH-43	Sae-Lee and Boonmee, 2014
9.	Hot spring	AR19M-311-2, AR19M-311-11 and, AR19M-311-21	Direct sequencing	NA	NA	5	50–90		GH- 1, GH-3, GH-31, and GH-43	Sato et al., 2017
10.	Compost starter mixer	deAX	Lambda ZAP II vector and packaged into lambda phage	Lambda ZAP II vector	NM	5.5–7.0	55	59.1	GH-43	Wagschal et al., 2009
11.	Cow rumen fluid metagenome	GH6284	Lambda ZAP Express and Gigapack II Gold Packaging Extract	Lambda phage library	1/50,000	5.8	45	110.8	GH-5 and GH-26	Palackal et al., 2007
12.	Goats ruminal liquid metagenome	Multiple xylanolytic enzymes	CopyControl fosmid library production kit	Fosmid	22/150,000 endo-xylanase 125/150,000 xylosidase	5–7	50	ND	GH-1, GH-5, GH-8, GH-14, and GH-43	Duque et al., 2018
13.	Kinema soyabean metagenome	NM	Whole metagenome sequencing	NA	NA	ND	ND	ND	GH39 and GH 43	Kumar et al., 2019
14.	Chinese holstein dairy cow rumen	UX66	CopyControl pCCIBAC vector and *E. coli*	BAC	18/15,360	6	50	63	GH-43	Zhao et al., 2010
15.	Lagoon of dairy farm	Xyn8	Lambda ZAP II Vector/lambda phage	Fosmid	–	6–7	20	45.9	GH-8	Lee et al., 2006

NM, not mentioned; ND, not determined; NA, not applicable.

three potential candidates for xylanolytic enzymes (MgrBr18, MgrBr61, and MgrBr135) out of the 12,960 fosmid clones having an average insert size of 40 Kb. Similarly, 17 novel positive xylanolytic clones were obtained out of 1,15,200 clones from a fosmid-based library (average inert size 30.5 kb). Of the 17 clones, one clone, KG42, revealed the highest xylanase (GH-10 family) activity under acidic conditions (Kim et al., 2018). In another report, a pCT3FK fosmid vector was used to construct 1,50,000 clones, and one positive clone (XylA3) has been identified for xylanase activity (Knapik et al., 2019). Rashamuse et al. (2016) identified several glycosyl hydrolase-positive clones from a fosmid-based metagenomic library of 40,000 clones. The high positive hit rate was also achieved by Liu et al. (2019) on large-scale screening of approximately 50,000 fosmid clones, in which 464 displayed polysaccharide hydrolyzing activity. Of these, 101 clones displayed endo-xylanase activity.

Plasmid-based metagenomic libraries have also emerged successfully in retrieving xylanolytic enzyme-encoding genes (Lee et al., 2006; Wang et al., 2015). One positive xylanolytic clone was attained from the p18GFP vector by screening 36,000 clones constructed from a soil–compost metagenomic library (Verma et al., 2013b). The success rate of plasmid-based metagenomic libraries is low as compared to the fosmid- and cosmid-based libraries; this is, however, one of the cost-effective processes of routine molecular biology laboratories in digging out novel bioactive molecules.

Mining Genes Using Degenerate Primers

PCR is another popular approach for exploring community DNA cost-effectively for retrieving hidden genes. Several such reports are available in which degenerate as well as specific sets of primers have been successfully used for amplifying novel xylanase-encoding genes from the metagenomes (Liu et al., 2005; Helianti, 2007; Sheng et al., 2015). The degenerate primers often enhance the possibility of fishing out novel candidates over the specific primers (Sheng et al., 2015). In a pioneering report, Sunna and Bergquist (2003) employed GW-PCR to amplify hidden xylanase genes from the hot pool metagenome in New Zealand. A total of 60 xylanolytic genes have been amplified from the gut metagenome of *Holotrichia parallela* larvae using degenerate primers, in which 19 gene products belonged to the GH-10 family, 14 were of GH-11 type, and the remaining genes showed similarity with the sequences of the GH-8 family (Sheng et al., 2015). Similarly, the metagenome of Alpine tundra soil revealed 96 different xylanases of the GH-10 family using a primer based approach, and 31 showed similarity with the members of the GH-11 family (Wang et al., 2010).

Metagenomes pose an issue for the occurrence of a low abundance of the representative genes, especially from extreme environments due to low microbial biomass. To overcome this, rolling circle amplification has been successfully employed to achieve 11 unique putative GH-10 xylanase genes (Yamada et al., 2008). One major drawback associated with PCR-based approaches is retrieving partial genes that chiefly remain uncharacterized.

Library Sequencing/Whole Metagenome Sequencing

Direct sequencing of the randomly picked clones from a metagenomic library is a straightforward approach to decode the hidden information in the sequences. This approach has also been employed in generating significant information for hundreds of xylanolytic enzyme-encoding genes (Sato et al., 2017; Ariaeenejad et al., 2019). The direct sequencing of a porcupine metagenome uncovered four genes having β-glucosidase, α-L-arabinofuranosidase, β-xylosidase, and endo-1,4-β-xylanase activities (Thornbury et al., 2019). Whole metagenome sequencing is a relatively more powerful technique over library sequencing to construe the hidden insights of the microbial communities as well as their functional characteristics (Sharpton, 2014; Wiseschart et al., 2019). This technique has been adopted to decipher xylan-depolymerizing enzyme-encoding genes from different environmental metagenomes (Wang et al., 2015; Haro-Moreno et al., 2018; Lee et al., 2018; Victorica et al., 2020). Wang et al. (2015) reported four β-xylosidases using shotgun sequencing of compost metagenome. A unique bifunctional cellulase–xylanase was identified by shotgun sequencing of a goat rumen metagenome (Lee et al., 2018). Although a sequence-based approach often emerges with a plethora of unique information, it demands extensive bioinformatics analysis over functional screening.

PROTEIN ENGINEERING OF METAGENOMIC ENDOXYLANASE

The metagenomic approach has been used to obtain the best-fit enzymes from metagenomes. Several reports focus on genetic manipulations in order to improve the characteristics of metagenomic xylanolytic enzymes (Verma et al., 2019). Unusual metagenomic xylanase (Mxyl) is further improved by site-directed mutagenesis in which four cumulative replacements of surface serine and threonine residues with arginine enhanced the $T_{1/2}$ of the muteins by 10 and 5 min at 80 and 90°C, respectively (Verma et al., 2013b). Similar findings have been reported by Turunen et al. (2002), who replaced multiple serine/threonine surface residues of *Trichoderma reesei* xylanase with arginine. The high pKa value of arginine enhances the overall hydrophobicity of the mutated xylanases (Verma et al., 2013a). Such multiple mutations have not been reported for enhancing the thermo-tolerance of endoxylanase II; this also shifted its optimum pH from neutral to the alkaline region. Sriprang et al. (2006) also report the improvement of xylanases using site-directed mutagenesis by multiple substitutions of arginine on protein surfaces. Another metagenome-derived xylanase (Xyn7) was engineered using error-prone PCR, wherein two mutants (XYL7-TC and XYL7-TS) showed a shift in optimum temperature by 10°C with an overall enhancement of 250-fold in their half-lives at 55°C (Qian et al., 2015). The generation of multiple hydrophobic interactions in mutated xylanase, especially between Met50 and other aromatic amino acids, improved

the thermostability significantly (Qian et al., 2015). A recently developed *in silico* web tool, TAXyl, assists in predicting the thermal attributes associated with thermophilic and hyperthermophilic xylanases (Shahraki et al., 2019). There are a few reports on the engineering of metagenomic xylanolytic enzymes; a majority of metagenomic xylanases are still awaiting their characterization.

SAMPLE SOURCES AND THE CHARACTERISTICS OF METAGENOMIC XYLANOLYTIC ENZYMES

Soil (Li et al., 2019), compost (Jeong et al., 2012), and gut (Shi et al., 2011) are the three most highly explored habitats for harvesting xylan-degrading enzyme-encoding genes in metagenomics. Compost soil grabbed attention due to the high load of microbial biomass and unique biochemical properties. Jeong et al. (2012) exploited a compost metagenome to retrieve xylanolytic enzyme-encoding genes. A majority of the xylan-hydrolyzing enzymes recovered from the compost displayed optimum activity in acidic/neutral conditions (Matsuzawa et al., 2015; Sun et al., 2015). Kwon et al. (2010) recovered five acidic and thermophilic xylanase-encoding genes from pig manure and mushroom cultivation residues. A novel xylanase-encoding gene (*Mxyl*) from a compost–soil metagenome was found suitable for bleaching of paper pulp samples. This is one of the rare xylanases that exhibits optimum activity at 80°C and alkaline pH of 9.0 with exceptional thermo-stability [$T_{1/2}$ (80°C) 120 min)] and alkalistability (pH 9.0), the desired conditions for bio-bleaching of paper pulp samples (Verma et al., 2013b). A majority of the existing xylanases do not exhibit such twin stabilities from culturable microbes (Shi et al., 2014; Yan et al., 2017). A few reports discuss the stabilities of xylanases at higher temperatures and pH, obtained through traditional cultural approaches (Mamo et al., 2006; Verma and Satyanarayana, 2012; Kumar and Satyanarayana, 2014). Several xylanases have been recovered from metagenomes having an optimum temperature of 70°C or beyond (Sunna and Bergquist, 2003; Kanokratana et al., 2014; Sun et al., 2015; Amel et al., 2016; Ellila et al., 2019; Fredriksen et al., 2019; Knapik et al., 2019). However, the majority of them exhibit their optimum pH from acidic to neutral range (5.0–7.0) (Mo et al., 2010; Ellila et al., 2019).

Thermostable and acid-stable xylanases have direct applicability in food and feed industries as well as extraction/clarification of fruit juices (Ngara and Zhang, 2018; Verma et al., 2019). Similarly, several alkaliphilic xylanases have been discovered from direct cloning of environmental DNA (Hu et al., 2008; Thidarat et al., 2012; Ariaeenejad et al., 2019; Liu et al., 2019). Camel rumen metagenome-derived xylanase (PersiXyn1) also shows an optimum pH of 8.0; it, however, exhibits optimum activity at 40°C (Ariaeenejad et al., 2019). Metagenomic xylanases reported by Thidarat et al. (2012) and Liu et al. (2019) share similar properties having optimum temperature in the range of 55–60°C and a pH of 8.0. Interestingly, both of these xylanases have been recovered from termite gut metagenomes.

The gut of insects and herbivores is the well-explored reservoir for a variety of lignocellulose-degrading enzymes (Leth et al., 2018; **Table 1**). Shi et al. (2011) stress exploring Orthoptera (grasshoppers) and Coleoptera (woodborers) over the Lepidoptera (leaf-consuming insects) to harvest more potent xylanases. It indicates that these insects degrade a more complex form of lignocellulosic materials. Comparative characteristics of the majority of the insect gut xylanases uncovered are acidic and thermophilic xylanases (Brennan et al., 2004; Qian et al., 2015). GH-11 xylanase (xyl7) from termite gut shows enzymatic activity in a broad range of pH (Qian et al., 2015). Similarly, Sheng et al. (2015) recovered numerous clones having xylanase activity from the gut of *H. parallela* larvae, in which a total of 19 GH-10, 14 GH-11, and 27 GH-8 xylanases were identified.

Halophilic/halotolerant xylanases show immense applications in the processing of seafood and clarification of juices and wine (Wang et al., 2016; Alokika et al., 2018). Metagenomics has dug out numerous halotolerant xylanolytic enzymes from different environmental sources (Al-Darkazali et al., 2017; Fredriksen et al., 2019; Ghadikolaei et al., 2019; Li et al., 2019; Alves et al., 2020). Metagenomic xylanase from alkaline-saline soil has shown the ability to retain 80% of its activity at 3M NaCl. On further characterization, this thermostable xylanase reveals that the surface glutamates (E137 and E139) play a key role in its halotolerance and halostability (Li et al., 2019). Chicken cecum metagenome is also shown to be the source of highly halotolerant xylanase that retains 96% activity at 4M NaCl (Al-Darkazali et al., 2017).

Other environmental samples, such as soil, sediment, water, and effluent, from extreme as well as normal habitats have also been explored for xylan-depolymerizing enzymes (Lee et al., 2006; Hu et al., 2008; Wang et al., 2010; Li et al., 2019; Alves et al., 2020). Several cold-active GH-10 and GH-11 xylanase-encoding genes have been retrieved from tundra soil; xylanases displayed cold adaptation (Wang et al., 2010). In a recent report, three potent xylanolytic clones were recovered from mangrove soil metagenomes (Alves et al., 2020). Overall, the majority of the uncovered metagenomic xylanases to date are acidophilic, and their optimum temperature ranges from mesophilic to thermophilic.

UNUSUAL METAGENOMIC XYLANOLYTIC ENZYMES

The culturable and metagenomic approaches exploit the same environment for mining the genes; therefore, comparing their properties does not look convincing. We cannot ignore that metagenomics also uncovered several hidden genes that we usually lose by routine microbiological culture-dependent methods. Therefore, it is quite interesting to figure out the characteristics of such xylanases and also identify the new additions in terms of their industrial prospects.

A hot pool environmental metagenome uncovered thermostable xylanases having optimum activity at 100°C; the rarest property among the existing xylanases (Sunna and Bergquist, 2003). Though several features of this xylanase are

comparable with that of *Thermotoga* spp., its metagenome did not amplify any of its representative sequences (Thermotogales) on 16S rRNA gene amplification. A majority of the xylanases of *Thermotoga* spp. are remarkable in exhibiting optimum xylanase activity at or above 80°C (Simpson et al., 1991; Sunna and Antranikian, 1997; Shi et al., 2014). Compost soil usually emerges with acidophilic xylanases (Kwon et al., 2010; Sae-Lee and Boonmee, 2014; Ellila et al., 2019). Several metagenomic xylanases display very similar properties and harmonize the acidic characteristics of the compost. However, metagenomic xylanase (Mxyl) is an exception for exhibiting notable thermostability at 80°C and alkalistability (pH 9.0–10.0) (Verma et al., 2013b). This is the only report from metagenomics for exhibiting dual stabilities at higher pH and temperature in an endoxylanase. Despite its high molecular weight, it has been categorized into the GH-11 family on the basis of hydrophobic cluster analysis. The endoxylanase of *Geobacillus thermoleovorans* (Sharma et al., 2007; Verma and Satyanarayana, 2012), *Bacillus halodurans* S7 (Mamo et al., 2006), *B. halodurans* TSEV1 (Kumar and Satyanarayana, 2014), *Microcella alkaliphila* JAM-AC0309 (Kuramochi et al., 2016), and *Geobacillus thermodenitrificans* TSAA1 (Anand et al., 2013) display such twin properties; however, their stabilities compromise at higher pH and temperature in the long-run processes and are suitable for industries, especially biobleaching of pulp. Therefore, Mxyl could be a potential candidate for the paper industry. Another unusual metagenomic xylanase (xyn8) from an environmental DNA library shows a distinct optimum temperature of 20°C, which is quite rare from the culture-dependent approach (Zhao et al., 2010).

A majority of metagenomic psychrophilic xylanases share similar properties with the cold-active xylanases of bacteria/fungi (Lee et al., 2006; Guo et al., 2009; Wang et al., 2010). More efforts are, therefore, needed that suit extreme cold environments for retrieving psychrophilic/psychrotolerant xylanolytic enzymes. Furthermore, a unique bifunctional xylanase/endoglucanase from a yak rumen metagenome has been revealed to exhibit a synergistic effect with β-1,4-xylosidase and β-1,4-glucosidase; these are considered ideal candidates for the bioethanol industry (Chang et al., 2011). The bifunctional xylanases also enhance feed nutrient digestibility multifold over the mono-functional xylanases for use in enzyme cocktails (Khandeparker and Numan, 2008).

The discovery of unique halophilic/halotolerant xylanolytic enzymes further enhances the significance of metagenomics. Camel rumen metagenome revealed a rare halophilic xylanase that gets stimulated by 132% in the presence of 5M NaCl, which is the highest reported among the various other existing xylanases from culturable approaches to date (Ghadikolaei et al., 2019). In a recent investigation, novel xylanase (AMOR_GH10A) was identified from the Arctic Mid-Ocean Ridge vent system that exhibited dual binding affinity on xylan as well as glycans. On sequencing analysis, its carbohydrate-binding domain (CBD) does not share any similarity with any of the existing sequences available in the databases (Fredriksen et al., 2019). This paved the way to search for new candidates of the CBM85 family that have promising substrate affinities. A unique feature has been reported for the first time from an exceptional metagenomic xylanase (UX66) that comprises two CBDs and two catalytic domains (Zhao et al., 2010). Another unique three catalytic domain multi-enzyme (a CE1 ferulic acid esterase, a GH62 α-L-arabinofuranosidase, and a GH10 β-D-1,4-xylanase) has been recovered from the metagenome of wastewater treatment sludge (Holck et al., 2019). Such attributes enhance the applicability of xylanase manifold over the xylanases having one single catalytic domain. Undoubtedly, metagenomics has emerged as an exceptional tool for adding several novel and unique xylan degrading biocatalysts that find potential in different industries.

ADDITION OF NOVEL METAGENOMIC XYLANASE-ENCODING GENES INTO THE CAZy DATABASE

Sequencing analysis of the retrieved metagenomic xylanase proteins uncovers a plethora of genes that were inaccessible earlier. The CAZy database comprises several representatives of metagenomic xylanolytic enzymes that are categorized into GH-10 and GH-11 families. Approximately 470 endo-acting xylanases are present in the GH-10 and GH-11 families of metagenomic origin; most of them are uncharacterized partial fragments of less than 100 amino acids. Similarly, β-xylosidases also account for 28 unclassified candidates. The blastp analysis of these sequences reveals that metagenome-originated xylanases share low/high percentage identity from their homologue sequences available in the databases (Mo et al., 2010; Bastien et al., 2013; Rashamuse et al., 2016; Ellila et al., 2019). The phylogenetic tree further depicts clustering of the sequences (full length/partial) into respective GH families (**Figure 2**); nine partially redundant endo-active xylanases and five arabinosidase/xylosidase metagenomic fragments with very high identity, in which most of the sequences display similarity with the homologue sequences of Bacteroidetes; an overrepresented group of termite gut (Bastien et al., 2013).

A gene derived from a Hu sheep rumen metagenome shares 97% identity with a contig1552 of an uncultured bacterium; its xylanase, however, shows very little similarity with the homologous proteins (Wang et al., 2019). Metagenomic xylanase Xyn10CD18 shares a maximum identity of 83% with thermostable xylanase of *Bacillus* sp. N16-5 (Sun et al., 2015). Similarly, the xylanase gene retrieved from a Pawan hot spring metagenome also shares an identity of 95% with its protein homologue (accession number: CAA84276) (Helianti, 2007). It may be due to the PCR approach to fetch the desired genes, when the specific and degenerate sets of primers are designed from the available gene sequences (Verma et al., 2019). In addition, numerous metagenomic xylanolytic enzymes have also been identified for showing low to very low percentage identity with their homologues (Ellila et al., 2019). Metagenomic xylanase (Xyn38) shares only 50.3% identity with the GH-10 xylanase of *Acidothermus cellulolyticus* 11B (Ellila et al., 2019). AMOR_GH10A shares a maximum of 42% identity with a hypothetical protein from a *Verrucomicrobia* (Fredriksen et al., 2019). However, its GH-10 domain shows an even lesser identity of 24–28% with the homologous xylanases. Similarly, KG42 (Kim

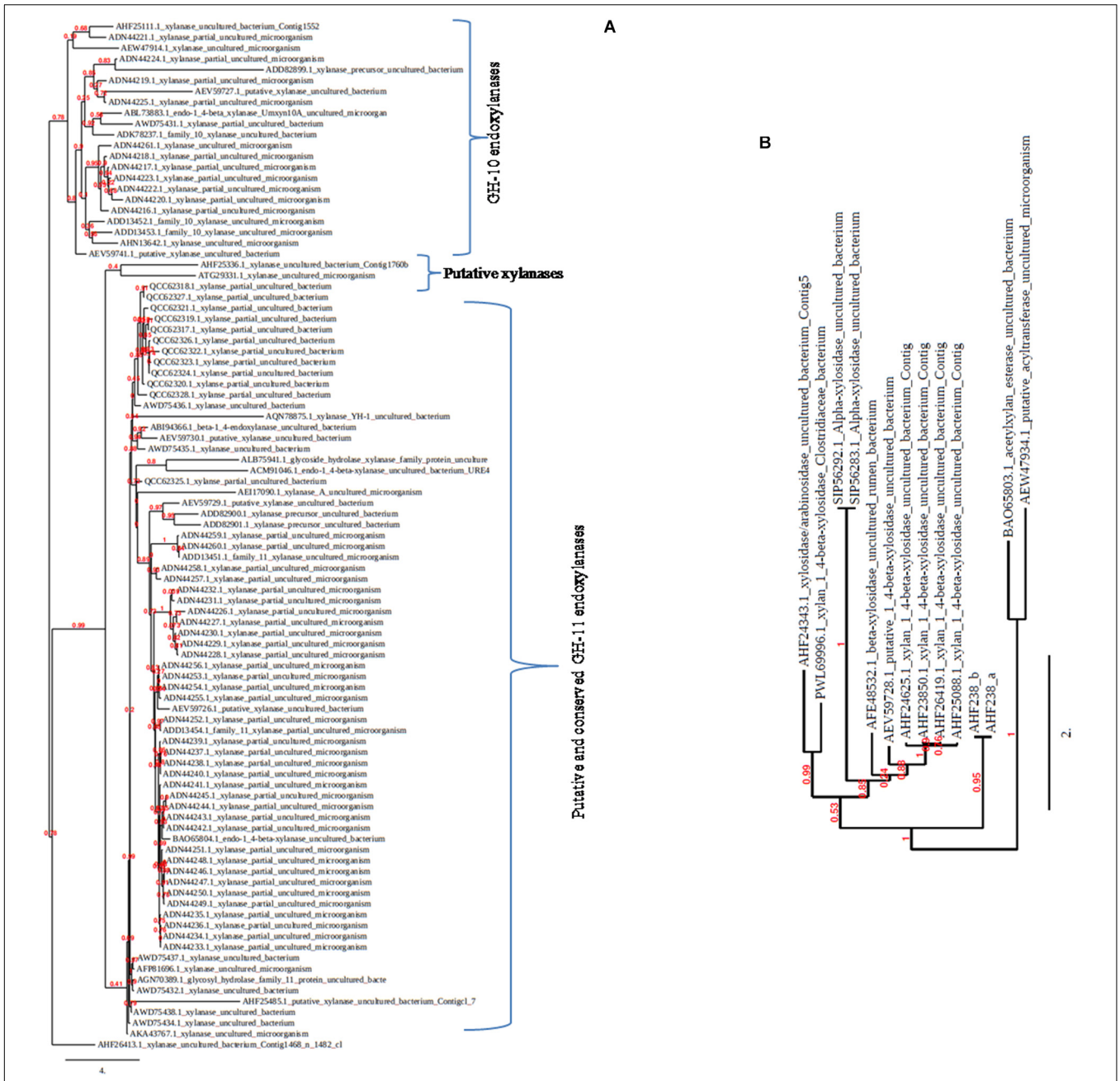


FIGURE 2 | The phylogenetic trees showing the diversity among metagenomic xylanolytic enzymes. The trees (**A:** GH-10 and GH-11 xylanases; **B:** xylanolytic enzymes other than GH-10 and GH-11 xylanases) were constructed by including the protein sequences (complete/partial) available in the NCBI database using the keyword xylanase/xylosidase/arabinofuranosidase/acetyltransferase from uncultured bacteria/microorganisms.

et al., 2018), XYL6806 (Brennan et al., 2004), and xyn10N18 (Gong et al., 2013) xylanases also share a very low identity of merely 40% with the homologue sequences. The discovery of bifunctional cellulase-xylanase also shows a low identity of 41% with Cel5 homologues (Lee et al., 2018).

On the contrary, the metagenomic xylanase (XynH) shares a higher identity of 56% and a similarity of 71% with the GH-10 family xylanase of *Cellvibrio mixtus* (accession number AF049493; Hu et al., 2008). Similarly, umxyn10A shares 58% identity and 73% similarity with xylanase from a bacterium *Thermobifida fusca* (Mo et al., 2010). Five novel xylanase genes having a 35–40% sequence identity have been retrieved from the PCR-based approach using a horse vermiform appendix metagenome, which is quite surprising (Yamada et al., 2008). Multiple sequence alignments, however, have detected the crucial catalytic residues (glutamate and aspartate) in the

signature sequences of xylanolytic enzymes of metagenomic origin (**Figure 3**).

METAGENOMIC β-XYLOSIDASES AND THEIR CHARACTERISTICS

Among xylanolytic enzymes, β-xylosidases (Exo-1, 4-β-D xylosidase E.C. 3.2.1.37) are the second most highly important xylan-depolymerizing enzymes that assist in complete degradation of xylans. These exotype glycosidases cleave the glycosidic linkages of short xylooligosaccharides (XOs) and liberate the pentaose sugar (xylose) as an end product (Shallom and Shoham, 2003; Jain et al., 2015). Pharmaceutical and food industries demand thermostable and halophilic β-xylosidases (Jain et al., 2014). Existing β-xylosidases are associated with several limitations, such as poor efficiency, low thermostability, salt sensitivity, and by-product inhibition (Bao et al., 2012; Anand et al., 2013). Therefore, efforts have been made to discover novel candidates of β-xylosidases using metagenomic approaches, with which little success has been achieved (Wagschal et al., 2009; Jordan et al., 2016; Cheng et al., 2017; Sato et al., 2017; Liu et al., 2018; Rohman et al., 2019; **Table 2**). Of the available metagenomic β-xylosidases, fewer than 20 have been extensively characterized to date. The majority of such β-xylosidases have been reported from the ruminal fluid of yak (Zhou et al., 2011, 2012; Bao et al., 2012), cow (Jordan et al., 2016), and buffalo (Singh et al., 2014) due to the abundance of hemicellulose-degrading microbes in their guts. Other promising environmental sources have also been explored, such as soil (Matsuzawa et al., 2016) and compost (Sae-Lee and Boonmee, 2014; Matsuzawa et al., 2015; Wang et al., 2015), for harnessing hidden β-xylosidase genes. On characterization, most of the β-xylosidases show

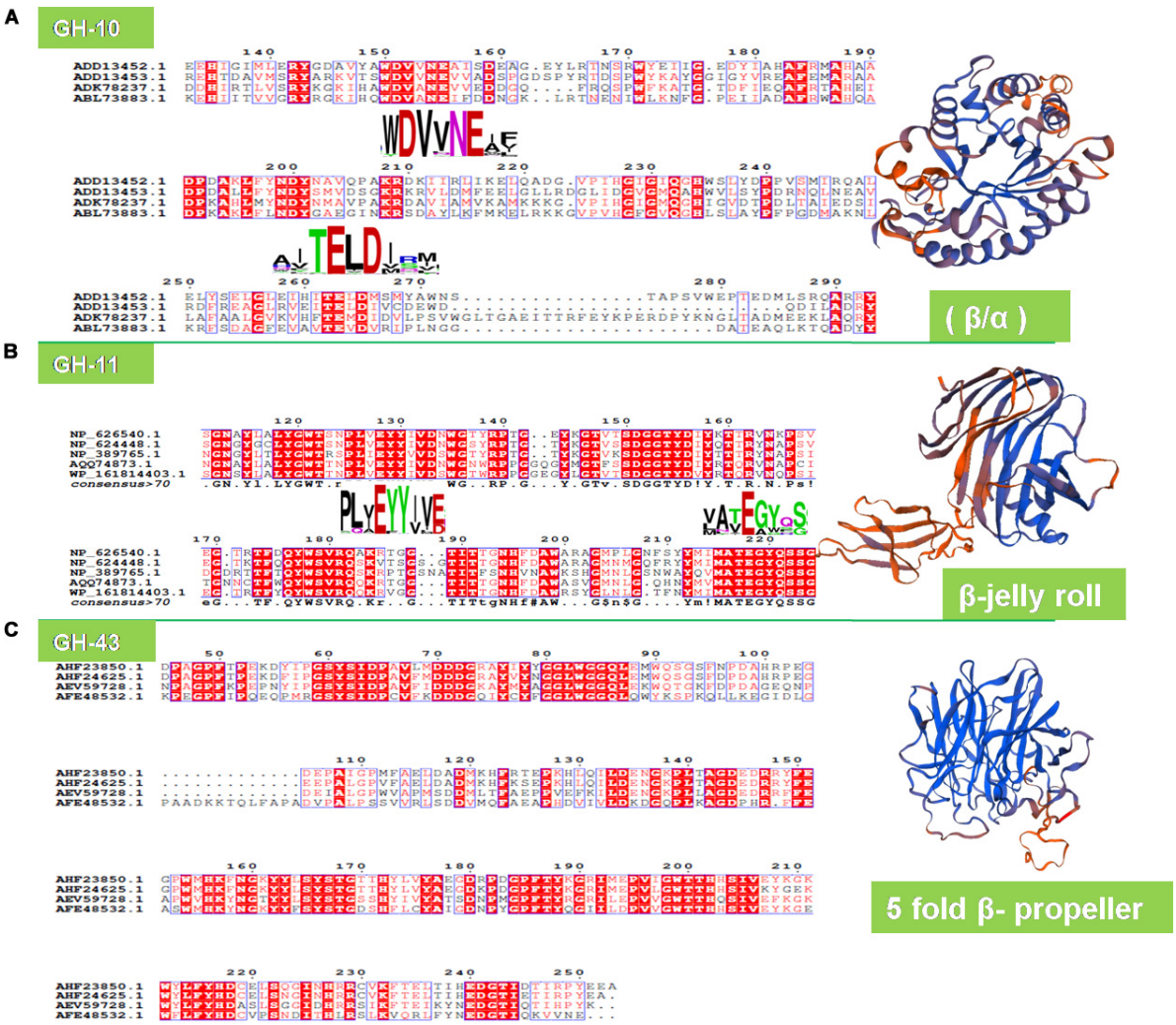


FIGURE 3 | Structure analysis of metagenomic xylanolytic enzymes. The ESPRipt analysis showing the conserved motifs for **(A)** GH-10 (WDVVNEI/A and TEL/MDM/I), **(B)** GH-11 (PLY/AEYYI/VE/D and ATEGYQ) endoxylanases, and **(C)** GH-43 (DDDG and LFYI). The GH-10 and GH-11 xylanases exhibited conserved motifs having highly conserved catalytic glutamate (E) and aspartate (D) residues. The signature structure of metagenomic GH-10 (Accession number: ADK78237.1), GH-11 (accession number: AQQ74873.1), and GH-43 (accession number: AFE48532.1) enzymes.

optimum pH from an acidic to neutral range (Wagschal et al., 2009; Jordan et al., 2016; Liu et al., 2018). Three highly acidic β-xylosidases (AR19M-311-2, AR19M-311-11, and AR19M311-21) derived from a hot spring metagenome show optimum pH of 5.0 (Sato et al., 2017). The yak rumen metagenome revealed many acidic β-xylosidases, and one of the β-xylosidases (RuBG3A) exhibited activity at pH 4.5 on pNPG. There is only one report of metagenomic β-xylosidase (xylM1989) that shows its optimum activity at an alkaline pH of 8.0 (Maruthamuthu et al., 2017).

Thermostable β-xylosidases are in high demand in industries to overcome microbial contamination and reduce the viscosity of the reaction mixture, which improves the overall reaction efficiency manifold (Haki and Rakshit, 2003). The β-xylosidases (AR19M-311-2, AR19M-311-11, and AR19M311-21) reported by Sato et al. (2017) are highly thermostable having optimum activity at 90°C with fair stability at 70°C for 1 h. It corroborates the characteristics of thermostable β-xylosidases of *Thermoanaerobacter ethanolicus, Thermotoga maritima*, and *Thermotoga thermarum* having the temperature optima at and above 90°C (Shao and Wiegel, 1992; Shi W. et al., 2013). In another report, four thermostable β-xylosidases have been reported from a compost–soil metagenome that show optimum activity at 60–75°C (Wang et al., 2015). In addition, these β-xylosidases are found to retain 80% of residual activity at 50°C after 2 h of incubation. A novel and first metagenomic α-xylosidase (MeXyl31) of the GH-31 family has been reported from a soil metagenome that shows an optimum temperature of 45°C (Matsuzawa et al., 2016).

Several β-xylosidases and metagenomic β-xylosidases have also been identified for their bifunctional enzymatic activities (DeCastro et al., 2016; Rohman et al., 2019). Such β-xylosidases find applicability in bioethanol production along with the endo-xylanases for efficient release of sugars from the hemicellulose component of agro-residues (Sae-Lee and Boonmee, 2014; Wang et al., 2015). The bifunctional β-glucosidase/xylosidase activity has been observed in Rubg3A, Rubg3B, and RubGX1 (Zhou et al., 2011; Bao et al., 2012). Many β-xylosidases derived from the functional screening of the metagenomic library also detected arabinofuranosidase activity (Wagschal et al., 2009; Wang et al., 2015; Jordan et al., 2016). On characterization, β-xylosidases/arabinofuranosidase (RUM630-BX) enhanced activity 84-fold due to the stimulation by divalent metal ions (Ca^{2+}, Co^{2+}, Fe^{2+}, Mg^{2+}, Mn^{2+}, and Ni^{2+}) (Jordan et al., 2016). Jordan et al. (2018) revealed the role of aspartate and histidine residues at the active site that chelates Ca^{2+}. Therefore, on supplementation of the divalent ions, the overall β-xylosidase activity was restored. The termite gut metagenome revealed a multimeric β-xylosidase having a catalytic activity of β-glucosidase or β-arabinosidases among four of the positive clones (Liu et al., 2018). A majority of the reported metagenomic β-xylosidases can be considered as novel due to their unique features along with very low sequence identity/similarity with their homologues.

METAGENOMIC ARABINOFURANOSIDASE AND ACETYL XYLAN ESTERASE

α-L-arabinofuranosidase (EC 3.2.1.55) represents another important class of hemicellulose-degrading enzymes that release arabinofuranosyl side moieties from a heteroxylan by cleaving α-1,2 and α-1,3 glycosidic bonds (Bouraoui et al., 2013; Kumar et al., 2013b). In addition, non-acetylated heteroxylans have direct access for their backbone-hydrolyzing enzymes, such as endoxylanases and β-xylosidases (Rennie and Scheller, 2014). More often, α-L-arabinofuranosidase acts synergistically along with the cocktail of different xylanolytic enzymes. It finds applications in biomass conversion, food and feed industries, and environmental waste management (Wilkens et al., 2017; Verma et al., 2019). Metagenomic approaches have brought out very few α-L-arabinofuranosidases from different environmental sources (Fortune et al., 2019). Recently, three α-L-arabinofuranosidases (AFaseH4, AFaseE3, and AFaseD3) have been reported from the high-temperature compost metagenome (Matsuzawa et al., 2015; Fortune et al., 2019). On characterization, all three display optimum activity at pH in the acidic range of pH 4.0–5.0 and at 40°C. However, sequence analysis reveals that AFaseH4- and AFaseE3-encoding genes share 100 and 99% identity with the existing sequences, respectively, while the ORF-encoding AFaseD3 shares 77% identity with the AFAse of a *Paenibacillus taihuensis* ORF (Kolinko et al., 2018). Many other metagenomic arabinofuranosidases have shown activity with β-xylosidases as bifunctional enzymes (Wagschal et al., 2009; Wang et al., 2015; Jordan et al., 2016).

The acetyl xylan esterases (AXEs) (EC 3.1.1.6) cleave the acetyl group from acetyl-xylose moieties and contribute to the complete degradation of the complex xylans. Recently, a few studies have uncovered a handful of AXEs using metagenomic approaches (Adesioye et al., 2018; Brito et al., 2018; **Table 2**). A metagenomic AXE (Axe1NaM1) was identified by using a hot desert hypolith metagenomic DNA sequence data set (Adesioye et al., 2018). On characterization, it was detected as a mesophilic (Topt. 30°C) and alkaliphilic AXE (pHopt. 8.5). Interestingly, a point mutation (N65S) improved its thermostability as well as catalytic efficiency; the crystal structure of this enzyme has also been solved. In another report, the gut microbiome of a shipworm also identified several AXEs that shared 50–75% similarity with the available carbohydrate esterases (Brito et al., 2018). Zhu et al. (2016) studied the association and relative abundance of *Firmicutes* along with lignocellulose-degrading enzymes.

UTILITY OF METAGENOMIC XYLANOLYTIC ENZYMES

Xylanolytic extremozymes find extensive applications in various industries due to their high performance under extreme conditions (Kumar et al., 2016; **Figure 1**). Unfortunately, a majority of the metagenomic xylanolytic enzymes are limited to

biophysical characteristics only (Gong et al., 2013; Kanokratana et al., 2014; Matsuzawa et al., 2015; Ghadikolaei et al., 2019). A handful of metagenome-derived xylan-degrading enzymes have been explored for their applicability in various industrial processes (Jeong et al., 2012; Wang et al., 2015; Berini et al., 2017). The metagenomic xylanase (Mxyl) was successfully employed for bleaching paper pulp samples in the paper industry, and an approximately 24% reduction in chlorine consumption was achieved (Verma et al., 2013a). Similarly, Ariaeenejad et al. (2019) also report the applicability of camel rumen metagenome–derived xylanase (PersiXyn1) in biobleaching of carton paper pulp. The recombinant enzyme (Mxyl) also finds application in saccharification of agro-residues (wheat bran, corn cob, sugarcane bagasse) for generating XOs of low DP of 2–4 xylose units (Verma et al., 2013b). Such XOs exhibit a prebiotic effect by promoting the growth of probiotic gut microbiota. The xylanase CoXyl43 shows synergistic action with cellulase of *T. reesei* in enhancing its saccharification efficiency on rice straw (Matsuzawa et al., 2015). Sun et al. (2015) achieved an overall 80% XOs (DP 2–4) from corn cob hemicellulose bioconversion by using metagenome-derived Xyn10CD.

The cocktail of β-xylosidases (RuBG3A/RuBG3B) along with endoxylanases facilitate hemicellulose saccharification efficiently (Bao et al., 2012). Another metagenome-derived cocktail of four thermostable xylanolytic enzymes (β-xylosidase and β-xylosidases/α-arabinofuranosidase) was able to hydrolyze approximately 55% of the steam-exploded corn cobs at 50°C in 48 h (Wang et al., 2015). Duque et al. (2018) claim that the metagenomic β-xylosidases are tenfold more efficient than the commercial fungal cocktails of the xylanolytic enzymes. However, Jeong et al. (2012) suggest supplementation of metagenomic xylanases (Xyn10J) with the commercial enzymes to improve their saccharification efficiency. In another report, a compost-derived xylanase (XYL40) exhibited higher hydrolysis than the xylanase of TrXyn11A (*T. reesei* xylanase). Sae-Lee and Boonmee (2014) propose that thermophilic Biof1_09 protein (with dual cellulase and xylanase) can be employed in stone washing and biopolishing industries that require thermostable and acidic extremozymes. Thermophilic xylanase and endoglucanase derived from sugarcane bagasse metagenome enhances the efficiency of a commercial cellulase Celluclast® (Novozymes, Bagsvaerd, Denmark) (Kanokratana et al., 2014). Kim et al. (2018) report the prebiotic effect of xylan hydrolysates on the growth of gut bacterial strains (*Bifidobacterium longum* and *Bifidobacterium lactis*). Interestingly, this metagenome-derived thermostable xylanase (KG42) significantly enhances the overall growth of these strains by 95–97%.

A few patents have also described various attributes of metagenome-derived xylan-degrading enzymes. Thermophilic xylanase from a hot spring metagenome has been patented (Patent no: EP2990482 A1) for its application in biofuel production from lignocellulosic biomass (Berini et al., 2017). Two Indian patents (Patent Appl. Nos. 201811041913 and 201711020622) were also filed in 2018 on process devolvement for XO production using metagenome-derived novel xylanases. Radomski et al. (1998) patented a method for retrieving the xylanase gene from soil metagenome (United States Patent US 5849491).

CONCLUSION AND FUTURE PERSPECTIVES

It has taken more than 15 years to report the first xylanase by Sunna and Bergquist (2003) using a culture-independent approach. Since then, the CAZy database is being enriched by several of the homologue sequences, where the count of GH-10 and GH-11 xylanases dominates the other xylan-degrading proteins. The supremacy of GH-10 xylanases clearly indicates their broad substrate specificity as compared to the GH-11 (true xylanases) members that are quite selective in their substrate range. A majority of the uncovered xylanolytic enzymes are of thermo-acidophilic type; this might be due to their selection from such environments. Better candidates can be recovered by employing a rational approach that includes high-throughput screening of fosmid-derived metagenomic libraries generated from extreme environmental metagenomes on suitable dye-conjugated xylan substrates. In addition, enrichment of cultures under desired physical conditions along with suitable substrates also enhances the success rate multifold to achieve xylanases with the requisite properties. Complementing shotgun sequencing and a PCR-based approach enhances the possibility of retrieving novel xylanolytic extremozymes from the environmental metagenomes.

Undoubtedly, metagenomics has uncovered several novel xylanolytic proteins having unusual properties that were earlier inaccessible. However, it cannot be denied that the recovery of metagenomic xylanolytic enzymes is comparatively higher, and their characterization and industrial applications must be stressed. A handful of reports discuss the feasibility of the retrieved enzymes at the industrial level. We have achieved very little success in utilizing these enzymes. Megazyme[2] commercialized an endo-1,4-β-xylanase (from a rumen microorganism) that finds applications in food/feed as well as in paper pulp industries. Similarly, Luminase, a commercial product of BASF[3] is also used for biobleaching in paper industries (Berini et al., 2017). We, therefore, emphasize more on characterizing the various available metagenomic xylanases for their applicability and screening for the new candidates.

AUTHOR CONTRIBUTIONS

Both authors wrote and finalized the manuscript.

ACKNOWLEDGMENTS

TS is grateful to the University Grant Commission, Government of India, New Delhi and DBT-IUSSTF, New Delhi for providing research grants while writing this review.

[2]www.megazyme.com

[3]www.basf.com

REFERENCES

Adesioye, F. A., Makhalanyane, T. P., Vikram, S., Sewell, B. T., Schubert, W. D., and Cowan, D. A. (2018). Structural characterization and directed evolution of a novel acetyl xylan esterase reveals thermostability determinants of the carbohydrate esterase 7 family. *Appl. Environ. Microbiol.* 84:e02695-17.

Al-Darkazali, H., Meevootisom, V., Isarangkul, D., and Wiyakrutta, S. (2017). Gene expression and molecular characterization of a xylanase from chicken cecum metagenome. *Int. J. Microbiol.* 2017:4018398.

Alokika, Singh, D., and Singh, B. (2018). Utility of acidic xylanase of *Bacillus subtilis* subsp. *subtilis* JJBS250 in improving the nutritional value of poultry feed. *3 Biotech* 8:503.

Alves, K. J., da Silva, M. C. P., and Cotta, S. R. (2020). Mangrove soil as a source for novel xylanase and amylase as determined by cultivation-dependent and cultivation-independent methods. *Braz. J. Microbiol.* 51, 217–228. doi: 10.1007/s42770-019-00162-7

Amel, B. D., Nawel, B., Khelifa, B., Mohammed, G., Manon, J., Salima, K. G., et al. (2016). Characterization of a purified thermostable xylanase from *Caldicoprobacter algeriensis* sp. nov. strain TH7C1T. *Carbohydr. Res.* 419, 60–68. doi: 10.1016/j.carres.2015.10.013

Anand, A., Kumar, V., and Satyanarayana, T. (2013). Characteristics of thermostable endoxylanase and β- xylosidase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1 and its applicability in generating xylooligosaccharides and xylose from agro-residues. *Extremophiles* 17, 357–366. doi: 10.1007/s00792-013-0524-x

Ariaeenejad, S., Hosseini, E., Maleki, M., Kavousi, K., Moosavi-Movahedi, A. A., and Salekdeh, G. H. (2019). Identification an characterization of a novel thermostable xylanase from camel rumen metagenome. *Int. J. Biol. Macromol.* 126, 1295–1302. doi: 10.1016/j.ijbiomac.2018.12.041

Badhai, J., Ghosh, T. S., and Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Front. Microbiol.* 6:1166. doi: 10.3389/fmicb.2015.01166

Bailey, M. J., Biely, P., and Poutanen, K. (1992). Inter-laboratory testing of methods for assay of xylanase activity. *J. Biotechnol.* 23, 257–270. doi: 10.1016/0168-1656(92)90074-j

Bao, L., Huang, Q., Chang, L., Sun, Q., Zhou, J., and Lu, H. (2012). Cloning and characterization of two β-glucosidase/xylosidase enzymes from yak rumen metagenome. *Appl. Biochem. Biotechnol.* 166, 72–86. doi: 10.1007/s12010-011-9405-x

Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Faure, R., et al. (2013). Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol. Biofuels* 6:78. doi: 10.1186/1754-6834-6-78

Benbelgacem, F. F., Isa, M. N. M., Abdelrahim, M. A. M., Tumian, A., Bellag, O. A., and Parman, A. (2018). Next generation sequencing-data analysis for cellulose- and xylan-degrading enzymes from POME metagenome. *Sains Malays.* 47, 2951–2960. doi: 10.17576/jsm-2018-4712-03

Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017). Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* 364:fnx211.

Bouraoui, H., Rebib, H., Ben Aissa, M., Touzel, J. P., O'donohue, M., and Manai, M. (2013). *Paenibacillus marinum* sp. nov., a thermophilic xylanolytic bacterium isolated from a marine hot spring in Tunisia. *J. Basic Microbiol.* 53, 877–883. doi: 10.1002/jobm.201200275

Brennan, Y. L., Callen, W. N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., et al. (2004). Unusual microbial xylanases from insect guts. *Appl. Environ. Microbiol.* 70, 3609–3617. doi: 10.1128/aem.70.6.3609-3617.2004

Brito, T. L., Campos, A. B., Bastiaan von Meijenfeldt, F. A., Daniel, J. P., Ribeiro, G. B., and Silva, G. G. Z. (2018). The gill-associated microbiome is the main source of wood plant polysaccharide hydrolases and secondary metabolite gene clusters in the mangrove shipworm *Neoteredo reynei*. *PLoS One* 13:e0200437. doi: 10.1371/journal.pone.0200437

Chang, L., Ding, M., Bao, L., Chen, Y., Zhou, J., and Lu, H. (2011). Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl. Microbiol. Biotechnol.* 90, 1933–1942. doi: 10.1007/s00253-011-3182-x

Cheng, F., Sheng, J., Dong, R., Men, Y., Gan, L., and Shen, L. (2012). Novel xylanase from a Holstein cattle rumen metagenomic library and its application in xylooligosaccharide and ferulic acid production from wheat straw. *J. Agric. Food Chem.* 60, 12516–12524. doi: 10.1021/jf302337w

Cheng, J., Romantsov, T., Engel, K., Doxey, A. C., Rose, D. R., and Neufeld, J. D. (2017). Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. *PLoS One* 12:e0172545. doi: 10.1371/journal.pone.0172545

DeCastro, M. E., Rodriguez-Belmonte, E., and Gonzalez-Siso, M. I. (2016). Metagenomics of thermophiles with a focus on discovery of novel thermozymes. *Front. Microbiol.* 7:1521. doi: 10.3389/fmicb.2016.01521

Duque, E., Daddaoua, A., Cordero, B. F., Udaondo, Z., Molina-Santiago, C., Roca, A., et al. (2018). Ruminal metagenomic libraries as a source of relevant hemicellulolytic enzymes for biofuel production. *Microb. Biotechnol.* 11, 781–787. doi: 10.1111/1751-7915.13269

Ellila, S., Bromann, P., and Nyyssonen, M. (2019). Cloning of novel bacterial xylanases from lignocellulose-enriched compost metagenomic libraries. *AMB Express* 9:124.

Fortune, B., Mhlongo, S., Van Zyl, L. J., Huddy, R., Smart, M., and Trindade, M. (2019). Characterisation of three novel α-L-arabinofuranosidases from a compost metagenome. *BMC Biotechnol.* 19:22. doi: 10.1186/s12896-019-0510-1

Fredriksen, L., Stokke, R., Jensen, M. S., Westereng, B., Jameson, J. K., Steen, I., et al. (2019). Discovery of a thermostable GH10 xylanase with broad substrate specificity from the Arctic Mid-Ocean Ridge vent system. *Appl. Environ. Microbiol.* 85:e02970-18.

Ghadikolaei, K. K., Sangachini, E. D., Vahdatirad, V., Noghabi, K. A., and Zahiri, H. S. (2019). An extreme halophilic xylanase from camel rumen metagenome with elevated catalytic activity in high salt concentrations. *AMB Express* 9:86.

Gong, X., Gruniniger, R. J., Forster, R. J., Teather, R. M., and McAllister, T. A. (2013). Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl. Microbiol. Biotechnol.* 97, 2423–2431. doi: 10.1007/s00253-012-4088-y

Gruppen, H., Hamer, R. J., and Voragen, A. G. J. (1992). Water unextractable cell wall material from wheat flour. 2. Fractionation of alkali extracted polymers and comparison with water extractable arabinoxylans. *J. Cereal Sci.* 16, 53–67. doi: 10.1016/s0733-5210(09)80079-9

Guo, B., Chen, X., Sun, C., Zhou, B., and Zhang, Y. (2009). Gene cloning, expression and characterization of a new cold-active and salt tolerant endo-β-1, 4-xylanase from marine *Glaciecola mesophila* KMM 241. *Appl. Microbiol. Biotechnol.* 84, 1107–1115. doi: 10.1007/s00253-009-2056-y

Haki, G. D., and Rakshit, S. K. (2003). Developments in industrially important thermostable enzymes: a review. *Biores. Technol.* 89, 17–34. doi: 10.1016/s0960-8524(03)00033-6

Haro-Moreno, J. M., Lopez-Perez, L., de la Torre, J. R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6:128.

Harris, A. D., and Ramalingam, C. (2010). Xylanases and its application in food industry: a review. *J. Exp. Sci.* 1, 1–11.

Helianti, I. (2007). Direct cloning of a xylanase gene from Pawan-riau hot spring. *HAYATI J. Biosci.* 14, 54–58. doi: 10.4308/hjb.14.2.54

Holck, J., Djajadi, D. T., Brask, J., Pilgaard, B., Krogh, K. B. R. M., Meyer, A. S., et al. (2019). Novel xylanolytic triple domain enzyme targeted at feruloylated arabinoxylan degradation. *Enzyme Microb. Technol.* 129:109353. doi: 10.1016/j.enzmictec.2019.05.010

Hu, Y., Zhang, G., Li, A., Chen, J., and Ma, L. (2008). Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Appl. Microbiol. Biotechnol.* 80, 823–830.

Jain, I., Kumar, V., and Satyanarayana, T. (2014). Applicability of recombinant β-xylosidase from the extremely thermophilic bacterium *Geobacillus thermodenitrificans* in synthesizing alkylxylosides. *Biores. Technol.* 170, 462–469. doi: 10.1016/j.biortech.2014.07.113

Jain, I., Kumar, V., and Satyanarayana, T. (2015). Xylooligosaccharides: an economical prebiotic from agroresidues and their health benefits. *Indian J. Exp. Biol.* 53, 131–142.

Jeong, Y. S., Na, H. B., Kim, S. K., Kim, Y. H., Kwon, E. J., Kim, J., et al. (2012). Characterization of xyn10J, a novel family 10 xylanase from a compost

metagenomic library. *Appl. Biochem. Biotechnol.* 166, 1328–1339. doi: 10.1007/s12010-011-9520-8

Jordan, D. B., Braker, J. D., and Wagschal, K. (2016). Isolation and divalent metal activation of a β-xylosidase, RUM630-BX. *Enzyme Microb. Technol.* 82, 158–163. doi: 10.1016/j.enzmictec.2015.10.001

Jordan, D. B., Stoller, R., Kibblewhite, R. E., Chan, V. J., Lee, C. C., and Wagschal, K. (2018). Absence or presence of metal ion activation in two structurally similar GH43 β-xylosidases. *Enzy. Microb. Technol.* 114, 29–32. doi: 10.1016/j.enzmictec.2018.03.007

Kanokratana, P., Eurwilaichitr, L., Pootanakit, K., and Champreda, V. (2014). Identification of glycosyl hydrolases from a metagenomic library of microflora in sugarcane bagasse collection site and their cooperative action on cellulose degradation. *J. Biosci. Bioeng.* 119, 384–391. doi: 10.1016/j.jbiosc.2014.09.010

Kaushal, G., Kumar, J., Sangwan, R. S., and Singh, S. P. (2018). Metagenomic analysis of geothermal water reservoir sites exploring carbohydrate related thermozymes. *Int. J. Biol. Macromol.* 119, 882–895. doi: 10.1016/j.ijbiomac.2018.07.196

Khandeparker, R., and Numan, M. H. (2008). Bi-functional xylanases and their potential use in biotechnology. *J. Ind. Microbiol. Biotechnol.* 35, 635–644. doi: 10.1007/s10295-008-0342-9

Kim, H. B., Leeb, K. T., Kim, M. J., Lee, J. S., and Kim, K. S. (2018). Identification and characterization of a novel KG42 xylanase (GH10 family) isolated from the black goat rumen-derived metagenomic library. *Carbohydr. Res.* 469, 1–9. doi: 10.1016/j.carres.2018.08.010

Knapik, K., Becerra, M., and Gonzqlez-Siso, M. (2019). Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios Hot Spring in Spain. *Sci. Rep.* 9:11195.

Kolinko, S., Wu, Y. W., Tachea, F., Denzel, E., Hiras, J., Gabriel, R., et al. (2018). A bacterial pioneer produces cellulase complexes that persist through community succession. *Nat. Microbiol.* 3, 99–107. doi: 10.1038/s41564-017-0052-z

Kormelink, F. J. M., and Voragen, A. G. J. (1993). Degradation of different [(glucurono) arabino] xylans by a combination of purified xylan-degrading enzymes. *Appl. Microbiol. Biotechnol.* 38, 688–695.

Kumar, J., Sharma, N., Kaushal, G., Samurailatpam, S., Sahoo, D., Rai, A. K., et al. (2019). Metagenomic insights into the taxonomic and functional features of *Kinema*, a traditional fermented soybean product of Sikkim Himalaya. *Front. Microbiol.* 10:1744. doi: 10.3389/fmicb.2019.01744

Kumar, V., Marin-Navarro, J., and Shukla, P. (2016). Thermostable microbial xylanases for pulp and paper industries: trends, applications and further perspectives. *World J. Microbiol. Biotechnol.* 32, 1–10.

Kumar, V., and Satyanarayana, T. (2014). Production of endoxylanase with enhanced thermostability by a novel polyextremophilic *Bacillus halodurans* TSEV1 and its applicability in waste paper deinking. *Process Biochem.* 49, 386–394. doi: 10.1016/j.procbio.2013.12.005

Kumar, V., Syal, P., and Satyanarayana, T. (2013a). Highly thermo-halo-alkali-stable β-1,4-endoxylanase from a novel polyextremophilic strain of *Bacillus halodurans*. *Bioprocess Biosyst. Eng.* 36, 555–565. doi: 10.1007/s00449-012-0811-4

Kumar, V., Verma, D., and Satyanarayana, T. (2013b). Extremophilic bacterial xylanases: production, characteristics and applications. *Curr. Biotechnol.* 2, 380–399. doi: 10.2174/18722083113076660027

Kuramochi, K., Uchimura, K., and Kurata, A. (2016). A high-molecular-weight, alkaline, and thermostable β-1,4-xylanase of a subseafloor *Microcella alkaliphila*. *Extremophiles* 20, 471–478. doi: 10.1007/s00792-016-0837-7

Kwon, E. J., Jeong, Y. S., Kim, Y. H., Kim, S. K., Na, H. B., Kim, J., et al. (2010). Construction of a metagenomic library from compost and screening of cellulase- and xylanase-positive clones. *J. Appl. Biol. Chem.* 53, 702–708. doi: 10.3839/jksabc.2010.106

Lee, C. C., Kibblewhite-Accinelli, R. E., Wagschal, K., Robertson, G. H., and Wong, D. W. S. (2006). Cloning and characterization of a cold-active xylanase enzyme from an environmental DNA library. *Extremophiles* 10, 295–300. doi: 10.1007/s00792-005-0499-3

Lee, K. T., Toushik, S. H., Baek, J. Y., Kim, J. E., Lee, J. S., and Kim, K. S. (2018). Metagenomic mining and functional characterization of a novel KG51 bifunctional cellulase/hemicellulase from black goat rumen. *J. Agric. Food Chem.* 66, 9034–9041. doi: 10.1021/acs.jafc.8b01449

Leth, M. L., Ejby, M., Workman, C., Ewald, D. A., Pedersen, S. S., Sternberg, C., et al. (2018). Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* 3, 570–580. doi: 10.1038/s41564-018-0132-8

Li, R., Kibblewhite, R., Orts, W. J., and Lee, C. C. (2009). Molecular cloning and characterization of multidomain xylanase from manure library. *World J. Microbiol. Biotechnol.* 25, 2071–2078. doi: 10.1007/s11274-009-0111-6

Li, Z., Li, X., Liu, T., Chen, S., Liu, H., Wang, H., et al. (2019). The critical roles of exposed surface residues for the thermostability and halotolerance of a novel GH11 xylanase from the metagenomic library of a saline-alkaline soil. *Int. J. Biol. Macromol.* 133, 316–323. doi: 10.1016/j.ijbiomac.2019.04.090

Liu, C., Zou, G., Yan, X., and Zhou, X. (2018). Screening of multimeric β-xylosidases from the gut microbiome of a higher termite, *Globitermes brachycerastes*. *Int. J. Biol. Sci.* 14, 608–615. doi: 10.7150/ijbs.22763

Liu, J. R., Yu, B., Lin, S. H., Cheng, K. J., and Chen, Y. C. (2005). Direct cloning of a xylanase gene from the mixed genomic DNA of rumen fungi and its expression in intestinal *Lactobacillus reuteri*. *FEMS Microbiol. Lett.* 251, 233–241. doi: 10.1016/j.femsle.2005.08.008

Liu, N., Li, H., and Chevrette, M. G. (2019). Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *ISME J.* 13, 104–117. doi: 10.1038/s41396-018-0255-1

Lorenz, P., and Eck, J. (2005). Metagenomics and industrial applications. *Nat. Rev. Microbiol.* 3, 510–516. doi: 10.1038/nrmicro1161

Mamo, G., Hatti-Kaul, R., and Mattiasson, B. (2006). A thermostable alkaline active endo-β-1-4-xylanase from *Bacillus halodurans* S7: purification and characterization. *Enzyme Microb. Technol.* 39, 1492–1498. doi: 10.1016/j.enzmictec.2006.03.040

Maruthamuthu, M., Jimenez, D. J., and van Elsas, J. D. (2017). Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach. *J. Appl. Microbiol.* 123, 145–158. doi: 10.1111/jam.13484

Matsuzawa, T., Kaneko, S., and Yaoi, K. (2015). Screening, identification and characterization of a GH43 family β-xylosidase/α-arabinosidase useful for biomass saccharification from a compost microbial metagenome. *Appl. Microbiol. Biotechnol.* 99, 8943–8954. doi: 10.1007/s00253-015-6647-5

Matsuzawa, T., Kimura, N., and Suenaga, H. (2016). Screening, identification, and characterization of α-xylosidase from a soil metagenome. *J. Biosci. Bioeng.* 122, 393–399. doi: 10.1016/j.jbiosc.2016.03.012

Mo, X. C., Chen, C. L., Pang, H., Feng, Y., and Feng, J. X. (2010). Identification and characterization of a novel xylanase derived from a rice straw degrading enrichment culture. *Appl. Microbiol. Biotechnol.* 87, 2137–2146. doi: 10.1007/s00253-010-2712-2

Mori, T., Kamei, I., Hirai, H., and Kondo, R. (2014). Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. *Springerplus* 3:365. doi: 10.1186/2193-1801-3-365

Ngara, T. R., and Zhang, H. (2018). Recent advances in function-based metagenomic screening. *Genomics Proteomics Bioinformatics.* 16, 405–415. doi: 10.1016/j.gpb.2018.01.002

Palackal, N., Lyon, C., Zaidi, S., Luginbuhl, P., Dupree, P., Goubet, F., et al. (2007). A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl. Microbiol. Biotechnol.* 74, 113–124. doi: 10.1007/s00253-006-0645-6

Qian, C., Liu, N., Yan, X., Wang, Q., Zhou, Z., and Wang, Q. (2015). Engineering a high-performance, metagenomic-derived novel xylanase with improved soluble protein yield and thermostability. *Enzyme Microb. Technol.* 70, 35–41. doi: 10.1016/j.enzmictec.2014.11.005

Radomski, C. C. A., Seow, K. T., Warren, R. A. J., and Yap, W. H. (1998). Method for isolating xylanase gene sequences from soil DNA, compositions useful in such method and compositions obtained thereby. U.S. Patent No 5,849,491. Washington, DC: U.S. Patent and Trademark Office.

Rashamuse, K., Sanyika Tendai, W., Mathiba, K., Ngcobo, T., Mtimka, S., and Brady, D. (2016). Metagenomic mining of glycoside hydrolases from the hindgut bacterial symbionts of a termite (*Trinervitermes trinervoides*) and the characterization of a multimodular β-1,4-xylanase (GH11). *Biotechnol. Appl. Biochem.* 64, 174–186. doi: 10.1002/bab.1480

Rennie, E. A., and Scheller, H. V. (2014). Xylan biosynthesis. *Curr. Opin. Biotechnol.* 26, 100–107. doi: 10.1016/j.copbio.2013.11.013

Rohman, A., Dijkstra, B. W., and Puspaningsih, N. N. T. (2019). β-Xylosidases: structural diversity, catalytic mechanism, and inhibition by monosaccharides. *Int. J. Mol. Sci.* 20:5524. doi: 10.3390/ijms20225524

Sae-Lee, R., and Boonmee, A. (2014). Newly derived GH43 gene from compost metagenome showing dual xylanase and cellulase activities. *Folia Microbiol.* 59, 409–417. doi: 10.1007/s12223-014-0313-7

Sato, M., Suda, M., Okuma, J., Kato, T., Hirose, Y., Nishimura, A., et al. (2017). Isolation of highly thermostable β-xylosidases from a hot spring soil microbial community using a metagenomic approach. *DNA Res.* 24, 649–656. doi: 10.1093/dnares/dsx032

Selvarajan, E., and Veena, R. (2017). Recent advances and future perspectives of thermostable xylanase. *Biomed. Pharmacol. J.* 10, 261–279. doi: 10.13005/bpj/1106

Shahraki, M. F., Farhadyar, K., Kavousi, K., Azarabad, M. A., Boroomand, A., Ariaeenejad, S., et al. (2019). TAXyl: an in-silico method for predicting the thermal activity for xylanases from GH10 and GH11 families. *bioRxiv [Preprint]* doi: 10.1101/826040

Shallom, D., and Shoham, Y. (2003). Microbial hemicellulases. *Curr. Opin. Microbiol.* 6, 219–228. doi: 10.1016/s1369-5274(03)00056-0

Shao, W., and Wiegel, J. (1992). Purification and characterization of a thermostable b-xylosidase from *Thermoanaerobacter ethanolicus*. *J. Bacteriol.* 174, 5848–5853. doi: 10.1128/jb.174.18.5848-5853.1992

Sharma, S., Adhikari, S., and Satyanarayana, T. (2007). Alkali-thermostable and cellulase free xylanase production by an extreme thermophile *Geobacillus thermoleovorans*. *World J. Microbiol. Biotechnol.* 23, 483–490. doi: 10.1007/s11274-006-9250-1

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5:209. doi: 10.3389/fpls.2014.00209

Sheng, P., Li, Y., Marshall, S. D. G., and Zhang, H. (2015). High genetic diversity of microbial cellulase and hemicellulase genes in the hindgut of *Holotrichia parallela* larvae. *Int. J. Mol. Sci.* 16, 16545–16559. doi: 10.3390/ijms160716545

Shi, H., Zhang, Y., Li, X., Huang, Y., Wang, L., Wang, Y., et al. (2013). A novel highly thermostable xylanase stimulated by Ca2+ from *Thermotoga thermarum*: cloning, expression and characterization. *Biotechnol. Biofuels* 6:26. doi: 10.1186/1754-6834-6-26

Shi, H., Zhang, Y., Zhang, H., Huang, Y., Li, X., and Wang, F. (2014). Cloning, over-expression and characterization of a thermo- tolerant xylanase from *Thermotoga thermarum*. *Biotechnol. Lett.* 36, 587–593. doi: 10.1007/s10529-013-1392-2

Shi, W., Ding, S. Y., and Yuan, J. S. (2011). Comparison of insect gut cellulase and xylanase activity across different insect species with distinct food sources. *Bioenergy Res.* 4, 1–10. doi: 10.1007/s12155-010-9096-0

Shi, W., Xie, S., Chen, X., Sun, S., Zhou, X., Liu, L., et al. (2013). Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genet.* 9:e1003131. doi: 10.1371/journal.pgen.1003131

Simpson, H. D., Haufler, U. R., and Daniel, R. M. (1991). An extremely thermostable xylanase from the thermophilic eubacterium Thermotoga. *Biochem. J.* 277(Pt 2), 413–417. doi: 10.1042/bj2770413

Singh, K. M., Reddy, B., and Patel, D. (2014). High potential source for biomass degradation enzyme discovery and environmental aspects revealed through metagenomics of Indian buffalo rumen. *Biomed Res. Int.* 2014:267189.

Sriprang, R., Asano, K., Gobsuk, J., Tanapongpipat, S., Champreda, V., and Eurwilaichitr, L. (2006). Improvement of thermostability of fungal xylanase by using site directed mutagenesis. *J. Biotechnol.* 12, 454–462. doi: 10.1016/j.jbiotec.2006.04.031

Sun, M., Zheng, H., and Meng, L. (2015). Direct cloning, expression of a thermostable xylanase gene from the metagenomic DNA of cow dung compost and enzymatic production of xylooligosaccharides from corncob. *Biotechnol. Lett.* 37, 1877–1886. doi: 10.1007/s10529-015-1857-6

Sunna, A., and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. *Crit. Rev. Biotechnol.* 17, 39–67. doi: 10.3109/07388559709146606

Sunna, A., and Bergquist, P. L. (2003). A gene encoding a novel extremely thermostable 1,4-β-xylanase isolated directly from an environmental DNA sample. *Extremophiles* 7, 63–70. doi: 10.1007/s00792-002-0296-1

Thidarat, N., Thongaram, T., Uengwetwanit, T., Pongpattanakitshote, S., and Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *J. Microbiol. Biotechnol.* 22, 462–469. doi: 10.4014/jmb.1108.08037

Thornbury, M., Sicheri, J., Slaine, P., Getz, L. J., Finlayson-Trick, E., and Cook, J. (2019). Characterization of novel lignocellulose-degrading enzymes from the porcupine microbiome using synthetic metagenomics. *PLoS One* 14:e0209221. doi: 10.1371/journal.pone.0209221

Turunen, O., Vuorio, M., Fenel, F., and Leisola, M. (2002). Engineering of multiple arginines into the Ser/Thr surface of *Trichoderma reesei* endo-1,4-beta-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH. *Protein Eng.* 15, 141–145. doi: 10.1093/protein/15.2.141

Uhl, A. M., and Daniel, R. M. (1999). The first description of an archaeal hemicellulase: the xylanase from *Thermococcus zilligii* strain AN1. *Extremophiles* 3, 263–267. doi: 10.1007/s007920050126

Verma, D., Anand, A., and Satyanarayana, T. (2013a). Thermostable and alkalistable endoxylanase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1: cloning, expression, characteristics and its applicability in generating xylooligosaccharides and fermentable sugars. *Appl. Biochem. Biotechnol.* 170, 119–130. doi: 10.1007/s12010-013-0174-6

Verma, D., Kawarabayasi, Y., Miyazaki, K., and Satyanarayana, T. (2013b). Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. *PLoS One* 8:e52459. doi: 10.1371/journal.pone.0052459

Verma, D., Kumar, R., and Satyanarayana, T. (2019). "Diversity in xylan-degrading prokaryotes and xylanolytic enzymes and their bioprospects," in *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications*, eds T. Satyanarayana, S. K. Das, and B. N. Johri (London: Springer Nature), 325–373. doi: 10.1007/978-981-13-8487-5_14

Verma, D., and Satyanarayana, T. (2012). Cloning, expression and applicability of thermo-alkali-stable xylanase of *Geobacillus thermoleovorans* in generating xylooligosaccharides from agro-residues. *Biores. Technol.* 107, 333–338. doi: 10.1016/j.biortech.2011.12.055

Verma, D., and Satynarayana, T. (2011). An improved protocol for DNA extraction from alkaline soil and sediment samples for constructing metagenomic libraries. *Appl. Biochem. Biotechnol.* 165, 454–464. doi: 10.1007/s12010-011-9264-5

Victorica, M. R., Soria, M. A., Batista-Garcia, R. A., Ceja-Navarro, J. A., and Vikram, S. (2020). Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes. *Sci. Rep.* 10:3864.

Wagschal, K., Heng, C., Lee, C. C., and Wong, D. W. (2009). Biochemical characterization of a novel dual-function α-arabinofuranosidase/β-xylosidase isolated from a compost starter mixture. *Appl. Microbiol. Biotechnol.* 81, 855–863. doi: 10.1007/s00253-008-1662-4

Wang, C., Dong, D., and Wang, H. (2016). Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9:22.

Wang, G., Wang, Y., Yang, P., Luo, H., Huang, H., Shi, P., et al. (2010). Molecular detection and diversity of xylanase genes in alpine tundra soil. *Appl. Microbiol. Biotechnol.* 87, 1383–1393. doi: 10.1007/s00253-010-2564-9

Wang, Q., Luo, Y., He, B., Jiang, L. S., Liu, J. X., and Wang, J. K. (2015). Characterization of a novel xylanase gene from rumen content of Hu sheep. *Appl. Biochem. Biotechnol.* 177, 1424–1436. doi: 10.1007/s12010-015-1823-8

Wang, W., Hu, H., and Zijlstra, R. T. (2019). Metagenomic reconstructions of gut microbial metabolism in weanling pigs. *Microbiome* 7:48.

Wilkens, C., Andersen, S., Dumon, C., Berrin, J. G., and Svensson, B. (2017). GH62 arabinofuranosidases: structure, function and applications. *Biotechnol. Adv.* 35, 792–804. doi: 10.1016/j.biotechadv.2017.06.005

Wiseschart, A., Mhuantong, W., and Tangphatsornruang, S. (2019). Shotgun metagenomic sequencing from Manao-Pee cave, Thailand, reveals insight into the microbial community structure and its metabolic potential. *BMC Microbiol.* 19:144. doi: 10.1186/s12866-019-1521-8

Yamada, K., Terahara, T., Kurata, S., Yokomaku, T., Tsuneda, S., and Harayama, S. (2008). Retrieval of entire genes from environmental DNA by inverse PCR with

pre-amplification of target genes using primers containing locked nucleic acids. *Environ. Microbiol.* 10, 978–987. doi: 10.1111/j.1462-2920.2007.01518.x

Yan, R., Vuong, T. V., Wang, W., and Master, E. R. (2017). Action of a GH115 α-glucuronidase from *Amphibacillus xylanus* at alkaline condition promotes release of 4-O methylglucopyranosyluronic acid from glucuronoxylan and arabinoglucuronoxylan. *Enzyme Microb. Technol.* 104, 22–28. doi: 10.1016/j.enzmictec.2017.05.004

Zhang, M., Chekan, J. R., Dodd, D., Hong, P. Y., Radlinski, L., Revindran, V., et al. (2014). Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3708–E3717.

Zhao, S., Bu, D., Wang, J., Liu, K., Zhu, Y., Dong, Z., et al. (2010). Novel glycoside hydrolases from a metagenome library of the rumen of Chinese holstein dairy cows. *Appl. Environ. Microbiol.* 76, 6701–6705. doi: 10.1128/aem.00361-10

Zhou, J., Bao, L., Chang, L., Zhou, Y., and Lu, H. (2012). Biochemical and kinetic characterization of GH43 β-D-xylosidase/α-L-arabinofuranosidase and GH30 α-L-arabinofuranosidase/β-D -xylosidase from rumen metagenome. *J. Ind. Microbiol. Biotechnol.* 39, 143–152. doi: 10.1007/s10295-011-1009-5

Zhou, L., Bao, L., Chang, L., Liu, Z., You, C., and Lu, H. (2011). Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. *Lett. Appl. Microbiol.* 54, 79–87. doi: 10.1111/j.1472-765x.2011.03175.x

Zhu, N., Yang, J., and Ji, L. (2016). Metagenomic and metaproteomic analyses of a corn stover-adapted microbial consortium EMSD5 reveal its taxonomic and enzymatic basis for degrading lignocellulose. *Biotechnol. Biofuels* 9:243.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/fmicb.2020.573482

Aspergillus sydowii: Genome Analysis and Characterization of Two Heterologous Expressed, Non-redundant Xylanases

*Sophie C. Brandt[1†], Bernhard Ellinger[2†], Thuat van Nguyen[1‡], Sönke Harder[3], Hartmut Schlüter[3], Richard L. Hahnke[4], Martin Rühl[5], Wilhelm Schäfer[1] and Martin Gand[1,5]**

[1]Department of Molecular Phytopathology, University of Hamburg, Hamburg, Germany, [2]Department ScreeningPort, Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Hamburg, Germany, [3]Mass Spectrometric Proteomics Group, Institute of Clinical Chemistry and Laboratory Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [4]Department of Microorganisms, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [5]Institute of Food Chemistry and Food Biotechnology, Justus Liebig University Giessen, Giessen, Germany

OPEN ACCESS

Edited by:
Sunil Khare,
Indian Institute of Technology Delhi, India

Reviewed by:
Satya P. Singh,
Saurashtra University, India
Digvijay Verma,
Babasaheb Bhimrao Ambedkar University, India

***Correspondence:**
Martin Gand
martin.gand@lcb.chemie.uni-giessen.de

[†]These authors have contributed equally to this work

[‡]Deceased 2015

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 17 June 2020
Accepted: 14 August 2020
Published: 18 September 2020

Citation:
Brandt SC, Ellinger B, van Nguyen T, Harder S, Schlüter H, Hahnke RL, Rühl M, Schäfer W and Gand M (2020) Aspergillus sydowii: Genome Analysis and Characterization of Two Heterologous Expressed, Non-redundant Xylanases. Front. Microbiol. 11:573482. doi: 10.3389/fmicb.2020.573482

A prerequisite for the transition toward a biobased economy is the identification and development of efficient enzymes for the usage of renewable resources as raw material. Therefore, different xylanolytic enzymes are important for efficient enzymatic hydrolysis of xylan-heteropolymers. A powerful tool to overcome the limited enzymatic toolbox lies in exhausting the potential of unexplored habitats. By screening a Vietnamese fungal culture collection of 295 undiscovered fungal isolates, 12 highly active xylan degraders were identified. One of the best xylanase producing strains proved to be an *Aspergillus sydowii* strain from shrimp shell (Fsh102), showing a specific activity of 0.6 U/mg. Illumina dye sequencing was used to identify our Fsh102 strain and determine differences to the *A. sydowii* CBS 593.65 reference strain. With activity based in-gel zymography and subsequent mass spectrometric identification, three potential proteins responsible for xylan degradation were identified. Two of these proteins were cloned from the cDNA and, furthermore, expressed heterologously in *Escherichia coli* and characterized. Both xylanases, were entirely different from each other, including glycoside hydrolases (GH) families, folds, substrate specificity, and inhibition patterns. The specific enzyme activity applying 0.1% birch xylan of both purified enzymes were determined with 181.1 ± 37.8 or 121.5 ± 10.9 U/mg for xylanase I and xylanase II, respectively. Xylanase I belongs to the GH11 family, while xylanase II is member of the GH10 family. Both enzymes showed typical endo-xylanase activity, the main products of xylanase I are xylobiose, xylotriose, and xylohexose, while xylobiose, xylotriose, and xylopentose are formed by xylanase II. Additionally, xylanase II showed remarkable activity toward xylotriose. Xylanase I is stable when stored up to 30°C and pH value of 9, while xylanase II started to lose significant activity stored at pH 9 after exceeding 3 days of storage. Xylanase II displayed about 40% activity when stored at 50°C for 24 h. The enzymes are tolerant toward mesophilic temperatures, while acting in a broad pH range. With site directed

mutagenesis, the active site residues in both enzymes were confirmed. The presented activity and stability justify the classification of both xylanases as highly interesting for further development.

Keywords: *Aspergillus sydowii*, xylanase, mass spectrometry, heterologous expression, glycoside hydrolases 10 and 11, biocatalysis, site directed mutagenesis

INTRODUCTION

Due to the rising population and the increasing global wealth, the demand for energy is constantly increasing (Sorrell, 2015). As a consequence, the petrochemical reserves are decreasing (Miller and Sorrell, 2014). As a result of the global climate change and the accelerated consumption of fossil fuels (Sommer, 2016), a general scientific and political rethink had taken place in the last decade. A transition to biobased products is, therefore, necessary [National Research Council (US) Committee on Biobased Industrial Products, 2000]. Renewable fuels (Sims et al., 2010) and other chemicals (Straathof, 2014) can be generated from plant materials. As a result of these efforts, the first generation of biofuels, which are produced directly from food crops, can substitute fossil fuels (Mohr and Raman, 2013). This first generation suffers in the limitations of the competition between food or feed and the biomass for the biofuels (Valentine et al., 2012). To overcome these limitations, the second generation biofuels based on plant waste material as a source for bioethanol were developed (Saini et al., 2015) and, at present, second-generation biofuels are still under development. The production of lignocellulosic plant waste material is estimated to be currently about 5 billion metric tons per year (Global Partnership on Waste Management, 2011). This amount of material has a stored energy equivalent to 1.2 billion metric tons of oil (Global Partnership on Waste Management, 2011). These numbers highlight that plant waste biomass is an important resource for renewable energy. The hydrolysis of biomass to monosaccharides is still challenging due to the complex plant cell-wall structure, being a mixed composition of cellulose, hemicellulose, and lignin (Gilbert, 2010; Agbor et al., 2011; Singhvi et al., 2014). Hemicellulose is the second most abundant polysaccharide in nature, in land plants the amount of xylan, the major constituent of hemicellulose, is 20–40% of the total dry weight (Bajpai, 2014). These hemicellulose macromolecules are loosely structured, consisting of polymers of pentoses (xylose and arabinose) and hexoses (mostly mannose) in varying ratios (Holtzapple, 2003). Xylan, the main component, consists of a linear backbone of β-D-xylopyranosyl sugars chemically linked by β-(1,4) glycosidic bonds. This backbone is substituted with 4-O-methyl-α-D-glucuronopyranosyl units, acetyl groups or α-L-arabinofuranosyl units, in different distribution (Bajpai, 2014). The complete biodegradation requires cooperation of several enzymes, namely endo-1,4-β-xylanases (EC 3.2.1.8), β-D-xylosidases (EC 3.2.1.37), α-L-arabinofuranosidases (EC 3.2.1.55), α-glucuronidases (EC 3.2.1.139), acetyl xylan esterases (EC 3.1.1.72), and ferulic acid esterases (EC 3.1.1.73; Shallom and Shoham, 2003). Sometimes exo-xylanases (EC 3.2.1.37) are completing the toolbox of hydrolytic hemicellulases. Among them, endo-1,4-β-xylanase plays an important role in the degradation of xylan by hydrolyzing the xylosyl backbone into smaller xylooligosaccharides (XOS). These enzymes belong to the GH10 and GH11 groups of glycoside hydrolases (GH; Henrissat, 1991). The GH family 10 was assigned to GH-A clan (Davies and Sinnott, 2008), because of their TIM barrel-$(\beta/\alpha)_8$ structure, whereas GH family 11 was allocated to GH-C clan (Davies and Sinnott, 2008), because of their β-jelly roll structure. The next step in the degradation cascade is the hydrolysis of the XOS into β-D-xylose residues by 1,4-β-xylosidases. After xylan degradation, the recalcitrant lignocellulose is more amendable to further degradation by other lignocellulosic enzymes (Chandra et al., 2007). Generally in industrial approaches, a chemical or physical pretreatment is used to reduce lignin content for an easier work flow (Chandra et al., 2007). Beside biofuels, xylanases are used for various other industrial processes such as food, feed, and pulp and paper industries (Beg et al., 2001). These current applications include prebleaching of pulp to reduce the amount of chemical bleaching (Walia et al., 2017), increasing the loaf volumes in baking (Butt et al., 2008), enhancing the weight gain of broiler chickens by incorporation of xylanase into the rye-based diet food (Bedford et al., 1991), clarifying juices of fruits and vegetables (Biely, 1985), and treating the wastewater from agricultural production (Biely, 1985).

In nature, the lignocellulosic biomass is mainly decomposed by fungi (Guerriero et al., 2016). The fungal kingdom includes over 5 million species, and it is estimated that only 5% are formally classified (Mueller and Bills, 2004; Blackwell, 2011). The Ascomycota are considered to have exceptional relevance for mankind (Bennett, 1998), in consequence of their potential to produce pharmaceuticals (Gibbs et al., 2000) and other chemicals (van der Straat et al., 2014), as well as their application in food and beverage industry (Ferreira et al., 2016). *Aspergillus* is one of the most abundant genus in that phylum (Klich, 2002). Depending on the growth media, they exhibit characteristic colors, from black, to gray or green up to white or milky. Over 250 species of the genera, *Aspergillus* are currently known. A less studied fungus of that genus is *Aspergillus sydowii*, which was first described in 1913 as *Sterigmatocystis sydowii* by Bainier and Sartory (1913). In 1926, the fungus was reclassified as *A. sydowii* (Thom and Church, 1926). The fungus is distributed all over the world and can occur in different environments, where it survives as a saprotroph (Raper and Fennell, 1965; Geiser et al., 1998; Klich, 2002). Beside its terrestrial appearance in soil, it can grow in the sea due to its salt tolerance (Hallegraeff et al., 2014). In addition, *A. sydowii* can cause human infections (Rinaldi, 1983) as well as contaminate food (Piotrowska, 2013; Biango-Daniels and Hodge, 2018). In marine ecosystems, it causes epizootic infections of sea fan corals (Geiser et al., 1998; Alker et al., 2001; Kim and Harvell, 2004).

Within this study, we aimed to investigate an *A. sydowii* strain, which originates from shrimp shell, and its endo-xylanases. This strain was part of a collection previously isolated from nine different substrates and environments, collected from the biodiverse country Vietnam. They were analyzed for their potential do degrade lipids, chitin, cellulose, and xylan (Brandt et al., 2018). Twelve highly active xylan degraders were identified, while two of them were *Aspergillus* strains with high activities in all four investigated assays. By genome sequencing, we compared the phylogenetic relationship and the potential CAZymes of our strain with the CBS 593.65 *A. sydowii* of unknown origin, which was recently sequenced by the Joint Genome Institute (JGI; de Vries et al., 2017). Using zymography coupled with mass spectrometry, we were able to identify three potential xylanases. Two of them were annotated as differing endo-xylanases highly active and heterologously expressed and characterized. These enzymes showed promising activities for sustainable industrial bioprocesses around the enzymatic degradation of xylan-containing biomass.

MATERIALS AND METHODS

Fungus Isolation and Growth Conditions

The *A. sydowii* strain was isolated from shrimp shells in Vietnam according to Brandt et al. (2018). Fungal mycelium was cultured in liquid potato-dextrose-broth (PDB) at 28°C with 150 rpm and used to isolate DNA.

For storage and further experiments, a conidia suspension was used. A 5 mm diameter mycelium piece from potato-dextrose-agar (PDA), was transferred to a new PDA plate and grown in the dark for 5–7 days at 28°C. The conidia were scraped with Drigalski cell spreader and sterile water centrifuged 4°C with 2,693 *g*, washed twice with sterile water, followed by filtering through a 40 µm mesh sieve. Centrifuged at 4°C with 2,693 *g* again for concentration and resuspended in sterile water, aliquoted and stored at −70°C. The fungus was deposited at the DSMZ under the number DSM105790. To investigate mycelial colors, fungal growth was determined on PDA, yeast extract peptone dextrose (YPD; CSH Protocols, 2010), complete media (CM; Leach et al., 1982), and malt extract agar (MEA; Raper and Thom, 1949; **Figure 1A**).

For the enzymatic evaluation and protein identification, a two-step mycelium cultivation was performed. Mycelium was pre-cultivated in YPD media at 28°C at 150 rpm for 3 days, then washed briefly and dried between filter papers (Whatman, Dassel, Germany). One tenth gram of semi-dried mycelia was incubated with 50 ml of inductive media for 3 days, at 28°C and 150 rpm. The inductive media consisted of a mineral salt media (0.35% $NaNO_2$, 0.15% K_2HPO_4, 0.05% $MgSO_4 \times 7\ H_2O$, 0.05% KCl, 0.001% $FeSO_4 \times 7\ H_2O$) supplemented with 1% (w/v) birch xylan (Sigma Aldrich, Steinheim, Germany).

Enzymatic Evaluation of the Fungal Supernatants

For the detection of the enzymatic activity of the 24 fungal strains of the collection identified by Brandt et al. (2018), 1.5 ml of the supernatant was incubated with 1.5 ml 2% (w/v) substrate (xylan from birch wood, Sigma-Aldrich, Steinheim, Germany) in 50 mM sodium acetate pH 6.5 in 37°C for 2 h under constant shaking at 300 rpm. Enzyme activity was

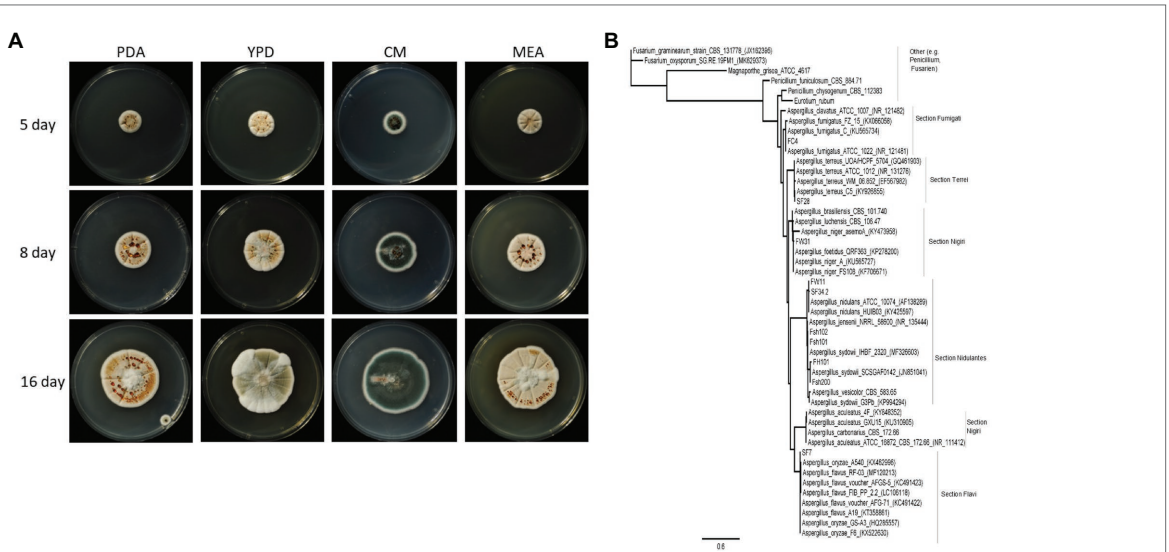


FIGURE 1 | (A) Overview of the growth of *Aspergillus sydowii* Fsh102 using four different media. The cultures were incubated at 28°C for 16 days. Pictures were taken after 5, 8, and 16 days. PDA, potato-dextrose-agar; YPD, yeast extract peptone dextrose; CM, complete media; and ME, malt extract agar. **(B)** Phylogenetic tree from *Aspergillus* clades of ITS sequences from the unique fungal strain collection from Vietnam (Brandt et al., 2018) with two strains of *Fusarium* and one for *Magnaporthe grisea*, as well as one for *Eurotium rubum*. Phylogenetic tree highlighting the position of *A. sydowii* isolate and closely related type strains. The tree was inferred from aligned internal transcribed spacers 1 and 2 without adjacent ribosomal genes. Phylogenetic tree was inferred under the maximum likelihood (ML) criterion using the PhyML function in geneious® applying the Jukes Cantor substitution model JC69.

determined by quantifying the resulting D-xylose originating from xylanase activity by DNS assay. DNS assay was done as described previously (Miller, 1959).

DNA Isolation and Genomic Sequencing

High molecular gDNA from *A. sydowii* strain were isolated with the CTAB method (Doyle and Doyle, 1990; Richards et al., 1994). The ITS-1/8S rRNA/ITS-2 region was amplified and sequenced with primers ITS1 [F] and ITS4 [F] (**Supplementary Table 5**) according to White et al. (1990). Sequencing reactions were performed using the ABI Dye Terminator technology according to the manufacturer's instructions (Applied Biosystems, Foster City, CA, United States). The ITS sequence have been deposited in GenBank with accession number MG098740. An amount 100 µl gDNA solution with a concentration of 2,600 ng/µl were sent to Beijing Genomics Institute (BGI) for sequencing, a 0.8% agarose gel of the isolated gDNA can be found in the Supplementary Material (**Supplementary Figure 1**). The gDNA for sequencing was prepared as describe by Meyer and Kircher (2010). The library products were sequenced *via* Illumina HiSeqTM 4000.

Bioinformatic, Phylogenetic, CAZyme, Xylanase Sequence, and Glycosylation Analysis

The genome is available at NCBI with the BioSample number of SAMN08024727. The SOAPaligner (version 2.21) was used to align reads to reference sequence (*A. sydowii* CBS 593.65) and to calculate the average depth and coverage ratio compared to the reference sequence as describe by Gu et al. (2013). Based on aligned results of filtered reads and reference sequence, the single nucleotide polymorphisms (SNPs) and an insertions or deletions of bases (InDels) are annotated.

From the unique fungal strain collection from Vietnam (Brandt et al., 2018), 10 strains from *Aspergillus* as well as ITS sequences from 33 *Aspergillus* spp., two *Penicillium* spp., two *Fusarium* spp., and one *Eurotium rubum* retrieved from MycoBank (Robert et al., 2013), have been used. The phylogenetic tree (**Figure 1B**) was inferred under the maximum likelihood (ML; Guindon et al., 2010), from sequences of the internal transcribed spacers 1 and 2, aligned with ClustalW (Thompson et al., 1994) with gap open cost of 15 and gap extend cost of 6.66 with free end gaps.

For the CAZyme analysis, the coded regions (CDS) resulting from genomic sequencing were compared with the CAZyme database (Cantarel et al., 2009; Lombard et al., 2014). The annotation of the CAZyme resulted from a combination of the RAPSearch2 search (Ye et al., 2011; Zhao et al., 2012) and HMMER scanning (Finn et al., 2014) described in Hahnke et al. (2015).

For the generation of protein trees the software Geneious version 9.1 was used. The alignment type was global with free end gaps, as cost matrix Blosum62 (Henikoff and Henikoff, 1992) was used with genetic distance model Jukes-Cantor (Jukes and Cantor, 1969). The tree build method was neighbor-joining with no outgroup and a gap open penalty of 12 and a gap extension penalty of 3. Proteins from GH10 or GH11 family, respectively, were downloaded manually from the JGI MycoCosm (Grigoriev et al., 2014), showing high similarity to Fsh102 strain, i.e., *Aspergillus brasiliensis*, *Aspergillus luchuensis*, *Aspergillus nidulans*, *A. sydowii*, *Aspergillus vesicolor*, as more distant strains *E. rubum* and *Penicillium chrysogenum* were used. For Fsh102 strain, the proteins of the GH10 and GH11 families were added. Furthermore, the highest similar structure (mmseqs2 method from the PDB) was downloaded from the Protein Data Bank (PDB) and added, namely 1TA3 an endo-1,4-β-xylanase GH10 from *A. nidulans* (Payan et al., 2004) for xylanase II or 1TE1 an endo-1,4-β-xylanase GH11 from *Penicillium funiculosum* (Payan et al., 2004) for xylanase I, respectively. Separate trees were created for the GH10 or the GH11 (**Figures 2B,C**). The GenBank files for xylanase I GH11 is BK013306, for the xylanase II GH10 is BK013307, and for a putative exo-1,4-β-xylosidase GH3 is BK013308.

For the O-glycosylation the tool NetOGlyc 4.0 Server (Steentoft et al., 2013) and for the N-glycosylation of the proteins the tool NetNGlyc 1.0 Server (Gupta et al., 2004) were used.

Zymography and Mass Spectrometry

For zymogram gels, 12.5% SDS-PAGE (Laemmli, 1970) mixed with 0.5% xylan. The supernatants were mixed in equal parts with SDS-PAGE sample buffer and separated in SDS-PAGE. The proteins in the gel were then renatured in refolding buffer [TRIS-HCl 50 mM pH 6.8, 2% (v/v) Triton X-100] for 30 min at room temperature. The gel was washed for a further 30 min in 50 mM TRIS-HCl (pH 6.8) and incubated for one hour in washing buffer (Sodium phosphate 100 mM pH 6.8) at 37°C. Subsequently, unreacted substrate was stained with Congo red and excess dye was removed with 1 M NaCl solution by changing the solution several times. In order to obtain a clear decoloration, the gel was incubated in 70% (v/v) ethanol for a few minutes. For the LC-MS/MS analysis, visible zones of the zymogram gel (**Figure 2A**) or the SDS gel (**Supplementary Figure 5**) were cut out and treated following the protocol of Shevchenko et al. (2007) prior to LC-MS/MS analysis. Protein identification was performed *via* analysis by subjecting the tryptic peptides to LC-MS/MS equipped with Acclaim PepMap RSLC C18 column (Thermo Fisher Scientific, Carlsbad, United States). Samples were injected onto a nano-liquid chromatography system (Dionex UltiMate 3000 RSLCnano, Thermo Scientific, Germany) coupled *via* electrospray-ionization (ESI) to a mass spectrometer (MS) equipped with a quadrupole, a linear trap, and an orbitrap (Orbitrap Fusion, Thermo Scientific, Germany). Identification of the proteins from the MS/MS spectra was performed using the search engine Sequest HT and the JGI *A. sydowii* database (de Vries et al., 2017) or the protein sequences of xylanase I or xylanase II from Fsh102. Peptides with a false discovery rate (FDR) of 1% were identified. At least two unique peptides per protein had been found for a reliable identification.

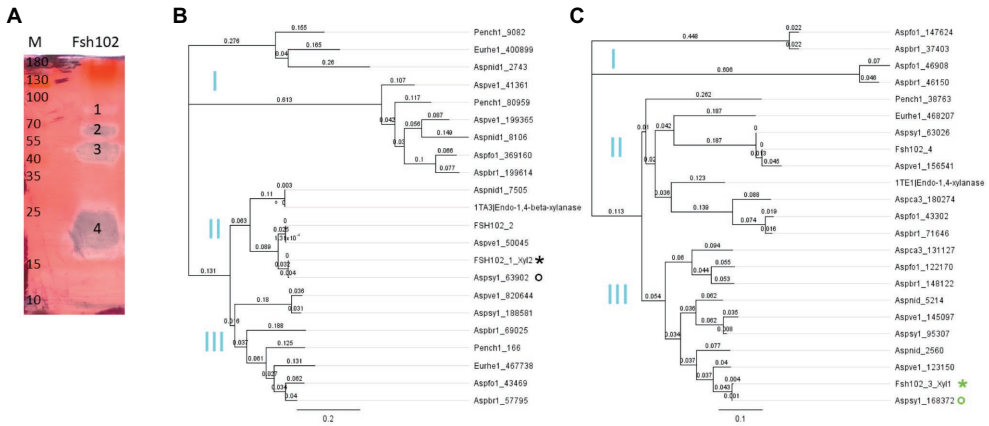


FIGURE 2 | (A) Zymogram of xylanase activity using the supernatant of the Fsh102 fungal strain separated by a 12.5% (w/v) SDS-PAGE containing 0.5% (w/v) xylan. The xylan in the gel was stained with 0.1% (w/v) Congo red. Fsh102: the total supernatant of the fungal isolate Fsh102 after 24 h incubation in 1% (w/v) xylan medium. The areas with xylanase activity (1–4) remain unstained. M: Page Ruler Plus Prestained Protein Ladder. The molecular weights of the standard are given in kDa. **(B,C)** Phylogenetic trees of amino acid sequences of selected members of the GH10 **(B)** or GH11 **(C)** families. The global alignment was performed with free end gaps and the cost matrix Blosum62 was used with genetic distance model Jukes-Cantor. The gene are from the JGI MycoCosm. Aspbr1, *Aspergillus brasiliensis*; Aspfo1, *Aspergillus luchuensis* formally known as *Aspergillus foetidus*; Aspnid1, *Aspergillus nidulans*; Aspsy1, *Aspergillus sydowii*; Aspve1, *Aspergillus versicolor*; Aspca3, *Aspergillus carbonarius* formally known as *Aspergillus aculeatus*; Eurhe1, *Eurotium rubum*; and Pench1, *Penicillium chrysogenum*. The gene identified are behind the underline. From the Protein Data Bank (PDB), a candidate for the GH10 or GH11, respectively, was used 1TA3 an 1,4-β-xylanase GH 10 from *Aspergillus nidulans* for the Xylanase II or 1TE1 an Endo-1,4-β-xylanase GH 11 from *Penicillium funiculosum*. For the Fsh102 two candidates are added per GH family (FSH102_1 to FSH102_4), while xylanase II = FSH102_1_Xyl2 (black star) and xylanase I = FSH102_3_Xyl1 (green star). The primers were deduced from the sequences marked with the circles, while the green one was used for the xylanase I or the black for the xylanase II, respectively. Subclades (I–III) are marked in cyan.

Cloning of Xylanase Encoding Genes

Mycelium (about 100 mg), grown 3 days in the inductive media, was harvested and grinded using liquid nitrogen. For isolation of total RNA, peqGOLD TriFast™ (Peqlab Biotechnologie GmbH, Erlangen, Germany) was applied. Integrity of the RNA was checked by agarose gel electrophoresis and ethidium bromide staining. cDNA was constructed using the RevertAid H Minus First Strand cDNA Synthesis Kit according to the manufacturer's instructions (Thermo Fisher Scientific, Carlsbad, United States), the reverse transcriptase was used for first-strand synthesis with random hexamer, as well as for the second-strand amplification of cDNA. PCR primers were deduced from amino acid sequences of xylanase I from the Aspsy1 168372|estExt_Genewise1.C_1_t60491 gene and for xylanase II from the Aspsy1 63902|estExt_fgenesh1_pg.C_1_t20379 gene, found by mass spectrometry detection. Primer design was assisted by the Oligo Calc tool[1], while the FastCloning method (Li et al., 2011) was used for direct cloning into the expression vector pET28b (BioCat, Heidelberg, Germany). The same method was used for the removal of the secretion signal (Petersen et al., 2011) for both recombinant xylanases enzymes.

Protein Biosynthesis and Purification

Kanamycin 50 μg/ml was used in the cultivations of *Escherichia coli* as antibiotic. In a typical procedure, *E. coli* BL21 (DE3) cells transformed with pET28b_Xylanase were inoculated in LB medium from an overnight culture (1:100 v/v) and grown at 37°C until an OD_{600} of 0.2–0.5. Protein expression was initiated by the addition of IPTG to a final concentration of 2 mM, while lowering the temperature to 22°C. 17 h after induction, the cells were harvested (4,000 g, 4°C, 10 min) and resuspended in 50 mM sodium phosphate buffer (pH 8). The cells were lysed with ultrasound, using Hielscher UP200H (Hielscher Ultrasonic GmbH, Teltow, Germany) twice for 2 min on ice, 60% amplitude and cycle 0.5, and centrifuged to separate cell debris from soluble protein. The His_6-tagged xylanases were purified by affinity chromatography using self-packed Nickel-NTA (PureCube Ni – NTA Agarose; Biozym, Hamburg, Germany) columns (Qiagen, Hilden, Germany). In a second step, the xylanases were desalted using Centricon filters of a molecular weight limit of 10 kDa (Merck Millipore, Billerica, USA). The fraction of the affinity chromatography was loaded in the Centricons and washed three times with sodium phosphate buffer (50 mM, pH 8) to remove imidazole. Finally, the purified protein was concentrated to ~2 ml using the same buffer. Purity was evaluated with SDS-PAGE, while protein content was determined using Roti®-Nanoquant protein detection kit (Carl Roth, Karlsruhe, Germany) following manufacturers' protocol.

Xylanase Assays

To examine substrate specificity 10 μl of appropriately diluted enzyme solutions was mixed with 40 μl buffer and added to 50 μl substrate solution. The enzymatic hydrolysis was performed in TGradient thermocycler (Biometra, Göttingen, Germany) at different temperatures for 15 min. Followed by the addition

[1]http://biotools.nubic.northwestern.edu/OligoCalc.html

of 100 µl DNS solution to generated color formation at 95°C for 10 min. One hundred microliter of these reaction mixtures were measured in the Mithras2 LB943 spectrophotometer (Berthold Technologies, Bad Wildbad, Germany), by the absorbance at 575 nm and correlated to a standard curve for quantification. For each measurement, a negative control without enzyme solution was included. One unit of activity was defined as the amount of enzyme required to release 1 µmol of product equivalent per min in the assay conditions.

To measure the substrate specificity, the xylan was replaced by xylobiose, xylotriose, xylotetraose, xylopentose and xylohexose, arabinan, arabinoxylan, CMC, galactan, and starch, respectively.

Inhibition or enhancement of xylanase activity was determined by applying a range of different compounds to the standard reaction. The following substances were evaluated: NaCl (10 mM, 1,000 mM), $MgCl_2$ (5 mM, 100 mM), $MnCl_2$ (5 mM), EDTA (2 mM, 50 mM), and DTT (10 mM, 50 mM). Furthermore 5 and 30% methanol, ethanol, acetonitrile, and DMSO, as well as Tween20 with 0.25 and 5%, Triton-X 100 with 0.25 and 5%, and SDS with 1% were evaluated.

To test the temperature range of the enzyme, the standard assay mixture activity was measured at temperatures between 10 and 90°C.

The pH range was determined with 0.1% of xylan from birch wood (Sigma-Aldrich, Steinheim, Germany) applying a broad range of pH from 2.5 to 10. Following buffers were used: glycine HCl (pH 2.5–3.0), citrate (pH 3.0–6.0), phosphate (pH 6.0–8.0), and glycine NaOH (pH 8.5–10.0), all 50 mM. Thermostability and buffer stability were determined by measuring residual activity with the standard assay after preincubation of enzyme samples under the following conditions: temperatures were 4, 30, 50, and 70°C for up to 8 days, in 50 mM sodium acetate buffer pH 4.8 and 50 mM glycine NaOH buffer pH 9 for up to 8 days.

To measure the substrate depended activity xylan concentration was changed in a range from 0.1 to 3.0% using standard conditions. The K_M, V_{max}, and k_{cat} values are calculated using the software GraphPad Prism 8.0.0 (GraphPad Software Inc., San Diego, United States).

The reported activities and standard deviations are the mean of three purifications (from independent cultivations), each measured in triplicates.

Analysis of Hydrolysis Products by Thin-Layer Chromatography

For the analysis of the hydrolysis products of birch xylan (Sigma Aldrich, Steinheim, Germany), XOS: xylose, xylobiose, xylotriose, xylotetraose, xylopentose, and xylohexose (Megazyme, Wicklow, Ireland) as solutions of 0.1% were used as standards. The 0.1% xylan (50 mM citrate buffer pH 4.8) was analyzed after applying 1 µg xylanase I or xylanase II, respectively, at 30°C. Aliquots from these incubation were taken after 15 min, 60 min, and 17 h, spotted and analyzed on silica-gel G-60 plates F_{254} (20 × 20 cm, Merck). For separation, ethyl acetate/acetic acid/formic acid/distilled water (9:3:1:4; v/v/v/v) was used as mobile phase. After separation, the plates were dried at 80°C for 10 min followed by visualization of the sugars by spraying freshly prepared 0.2% (w/v) orcinol in sulfuric acid/methanol (1:9; v/v) solution and incubation at 80°C for 2 h.

Creation of Homology Models and Site-Directed Mutagenesis

Homology models for the two xylanases were created using YASARA (ver. 13.9.8). These models were used to determine the active site residues.

The site-directed mutagenesis to prove the active site residues was performed using a modified QuikChange® PCR protocol (Nobili et al., 2013), and the products were transformed in *E. coli* NEB 10 β cells to repair the nicked ends, followed by plasmid isolations to finally transform into BL21 (DE3) for protein expression.

RESULTS

Isolation and Characterization of the Fungal Strain Fsh102

By screening a collection of 295 fungi from nine different habitats in Vietnam, prolific universal degrader of lipids, chitin, cellulose, and xylan (Brandt et al., 2018) were identified. We investigated 24 different isolates of these fungi, while the specific enzymatic activity of the cultural supernatant of fungus Fsh102 was 0.61 ± 0.01 U/mg (**Supplementary Table 1**). For this fungus, isolated from shrimp shells, mycelial growth and pigmentation on different media were monitored (**Figure 1A**). Depending on the media, pale-white, to orange and up to green colored colonies were identified, which are typical for *Aspergillus* species. Genomic DNA was isolated (**Supplementary Figure 1**) and the internal transcribed spacer regions one and two were amplified and sequenced (**Supplementary Data**). Sequencing revealed Fsh102 to be an *A. sydowii* strain (now stored under the number: DSM105790) and the phylogenetic relationship fortifies this results (**Figure 1B**). The three nearest relatives are all different *A. sydowii* strains, while *A. sydowii* IHB_2320 showing 100% identity with Fsh102. In general, *A. sydowii* is known to have a marine lifestyle, which is in line with the discovery on shrimp shells (Liu et al., 2019).

Evaluation of the Enzymatic Activity and Identification of Xylanases

By cultivating the fungus in liquid inductive media containing 1% xylan as only carbon source, xylan biopolymer degradation were be observed (**Supplementary Figure 2**). We, therefore, performed a zymogram, of supernatant from inductive media, to identify the responsible proteins (**Figure 2A**). The separated zymogram bands were digested and analyzed by protein mass spectrometry (**Supplementary Tables 2** and **3**). Analyzing the four activity spots from the zymogram by mass spectrometry, 24 proteins were detected in total, from which five of them occurred twice. For the 19 unique proteins, only six putative enzymes were assigned. Five of them are GH family members of which three are predicted enzymes acting on xylan, one

exo-xylanase, and two endo-xylanase. We focused on the two putative endo-xylanases, designated as xylanase I (**Supplementary Table 4**, similar to the JGI automated annotation Aspsy1 168372|estExt_Genewise1.C_1_t60491) and xylanase II (**Supplementary Table 4**, similar to the JGI automated annotation Aspsy1 63902|estExt_fgenesh1_pg.C_1_t20379). Xylanase II was identified from zymogram band 3, and has a length of 328 amino acids and a predicted mass of 35.43 kDa. Xylanase I was detected from zymogram band 4 with a length of 220 amino acids and a predicted mass of 23.52 kDa. The clearing zone in band 1 was caused by the putative exo-xylanase (**Supplementary Table 4**, similar to the JGI automated annotation Aspsy1 84747|gm1.1430_g), while the zone 2 presumably originated from a promiscuous endo-glucosidase, because no other enzymes was found inside this zone.

Bioinformatic Analysis of *Aspergillus sydowii* Fsh102

We used the JGI sequenced strain *A. sydowii* CBS 593.65 (taxid: 1036612) for comparison of with isolate Fsh102. The assembled genome of the *A. sydowii* strain CBS 593.65 has a size of 34.38 Mbp (reference genome). In comparison to this, a sequencing coverage rate of 88.65% was obtained, resulting in a 30.48 Mbp genome of *A. sydowii* strain Fsh102. Compared to the *A. sydowii* CBS 593.65, the Fsh102 has 91,709 total SNPs (0.267%) and 2,829 total InDels (0.008%).

CAZyme Analysis

The CAZyme analysis of the Fsh102 *A. sydowii* in comparison to selected *Aspergillus* strains and some other ascomycetes can be found in the **Supplementary Table 6**. In general, it can be stated that the set of extracellular enzymes predicted in Fsh102 was higher compared to the eight *Aspergillus* strains from the database. In the Fsh102 genome, 640 genes coding for different CAZymes were found compared to 612 in the *A. sydowii* reference strain (CBS 593.65). The lowest number was identified for *Aspergillus clavatus* NRRL1 with 384, while the other species rank between 522, for the *Aspergillus aculeatus* ATCC1687, and 597 for *Aspergillus flavus* NRRL3357. The highest number was identified in a *Fusarium oxysporum* strain with 777 putative CAZymes, while *Fusarium graminearum* PH_1 harbors 533 and *Magnaporthe grisea* DS9461 575, respectively. A comparison of Fsh102 and the *A. sydowii* strain sequenced by JGI showed a similar distribution of the CAZymes, reflecting their close phylogenetic relationship. For the Fsh102 strain 310 GHs and 35 GH with additional carbohydrate binding modules (CBMs) were found, while in the JGI strain 297 GHs and 29 GHs with CBMs were present. A difference between both *A. sydowii* strains was found in the GH18 family, which harbors different chitin degrading enzymes, such as chitinases or endo-β-N-acetylglucosaminidases. Enzymes of this family are represented by 20 genes in Fsh102, whereas only 11 GH18 coding genes have been found in the JGI reference strain. Similar differences are found in the GH18 family with additional CBMs, where 13 are predicted in the Fsh102 and only eight in the JGI strain, respectively. In addition, 25 and 29 members of the GH3, family, which consists mostly of exo-acting β-D-glucosidases were identified in Fsh102 and the JGI reference strain, respectively. The Fsh102 has two putative of GH11 candidates, while the JGI strain has three GH11 and both have two GH10 proteins.

Comparisons of the Xylanases Inside the *Aspergillus* Family

Sequence analysis of xylanase I indicated a membership to GH family 11, while xylanase II belongs to GH family 10. In total, 46 sequences were used for both the protein tress of the GH10 and GH11 families and both families form three subclades (I to III; **Figures 2B,C**). When considering the protein sequence, both Fsh102 xylanases are very similar to *A. sydowii* CBS 593.65 gene Aspsy1_63902 (xylanase II) or the JGI reference strain gene Aspsy1_168372 (xylanase I). They have a score of 0.004 for both enzymes using the cost matrix Blosum62, indicating that these enzymes are very similar but not identical to the ones in *A. sydowii* CBS 593.65. The differences are 0.45% for xylanase I or 0.3% for xylanase II, when directly compared to the corresponding proteins of the *A. sydowii* CBS 593.65. This difference is higher than the SNPs rate of 0.267%. Xylanase II seems to fall in the same subclade (II), as the nearest crystallized protein, 1TA3 from *A. nidulans*. Xylanase I is found in a different subclade (III) then their nearest crystallized protein 1TE1 from *P. funiculosum* which belongs to subclade II, both according to the PDB. Furthermore, the GH10 candidates of the Fsh102 are more similar than the GH11 candidates, as both proteins of the GH10 belongs to the same subclade II, while the GH11 do not (subclade II and III).

Glycosylation Prediction

The prediction of the glycosylation of xylanase I indicates that the enzyme is N-glycosylated at one site and O-glycosylated at two positions. For xylanase II, two N-glycosylations and two O-glycosylations, both O-glycosylations within the sequence of the signal peptide, were predicted (**Supplementary Figure 11**).

Cloning of the Xylanase Genes and Purification of the Two Endo-Xylanases

The ORF of xylanase I has a length of 729 bp and is organized in two exons found on contig 108. The ORF of xylanase II is considerably longer, consisting of 1,541 bp, which are organized in 11 exons, found on contig 55 (**Supplementary Table 7**). Xylanase I has a putative 19 amino acid long signal peptide, which was predicted from sequence analysis using the SignalP 4.1 Server (Petersen et al., 2011; **Supplementary Figure 3**). Not so obvious was a putative 21 amino acid long signal peptide for the xylanase II, because the S-score, indicating the end of the putative signal peptide after position 20, whereas the Y-score, which indicate the cleave position, was below the given threshold (Nielsen, 2017; **Supplementary Figure 4**). The expression of both enzymes including their putative signal

peptides resulted in the production of inactive proteins. Therefore, FastCloning (Li et al., 2011) was used to remove the putative signal peptides from both enzymes and expression was performed using an inducible system. The specific activity of the crude extract of xylanase I was 33.7 U/mg, while the elution fraction 1 had a specific activity of 181.1 U/mg, resulting in 55% yield and a purification factor of 5.4 (**Supplementary Table 8**). For xylanase II, the activity of the crude extract was 57.1 U/mg, already almost twice as high as xylanase I crude extract. The specific activity of the elution fraction 1 of xylanase II was 106.1 U/mg resulting in a yield of 27% and a purification factor of about two, but the elution fraction 2 corresponded to a specific activity of 121.5 U/mg. Further analysis was performed using the elution fraction 1 of both enzymes. In **Supplementary Figure 5**, the electrophoretic separation of different protein fractions is displayed, showing that a large portion of the total protein is found in the insoluble part for both proteins. Additionally, it was observed that the binding of xylanase I to the Ni-NTA agarose was tighter, if compared to xylanase II, as about 60% of the xylanase II are lost during cleaning procedure (flow through and wash fraction), while only about 36% of xylanase I was lost under the same conditions. The heterologous expressed xylanases were verified by liquid chromatography with coupled mass spectrometry (LC/MS-MS) of the corresponding protein bands from the SDS gels (**Supplementary Figure 5**). For xylanase I 47.68% (three different peptides) and for xylanase II 48.83% (five different peptides) of their total sequences were found (**Supplementary Figure 9**). These prove the expression of the targeted enzymes.

Substrate Specificity of Xylanase I and Xylanase II

Both xylanases showed high specificity against xylan compared to other substrates (**Supplementary Figure 6A**). Other hemicellulose-related substrates such as arabinoxylan or arabinan were hardly converted (less than 8% relative activity) and similar weak activities were observed using starch, galactan, or carboxymethylcellulose as substrate. Moreover, the enzymes showed little activity against xylopentose, 21.5 ± 11.0% for xylanase I and 13.5 ± 1.3% for xylanase II, and xylohexose with relative activities ranging from 16 ± 5.4 to 15 ± 1.0% for xylanase I and xylanase II, respectively. Interestingly, xylanase II showed low (12.4 ± 1.4%) but significant activity against xylotriose, but no activity in case of xylobiose or xylotetraose. Xylanase I was not able to hydrolyze XOS shorter than xylopentose or their conversion was below detection limit.

For both enzymes, we found that higher substrate concentrations resulted in higher specific activities (**Supplementary Figure 6B**), with an even more pronounced effect for xylanase II. The enzyme xylanase I showed about 12-fold higher activity at 3% xylan compared to 0.1% xylan, whereas xylanase II increased its activity roughly 8-fold. The K_M values are 20.68 ± 8.53 and 6.71 ± 3.14 g/L for xylanase I or xylanase II, respectively, while V_{max} are 3.12 ± 0.74 and 0.94 ± 0.16 mM/min. Turnover number, k_{cat}, was found to be 439.98 s^{-1} for xylanase I and 199.44 s^{-1} for xylanase II, resulting in an catalytic efficiency, k_{cat}/K_M, of 21.28 $L^*g^{-1}*sec^{-1}$ and 29.72 $L^*g^{-1}*sec^{-1}$.

Influence of Temperature and pH on Xylanase Activities

Xylanase I was completely functional in the range of 30–50°C (100 ± 11 to 109 ± 12%) and retained roughly half of its activity at 70°C (47 ± 3%; **Supplementary Figure 7A**). Xylanase II almost doubled its activity at 50°C compared to 30°C, and retained 27 ± 1% activity at 70°C (**Supplementary Figure 7A**). The best assay temperature would be 50°C, nevertheless for better comparability 30°C was used. The optimal pH for xylanase I activity was 4.8, and this enzyme showed more than 70% relative activity in a mildly acidic to neutral pH range (pH 4.5–7; **Supplementary Figure 7B**). Xylanase II activity at pH 4.8 was set to 100% because all measurements were performed at pH 4.8, but highest activity were detected at pH 7.5 with 123 ± 12%. This enzyme has also a much broader range, from 4.5 to 8.5, at which activity remained higher than 85% (**Supplementary Figure 7B**).

Temperature, pH, and Solvent Stability of the Xylanases

Storing at 4°C did not affect the activity of xylanase I, even after 8 days. If the enzyme was stored at 30°C only a small impact on activity was observed after the third day (>80%; **Figure 3A**). The enzyme showed 10–20% residual activity after heating it at 50 or 70°C for 15 min. Xylanase II maintained clearly more than 85% activity after incubation at 4, 30, or 50°C for 2 h. Heating the enzyme at 70°C for 2 h resulted in 15% residual activity. After prolonged incubation of 24 h at 50°C, the remaining activity was 40%, while storing at 4 or 30°C more than 85% of the activity was preserved for up to 72 h (**Figure 3B**). Additionally, the stability in two buffers (pH 4.8 and pH 9) at 4°C was investigated. At pH 4.8, the activity of both xylanases was stable for 2 h and then gradually decreasing to about 20% after 192 h (8 days; **Figure 3C**). Both enzymes are more stable, if they are stored at pH 9, displaying 80–100% activity for up to 72 h (**Figure 3D**). The activity of xylanase II was decreasing after that time point, while xylanase I activity was stable (at about 100%) until the end of the measurement (192 h).

The impact of salt ions and modifying reagents on enzyme activity were determined, too (**Figures 3E,F**). Concentrations up to 5% (v/v) of the different organic solvents used in this study were not affecting xylanase I activity (80–110% residual activity), while 30% (v/v) of the solvents decrease its activity to 10, 40, 40, and 30% for ACN, EtOH, MeOH, and DMSO, respectively. Low concentrations of NaCl (10 mM) and $MgCl_2$ (5 mM) did not impair activity, while high concentrations (1 or 0.1 M, respectively) resulted in an activity loss of up to 30%. Almost no effect was observed when using up to 5% (v/v) Tween 20 or 5% (v/v) Triton-X 100. High concentrations of EDTA (50 mM) inhibited the activity as SDS did (**Figure 3E**). The enzyme activity of xylanase II displayed also a high stability

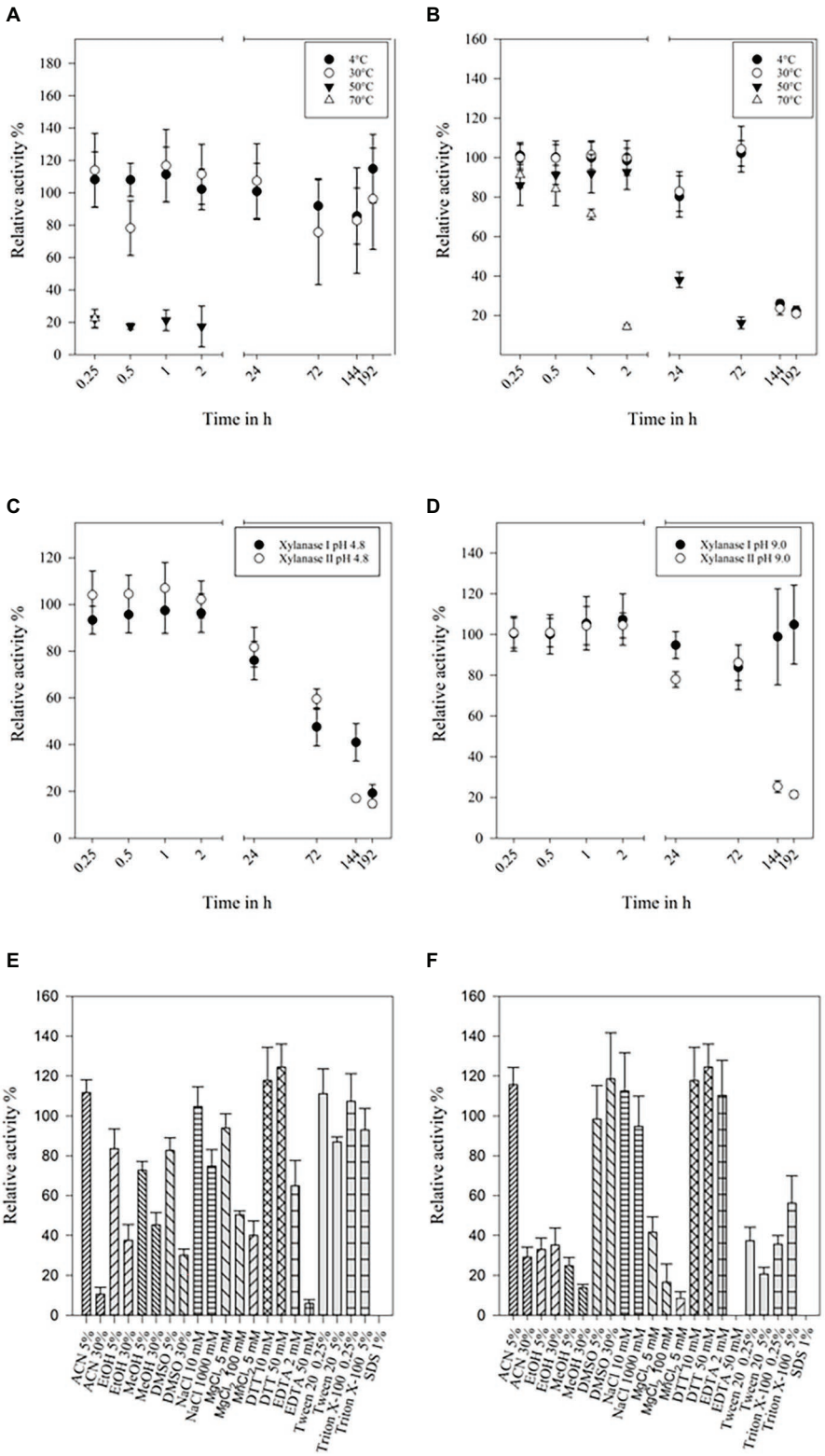


FIGURE 3 | Temperature stability of xylanase I **(A)** and xylanase II **(B)**, buffer stability of xylanase I and xylanase II at pH-values 4.8 **(C)** and 9.0 **(D)**. **(A)** Relative activity of xylanase I at 4, 30, 50, and 70°C over a time of 192 h with 30°C set to 100%, **(B)** relative activity of xylanase II at 4, 30, 50, and 70°C over a time of 192 h with 30°C set to 100%, **(C)** relative activities of xylanase I and xylanase II stored in citrate buffer pH 4.8 at 4°C over a time of 192 h with 0.25 h are set to 100%, and

(Continued)

FIGURE 3 | (D) relative activities of xylanase I and xylanase II stored in glycine NaOH buffer pH 9.0 at 4°C over a time of 192 h with 0.25 h set to 100%. Relative activity of the xylanase I **(E)** and xylanase II **(F)** applying solvents and salts. The following solvents and salts were used: 5 and 30% ACN, 5 and 30% EtOH, 5 and 30% MeOH, 5 and 30% DMSO, 10 and 1,000 mM NaCl, 5 and 100 mM $MgCl_2$, 10 and 50 mM DTT, 2 and 50 mM EDTA, 0.25 and 5% Tween 20, 0.25 and 5% Triton X-100, and 1% SDS. One hundred percent are set applying no salts or solvents.

in different solvents, such as 5% ACN, 5% or 30% DMSO, or different concentration of additives like NaCl (10 or 1,000 mM), DTT (10 or 50 mM), and ETDA (2 mM), did not affect the enzymes activity (**Figure 3F**).

Analysis of Hydrolyzed Products

In order to analyze the reaction products of the depolymerization of xylan by both xylanases, we used TLC on a silica gel plate and stopped the reaction at different time points (**Figure 4**). The main products of the hydrolysis of xylan by xylanase I (**Figure 4A**) were xylobiose, xylotriose, and xylohexose, which are visible after 15 min of incubation. Additional spots between the starting point and the spot for the xylohexose might derive from higher polymerized XOS. During the course of the conversion, the product profile changed slightly, visible by a small spot corresponding to xylose after 17 h. Interestingly, xylanase II generated a different reaction profile (**Figure 4B**). Directly after the start of the reaction the monosaccharide xylose can be detected, as well as xylobiose, xylotriose, xylopentose, and some higher polymerized XOS. Interestingly, the reaction of xylanase II does not yield xylohexose, and the spot corresponding to xylotriose is disappearing over the course of the reaction. Xylotetraose is neither in the enzymatic reaction of xylanase I nor xylanase II observable, indicating that either no xylotetraose is formed during the reaction or a very rapid turnover of this tetrameric sugar takes place. Due to the determination of the substrate spectra, it is unlikely that a very rapid turnover of this sugar tetramer is performed. These findings underlining the need for cooperation between the two non-redundant xylanases for an efficient degradation.

Active Site Variants of Xylanase I and Xylanase II

An acid-based catalyzed mechanism is assumed for both enzymes. One residue is acting as a nucleophile, while the other serves as a general acid/base residue. This activity follows a Koshland double-displacement mechanism (Withers et al., 1986; Gebler et al., 1992). To verify this theory, we mutated the identified residues and generated two single mutants and one double mutant for each enzyme. As a replacement for the two glutamate residues, methionine was chosen, based on its similar steric and spatial needs. In the case of xylanase I, the glutamic acid on positions 97 and 188 and for xylanase II the glutamic acids on the positions 155 and 241 were replaced by methionine. For both enzymes, two single and one double mutant were created. The enzyme xylanase I is predicted to exhibit a typical β-jelly roll fold (**Figure 5A**), while xylanase II has a TIM barrel-$(\beta/\alpha)_8$ fold (**Figure 5B**), which is in line with the GH family and GH-A clan assignment. The active sites of the enzymes are depicted in the magnified section of **Figure 5** (right part). For the xylanase I the distance between the active site glutamic acids is 3.7 and 4.7 Å and for xylanase II we observed 3.0 and 5.4 Å. The activities of the mutated xylanase I and xylanase II were determined to verify that the active site residues were matched correctly. The xylanase I variant E87M displayed 4 ± 0.5%, the variant E188M 13 ± 4.0% and the double

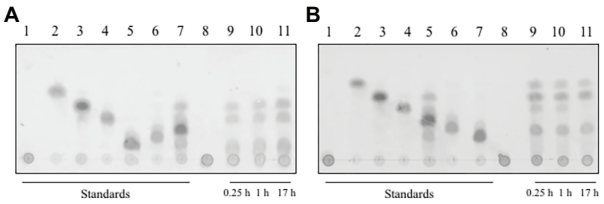


FIGURE 4 | Thin layer chromatography of the reaction products of **(A)** xylanase I and **(B)** xylanase II on a silica gel plate, with a running time of 40 min in organic solvent. The plate was incubated with 0.2% (w/v) oricin at 80°C for 40 min. The produced sugars are visualized as dark spots. **(A)** Lane 1: xylan, lane 2: xylose, lane 3: xylobiose, lane 4: xylotriose, lane 5: xylohexose, lane 6: xylopentose, lane 7: xylotetraose, lane 8: xylanase I + xylan after 0 h incubation, lane 9: xylanase I + xylan after 0.25 h incubation, lane 10: xylanase I + xylan after 1 h incubation, and lane 11; xylanase I + xylan after 17 h incubation. **(B)** Lane 1: xylan, lane 2: xylose, lane 3: xylobiose, lane 4: xylotriose, lane 5: xylotetraose, lane 6: xylopentose, lane 7: xylohexose, lane 8: xylanase I + xylan after 0 h incubation, lane 9: xylanase I + xylan after 0.25 h incubation, lane 10: xylanase I + after 1 h incubation, and lane 11: xylanase I + xylan after 17 h incubation.

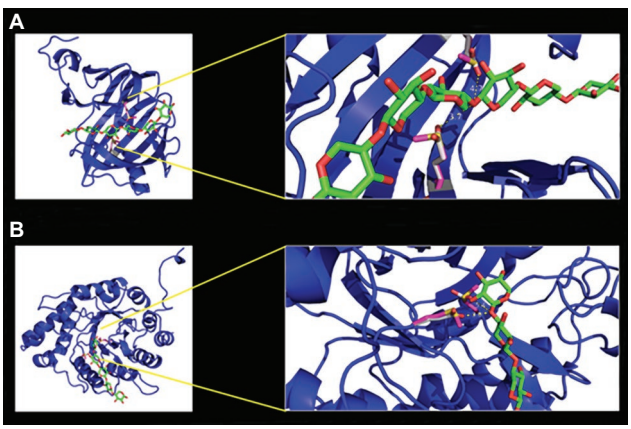


FIGURE 5 | Homology models of xylanase I **(A)** and xylanase II **(B)** with enlargement of the active center. The models were calculated with the program YASARA using the protein database PDB. The models were displayed with the program PyMol. The enzyme backbones are colored in blue, the xylooligosaccharides (XOS) carbon atoms in green, oxygen in red, active site residues glutamic acid carbon atoms in gray, methionine carbon atoms of the variants in pink, and in yellow is the sulfur.

mutant E97M/E188M 4.0 ± 0.5% relative activity (**Supplementary Figure 8A**). The xylanase II variant E155M displays 9.0 ± 2.4%, the variant E241M 7.0 ± 0.6% and the double mutant E155M/E241M 10.0 ± 0.9% relative activity (**Supplementary Figure 8B**). The purification of different heterologous expressed variants can be found in **Supplementary Figure 10**.

DISCUSSION

In the present study, we report the assignment of Fsh102, DSMZ 105790, to the genus *Aspergillus* of the section Nidulantes and classify it as *A. sydowii*, by comparing our genomic data with available genomic information of the *A. sydowii* CBS 593.65, as well as an ITS based sequences analysis with other *Aspergillus* spp.

In comparison to our data, similar coverage and identity were found by a report of two strains of *A. carbonarius*, which had a sequence similarity higher than 97.2% and a coverage rate of more than the 84% compared to the non-repetitive reference genome (Cabañes et al., 2015). The genome size of Fsh102, with approximately 34 Mbp, is in the range of *Aspergillus* species (29–37 Mbp) and most of the genomes of Eurotiales reported to date are in that range (Yang et al., 2016). Especially when comparing the *A. sydowii* genome the section Nidulantes strains, like *A. nidulans* with 30.48 Mbp or *Aspergillus versicolor* with 33.13 Mbp, a slightly larger genome size become obvious. One possible explanation for the large genome size may be the variability of *A. sydowii* fungus to grow pathogenic on animals like corals or humans, as well as to live saprophytic on dead plant matter (Alker et al., 2001; Hallegraeff et al., 2014; Liu et al., 2019). A close relative, *A. carbonarius* ITEM 5010 has a genome size of 36.3 Mbp, making it slightly larger than the *A. sydowii* genome CBS 593.65 genome. Next to the size, the frequency of SNPs and InDels in a genome is an interesting marker. By comparing different *A. carbonarius* strains with our reported analysis of Fsh102, which is based on the *A. sydowii* reference genome, the total number of SNPs was found to be lower (52,661), while the number of InDels was increased (7,567) if the *A. carbonarius* strains were compared to the *A. sydowii* strain analysis.

The wild type extracellular xylan degrading activity of the fungal strain Fsh102 is comparable to published *A. sydowii* SBS 45 (Nair et al., 2008). Other fungi are able to generated similar or higher xylanase activity, like a *F. oxysporum* isolate with 0.41 U/mg (Christakopoulos et al., 1996) or *Trichoderma viride* Fd18, which was reported to be 1.80 U/mg (Ja'afaru, 2013), but the wild type activity of our strain is very comparable to the other reported fungi.

By analyzing the CAZyme prediction in more detail, it was found that Fsh102 is more similar to *A. sydowii* CBS 593.65, while that the total number of CAZyme genes is higher, if compared to other species of this genus. This underlines the general hypothesis that *A. sydowii* needs a broad portfolio of extra-cellular enzymes for its variable lifestyle. These differences are most likely an adaptation to the habitat, while the origin of the JGI *A. sydowii* is not known, Fsh102 was isolated from shrimp shells, which is mirrored in the distribution of the CAZymes. As can be seen in the higher number of chitin degradation enzymes of the Fsh102, which was isolated from aquatic source, compared to the *A. sydowii* reference strain, which has more enzymes responsible for the degradation of (hemi-) cellulose, so it could be speculated that the CBS strain originated from a terrestrial source. Until now, xylanase activities were found in six GH families (5, 7, 8, 10, 11, and 43), but the main focus of biotechnological research is centered on two of the xylanase containing GH families 10 and 11 (Collins et al., 2005). The exo-xylanase (GH3) was not investigated in this work, as we focused on the first steps of xylan degradation. The analyzed xylanase I belongs to the GH11 family, while xylanase II is an enzyme of the GH10 family. The JGI strain has two GH10 and three GH11 enzymes, Fsh102 encodes two members of both. It is striking that only one of these two GHs was found in the zymogram using Fsh102 supernatant, as activities in both families are predominantly related to xylan depolymerization. The lack of observed enzymes of GH10 or GH11 family members can have different reasons. The four most likely are

I. The different members are induced differently dependent on growth conditions like various kinds of xylan containing substrates or temperatures. This behavior was observed for *Aspergillus niger* E-1, which encodes different xylanase (GH10 and GH11) in the genome, while only three of them were identified by MS if this fungus was grown on beech wood xylan (Takahashi et al., 2013).
II. The members of the GH10 and GH11 family may have different enzymatic activities, but this is unlikely because both classes are extensively studied. There are, at least, 127 GH10 enzymes and 173 GH11 enzymes with a demonstrated activity on xylan described in the literature and only in the case of GH10 different activities were reported. In the GH10 family also endo-β-1,4-xylanases (EC 3.2.1.8), endo-1,3-β-xylanases (EC 3.2.1.32), and cellobiohydrolases (EC 3.2.1.91) are known, while GH11 members are considered to be monospecific (Davies and Henrissat, 1995; Paës et al., 2012; Chakdar et al., 2016).
III. The peptides generated by the tryptic digestion were not ionisable. Although this is unlikely, as ESI method is capable of ionizing the majority of peptides and the MS/MS device has a very high mass resolution thereby lowering the detection limit further, it cannot be ruled out.
IV. The variation within the GH protein families is high in general. Different members of the GH10 subclass belong to different subclade of the protein trees (**Figure 2C**) and the GH11 member (FSH102_2) showed more similarity toward an enzyme from *A. versicolor* (**Figure 2B**) than the JGI counterpart. It can be speculated that the predicted GH10 and GH11 enzymes have with different substrate specificities, because this part of the protein tree is more widely dispersed.

To the best of your knowledge, this report describes for the first time that enzymes of the *A. sydowii* were heterologous

expressed in high yield, by using a prokaryotic *E. coli* system. An activity of xylanase I and xylanase II with 181.1 and 121.5 U/mg was obtained in their N-terminal truncated versions, while only insoluble proteins were formed, if expressed with the corresponding fungal N-terminal secretion signal. Frequently, more complex eukaryotic expression systems, like *Komagataella phaffii* (formerly known as *Pichia pastoris*) or *Yarrowia lipolytica*, have to be used to express eukaryotic xylanases from *A. niger*, *Trichioderma harzianum*, *Hypocrea orientalis*, or *Trichioderma reesei* (He et al., 2009; Wang et al., 2014; Li et al., 2018) as they have the advantage of correct N- and O-glycosylation. The glycosylation can have different effects like enhancement of solubility (Solá and Griebenow, 2009), increased activity, as shown for the endo-cellulase IIa from *Penicillium verrucosum* (Dotsenko et al., 2016), or enhanced stability (Sarkar and Wintrode, 2011). On the other hand, there are a number of reports showing the expression of xylanases in *E. coli* (Basaran et al., 2001; Le et al., 2011) but having the disadvantage of either being periplasm bound to the *E. coli* or having low activity (about 30 U/mg). The two presented xylanases exhibit comparable activity to other heterologous expressed xylanases. Two reports from *Komagataella phaffii* expression studies yielded 175 U/mg from a xylanase originating from an *A. niger* strain (Liu et al., 2006) or 123 U/mg from a strain of *Bacillus licheniformis* (Liu and Liu, 2008), but there are also studies reporting exceptional high enzymatic activity like a *T. reesei* xylanase with 746 U/mg (He et al., 2009). It, therefore, cannot be ruled out that the activity, of the xylanases reported in this study, could be much higher if a eukaryotic expression system would have been used. The xylanase from previously mentioned *T. reesei* had about 12-fold reduced activity if expression was performed in *E. coli* instead of *P. pastoris* (He et al., 2009).

In the publication of Nair et al. (2008), two xylanases of a crude culture filtrate of *A. sydowii* SBS 45 were partially characterized, but no heterologous expression was performed. Comparing the molecular weights of the enzymes with our study, a good match can be found for both of the heterologous expressed xylanases of the *A. sydowii* Fsh102. Xylanase I, reported in this study, shows a clearing zone in the zymogram in the region of 15 and 25 kDa (**Figure 2A** Band 4), which is similar to the xylanase I of SBS 45 which has a molecular weight of 20.1 kDa. For xylanase II of Fsh102, a weight of 37.1 kDa was calculated (**Supplementary Figure 5B**, 39.2 kDa, including signal peptide), but the zymogram showed a clearing zone in the region of 45 kDa (**Figure 2A** Band 3). This shift can be explained by the glycosylation of the enzyme (2 N-glycosylations shown in **Supplementary Figure 11**). The xylanase II from *A. sydowii* SBS 45 was reported with 43 kDa based on an SDS-PAGE analysis, displaying a similar observed size as xylanase II of Fsh102. The purification quotient of 5.41 for xylanase I and 1.96 for xylanase II is comparable to the purification quotient of 5.3 and 1.9 for two xylanases of *Trichoderma inhamatum* (Silva et al., 2015). To increase the protein yield during the purification process, it is possible to use an extended His_6-tag, a sequence extended by six histidine molecules. This led to a better binding of the neurotensin membrane receptor to the column and thus to a better purification performance of the protein (Grisshammer and Tucker, 1997).

The substrate specificity study of the purified enzymes reveals that both enzymes have next to their xylanolytic activity, a weak promiscuity. An activity of around 8%, compared to the xylanase activity, was observed for the action on other (bio) polymers like cellulose, identified by carboxymethylcellulose as substrate or arabinan, which consists of about 70% arabinose α-(1–5) linked with 20% galactose and other sugars (**Supplementary Figure 6**). Generally, xylanases are very specific as shown for the *T. reesei* strain Rut C-30 endo-xylanase Xyn2 of the GH family 11 (He et al., 2009), which has about 1% activity for CMC and the same was shown for a GH10 xylanase from *Caldicellulosiruptor bescii* (An et al., 2015). This weak but existing promiscuity of our xylanases is advantageous, because the use of our xylanases could reduce the need to add other enzymes in a rye-based diet, in food preparation or during the treatment of agricultural wastewater, which both contains different biopolymers. Often, higher xylanase activities are observed if higher xylan concentration were applied during the fermentation by wild type organism like for *Anoxybacillus kamchatkensis* (Yadav et al., 2018). For a better comparison, 0.1% substrate concentration was used for almost all experiments, but higher concentrations yielded higher activities. Around 12- and 7-fold higher activities were observed for xylanase I and xylanase II, if 3% substrate was used. Based on the data available, it can be assumed that for xylanase I, a substrate concentration of 2.5% provides the highest activities, while for xylanase II no possible substrate inhibition was observed in the measured range. Similar behavior could be observed for the heterologous expressed XynDZ5 GH10 xylanase from *Thermoanaerobacterium* sp., where three different xylan sources were used, showing about 10-fold increase of activity if 30 g/L (3%) was used instead 0.3% (Zarafeta et al., 2020). Substrate affinity of xylanase II 6.71 ± 3.14 g/L was comparable to literature values of 0.7–6.6 g/L (Li et al., 2006, 2010; Sharma et al., 2010; Ding et al., 2018; Yadav et al., 2018), while xylanase I showed a reduced affinity with a K_M of 20.68 ± 8.53 g/L. Although the substrate affinity of xylanase I is reduced, the maximal velocity 3.12 ± 0.74 in mM/min was quite good compared to literature values of 0.067, 1.29, 1.64, and 9.17mM/min. Xylanase II exhibits a maximal velocity of only 0.94 ± 0.16 mM/min, which is also reflected in a relative low turnover number of around 200 per second, while literature values reach up to 7,000 per second. These numbers depend entirely on the substrate used and others substrates might have increased the values what would be especially relevant for the catalytic efficiency, which was found to be between 20 and 30 for xylanase I and II. Here, literature values of around 150–1,000 show an optimization potential for the identified xylanases (Li et al., 2006, 2010; Sharma et al., 2010; Ding et al., 2018).

When comparing the cleavage products of both xylanases, it is noticeable that both endo-xylanases are able to cleave short XOS. Both enzymes have 10–20% activity comparing to birch wood xylan, when substrates with a chain length

from five xylose units to six units are used. The enzyme xylanase II has a rather uncommon exception; it is able to degrade xylotriose with 9% activity but not xylotetraose. While other researchers found that, the shortest XOS converted by GH10 or GH 11 xylanases are the xylotriose or xylotetraose longer XOS (**Supplementary Table 10**; Kolenová et al., 2006; Takahashi et al., 2013; Li et al., 2018; Han et al., 2019).

Both xylanases have mesophilic temperature preferences and show an optimum temperature at 50°C (**Supplementary Figure 7A**). In addition to their broad temperature profile, both enzymes have a broad pH range, xylanase I 4.5–7.0 and xylanase II 4.5–8.0 (**Supplementary Figure 7B**). This is similar to different endo-xylanases from other *Aspergillus* strains which have their maximal activities between 42 and 60°C, and a pH range of 4.0–7.0 (Teixeira et al., 2010), or more specifically with *A. sydowii* SBS 45 (Nair et al., 2008) or a GH10 xylanase of *A. niger* E-1 (Takahashi et al., 2013) both having an optimum between 50 and 55°C. The enzymes from Fsh102 show moderate to good activities at 10°C, granting the possibility to degrade xylan at low temperatures for food applications in fruit juice clarification, as it is nowadays mostly performed at mesophilic temperatures between 27 and 43°C (Sharma et al., 2017).

When comparing the temperature stability of both xylanases, it is noticeable that both xylanases have a high stability for up to at least 3 days at lower temperatures. Especially, xylanase II has also a good stability at high temperatures making it competitive toward some of the best xylanases identified so far (**Supplementary Table 9**). Similar is true when comparing pH stabilities although long time data are scarcely reported (**Supplementary Table 9**).

The temperature stability results are in line with the investigations of the effect of additional salts and solvents to the enzymatic reaction. In general, xylanase I is more resilient than xylanase II. By comparing the salt tolerance to GH10 xylanases from *A. niger* strain C3486, similarities to the studied xylanases were found. In both cases, low concentrations of Mg^{2+} ions cause no inhibition, but low Mn^{2+} concentrations do (Yang et al., 2010). Interestingly, Mn^{2+} ions are not inhibiting the two *A. sydowii* SBS 45 xylanases, but EDTA is reducing the activity to only 9 and 30%, while both Fsh102 xylanases are more tolerant to EDTA retaining 70 to 110% activity, respectively. The SBS 45 xylanase 2 is moderately inhibited by SDS to 30%, while the Fsh102 xylanase II did not show any residual activity after SDS addition. Reducing agents, like DTT (for Fsh102) or β-Mercaptoethanol (for SBS 45 xylanases), do not affect the enzymes as no disulfide bonds are present. By comparing the enzymes to a xylanase from *A. niger* DSM 1957 (Do et al., 2009), which is stated to be resistant to methanol, ethanol, isopropanol, and acetone up to 30 (v/v)%, or to *Aspergillus awamori* VTCC-F312 xylanase with has a residual activity of 63–86% by 30 (v/v)% solvents (Do et al., 2012), the Fsh102 xylanases can be classified as moderate solvent stable, which is necessary for the use some pulp and paper processing like the organosolve process (Brosse et al., 2019).

The enzymatic reaction for both xylanases is performed using a nucleophile and acid-base catalyst, as two glutamate residues are essential as shown by mutation analysis performed in the presented study. All single and double mutants had a residual activity between 5 and 10%, which is relatively high, if compared to other studies, which are mostly performed with bacterial xylanases like *Streptomyces lividans* or *Bacillus circulans*, whereas the reduction is higher than 10^3 with up to no xylanolytic activity (Ko et al., 1992; Moreau et al., 1994; Wakarchuk et al., 1994). It might be the case that acidic residues structurally close to the catalytic active residues can replace those originally catalytic active residues to perform hydrolysis, although with lower efficiency. If this taken into account, in a reverse study, were amino acids in 12 Å distance from either of the two catalytic amino acids were replaced, resulting in about 12% residual activity for the worst variant, gives one possible hint that structural close residues to the active site residues have an influence on the catalysis (Bai et al., 2015). A similar argument is given by the fact that in a xylanase originating from *Bispora* sp. MEY-1, two non-polar amino acids near the active site residues were mutated to acidic residues resulted in an increase of the specific activity up to 1.3 fold (Wang et al., 2017).

CONCLUSION

In this study, we trace the whole biotechnology pipeline, from strain isolation, activity-based identification of the enzymes by mass spectrometry, heterologous expression to characterization of two xylanase from *Aspergillus sydowii*. Due to their differences in activity, hydrolysis products and stability, these enzymes are non-redundant xylanases and their characteristics provide an interesting starting point for directed evolution. Those robust enzymes, are tolerant toward mesophilic temperatures, acting in a broad pH range and are stable toward organic solvents, detergents and salts, which granting the possibility for food applications in fruit juice clarification or in the bakery or the pulp and paper processing industry.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SB performed DNA isolation, ITS-PCR, BLAST ITS sequences, plate-based screening, and pre cultivation of the fungi. BE measured the activities of the supernantants. TN collected the fungal sample. RH analyzed the ITS-data and generated the phylogentic tree and made the CAZyme analysis. MR annotated the Fsh102 genes. SH and HS performed and analysis the mass spectrometry. WS started the project.

MG organized the lab work, coordinated the progress between the different institutions, and evaluated the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the German federal ministry BMBF under contract no. 03A0150B within the framework of Bioeconomy International 2014.

ACKNOWLEDGMENTS

We thank Lea Vaas and Hévila Brognaro for valuable discussion and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/article/10.3389/fmicb.2020.573482/full#supplementary-material

REFERENCES

Agbor, V., Blunt, W., Cicek, N., Sparling, R., Berlin, A., and Levin, D. B. (2011). Processing of cellulosic feedstocks for ethanol and hydrogen production. *J-FOR* 1, 54–61.

Alker, A. P., Smith, G. W., and Kim, K. (2001). Characterization of *Aspergillus sydowii* (Thom et Church), a fungal pathogen of Caribbean sea fan corals. *Hydrobiologia* 460, 105–111. doi: 10.1023/A:1013145524136

An, J., Xie, Y., Zhang, Y., Tian, D. S., Wang, S. H., Yang, G. Y., et al. (2015). Characterization of a thermostable, specific GH10 xylanase from *Caldicellulosiruptor bescii* with high catalytic activity. *J. Mol. Catal. B Enzym.* 117, 13–20. doi: 10.1016/j.molcatb.2015.04.003

Bai, W. Q., Xue, Y. F., Zhou, C., and Ma, Y. H. (2015). Cloning, expression, and characterization of a novel alkali-tolerant xylanase from alkaliphilic *Bacillus* sp SN5. *Biotechnol. Appl. Biochem.* 62, 208–217. doi: 10.1002/bab.1265

Bainier, G., and Sartory, A. (1913). Étude d'une espèce nouvelle de Sterigmatocystis, *Sterigmatocystis sydowi* (n.sp.). *Ann. Mycol.* 11, 25–29.

Bajpai, P. (ed.) (2014). "Chapter 2-xylan: occurrence and structure" in *Xylanolytic enzymes*. Amsterdam: Academic Press, 9–18.

Basaran, P., Hang, Y. D., Basaran, N., and Worobo, R. W. (2001). Cloning and heterologous expression of xylanase from *Pichia stipitis* in *Escherichia coli*. *J. Appl. Microbiol.* 90, 248–255. doi: 10.1046/j.1365-2672.2001.01237.x

Bedford, M. R., Classen, H. L., and Campbell, G. L. (1991). The effect of pelleting, salt, and pentosanase on the viscosity of intestinal contents and the performance of broilers fed rye. *Poult. Sci.* 70, 1571–1577. doi: 10.3382/ps.0701571

Beg, Q. K., Kapoor, M., Mahajan, L., and Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol.* 56, 326–338. doi: 10.1007/s002530100704

Bennett, J. W. (1998). Mycotechnology: the role of fungi in biotechnology. *J. Biotechnol.* 66, 101–107. doi: 10.1016/S0168-1656(98)00133-3

Biango-Daniels, M. N., and Hodge, K. T. (2018). Sea salts as a potential source of food spoilage fungi. *Food Microbiol.* 69, 89–95. doi: 10.1016/j.fm.2017.07.020

Biely, P. (1985). Microbial xylanolytic systems. *Trends Biotechnol.* 3, 286–290. doi: 10.1016/0167-7799(85)90004-6

Blackwell, M. (2011). The fungi: 1, 2, 3 … 5.1 million species? *Am. J. Bot.* 98, 426–438. doi: 10.3732/ajb.1000298

Brandt, S. C., Ellinger, B., van Nguyen, T., Thi, Q. D., Van Nguyen, G., Baschien, C., et al. (2018). A unique fungal strain collection from Vietnam characterized for high performance degraders of bioecological important biopolymers and lipids. *PLoS One* 13:e0202695. doi: 10.1371/journal.pone.0202695

Brosse, N., Hussin, M. H., and Rahim, A. A. (2019). "Organosolv processes" in *Biorefineries*. eds. K. Wagemann and N. Tippkötter (Cham: Springer International Publishing), 153–176.

Butt, M. S., Tahir-Nadeem, M., Ahmad, Z., and Sultan, M. T. (2008). Xylanases and their applications in baking industry. *Food Technol. Biotechnol.* 46, 22–31.

Cabañes, F. J., Sanseverino, W., Castellá, G., Bragulat, M. R., Cigliano, R. A., and Sánchez, A. (2015). Rapid genome resequencing of an atoxigenic strain of *Aspergillus carbonarius*. *Sci. Rep.* 5:9086. doi: 10.1038/Srep09086

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663

Chakdar, H., Kumar, M., Pandiyan, K., Singh, A., Nanjappan, K., Kashyap, P. L., et al. (2016). Bacterial xylanases: biology to biotechnology. *3 Biotech* 6:150. doi: 10.1007/S13205-016-0457-Z

Chandra, R. P., Bura, R., Mabee, W. E., Berlin, A., Pan, X., and Saddler, J. N. (2007). "Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics?" in *Biofuels*. ed. L. Olsson (Berlin, Heidelberg: Springer Berlin Heidelberg), 67–93.

Christakopoulos, P., Nerinckx, W., Kekos, D., Macris, B., and Claeyssens, M. (1996). Purification and characterization of two low molecular mass alkaline xylanases from *Fusarium oxysporum* F3. *J. Biotechnol.* 51, 181–189. doi: 10.1016/0168-1656(96)01619-7

Collins, T., Gerday, C., and Feller, G. (2005). Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* 29, 3–23. doi: 10.1016/j.femsre.2004.06.005

CSH Protocols (2010). YPD media. *Cold Spring Harb. Protoc.* 2010:wpdb.rec12315. doi: 10.1101/pdb.rec12315

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859. doi: 10.1016/S0969-2126(01)00220-9

Davies, G. J., and Sinnott, M. L. (2008). Sorting the diverse: the sequence-based classifications of carbohydrate-active enzymes. *Biochemist* 30, 26–32. doi: 10.1042/BIO03004026

de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., et al. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* 18:28. doi: 10.1186/s13059-017-1151-0

Ding, C. H., Li, M. X., and Hu, Y. Q. (2018). High-activity production of xylanase by *Pichia stipitis*: purification, characterization, kinetic evaluation and xylooligosaccharides production. *Int. J. Biol. Macromol.* 117, 72–77. doi: 10.1016/j.ijbiomac.2018.05.128

Do, T. T., Dam, T. H., and Quyen, D. T. (2009). Purification and biophysical characterization of xylanase from *Aspergillus niger* DSM 1957. *Sci. Technol. J. Agric. Rural Dev.* 6, 16–21.

Do, T. T., Quyen, D. T., and Dam, T. H. (2012). Purification and characterization of an acid-stable and organic solvent-tolerant xylanase from *Aspergillus awamori* VTCC-F312. *Sci. Asia* 38, 157–165. doi: 10.2306/scienceasia1513-1874.2012.38.157

Dotsenko, A. S., Gusakov, A. V., Rozhkova, A. M., Sinitsyna, O. A., Nemashkalov, V. A., and Sinitsyn, A. P. (2016). Effect of N-linked glycosylation on the activity and other properties of recombinant endoglucanase IIa (Cel5A) from *Penicillium verruculosum*. *Protein Eng. Des. Sel.* 29, 495–501. doi: 10.1093/protein/gzw030

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

Ferreira, J. A., Mahboubi, A., Lennartsson, P. R., and Taherzadeh, M. J. (2016). Waste biorefineries using filamentous ascomycetes fungi: present status and future prospects. *Bioresour. Technol.* 215, 334–345. doi: 10.1016/j.biortech.2016.03.018

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., et al. (1992). Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases. *J. Biol. Chem.* 267, 12559–12561.

Geiser, D. M., Taylor, J. W., Ritchie, K. B., and Smith, G. W. (1998). Cause of sea fan death in the West Indies. *Nature* 394, 137–138. doi: 10.1038/28079

Gibbs, P. A., Seviour, R. J., and Schmid, F. (2000). Growth of filamentous fungi in submerged culture: problems and possible solutions. *Crit. Rev. Biotechnol.* 20, 17–48. doi: 10.1080/07388550091144177

Gilbert, H. J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.* 153, 444–455. doi: 10.1104/pp.110.156646

Global Partnership on Waste Management (2011). Waste agricultural biomass to energy (WAB2E). United Nations Environment Programme.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704. doi: 10.1093/nar/gkt1183

Grisshammer, R., and Tucker, J. (1997). Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expr. Purif.* 11, 53–60. doi: 10.1006/prep.1997.0766

Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for short reads alignment. *Curr. Protoc. Bioinformatics* 44, 11.11.11–11.11.17. doi: 10.1002/0471250953.bi1111s44

Guerriero, G., Hausman, J. -F., Strauss, J., Ertan, H., and Siddiqui, K. S. (2016). Lignocellulosic biomass: biosynthesis, degradation, and industrial utilization. *Eng. Life Sci.* 16, 1–16. doi: 10.1002/elsc.201400196

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gupta, R., Jung, E., and Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. *In preparation*. Available at: http://www.cbs.dtu.dk/services/NetNGlyc/ (Accessed May 13, 2018).

Hahnke, R. L., Stackebrandt, E., Meier-Kolthoff, J. P., Tindall, B. J., Huang, S. X., Rohde, M., et al. (2015). High quality draft genome sequence of *Flavobacterium rivuli* type strain WB 3.3-2(T) (DSM 21788(T)), a valuable source of polysaccharide decomposing enzymes. *Stand. Genomic Sci.* 10:46. doi: 10.1186/s40793-015-0032-y

Hallegraeff, G., Coman, F., Davies, C., Hayashi, A., McLeod, D., Slotwinski, A., et al. (2014). Australian dust storm associated with extensive *Aspergillus sydowii* fungal "bloom" in coastal waters. *Appl. Environ. Microbiol.* 80, 3315–3320. doi: 10.1128/AEM.04118-13

Han, Z. G., Shang-Guan, F., and Yang, J. K. (2019). Molecular and biochemical characterization of a bimodular xylanase from Marinifilaceae bacterium strain SPP2. *Front. Microbiol.* 10:1507. doi: 10.3389/fmicb.2019.01507

He, J., Yu, B., Zhang, K. Y., Ding, X. M., and Chen, D. W. (2009). Expression of endo-1, 4-beta-xylanase from *Trichoderma reesei* in *Pichia pastoris* and functional characterization of the produced enzyme. *BMC Biotechnol.* 9:56. doi: 10.1186/1472-6750-9-56

Henikoff, S., and Henikoff, J. G. (1992). Amino-acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10915–10919. doi: 10.1073/pnas.89.22.10915

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino-acid-sequence similarities. *Biochem. J.* 280, 309–316. doi: 10.1042/bj2800309

Holtzapple, M. T. (2003). "Hemicelluloses" in *Encyclopedia of food sciences and nutrition. 2nd Edn.* eds. B. Caballero, P. Finglas and L. Trugo (Oxford: Academic Press), 3060–3071.

Ja'afaru, M. I. (2013). Screening of fungi isolated from environmental samples for xylanase and cellulase production. *ISRN Microbiol.* 2013:283423. doi: 10.1155/2013/283423

Jukes, T., and Cantor, C. (1969). *Evolution of protein molecules*. New York: Academic Press, 22–132.

Kim, K., and Harvell, C. D. (2004). The rise and fall of a six-year coral-fungal epizootic. *Am. Nat.* 164, S52–S63. doi: 10.1086/424609

Klich, M. A. (ed.) (2002). "Introduction: economic and medical importance of Aspergillus" in *Identification of common Aspergillus species*. Utrecht: Centraalbureau voor Schimmelcultures, 1–16.

Ko, E. P., Akatsuka, H., Moriyama, H., Shinmyo, A., Hata, Y., Katsube, Y., et al. (1992). Site-directed mutagenesis at aspartate and glutamate residues of xylanase from *Bacillus pumilus*. *Biochem. J.* 288, 117–121. doi: 10.1042/bj2880117

Kolenová, K., Vrsanska, M., and Biely, P. (2006). Mode of action of endo-beta-1,4-xylanases of families 10 and 11 on acidic xylooligosaccharides. *J. Biotechnol.* 121, 338–345. doi: 10.1016/j.jbiotec.2005.08.001

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0

Le, Y. L., Peng, J. J., Wu, H. W., Sun, J. Z., and Shao, W. L. (2011). An approach to the production of soluble protein from a fungal gene encoding an aggregation-prone xylanase in *Escherichia coli*. *PLoS One* 6:e18489. doi: 10.1371/journal.pone.0018489

Leach, J., Lang, B. R., and Yoder, O. C. (1982). Methods for selection of mutants and *in vitro* culture of *Cochliobolus heterostrophus*. *J. Gen. Microbiol.* 128, 1719–1729. doi: 10.1099/00221287-128-8-1719

Li, X. T., She, Y. L., Sun, B. G., Song, H. L., Zhu, Y. P., Lv, Y. G., et al. (2010). Purification and characterization of a cellulase-free, thermostable xylanase from *Streptomyces rameus* L2001 and its biobleaching effect on wheat straw pulp. *Biochem. Eng. J.* 52, 71–78. doi: 10.1016/j.bej.2010.07.006

Li, L. T., Tian, H. M., Cheng, Y. Q., Jiang, Z. Q., and Yang, S. Q. (2006). Purification and characterization of a thermostable cellulase-free xylanase from the newly isolated *Paecilomyces themophila*. *Enzym. Microb. Technol.* 38, 780–787. doi: 10.1016/j.enzmictec.2005.08.007

Li, C. K., Wen, A. Y., Shen, B. C., Lu, J., Huang, Y., and Chang, Y. C. (2011). FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* 11:92. doi: 10.1186/1472-6750-11-92

Li, H. L., Wu, H. L., Jiang, F. J., Wu, J. L., Xue, Y., Gan, L. H., et al. (2018). Heterologous expression and characterization of an acidic GH11 family xylanase from *Hypocrea orientalis*. *Appl. Biochem. Biotechnol.* 184, 228–238. doi: 10.1007/s12010-017-2532-2

Liu, M. Q., and Liu, G. F. (2008). Expression of recombinant *Bacillus licheniformis* xylanase A in *Pichia pastoris* and xylooligosaccharides released from xylans by it. *Protein Expr. Purif.* 57, 101–107. doi: 10.1016/j.pep.2007.10.020

Liu, N. Z., Peng, S., Yang, J., Cong, Z. W., Lin, X. P., Liao, S. R., et al. (2019). Structurally diverse sesquiterpenoids and polyketides from a sponge-associated fungus *Aspergillus sydowii* SCSI041301. *Fitoterapia* 135, 27–32. doi: 10.1016/j.fitote.2019.03.031

Liu, M. Q., Weng, X. Y., and Sun, J. Y. (2006). Expression of recombinant *Aspergillus niger* xylanase A in *Pichia pastoris* and its action on xylan. *Protein Expr. Purif.* 48, 292–299. doi: 10.1016/j.pep.2006.04.007

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010:pdb.prot5448. doi: 10.1101/pdb.prot5448

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428. doi: 10.1021/ac60147a030

Miller, R. G., and Sorrell, S. R. (2014). The future of oil supply. *Philos. Trans. A Math. Phys. Eng. Sci.* 372:20130179. doi: 10.1098/rsta.2013.0179

Mohr, A., and Raman, S. (2013). Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. *Energy Policy* 63, 114–122. doi: 10.1016/j.enpol.2013.08.033

Moreau, A., Roberge, M., Manin, C., Shareck, F., Kluepfel, D., and Morosoli, R. (1994). Identification of 2 acidic residues involved in the catalysis of xylanase A from *Streptomyces lividans*. *Biochem. J.* 302, 291–295. doi: 10.1042/bj3020291

Mueller, G. M., and Bills, G. F. (eds.) (2004). "Introduction" in *Biodiversity of Fungi* (Burlington: Academic Press), 1–4.

Nair, S. G., Sindhu, R., and Shashidhar, S. (2008). Purification and biochemical characterization of two xylanases from *Aspergillus sydowii* SBS 45. *Appl. Biochem. Biotechnol.* 149, 229–243. doi: 10.1007/s12010-007-8108-9

National Research Council (US) Committee on Biobased Industrial Products (ed.) (2000). "Making the transition to biobased products" in *Biobased industrial products: Priorities for research and commercialization*. Washington: National Academy Press.

Nielsen, H. (2017). "Predicting secretory proteins with SignalP" in *Protein function prediction: Methods and protocols*. ed. D. Kihara (New York, NY: Springer New York), 59–73.

Nobili, A., Gall, M. G., Pavlidis, I. V., Thompson, M. L., Schmidt, M., and Bornscheuer, U. T. (2013). Use of "small but smart" libraries to enhance the enantioselectivity of an esterase from *Bacillus stearothermophilus* towards tetrahydrofuran-3-yl acetate. *FEBS J.* 280, 3084–3093. doi: 10.1111/febs.12137

Paës, G., Berrin, J. G., and Beaugrand, J. (2012). GH11 xylanases: structure/function/properties relationships and applications. *Biotechnol. Adv.* 30, 564–592. doi: 10.1016/j.biotechadv.2011.10.003

Payan, F., Leone, P., Porciero, S., Furniss, C., Tahir, T., Williamson, G., et al. (2004). The dual nature of the wheat xylanase protein inhibitor XIP-I-structural basis for the inhibition of family 10 and family 11 xylanases. *J. Biol. Chem.* 279, 36029–36037. doi: 10.1074/jbc.M404225200

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Piotrowska, M. (2013). Contamination of breakfast cereal products by fungi and mycotoxins – a potential risk for consumer's health. *Biotechnol. Food Sci.* 77, 3–10.

Raper, K. B., and Fennell, D. I. (1965). *The genus Aspergillus*. Baltimore: The Williams and Wilkins Co.

Raper, K. B., and Thom, C. (1949). *A manual of the Penicillia*. Baltimore: Williams and Wilkons.

Richards, E., Reichardt, M., and Rogers, S. (1994). Preparation of genomic DNA from plant tissue. *Curr. Protoc. Mol. Biol.* 27, 2.3.1–2.3.7. doi: 10.1002/0471142727.mb0203s27

Rinaldi, M. G. (1983). Invasive aspergillosis. *Rev. Infect. Dis.* 5, 1061–1077.

Robert, V., Vu, D., Amor, A. B., van de Wiele, N., Brouwer, C., Jabas, B., et al. (2013). MycoBank gearing up for new horizons. *IMA Fungus* 4, 371–379. doi: 10.5598/imafungus.2013.04.02.16

Saini, J. K., Saini, R., and Tewari, L. (2015). Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 5, 337–353. doi: 10.1007/s13205-014-0246-5

Sarkar, A., and Wintrode, P. L. (2011). Effects of glycosylation on the stability and flexibility of a metastable protein: the human serpin alpha(1)-antitrypsin. *Int. J. Mass Spectrom.* 302, 69–75. doi: 10.1016/j.ijms.2010.08.003

Shallom, D., and Shoham, Y. (2003). Microbial hemicellulases. *Curr. Opin. Microbiol.* 6, 219–228. doi: 10.1016/S1369-5274(03)00056-0

Sharma, M., Chadha, B. S., and Saini, H. S. (2010). Purification and characterization of two thermostable xylanases from *Malbranchea flava* active under alkaline conditions. *Bioresour. Technol.* 101, 8834–8842. doi: 10.1016/j.biortech.2010.06.071

Sharma, H. P., Patel, H., and Sugandha, (2017). Enzymatic added extraction and clarification of fruit juices-a review. *Crit. Rev. Food Sci. Nutr.* 57, 1215–1227. doi: 10.1080/10408398.2014.977434

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860. doi: 10.1038/nprot.2006.468

Silva, L. A. O., Terrasan, C. R. F., and Carmona, E. C. (2015). Purification and characterization of xylanases from *Trichoderma inhamatum*. *Electron. J. Biotechnol.* 18, 307–313. doi: 10.1016/j.ejbt.2015.06.001

Sims, R. E. H., Mabee, W., Saddler, J. N., and Taylor, M. (2010). An overview of second generation biofuel technologies. *Bioresour. Technol.* 101, 1570–1580. doi: 10.1016/j.biortech.2009.11.046

Singhvi, M. S., Chaudhari, S., and Gokhale, D. V. (2014). Lignocellulose processing: a current challenge. *RSC Adv.* 4, 8271–8277. doi: 10.1039/c3ra46112b

Solá, R. J., and Griebenow, K. (2009). Effects of glycosylation on the stability of protein pharmaceuticals. *J. Pharm. Sci.* 98, 1223–1245. doi: 10.1002/jps.21504

Sommer, A. (2016). Burning fossil fuels: impact of climate change on health. *Int. J. Health Serv.* 46, 48–52. doi: 10.1177/0020731415625253

Sorrell, S. (2015). Reducing energy demand: a review of issues, challenges and approaches. *Renew. Sust. Energ. Rev.* 47, 74–82. doi: 10.1016/j.rser.2015.03.002

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., et al. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32, 1478–1488. doi: 10.1038/emboj.2013.79

Straathof, A. J. J. (2014). Transformation of biomass into commodity chemicals using enzymes or cells. *Chem. Rev.* 114, 1871–1908. doi: 10.1021/cr400309c

Takahashi, Y., Kawabata, H., and Murakami, S. (2013). Analysis of functional xylanases in xylan degradation by *Aspergillus niger* E-1 and characterization of the GH family 10 xylanase XynVII. *Springerplus* 2:447. doi: 10.1186/2193-1801-2-447

Teixeira, R. S. S., Siqueira, F. G., de Souza, M. V., Ferreira, E. X., and Bon, E. P. D. (2010). Purification and characterization studies of a thermostable beta-xylanase from *Aspergillus awamori*. *J. Ind. Microbiol. Biotechnol.* 37, 1041–1051. doi: 10.1007/s10295-010-0751-4

Thom, C., and Church, M. (1926). *The Aspergilli*. Baltimore: The Williams and Wilkins Co.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Valentine, J., Clifton-Brown, J., Hastings, A., Robson, P., Allison, G., and Smith, P. (2012). Food vs. fuel: the use of land for lignocellulosic next generation' energy crops that minimize competition with primary food production. *Glob. Change Biol. Bioenergy* 4, 1–19. doi: 10.1111/j.1757-1707.2011.01111.x

van der Straat, L., Vernooij, M., Lammers, M., van den Berg, W., Schonewille, T., Cordewener, J., et al. (2014). Expression of the *Aspergillus terreus* itaconic acid biosynthesis cluster in *Aspergillus niger*. *Microb. Cell Fact.* 13:11. doi: 10.1186/1475-2859-13-11

Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994). Mutational and crystallographic analyses of the active-site residues of the *Bacillus circulans* xylanase. *Protein Sci.* 3, 467–475. doi: 10.1002/pro.5560030312

Walia, A., Guleria, S., Mehta, P., Chauhan, A., and Parkash, J. (2017). Microbial xylanases and their industrial application in pulp and paper biobleaching: a review. *3 Biotech* 7:11. doi: 10.1007/s13205-016-0584-6

Wang, W., Wei, H., Alahuhta, M., Chen, X. W., Hyman, D., Johnson, D. K., et al. (2014). Heterologous expression of xylanase enzymes in lipogenic yeast *Yarrowia lipolytica*. *PLoS One* 9:e111443. doi: 10.1371/journal.pone.0111443

Wang, X. Y., Zheng, F., Wang, Y., Tu, T., Ma, R., Su, X. Y., et al. (2017). Improvement of the catalytic efficiency of a hyperthermophilic xylanase from *Bispora* sp MEY-1. *PLoS One* 12:e0189806. doi: 10.1371/journal.pone.0189806

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics" in *PCR protocols: A guide to methods and applications*. eds. M. Innis, D. Gelfand, J. Sninsky and T. White (Orlando, Florida: Academic Press), 315–322.

Withers, S. G., Dombroski, D., Berven, L. A., Kilburn, D. G., Miller, R. C., Warren, R. A. J., et al. (1986). Direct H-1-NMR determination of the stereochemical course of hydrolyzes catalyzed by glucanase components of the cellulase complex. *Biochem. Biophys. Res. Commun.* 139, 487–494. doi: 10.1016/S0006-291X(86)80017-1

Yadav, P., Maharjan, J., Korpole, S., Prasad, G. S., Sahni, G., Bhattarai, T., et al. (2018). Production, purification, and characterization of thermostable alkaline xylanase from *Anoxybacillus kamchatkensis* NASTPD13. *Front. Bioeng. Biotech.* 6:65. doi: 10.3389/fbioe.2018.00065

Yang, Y., Chen, M., Li, Z. W., Al-Haimi, A. M. S., de Hoog, S., Pan, W. H., et al. (2016). Genome sequencing and comparative genomics analysis revealed pathogenic potential in *Penicillium capsulatum* as a novel fungal pathogen belonging to Eurotiales. *Front. Microbiol.* 7:1541. doi: 10.3389/fmicb.2016.01541

Yang, Y. L., Zhang, W., Huang, J. D., Lin, L., Lian, H. X., Lu, Y. P., et al. (2010). Purification and characterization of an extracellular xylanase from *Aspergillus niger* C3486. *Afr. J. Microbiol. Res.* 4, 2249–2256.

Ye, Y. Z., Choi, J. H., and Tang, H. X. (2011). RAPSearch: a fast protein similarity search tool for short reads. *BMC Bioinformatics* 12:159. doi: 10.1186/1471-2105-12-159

Zarafeta, D., Galanopoulou, A. P., Leni, M. E., Kaili, S. I., Chegkazi, M. S., Chrysina, E. D., et al. (2020). XynDZ5: a new thermostable GH10 xylanase. *Front. Microbiol.* 11:545. doi: 10.3389/fmicb.2020.00545

Zhao, Y. A., Tang, H. X., and Ye, Y. Z. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 28, 125–126. doi: 10.1093/bioinformatics/btr595

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 October 2020
doi: 10.3389/fmicb.2020.567863

MCIC: Automated Identification of Cellulases From Metagenomic Data and Characterization Based on Temperature and pH Dependence

Mehdi Foroozandeh Shahraki[1], Shohreh Ariaeenejad[2], Fereshteh Fallah Atanaki[1], Behrouz Zolfaghari[3], Takeshi Koshiba[4], Kaveh Kavousi[1] and Ghasem Hosseini Salekdeh[2,5]**

[1] Laboratory of Complex Biological Systems and Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran, [2] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran, Agricultural Research Education and Extension Organization, Karaj, Iran, [3] Computer Science and Engineering Department, Indian Institute of Technology Guwahati, Guwahati, India, [4] Department of Mathematics, Faculty of Education and Integrated Arts and Sciences, Waseda University, Tokyo, Japan, [5] Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Digvijay Verma,
Babasaheb Bhimrao Ambedkar University, India
Yu-Wei Wu,
Taipei Medical University, Taiwan

***Correspondence:**
Kaveh Kavousi
kkavousi@ut.ac.ir;
kkavousi@yahoo.com
Ghasem Hosseini Salekdeh
h_salekdeh@abrii.ac.ir;
hsalekdeh@yahoo.com

Specialty section:
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 30 May 2020
Accepted: 30 September 2020
Published: 23 October 2020

Citation:
Foroozandeh Shahraki M, Ariaeenejad S, Fallah Atanaki F, Zolfaghari B, Koshiba T, Kavousi K and Salekdeh GH (2020) MCIC: Automated Identification of Cellulases From Metagenomic Data and Characterization Based on Temperature and pH Dependence. Front. Microbiol. 11:567863. doi: 10.3389/fmicb.2020.567863

As the availability of high-throughput metagenomic data is increasing, agile and accurate tools are required to analyze and exploit this valuable and plentiful resource. Cellulose-degrading enzymes have various applications, and finding appropriate cellulases for different purposes is becoming increasingly challenging. An *in silico* screening method for high-throughput data can be of great assistance when combined with the characterization of thermal and pH dependence. By this means, various metagenomic sources with high cellulolytic potentials can be explored. Using a sequence similarity-based annotation and an ensemble of supervised learning algorithms, this study aims to identify and characterize cellulolytic enzymes from a given high-throughput metagenomic data based on optimum temperature and pH. The prediction performance of MCIC (metagenome cellulase identification and characterization) was evaluated through multiple iterations of sixfold cross-validation tests. This tool was also implemented for a comparative analysis of four metagenomic sources to estimate their cellulolytic profile and capabilities. For experimental validation of MCIC's screening and prediction abilities, two identified enzymes from cattle rumen were subjected to cloning, expression, and characterization. To the best of our knowledge, this is the first time that a sequence-similarity based method is used alongside an ensemble machine learning model to identify and characterize cellulase enzymes from extensive metagenomic data. This study highlights the strength of machine learning techniques to predict enzymatic properties solely based on their sequence. MCIC is freely available as a python package and standalone toolkit for Windows and Linux-based operating systems with several functions to facilitate the screening and thermal and pH dependence prediction of cellulases.

Keywords: cellulase, machine learning, metagenomics, enzyme screening, optimum temperature, optimum pH, MCIC

INTRODUCTION

Lignocellulosic biomass is a great potential resource for the production of bio-energy and bio-based material since it is largely abundant, inexpensive, and renewable (Demain et al., 2005). Cellulose, hemicellulose, and lignin together comprise the lignocellulosic biomass. Only a small volume of these biopolymers is currently used, with the rest being considered waste (Sánchez, 2009). Endoglucanases (EC 3.2.1.4), cellobiohydrolases or exoglucanases (EC 3.2.1.91), and beta-glucosidases (EC 3.2.1.21) are three major groups of cellulolytic enzymes and the process of cellulose degradation depends on their collaborative activity (Kumar et al., 2008). For many years, much attention has been paid to cellulases due to their potential of being utilized for various purposes including biofuel production, pulp, and paper, textile, food processing, animal feed, detergent, and agricultural industries as well as waste management (Kuhad et al., 2011). Evidently, each of these applications requires cellulases with specific characteristics and capabilities. The ability of enzymes to operate at a certain temperature or pH range is among the most critical attributes which make an enzyme suitable for a particular usage (Kirk et al., 2002).

Metagenomics is the culture-independent analysis of an environmental sample by a combination of molecular biology and genetics to extract, identify, characterize, and utilize almost all the genetic information embedded within that sample (Handelsman, 2005; Gharechahi and Salekdeh, 2018). The emergence of metagenomics has aided the isolation of various novel enzymes, cellulases included, from the uncultured microbiota of an environment. Unlike culture-dependent methods that are unable to present an inclusive understanding of microbial communities, their properties, and enzymatic capabilities, metagenomics offers a rich and valuable information resource to explore different environmental niches. Each microbial community has particularly evolved to meet the criteria of their ecosystem. As an instance, rumen microbiota has a rich hydrolyzing enzyme profile adapted to augment the digestion of the plant matter and plant-derived complex polysaccharides, such as cellulose, which dominate the ruminant diet (Stewart et al., 2019; Ariaeenejad et al., 2020c; Motahar et al., 2020). Termite has been known for its remarkable ability in the rapid deconstruction of recalcitrant woody biomass. This ability stems from complex relationships among termite intestinal symbionts (Liu et al., 2019).

Numerous enzymes have been isolated from metagenomic samples without the employment of computational methods including some thermophilic cellulases that were derived from environments such as biogas digester, sugarcane bagasse, rice straw compost, and hot springs (Geng et al., 2012; Yeh et al., 2013; Schröder et al., 2014; Kanokratana et al., 2015). However, the process of mining for novel enzymes solely through functional screening techniques and without the utilization of *in silico* methods can be laborious and time-consuming. For instance, to isolate highly thermostable beta-xylosidases from hot spring soil microbiota, the researchers had to express, examine, and screen 269 candidate proteins based on their xylosidase activity (Sato et al., 2017). Doubtlessly, powerful *in silico* approaches could effectively facilitate this process and reduce experimental expenses and time consumption.

As the availability of metagenomic data is rapidly growing, due to the advent of next-generation sequencing technologies and their constant advancements, employing computational methods instead of wet-lab experiments to identify new enzymes with specific properties, can significantly reduce the costs and accelerate the process. In this context, several sequence-based enzyme analysis tools have been developed and extensively used in recent years. These tools are designed to tackle several key problems in this field such as *in silico* prediction of enzymatic functions (Dalkiran et al., 2018; Li et al., 2018; Zhang et al., 2020), protein structure homology-modeling (Waterhouse et al., 2018) and docking (Grosdidier et al., 2011), etc. Moreover, some studies have been dedicated to the sequence-based prediction of thermostability (Zahiri, 2016), and computational engineering of enzymes toward specific characteristics of interest (Mazurenko et al., 2020).

One of the major computational approaches that has been rigorously employed in bioinformatics is machine learning, which as the literature can prove, is capable of mapping the relationships between the primary structure of proteins and their different properties and make reasonable predictions based upon them (Shastry and Sanjay, 2020). Machine learning techniques have been successfully applied so predict various properties of proteins such as activity (Ariaeenejad et al., 2018), tertiary structure (Cheng et al., 2008), subcellular localization (Almagro Armenteros et al., 2017), stability at different environmental conditions (Wu et al., 2009), etc. More specifically, Yan and Wu (2012) used a neural network to predict optimum pH and temperature of endoglucanases (EC 3.2.1.4) from their primary structure. AcalPred is another study that utilizes support vector machines to discriminate between acidic and alkaline enzymes based on their amino acid sequences (Lin et al., 2013). Moreover, several research studies have focused on the sequence-based prediction of protein thermostability (Ebrahimi and Ebrahimie, 2010; Pucci et al., 2014). In a 2019 study, Li et al. (2019) utilized machine learning to predict the optimum growth temperature (OGT) for microorganism based in their proteome and subsequently, used the predicted OGT alongside with the enzymes' amino acid compositions to predict their catalytic optima. In another research, Ariaeenejad et al. (2018) applied a regression model based on pseudo amino acid composition to predict the optimum temperature and pH of xylanase in strains of *Bacillus subtilis* enzymes.

Herein, we present an automated pipeline for sequence-based identification of cellulose-degrading enzymes, as well as a machine learning approach aimed at the prediction of their thermal and pH dependence. MCIC (metagenome cellulase identification and characterization) can explore metagenomic assembled contigs and screens them to find probable cellulolytic enzymes and classifies them based on their optimum pH and temperature. Furthermore, four metagenomic data from the soil, termite gut, sheep, and cattle rumen sources were analyzed by MCIC and their cellulolytic profiles are compared. Two computationally predicted enzymes from cattle rumen

data were cloned, expressed, and tested to validate the tool's competence.

MATERIALS AND METHODS

Development and Evaluation of the Prediction Model

Dataset Preparation

The first step toward training a model is data collection and since there was not any previous research on the prediction of thermal and pH dependence of three enzyme families involved in cellulose degradation, two new datasets for temperature and pH optima had to be collected. The BRENDA (Jeske et al., 2019) and UniProt (Bateman, 2019) databases were explored for enzyme families with EC 3.2.1.4, EC 3.2.1.21, and EC 3.2.1.91 and samples with reported optima were extracted. Since the existence of any character that does not represent an amino acid residue will interfere with the process of feature generation and learning, therefore any character other than amino acid symbols were removed from extracted sequences. Redundant or highly similar enzymes' were removed using the CD-Hit tool with a cut-off value of 0.9 (Huang et al., 2010). To reduce the redundancy, CD-Hit clusters highly homologous sequences and keeps one sample from each cluster. The final datasets consisted of 155 and 145 samples with optimum pH and temperature labels, respectively, and the non-redundant union of both datasets had 163 instances. Samples were labeled according to their reported optima. For temperature dataset, samples were labeled into mesophilic ($T_{opt} < 50°C$), thermophilic ($50°C < T_{opt} < 75°C$) and hyper-thermophilic ($75°C < T_{opt}$). Likewise, samples in the pH dataset were labeled into acidic ($pH_{opt} < 5$), neutral ($5 < pH_{opt} < 8$), and alkaline ($8 < pH_{opt}$). **Figure 1** presents the ratio of enzyme families, and classes in the datasets.

Data imbalance is a common obstacle in many classification problems. As illustrated in **Figure 1**, the collected datasets are relatively imbalanced. More severely, pH dataset lacks sufficient numbers of alkaline cellulase samples, with seven alkaline instances comprising a ratio of less than five percent of the whole dataset. This insufficiency of samples leads to decreased ability of the final prediction model in detecting alkaline cellulases. In order to address the data imbalance problem, synthetic minority over-sampling technique (SMOTE) was implemented on training datasets (Chawla et al., 2002). SMOTE tackles this problem by adding new synthetic instances of the minority classes, to the point where it is in balance with other majority classes.

Feature Generation

Utilization of a proper set of features is one of the most critical steps in the effort to train an accurate predictor. Due to the lack of precise evidence on the most related protein features to its thermal and pH dependence, in this study, numerous protein features were calculated and selected in order to find the most appropriate set of features for the current classification problem. In the feature generation step, we utilized iFeature and Pfeature that are two publicly available tools capable of generating a comprehensive spectrum of descriptors to facilitate the numerical representation of biological sequences for machine learning purposes (Chen et al., 2018; Pande et al., 2019). These sequence-based and length-independent features incorporate various properties of proteins such as composition of atoms, bonds, amino acids, and dipeptides as well as physico-chemical properties, repeats, distribution, etc. Many of the generated descriptors were duplicates and thus were removed. **Supplementary Table 1** presents a summary of computed protein features and their dimensionality after removing duplicate descriptors.

Consequently, a feature vector with a dimension of 6524 unique descriptors was generated for each enzyme sample in the datasets. Due to the fact that different descriptors have dissimilar ranges, all raw values had to be transformed into the same scale. MinMaxScaler method was employed to rescale features' value ranges between 0 and 1.

Feature and Model Selection

Enzymatic attributes, including temperature and pH dependence, are under the influence of various sequential, structural, and physiochemical features; nonetheless, not every generated descriptor is equally related to the attributes of interest. Therefore, a process of selecting the most relevant features to the optimum pH and temperature was required to prune features with lower importance and predictive ability. Filter feature selection methods use a statistical measure to score and rank features. For this step, Chi-square and ANOVA *F*-test were tested, and due to better results, the *F*-test was finally applied as the filter method, and the best descriptors were nominated by the SelectKBest method, with K-values of 253 and 93 for temperature and pH models, respectively (number of features with P-value < 0.05). Since multiple tests are being performed, Benjamini–Hochberg (Ferreira and Zwinderman, 2006) method was used as a false discovery rate controlling procedure to obtain realistic P-values in order to find the number of features with significant relationship with modeled responses.

Since there are numerous classification methods with different strengths and weaknesses, we had to find those that are most suited for this problem. Various classifiers were tested including, Multilayer perceptron (MLP), Decision Tree, Random Forest (RF), NaÏve Bayes, Gaussian Process, Bagging Classifier, AdaBoost, K-Nearest Neighbors, Support Vector Machine (SVM), XGBoost, and Gradient Boosting, all of which were implemented from sci-kit learn python package (Pedregosa et al., 2011). Among these methods, MLP, SVM, and RF depicted considerably better prediction performance. MLPs are a subset of artificial neural networks (ANN) that are inspired by the structure and function of actual biological neural networks and are widely used supervised learning algorithms (Tadeusiewicz, 1995). RF is consisted of multiple decision trees which result in reduced variance and better generalization in comparison to single decision trees (Breiman, 2001). For classification problems, support vector machines construct hyperplanes using a variety of kernel functions (Cortes and Vapnik, 1995). RF and SVM are

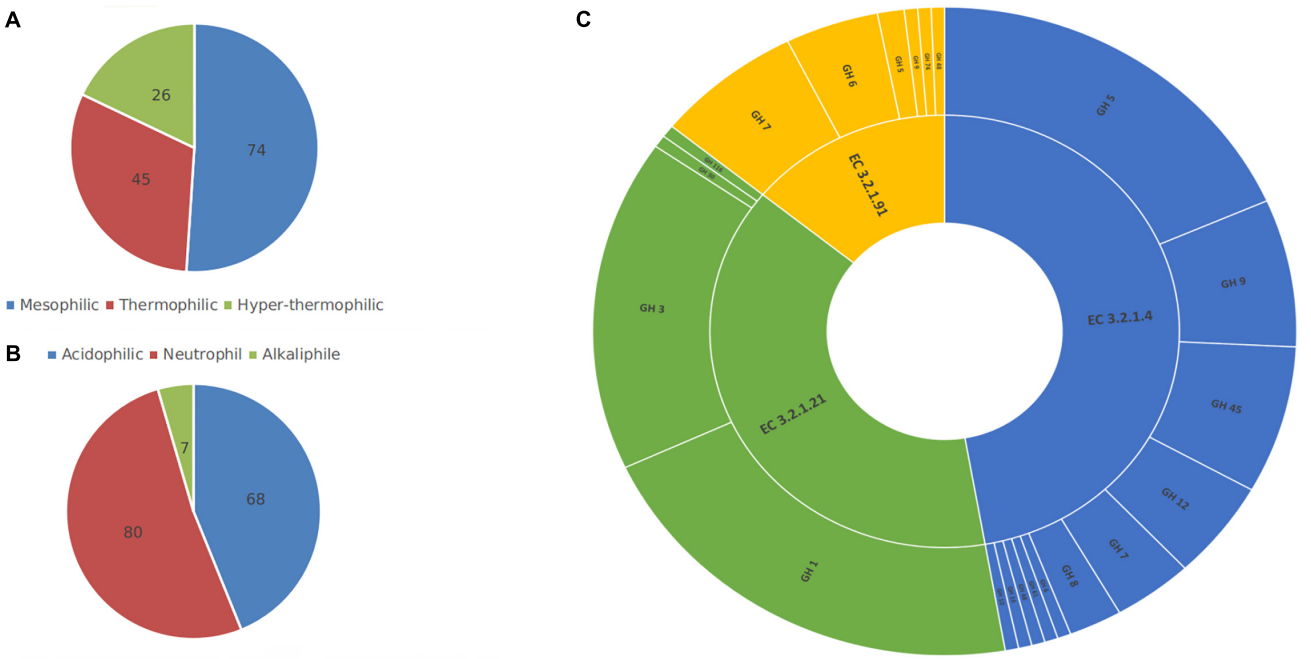


FIGURE 1 | Samples in temperature **(A)** and pH **(B)** datasets were divided into three classes based on their reported optima. The non-redundant union **(C)** of both datasets, had 163 instances with 3 different EC numbers and from 16 different glycoside hydrolase (GH) families. This figure shows the composition of each class or family in the collected datasets.

popular methods in computational biology due to their ability in dealing with high-dimensional feature space, small number of samples, and complex data structures (Ben-Hur et al., 2008; Qi, 2012).

After testing various combinations of different classification algorithms, an ensemble classifier was built from a MLP (with two hidden layers, 200 nodes in each layer, ReLU as activation function, and Adam optimizer), a RF (with 200 decision trees and information gain as splitting criteria), and a SVM (with radial basis function kernel). The ensemble method decides the final output from weighted soft voting between three mentioned classifiers. In the process of soft voting, each model returns an array representing its computed probability of occurrence for each class.

The workflow mentioned above required several hyper-parameter tuning steps, all of which were performed using the GridSearchCV method (Yu and Zhu, 2020). There are several approaches to detect the best combination of hyperparameters for a machine learning model. The GridSearchCV is an algorithm which given multiple options for each hyper-parameter, tries every possible combination to build the desired model and evaluates that model through cross-validation tests. This method can thoroughly investigate the hyper-parameter space to find the best configuration for the model which achieves the best performance.

Evaluation

Among evaluation strategies for machine learning models, multiple iterations of random train-test splitting and single cross-validation tests are two most commonly used approaches. In this study, for the purpose of evaluation, 100 iterations of sixfold cross-validation tests with different random seeds were performed to assure a thorough assessment of the prediction models by training and testing them on various random combinations of data. In the sixfold cross-validation, the dataset was randomly split into six equal subsamples, five of which were used as training sets that were subjected to feature selection, oversampling, and one subsample was then used for testing with different evaluation metrics. This validation is executed six times leaving out one subsample each time as the test set to ensure that all samples in the set were tested.

Accuracy, macro-recall, macro-precision, macro-f1 scores are among the most commonly used metrics and are calculated through the following formulae:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Precision} = \frac{TN}{TN + FP} \quad (3)$$

$$F1 = 2.\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Here, TP (true positive) and TN (true negative) are positive and negative examples, respectively, that were correctly predicted. Accordingly, FP (false positive) and FN (false

negative) were mistakenly classified. "macro" prefix refers to the unweighted average of each metric among different classes. Sci-kit learn python package was used several times during the above mentioned pipeline of development and evaluation of the prediction model (Pedregosa et al., 2011). **Figure 2** illustrates the graphical workflow of developing MCIC's prediction models.

Comparative Metagenome Analysis

Assembly of Metagenomic Data

Firstly, raw data was prepared for samples in four different environments, including termite gut, soil, sheep rumen, and cattle rumen. Alongside a cattle rumen raw reads that was obtained from authors' another study (Gharechahi et al., 2020), three other samples were downloaded from NCBI's Short Read Archive (SRA) including termite sample (Whole genome shotgun sequencing of *Macrotermes natalensis* soldier gut metagenome) under accession number: SRR797686 (Hu et al., 2019), soil sample raw reads downloaded under accession number: ERR1939274 (Orellana et al., 2018), and sheep rumen with accession number SRR1222429 (Kamke et al., 2016). Reads quality was checked using FastQC tools, and reads with relatively poor quality (Phred score < 20) were discarded for further analysis. The subsequent cleaned reads were given to MEGAHIT as inputs (Li et al., 2015). MEGAHIT is a standalone software available as both CPU and GPU versions. This tool allows users to perform large and complex metagenomics assembly up to hundreds of GB data. Metagenome assembly was carried out using options --kmin-1pass, --k-list 27,37,47,57,67,77,87, --m 60e + 10, --min-contig-len 300, -t 16. The results from *de novo* assembly of metagenomic data contained millions of assigned contigs.

Identification and Characterization of Cellulases From Assembled Contigs

In order to identify putative cellulase genes among assembled contigs, the sequence similarity-based method of NCBI's BLASTx (Altschul et al., 1990) was executed against the cellulase sequences dataset (163 samples), which were used to train the prediction machine. These 163 sequences are the non-redundant union of both pH and temperature datasets. The BLAST results were then filtered on the basis of bit-score (Pearson, 2013). The minimum bit-score cut-off was chosen to be 50 which indicates homology with the known cellulase samples. The screened cellulase genes are then translated and the resulting amino acid sequences are then given to the prediction

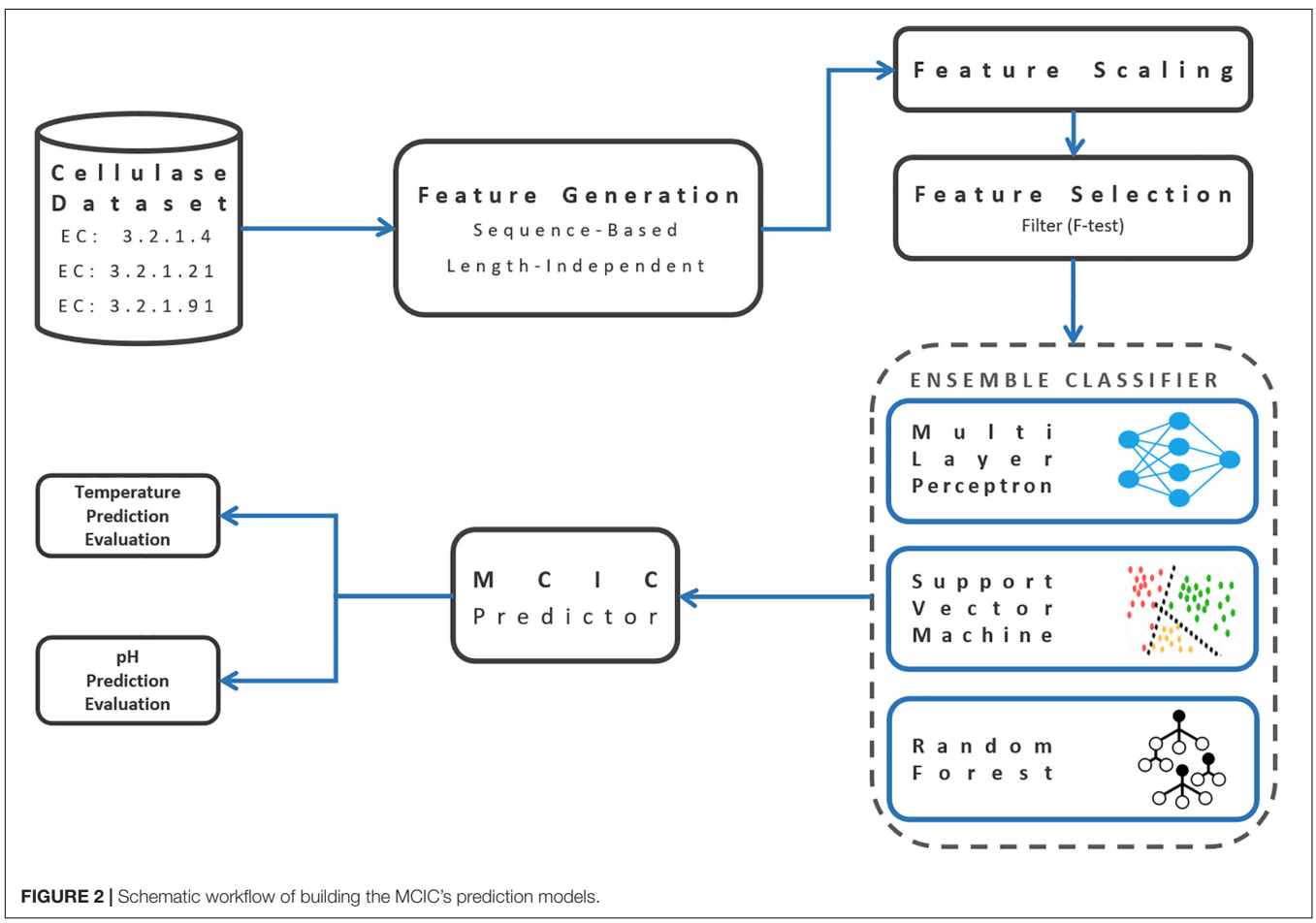


FIGURE 2 | Schematic workflow of building the MCIC's prediction models.

module as inputs to be characterized in terms of thermal and pH dependence.

Identification, Cloning, Expression, Purification, and Characterization of Two Cellulolytic Enzymes' Genes

With the objective of experimental validation of MCIC's predictive performance and demonstration of its potential applicability for mining metagenomic big data for cellulases, MCIC was exploited to computationally identify and characterize putative cellulases within cattle rumen's metagenomic data. As the process of gene cloning and production of enzymes has lower success-rate when the abundance of the target gene in the sample is not high, and additionally since this process is costly and time-consuming, therefore some strict measures were taken to refine the final list of candidates as much as possible. Two main criteria were considered for selecting the final set of target candidates for further experiments. First, by mapping reads back to assembled contigs using the BWA (Li, 2013), only contigs that were in the top 25% of the mapped reads were retained. Bit-scores assigned to each computationally identified cellulase were considered as the secondary selection filter, and sequences with lower than 300 bit-scores were removed. The remaining putative cellulase genes in the filtered shortlist were nominated for cloning, expression and further experiments in order to verify the MCIC's ability in correct cellulase identification and characterization. These wet-lab analysis were aimed at experimental validation of MCIC's predictive performance.

In order to acquire cellulase genes, metagenomic DNA templates from cattle rumen were used for polymerase chain reaction (PCR) amplification with two pairs of primers. For PersiCel5 a forward (5′-TAATAGGCTAGCATGAAGAAGTC CTTTGTATTTGT-3′) primer with NheI restriction site and a reverse (5′-TGATAGGTCGACTTATTTTATATCTATCTCATT GCG-3′) primers with SalI restriction site was used for amplification. Similarly, PersiCel6 was PCR-amplified using a forward (5′-TAATAGGCTAGCATGAATAAGAAGCATTTGC GG-3′) primer with NheI site and reverse (5′-TGATAGGCG GCCGCCTATTTTCCAGCCTTCTCCT-3′) primers containing NotI restriction site. The resulting PCR products were detected on agarose gel 1.5% (w/v) and purified using the gel extraction kit (BioRon, Germany). Purified DNA fragments were cloned and digested into the pET28a.

The resulting plasmids were then transformed into the *Escherichia coli* BL21 (DE3) and correct insertion was confirmed by sequencing. In the Luria-Bertoni (LB) medium, the recombinant strain pET28a was cultivated at the temperature of 37°C. Adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM for 20 h at 20°C, expression of enzymes was induced. By utilizing Ni-NTA Fast Start Kit (Qiagen, Hilden, Germany), N-terminal Histidine-tagged recombinant protein was purified and evaluated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Two candidate enzymes produced from successful cloning, expression, and purification were named PersiCel5 and PersiCel6. Both enzymes were subjected to further biophysical experiments. Nucleotide sequences of both PersiCel5 and PersiCel6 were submitted to GenBank and are available in the **Supplementary Material**. Protein concentrations were determined through the Bradford method using the bovine serum albumin as the standard. By measuring the optical density (OD) of chromatography eluent at 280 nm, the protein concentration was estimated.

To determine the optimum pH of enzyme activity, 10 mM phosphate buffer was prepared at different pH (4–11), and after added enzyme was incubated with a substrate for 20 min at room temperature. The DNS was used to measure activity. In order to determine the optimum temperature of enzyme activity, enzyme solution in 10 mM phosphate buffer (pH 8) with substrate was incubated in the different temperature (30–90°C) for 20 min and DNS was used to measure activity (Ariaeenejad et al., 2019, 2020b,a). For reporting, relative activity was considered as percentage of the highest activity.

RESULTS AND DISCUSSION

Prediction Model's Performance

Generated protein descriptors incorporate various molecular and sequential information with different degrees of relevance to the enzymes' thermal or pH dependence. With the purpose of better illustration, the overall importance scores of different features, at the level of feature category, were calculated and shown in **Supplementary Figure 1**. The feature category importance score is calculated as the average of all F-scores corresponding to single descriptors in that particular feature category.

Even though models demonstrated agreement upon correct predictions, in case of mis-classification their outputs were mostly different. This diversity in predictions stems from the dissimilarity in the basis of these classification algorithms. Hence, employing an ensemble method such as a voting classifier, could make us capable of exploiting three different machine learning approaches for a single task of prediction. The ensemble voting classifier, enhanced the overall performance of the model compared to single classifiers therefore it was the method of choice. Reported performance metrics are computed through 100 iterations of sixfold cross-validation tests. In each CV iteration, to assure the unbiased and effective evaluation, datasets were shuffled with different random seeds before splitting. **Figure 3** and **Table 1** represent the evaluation results of final model through 100 iterations sixfold CV tests.

To elaborate, in the prediction of temperature dependence, the model showed noticeably higher recall score (∼0.84) in the hyper-thermophilic class, in comparison to mesophilic and moderate-thermophilic classes (∼0.79 and ∼0.67, respectively), which suggest better applicability of this model in *in silico* characterization of hyper-thermophilic cellulases. In the prediction of cellulases' pH dependence, while the model had higher than 0.7 recall score in both acidic and neutral categories, it depicted a weak predictive performance in alkaline category (∼0.28). This poor performance on prediction of the alkaline cellulases is resulting from very small number of samples (seven) and lack of enough instances in the training

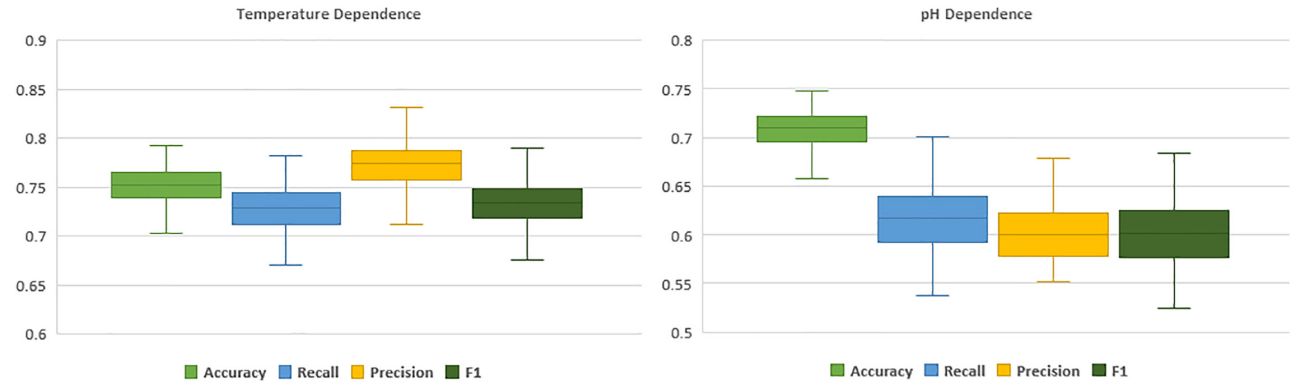


FIGURE 3 | Boxplots of temperature (left) and pH (right) dependence prediction performance computed through 100 sixfold CV tests.

dataset that was collected from the available cellulases in the literature with reported pH and temperature optima. Moreover, the relatively low variation in overall performance and the small error bars in the boxplots, imply the robustness of the models in the evaluation through 100 iterations of sixfold cross-validation test.

Comparative Metagenomic Analysis of Cellulolytic Profiles From Different Sources

After the assembly of raw metagenomic data for all four samples, millions of contigs were acquired. The total number of contigs obtained from *de novo* assembly of high-quality clean reads in each sample is presented in **Supplementary Table 2**. It was anticipated that undoubtedly copious enzyme genes were among these contigs. Therefore, a BLASTx against our dataset of cellulases could address the challenge of identifying cellulolytic enzymes. Numbers of identified cellulases from each metagenome source and the ratio of cellulases to all assembled contigs are also presented at **Supplementary Table 2**. After identification and translation of target cellulase genes, amino-acid sequences were used as inputs to the prediction tool for determination of thermal and pH dependence. Rapid and accurate analysis of the prevalence, pH, and temperature dependence of cellulose degrading enzymes found within a sample can be considered a significant step to obtain a comprehensive understanding of cellulolytic profile and abilities of different environments. **Supplementary Figure 2** illustrates the detailed results of the comparative analysis of four metagenome samples by using MCIC.

The comparative analysis results indicate that the sheep rumen had the highest number of cellulase genes while the ratio of cellulases to all assembled contigs was higher in cattle rumen. In comparison, fewer cellulases were found within soil and termite samples. The significantly higher abundance of cellulolytic enzymes in the rumen environments proves the cellulolytic abilities of ruminants' digestive system (Stewart et al., 2019). Not surprisingly, the majority of identified cellulases in all four samples were neutral and mesophilic. However, comparatively, the soil environmental sample had more acidic enzymes.

Identified, Cloned, Expressed, and Characterized Cellulases

Aiming to experimentally validate the MCIC's ability to facilitate the process of targeted mining for cellulases in metagenomic libraries, this tool was utilized to isolate and produce two novel cellulases from cattle rumen metagenome. After applying a two-step filtration, based on number of mapped reads and the bit-score, a shortlist of computationally identified candidates were obtained to be subjected to further experimental analysis. Consequently, two novel cellulases named PersiCel5 and PersiCel6 were produced and characterize through wet-lab experiments and both enzymes are used for other studies on enhancing lignocellulose degradation. **Table 2** presents the predicted classes of thermal and pH dependence for two enzymes in addition to the real values of pH and temperature optima.

The successful results for the targeted enzyme mining by the help of MCIC and its predictions indicate the potential applicability of this tool for a better exploration of different metagenomic data. To the best of our knowledge, this is the first study that provides software and descriptions of an automated pipeline for computational identification and pH and temperature dependence characterization of cellulose-degrading enzymes from high-throughput data and its ability was verified by wet-lab experiments and comparative analysis of different metagenome samples. As generally developing powerful machine learning-based models requires large training datasets, one of the greatest challenges of this study was the relatively limited number of cellulases existing in public databases with reported temperature and pH optima. Therefore, to compensate for the relatively small size of training datasets, in a very time-consuming and computationally intensive stage of this study, several classification algorithms with various hyper-parameters were rigorously explored, both individually and in different combinations, to find the optimal configuration with best

TABLE 1 | MCIC's prediction performance during 100 iterations of sixfold CV tests.

	Accuracy		Macro-recall		Macro-precision		Macro-F1	
	T_{opt}	pH_{opt}	T_{opt}	pH_{opt}	T_{opt}	pH_{opt}	T_{opt}	pH_{opt}
CV performance	0.75	0.71	0.73	0.62	0.77	0.60	0.73	0.60

TABLE 2 | Two novel cellulolytic enzymes were discovered from camel rumen metagenome and are being investigated in other studies. This table represents the detected enzymatic function of each enzyme and the comparison of predicted dependence characteristics with real pH and temperature optima values. MCIC was able to correctly predict the desired attributes.

Enzyme name	Enzymatic function	Predicted thermal dependence	Predicted pH dependence	Optimum temperature	Optimum pH
PersiCel5	Endo-glucanase	Mesophilic	Neutral	50	6.5
PersiCel6	Endo-glucanase	Thermophilic	Neutral	70	7.5

performance. Moreover, detailed analysis of the classification results suggested that, despite the utilization of SMOTE and its contribution in over-sampling of the minority classes, the models' performances were generally higher in prediction of classes with more training data in comparison to ones with fewer training instances. This implies that in future studies, when more experimentally identified and characterized cellulase enzymes are available, development of more accurate machine learning-based models for this task will be possible. Undoubtedly, this methodology can be generalized to other enzymatic attributes and for other enzyme families. Through the implementation of such automated pipelines with various predictive models, the challenge of targeted enzyme identification as well as agile, effective, and inexpensive screening of high-throughput data can be effectively addressed. By this means, the number of potential candidate enzymes with specific characteristics of interest could be significantly reduced prior to engaging the wet-lab experiments. From the prediction perspective, unlike Yan and Wu (2012) study which was only capable of performing the prediction for endoglucanases (EC 3.2.1.4), our method could extend the ability to other cellulolytic-enzymes cellobiohydrolases (EC 3.2.1.91), and beta-glucosidases (EC 3.2.1.21).

MCIC Standalone Toolkit

Metagenome cellulase identification and characterization is freely accessible in the form of a standalone toolkit and a python package. This tool helps users with screening of high-throughput metagenomic contigs in order to find probable cellulases and classify them on the basis of their pH and thermal dependence. Furthermore, both MCIC's prediction and screening tools are individually operational. The predictor accepts single and multiple cellulases in form of amino-acid sequence, while the screening tool, can identify and screen cellulases among millions of metagenome-derived contigs. The standalone software is available for Windows and Linux-based operating systems. MCIC is downloadable from https://cbb.ut.ac.ir/MCIC, and MCIC's repository on GitHub[1]. Further description and help about different functions of MCIC and their usage is available in the tool's manual.

[1]https://github.com/mehdiforoozandeh/MCIC

The MCIC software is particularly aimed at facilitating and automation of the process for getting access to cellulolytic enzymes existing in extensive metagenomic data in a targeted manner. As the main criterion for identification of cellulase-coding genes, the bit-score cut-off, is configurable by the users therefore, if this criterion is set more strictly to higher values, only sequences with more similarity to sample sequences in the reference database will be retained. This way, although potentially numerous cellulases with less sequence-similarity to the database may be omitted, users can obtain more accurate predictions of pH and temperature dependence due to the similarity of identified sequences to the predictive models' training datasets.

CONCLUSION

In this study, MCIC tool was designed and implemented in order to address two major objectives, one being the identification of probable cellulose degrading enzymes among metagenomic contigs and the other being the sequence-based characterization of their temperature and pH dependence. For the first purpose, a sequence similarity-based method, was employed to screen cellulase enzymes and for the prediction task, an ensemble of three supervised learning classifiers, MLP, RF, and SVM, were trained by a combination and selection of various sequence-based descriptors to classify cellulases on the basis of their temperature and pH dependence.

Moreover, MCIC was used for the analysis and comparison of cellulolytic profiles of four metagenome samples. This comparative analysis highlighted the extended enzymatic diversity and cellulose-hydrolytic ability of ruminants' rumen microbial community. Two identified cellulases were cloned, produced, and experimentally tested in order to validate the tool's abilities. The results of this study imply the competence of both machine learning techniques and sequence-based protein features for the prediction of enzymatic attributes. The MCIC toolkit has been made available as a standalone software and a python package and offers a variety of services. Since a plethora of enzymes are in constant demand and discovery of novel enzymes for various purposes is of great importance, similar analytic and predictive tools are needed for other enzyme families and this may be a possible direction for future studies. Moreover,

different machine learning techniques are being developed and advanced and they have proved effective in solving various biological problems.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: accession number SRR797686, https://www.ebi.ac.uk/biosamples/samples/SAMN08019258; accession number ERR1939274, https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR1939274; accession number SRR1222429, https://www.ebi.ac.uk/biosamples/samples/SAMN06452595.

AUTHOR CONTRIBUTIONS

MFS performed software development, evaluation, implementation, and comparative analysis. KK performed the theoretical design and provided significant advice on the method design. FFA performed the acquisition and assembly of metagenomic data and assisted in software development. SA executed the gene cloning, expression, purification, and biochemical characterization. BZ helped to implement the packages, data analysis, review, and editing. TK performed complete review, redesign, and check the machine learning models and pipeline according to the reviewers' comments, writing, review, and editing. GHS supervised the study and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was supported by grants from Agricultural Biotechnology Research Institute of Iran (ABRII).

ACKNOWLEDGMENTS

We give thanks to the supports from members of the Laboratory of Complex Biological Systems and Bioinformatics (CBB) who contributed to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.567863/full#supplementary-material

REFERENCES

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395. doi: 10.1093/bioinformatics/btx431

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ariaeenejad, S., Lanjanian, H., Motamedi, E., Kavousi, K., Moosavi-Movahedi, A. A., and Hosseini Salekdeh, G. (2020a). The stabilizing mechanism of immobilized metagenomic xylanases on bio-based hydrogels to improve utilization performance: computational and functional perspectives. *Bioconjug. Chem.* 31, 2158–2171. doi: 10.1021/acs.bioconjchem.0c00361

Ariaeenejad, S., Nooshi-nedamani, S., Rahban, M., Kavousi, K., Pirbalooti, A. G., Mirghaderi, S., et al. (2020b). A novel high glucose-tolerant β -Glucosidase : targeted computational approach for metagenomic screening. *Front. Bioeng. Biotechnol.* 8:813. doi: 10.3389/fbioe.2020.00813

Ariaeenejad, S., Sheykhabdolahzadeh, A., Maleki, M., Kavousi, K., Shahraki, M. F., and Salekdeh, G. H. (2020c). A novel high performance metagenome-derived alkali-thermostable endo-β-1,4-glucanase for lignocellulosic biomass hydrolysis in the harsh conditions. *Res. Square* 54, 1–26. doi: 10.21203/rs.2.23802/v1

Ariaeenejad, S., Maleki, M., Hosseini, E., Kavousi, K., Moosavi-Movahedi, A. A., and Salekdeh, G. H. (2019). Mining of camel rumen metagenome to identify novel alkali-thermostable xylanase capable of enhancing the recalcitrant lignocellulosic biomass conversion. *Bioresour. Technol.* 281, 343–350. doi: 10.1016/j.biortech.2019.02.059

Ariaeenejad, S., Mousivand, M., Dezfouli, P. M., Hashemi, M., Kavousi, K., and Salekdeh, G. H. (2018). A computational method for prediction of xylanase enzymes activity in strains of *Bacillus subtilis* based on pseudo amino acid composition features. *PLoS One* 13:e0205796. doi: 10.1371/journal.pone.0205796

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., and Rätsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Comput. Biol.* 4:e1000173. doi: 10.1371/journal.pcbi.1000173

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 341–378.

Chen, Z., Zhao, P., Li, F., Leier, A., Marquez-Lago, T. T., Wang, Y., et al. (2018). IFeature: a python package and web server for features extraction and selection from protein and peptide sequences. *Bioinformatics* 34, 2499–2502. doi: 10.1093/bioinformatics/bty140

Cheng, J., Tegge, A. N., and Baldi, P. (2008). Machine learning methods for protein structure prediction. *IEEE Rev. Biomed. Eng.* 1, 41–49. doi: 10.1109/RBME.2008.2008239

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1023/A:1022627411411

Dalkiran, A., Rifaioglu, A. S., Martin, M. J., Cetin-Atalay, R., Atalay, V., and Doğan, T. (2018). ECPred: a tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature. *BMC Bioinform.* 19:334. doi: 10.1186/s12859-018-2368-y

Demain, A. L., Newcomb, M., and Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124–154. doi: 10.1128/MMBR.69.1.124-154.2005

Ebrahimi, M., and Ebrahimie, E. (2010). Sequence-based prediction of enzyme thermostability through bioinformatics algorithms. *Curr. Bioinform.* 5, 195–203. doi: 10.2174/157489310792006693

Ferreira, J. A., and Zwinderman, A. H. (2006). On the Benjamini–Hochberg method. *Ann. Stat.* 34, 1827–1849. doi: 10.1214/009053606000000425

Geng, A., Zou, G., Yan, X., Wang, Q., Zhang, J., Liu, F., et al. (2012). Expression and characterization of a novel metagenome-derived cellulase Exo2b and its application to improve cellulase activity in *Trichoderma reesei*. *Appl. Microbiol. Biotechnol.* 96, 951–962. doi: 10.1007/s00253-012-3873-y

Gharechahi, J., and Salekdeh, G. H. (2018). A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation. *Biotechnol. Biofuels.* 11:216. doi: 10.1186/s13068-018-1214-9

Gharechahi, J., Vahidi, M. F., Ding, X.-Z., Han, J.-L., and Salekdeh, G. H. (2020). Temporal changes in microbial communities attached to forages with different lignocellulosic compositions in the cattle rumen. *FEMS Microbiol. Ecol.* 96, 1–22. doi: 10.1093/femsec/fiaa069

Grosdidier, A., Zoete, V., and Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* 39, W270–W277. doi: 10.1093/nar/gkr366

Handelsman, J. (2005). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi: 10.1128/mmbr.69.1.195.2005

Hu, H., da Costa, R. R., Pilgaard, B., Schiøtt, M., Lange, L., and Poulsen, M. (2019). Fungiculture in termites is associated with a mycolytic gut bacterial community. *mSphere* 4, 1–22. doi: 10.1128/msphere.00165-19

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47, D542–D549. doi: 10.1093/nar/gky1048

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4:56. doi: 10.1186/s40168-016-0201-2

Kanokratana, P., Eurwilaichitr, L., Pootanakit, K., and Champreda, V. (2015). Identification of glycosyl hydrolases from a metagenomic library of microflora in sugarcane bagasse collection site and their cooperative action on cellulose degradation. *J. Biosci. Bioeng.* 119, 384–391. doi: 10.1016/j.jbiosc.2014.09.010

Kirk, O., Borchert, T. V., and Fuglsang, C. C. (2002). Industrial enzyme applications. *Curr. Opin. Biotechnol.* 13, 345–351. doi: 10.1016/S0958-1669(02)00328-2

Kuhad, R. C., Gupta, R., and Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Res.* 2011, 1–10. doi: 10.4061/2011/280696

Kumar, R., Singh, S., and Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* 35, 377–391. doi: 10.1007/s10295-008-0327-8

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, G., Rabe, K. S., Nielsen, J., and Engqvist, M. K. M. (2019). Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS Synth. Biol.* 8, 1411–1420. doi: 10.1021/acssynbio.9b00099

Li, H. (2013). [Heng Li - compares BWA to other long read aligners like CUSHAW2] Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. Available online at: https://arxiv.org/abs/1303.3997

Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L., et al. (2018). DEEPre: sequence-based enzyme EC number prediction by deep learning. *Bioinformatics* 34, 760–769. doi: 10.1093/bioinformatics/btx680

Lin, H., Chen, W., and Ding, H. (2013). AcalPred: a sequence-based tool for discriminating between acidic and alkaline enzymes. *PLoS One* 8:e75726. doi: 10.1371/journal.pone.0075726

Liu, N., Li, H., Chevrette, M. G., Zhang, L., Cao, L., Zhou, H., et al. (2019). Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *ISME J.* 13, 104–117. doi: 10.1038/s41396-018-0255-1

Mazurenko, S., Prokop, Z., and Damborsky, J. (2020). Machine learning in enzyme engineering. *ACS Catal.* 10, 1210–1223. doi: 10.1021/acscatal.9b04321

Motahar, S. F. S., Khatibi, A., Salami, M., Ariaeenejad, S., Emam-Djomeh, Z., Nedaei, H., et al. (2020). A novel metagenome-derived thermostable and poultry feed compatible α-amylase with enhanced biodegradation properties. *Int. J. Biol. Macromol.* 164, 2124–2133. doi: 10.1016/j.ijbiomac.2020.08.064

Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E., and Konstantinidis, K. T. (2018). Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Appl. Environ. Microbiol.* 84:e01646. doi: 10.1128/AEM.01646-17

Pande, A., Patiyal, S., Lathwal, A., Arora, C., Kaur, D., Dhall, A., et al. (2019). Computing wide range of protein/peptide features from their sequence and structure. *bioRxiv* doi: 10.1101/599126

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinforma.* Chapter 3, Unit3.1. doi: 10.1002/0471250953.bi0301s42

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pucci, F., Dhanani, M., Dehouck, Y., and Rooman, M. (2014). Protein thermostability prediction within homologous families using temperature-dependent statistical potentials. *PLoS One* 9:e91659. doi: 10.1371/journal.pone.0091659

Qi, Y. (2012). "Random forest for bioinformatics," in *Ensemble Machine Learning: Methods and ApplicatiOns*, eds Z. Cha and M. Yunqian (Springer).

Sánchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnol. Adv.* 27, 185–194. doi: 10.1016/j.biotechadv.2008.11.001

Sato, M., Suda, M., Okuma, J., Kato, T., Hirose, Y., Nishimura, A., et al. (2017). Isolation of highly thermostable β-xylosidases froma hot spring soil microbial community using a metagenomic approach. *DNA Res.* 24, 649–656. doi: 10.1093/dnares/dsx032

Schröder, C., Elleuche, S., Blank, S., and Antranikian, G. (2014). Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. *Enzyme Microb. Technol.* 57, 48–54. doi: 10.1016/j.enzmictec.2014.01.010

Shastry, K. A., and Sanjay, H. A. (2020). "Machine learning for bioinformatics," in *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications. Algorithms for Intelligent Systems*, eds K. Srinivasa, G. Siddesh, and S. Manisekhar (Singapore: Springer).

Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., and Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* 37, 953–961. doi: 10.1038/s41587-019-0202-3

Tadeusiewicz, R. (1995). Neural networks: a comprehensive foundation. *Control Eng. Pract.* 3, 746–747. doi: 10.1016/0967-0661(95)90080-2

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427

Wu, L. C., Lee, J. X., Huang, H., Da, Liu, B. J., and Horng, J. T. (2009). An expert system to predict protein thermostability using decision tree. *Expert Syst. Appl.* 36, 9007–9014. doi: 10.1016/j.eswa.2008.12.020

Yan, S.-M., and Wu, G. (2012). Prediction of optimal pH and temperature of cellulases using neural network. *Protein Pept. Lett.* 19, 29–39. doi: 10.2174/092986612798472794

Yeh, Y. F., Chang, S. C., yu, Kuo, H. W., Tong, C. G., Yu, S. M., and Ho, T. H. D. (2013). A metagenomic approach for the identification and cloning of an endoglucanase from rice straw compost. *Gene* 519, 360–366. doi: 10.1016/j.gene.2012.07.076

Yu, T., and Zhu, H. (2020). Hyper-parameter optimization: a review of algorithms and applications. *arXiv*. Available online at: https://arxiv.org/abs/2003.05689

Zahiri, J. (2016). An overview of the protein thermostability prediction: databases and tools. *J. Nanomed. Res.* 3:00072. doi: 10.15406/jnmr.2016.03.00072

Zhang, T., Tian, Y., Yuan, L., Chen, F., Ren, A., and Hu, Q. N. (2020). Bio2Rxn: sequence-based enzymatic reaction predictions by a consensus strategy. *Bioinformatics* 36, 3600–3601. doi: 10.1093/bioinformatics/btaa135

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fmicb.2020.592353

A 'Split-Gene' Transketolase From the Hyper-Thermophilic Bacterium *Carboxydothermus hydrogenoformans*: Structure and Biochemical Characterization

Paul James†, Michail N. Isupov, Simone Antonio De Rose, Christopher Sayer, Isobel S. Cole† and Jennifer A. Littlechild**

Henry Wellcome Building for Biocatalysis, Biosciences, University of Exeter, Exeter, United Kingdom

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Woody Fessner,
Darmstadt University of Technology, Germany
Willem J. H. Van Berkel,
Wageningen University and Research, Netherlands

***Correspondence:**
Jennifer A. Littlechild
J.A.Littlechild@exeter.ac.uk
Paul James
P.B.C.James@exeter.ac.uk;
Paul.B.C.James@Northumbria.ac.uk

†Present address:
Isobel S. Cole,
School of Biological and Marine Sciences, University of Plymouth, Plymouth, United Kingdom
Paul James,
Hub for Biotechnology in the Built Environment, Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *06 August 2020*
Accepted: *12 October 2020*
Published: *30 October 2020*

Citation:
James P, Isupov MN, De Rose SA, Sayer C, Cole IS and Littlechild JA (2020) A 'Split-Gene' Transketolase From the Hyper-Thermophilic Bacterium Carboxydothermus hydrogenoformans: Structure and Biochemical Characterization. Front. Microbiol. 11:592353. doi: 10.3389/fmicb.2020.592353

A novel transketolase has been reconstituted from two separate polypeptide chains encoded by a 'split-gene' identified in the genome of the hyperthermophilic bacterium, *Carboxydothermus hydrogenoformans*. The reconstituted active $\alpha_2\beta_2$ tetrameric enzyme has been biochemically characterized and its activity has been determined using a range of aldehydes including glycolaldehyde, phenylacetaldehyde and cyclohexanecarboxaldehyde as the ketol acceptor and hydroxypyruvate as the donor. This reaction proceeds to near 100% completion due to the release of the product carbon dioxide and can be used for the synthesis of a range of sugars of interest to the pharmaceutical industry. This novel reconstituted transketolase is thermally stable with no loss of activity after incubation for 1 h at 70°C and is stable after 1 h incubation with 50% of the organic solvents methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone. The X-ray structure of the holo reconstituted $\alpha_2\beta_2$ tetrameric transketolase has been determined to 1.4 Å resolution. In addition, the structure of an inactive tetrameric β_4 protein has been determined to 1.9 Å resolution. The structure of the active reconstituted $\alpha_2\beta_2$ enzyme has been compared to the structures of related enzymes; the E1 component of the pyruvate dehydrogenase complex and D-xylulose-5-phosphate synthase, in an attempt to rationalize differences in structure and substrate specificity between these enzymes. This is the first example of a reconstituted 'split-gene' transketolase to be biochemically and structurally characterized allowing its potential for industrial biocatalysis to be evaluated.

Keywords: hyperthermophilic, 'split-gene', transketolase, thermal stability, industrial applications

INTRODUCTION

Transketolase (TK, EC 2.2.1.1) is a thiamine diphosphate-dependant (TPP) enzyme which plays an important role in the pentose phosphate pathway. It catalyses the rearrangement of sugar molecules by the transfer of a C2 unit from D-xylulose-5-phosphate to erythrose-4-phosphate, resulting in the formation of fructose-6-phosphate and glyceraldehyde-3-phosphate which are fed back into the glycolysis pathway (Racker et al., 1954).

TK enzymes have been found to accept a broad range of donor and acceptor substrates including xylulose 5-phosphate, ribose 5-phosphate, fructose 6-phosphate, glyceraldehyde 3-phosphate, erythrose 4-phosphate and hydroxypyruvate (HPA) (Schenk et al., 1998). Use of HPA as the ketol donor allows the release of the volatile reaction product CO_2 which drives the reaction to completion (**Scheme 1**). The ability of TKs to form enantioselective carbon-carbon bonds has generated increasing interest for their use as biocatalysts in industrial synthetic reactions (Hibbert et al., 2008). This reaction has previously been described as irreversible due to the production and evolution of CO_2 but was shown by Marsden et al. (2017) to be reversible over a period of weeks because while the decarboxylation of hydroxypyruvate is virtually irreversible, the carbon-carbon bond formation is not. The *Escherichia coli* TK (EcTK) has previously been used to demonstrate the potential synthesis of a range of pharmaceutical relevant sugars with HPA as the ketol donor and glycolaldehyde as the acceptor to produce the sugar erythrulose at a 100% conversion (Lilly et al., 1996).

TK enzymes are found throughout nature and have been isolated and structurally characterized from a number of different organisms including *E. coli* (Littlechild et al., 1995; Martin, 2008; Ludtke et al., 2013), *Thermus thermophilus* (PDB: 2E6K), *Saccharomyces cerevisiae* (Lindqvist et al., 1992; Nikkola et al., 1994), maize (Gerhardt et al., 2003), *Leishmania mexicana* (Veitch et al., 2004), *Mycobacterium tuberculosis* (Fullam et al., 2012), and human (Mitschke et al., 2010). Most TKs have a monomeric molecular mass ∼70 kDa and are active as homodimers with amino acid residues from each monomer contributing to the two active sites. Each of these active sites contains a molecule of TPP and a Mg^{2+} or Ca^{2+} ion which plays an important role in the co-factor binding (Selivanov et al., 2003).

TPP dependent enzymes are similar in that they all bind the cofactor at the interface of the two domains of each monomer: the pyrophosphate binding domain (PP) and the pyrimidine binding (Pyr) domain (Muller et al., 1993). The TK-like enzymes are a subgroup of homo-dimeric TPP enzymes which contain the N-terminal PP domain, followed by the Pyr domain and a transketolase C-terminal domain (TKC). The TKC domain is thought to have been introduced along the evolutionary route (Costelloe et al., 2008). The TK-like group contains related enzymes such as, D-xylulose-5-phosphate synthase (DXPS) and the E1 component of the large multi-subunit pyruvate dehydrogenase (PDH). Both DXPS and bacterial PDH E1 have the same order of domains (PP-, Pyr-, TKC) along the polypeptide chain (Xiang et al., 2007). Although the DXPS and PDHs E1 have limited sequence identity to TKs (around 26–30%) the arrangement of domains within the dimer and TPP binding sites of these enzymes are very similar (Muller et al., 1993; Arjunan et al., 2002). The major difference between TK and these related enzymes is that the TPP binds on the interface of the PP domain of one monomer and the Pyr domain of the adjacent one to form the TK dimer whereas in DXPS and bacterial PDH E1 the active sites are formed between the PP and Pyr domains within the same subunit (Costelloe et al., 2008).

Interestingly the mammalian PDH E1 (Kato et al., 2008) component proteins are hetero-tetramers with mutual positions of domains quite similar to those observed in bacterial dimeric PDHs, however they are built up from two polypeptide chains of comparable size with the PP domain located on one chain and the Pyr and the TKC domains on the second chain.

It has been previously reported that many archaea are lacking some, if not all of the enzymes, that constitute the pentose phosphate pathway, and the TK enzyme is either absent in the archaeal genomes or is encoded by two separate genes, referred to as a 'split-gene', which may or may not be located next to each other on the genome (Bräsen et al., 2014). The thermophilic and hyper-thermophilic bacteria have been reported to contain either a full-length or both a full-length and 'split-gene' TK enzymes, where the latter has been proposed to be acquired by horizontal gene transfer from the archaea (Koonin and Galperin, 2013).

To date only a small number of thermostable TK enzymes have been biochemically characterized. The TK from *Geobacillus stearothermophilus* (GsTK) has been shown to have an optimum temperature of 70°C and has the ability to use a range of aldehydes as the acceptor in the potential commercial reaction with hydroxypyruvate as the ketol donor (Abdoul-Zabar et al., 2013). Also characterized are two TK enzymes from *Deinococcus geothermalis* (DgTK) and *Deinococcus radiodurans* (DrTK) that were shown to have an optimal temperature of 50°C and have been used in combination with a thermostable transaminase enzyme to produce L-gluco-heptulose from L-arabinose (Bawn et al., 2018). The TK from the thermophilic bacterium *T. thermophilus* has provided structural insights that have been used to guide the design of mutants of the *E. coli* TK (EcTK) to increase its thermal stability (Morris et al., 2016).

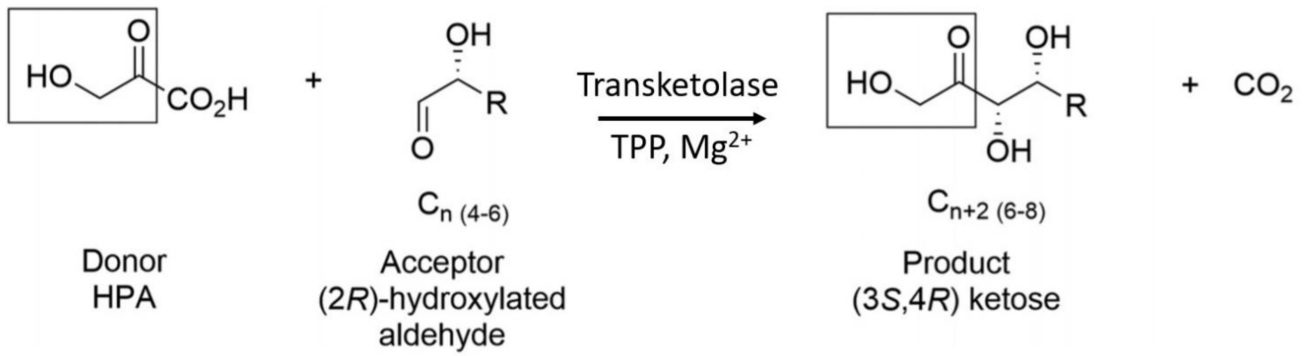
Carboxydothermus hydrogenoformans Z-2901 is a hyper-thermophilic, anaerobic bacterium isolated from a hot swamp of Kunashir Island, Russia (Svetlichny et al., 1991). This organism has an optimal growth temperature of 78°C and is believed to be one of the fastest growing carbon monoxide (CO) utilizing bacteria known and has five highly differentiated anaerobic CO dehydrogenase complexes (Wu et al., 2005).

Here, we report the biochemical and structural characterization of the 'split-gene' TK from this organism (ChTK-F) which has been reconstituted by combining the two proteins encoded on the *C. hydroxydothermus* genome. This is the first example of a 'split-gene' TK that has been reconstituted and found to be active. This enzyme has been biochemically and structurally characterized which has allowed its potential application for industrial biocatalysis to be evaluated.

RESULTS AND DISCUSSION

Gene Identification, Protein Expression, and Purification

The search for a TK enzyme in thermophilic bacteria using the BLAST database with the EcTK as the query sequence revealed the genes ChTK-N (Accession number: ABB15544) and ChTK-C (Accession number: ABB16214) in the bacterium *C. hydrogenoformans*. These genes which make up the reconstituted ChTK-F have 32% sequence identity to the



SCHEME 1 | TK catalyzed conversion of hydroxypyruvate and an acceptor aldehyde to a ketose product and carbon dioxide.

EcTK (PDB: 1QGD) and 40% sequence identity to the human TK (PDB: 3MOS). The ChTK-F also has 27% sequence identity to the DXPS from *E. coli* (PDB: 2O1S) and 27% sequence identity to the E1 subunit of the multi-subunit PDH from *E. coli* (PDB: 1L8A).

The ChTK-N and ChTK-C genes were successfully cloned into the pLATE51 (Thermo Scientific) expression vector and over expressed in *E. coli* BL21 DE3* using the auto induction ZYM-5052 medium (Studier, 2005). After purification by Ni-NTA affinity chromatography and size exclusion chromatography (GF-200) the purified recombinant proteins ran as single bands on SDS-PAGE at the calculated molecular mass of the His-tagged proteins (31.0 kDa and 33.2 kDa). The apparent molecular mass of the native enzyme after gel filtration chromatography indicated that ChTK-N eluted as a dimer (∼62 kDa) whereas ChTK-C was purified as a tetramer (∼132 kDa) (**Figure 1**). The active ChTK-F was reconstituted by mixing the two proteins ChTK-N (dimer) and ChTK-C (tetramer) at 4°C overnight (with an excess of ChTK-N) and was purified by size exclusion chromatography. The formation of the complex was analyzed using size exclusion chromatography (GF-200) (**Figure 1**) in which a slight decrease in MW can be observed between the ChTK-F and ChTK-C peaks as well as small ChTK-N peak corresponding to the excess ChTK-N used in the reaction. The purity of the collected ChTK-F fraction was sufficient for crystallization of the complex.

A 'Split-Gene' TK

TK enzymes are widespread in nature with most eukaryotic and bacterial genomes containing at least one TK gene encoding a single protein of 600–700 amino acids. The presence of a full-length or 'split-gene' TK in some bacterial and archaeal genomes has been investigated using a BLAST search with the EcTK or ChTK protein sequence as a reference. Previous studies have reported that archaea are either missing the TK enzyme altogether or the enzyme is encoded by two different genes that can either be situated next to each other on the genome (*Sulfolobus solfataricus)* (She et al., 2001) or can be spread across the genome (*Methancaldococcus jannaschii*) (Soderberg, 2005). The thermophilic and hyper-thermophilic bacteria have been reported to possess either both the 'split-gene' and the full-length TK enzyme (*Thermotoga maritima*) (Rodionova et al., 2012) or just the full-length TK enzyme (*G. stearothermophilus*) (Egan et al., 2017). Mesophilic bacteria seem to only have the full-length TK enzyme (*E. coli*, *Lactobacillus plantarum*). However, the *C. hydrogenoformans* has been found in this study to not contain a full-length TK encoding gene and to only code for a 'split-gene' which is novel for any organism outside of the archaeal kingdom. The genome of the hyperthermophilic archaeon *Nanoarchaeum equitans* contains numerous examples of 'split-gene' proteins that are encoded by single genes in other archaea and it has been suggested that multi-domain proteins such as TK might have evolved from the fusion of these different 'split-genes'. It has been proposed that the presence of such 'split-genes' in a microorganism could reflect its ancestral state (Waters et al., 2003) and also be involved in the evolutionary path of new enzymes.

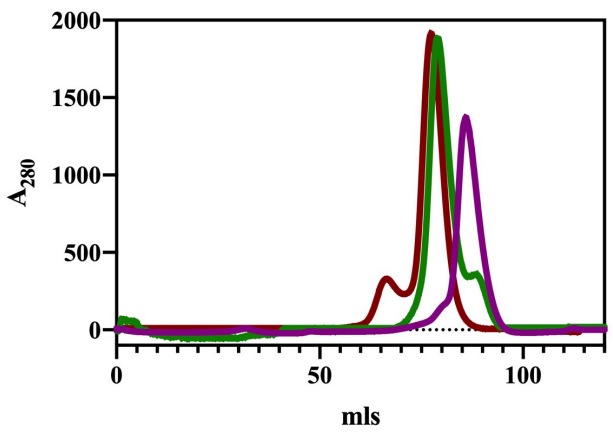


FIGURE 1 | Gel filtration chromatography elution (Superdex 200) of ChTK-C, ChTK-N and ChTK-F. ChTK-C with apparent molecular mass suggesting a tetramer (∼132 kDa) (red); ChTK-N was eluted as a dimer (∼62 kDa) (purple); and ChTK-F was eluted as a heterotetramer (green). A slight decrease in the MW can be observed between the ChTK-F and ChTK-C peak as well as a small ChTK-N peak in the ChTK-F trace corresponding to the excess ChTK-N used in the reaction.

Enzyme Activity

Initial activity measurements of the gene products ChTK-N and ChTK-C and reconstituted enzyme, ChTK-F were carried out using a colorimetric tetrazolium red assay (Smith et al., 2006) that showed that the individual ChTK-N and ChTK-C did not possess TK activity but the reconstituted ChTK-F was active. This activity was confirmed using a more accurate HPLC assay (**Supplementary Figure S1**) and this subsequently has led to the ChTK-F to be biochemically characterized.

Using the substrate glycolaldehyde as the aldehyde acceptor and HPA (100 mM) as the ketol donor the ChTK-F was found to have a Km of 41 ± 4 mM and a kcat/Km of 0.25 ± 0.01 s^{-1} mM^{-1}. These kinetic constants show that the enzyme has a higher Km and lower kcat/Km toward these substrates in comparison to both the yeast and EcTK enzyme (Yi et al., 2012) which is expected for a hyperthermophilic enzyme assayed at room temperature. The kinetic data is similar to that previously reported for a mutant EcTK (D469E) enzyme (Yi et al., 2012) but this amino acid substitution is not present in the ChTK-F enzyme. The reconstituted ChTK-F enzyme was assayed using a range of aldehyde acceptors and was shown to be active toward the bulkier substrates phenylacetaldehyde and cyclohexanecarboxaldehyde (**Supplementary Figure S5**). EcTK has been shown to have some activity toward these substrates however the yields were low and by-products were formed, which was attributed to the steric hindrance for binding of these bulkier substrates in the enzyme active site (Hobbs et al., 1993; Morris et al., 1996).

Temperature and pH Stability

To test the enzyme stability and activity at elevated temperatures the ChTK-F enzyme was incubated at increasing temperatures for 1 hour and then cooled to room temperature before being assayed. No loss of activity was observed for the ChTK-F after incubation for 1 hour at temperatures up to 70°C. The enzyme activity was reduced to ∼50% after incubation for 1 hour at 80°C and became denatured with a complete loss of activity after incubation at 90°C (**Figure 2A**), This is a small improvement in stability as measured from these experiments from that reported with the on the GsTK enzyme (Abdoul-Zabar et al., 2013).

Thermally stable proteins are often able to withstand a number of other denaturing conditions such as extremes of pH. The pH stability of the ChTK-F protein was tested by incubation for one hour in the range pH 5.0 – 12.0 (**Figure 2B**). The enzyme showed the highest activity at pH 8.0 and was able to retain ∼ 50% activity after incubation at pH 10. The ChTK-F pH at which the enzyme retained maximum activity is similar to that reported in the literature for the TK from the moderate thermophile *G. stearothermophilus*, pH 7.0–8.0 (Abdoul-Zabar et al., 2013), *D. geothermalis,* pH 8.0 (Bawn et al., 2018) and the mesophilic EcTK, pH 8.0–8.5 (Littlechild et al., 1995; Martin, 2008).

Solvent Stability

ChTK-F was incubated in a range of common organic solvents and its residual activity measured (**Figure 3**). The enzyme

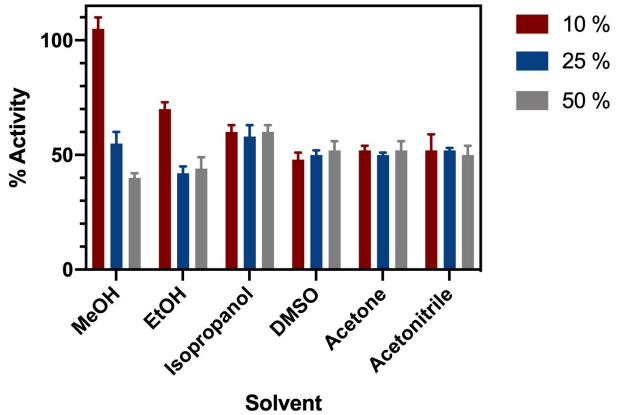


FIGURE 3 | Solvent Stability of ChTK-F. The percentage of relative enzyme activity for ChTK-F after incubation for 1 h in buffer and with 10% (Black Bar), 25% (Blue Bar), and 50% (Yellow Bar) of the organic solvents methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone.

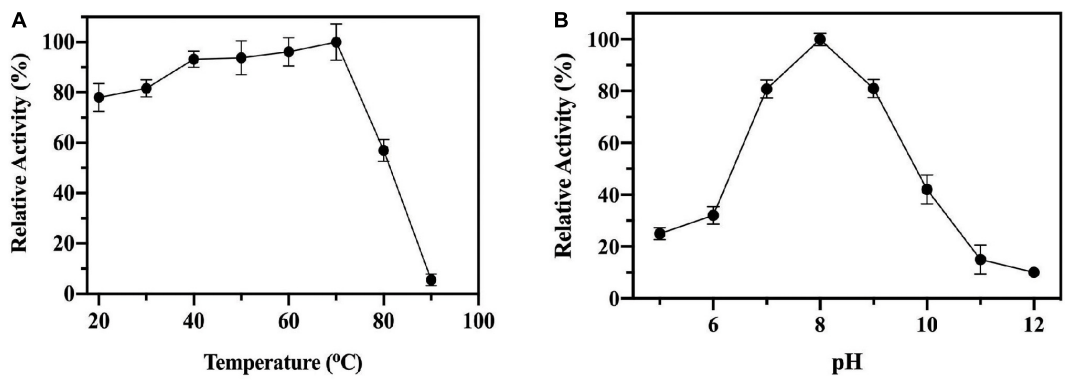


FIGURE 2 | The biochemical characterization of ChTK-F. **(A)** The percentage of relative enzyme activity for ChTK-F after incubation for 1 h at a range of temperatures indicated against a control kept at room temperature in a reaction with HPA as a ketol donor and glycolaldehyde as an acceptor. **(B)** The percentage of relative activity for ChTK-F after incubation for 1 h at different pHs.

retained ∼ 60% activity after incubation for 1 hour in buffer containing either 10%, 25% or 50% methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone. The lowest enzyme activity was in 25 – 50% ethanol where it retained ∼ 40% of its relative activity compared to the control. The final concentration of solvent in the enzyme assay was no greater than 7.5%. Such solvent stability agrees with the higher thermal stability of ChTK-F.

Crystallization and Structural Determination

Initial attempts to crystallize the reconstituted active fraction of ChTK-F resulted in crystals of a tetrameric ChTK-C. It appeared that 200 mM ammonium sulfate present in the crystallization conditions had disrupted the formation of the ChTK-F heterotetramer. Later crystallization trials using PEG based conditions and an excess of TPP and Ca^{2+} ions resulted in crystals of the ChTK-F α2β2 protein. These crystals belonged to the space group I222 with cell dimensions of $a = 123.0$, $b = 130.0$, $c = 165.9$ Å and diffracted to 2.1 Å resolution and contained TPP at partial occupancy. Changing the precipitant to malic acid (2.1 M DL-malic acid pH 7.0) produced crystals in the same space group but which diffracted to an improved resolution of 1.4 Å and had full occupancy of the TPP cofactor in the active site.

The CHTK-F structure shows that two ChTK-N subunits (**Figure 4A**) and two ChTK-C subunits (**Figure 4B**) had come together to form a structure similar to the dimeric structure seen for full length transketolases such as EcTK (PDB 1QGD) (Littlechild et al., 1995; Martin, 2008). The division of the ChTK-F heterotetramer into polypeptide chains resembles the structure of the mammalian PDH heterotetramer. As other TK-like group enzymes the ChTK-F is formed by 3 domains of the α/β type, the PP-domain (ChTK-N), the Pyr-domain (ChTK-C residues 1–162) and the C-terminal domain (ChTK-C residues 164–312). The PP and Pyr domains are usually connected by a long flexible linker region in other transketolases and much shorter linkers in DXPSs and bacterial PDHs, however in the ChTK-F structure the linker between the PP and Pyr domain is missing. The linker between the Pyr and the C- terminal domain remains intact (**Figure 5**). These features potentially allow more flexibility in the reconstituted ChTK-F structure than seen for other full length TKs.

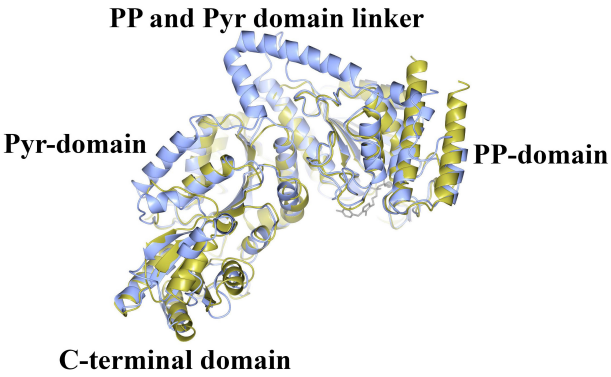


FIGURE 5 | A comparison of the structure of ChTK αβ heterodimer (gold) superimposed onto the monomer of EcTK (ice blue) showing the equivalent positions of the PP-domain, the Pyr-domain and the C-terminal domain. The linker between the PP and Pyr domains is present in EcTK but not in ChTK-F.

Hetero-Tetrameric Structure

The 'split-gene' ChTK-F protein forms a heterotetramer with a ChTK-C dimer and a ChTK-N dimer coming together to form a similar overall arrangement to that observed in full length TK and TK-like group enzymes (**Figure 6**). The overall shape of the ChTK-F tetramer can be represented as a trigonal prism with the dimensions 98 × 86 × 76 Å. The heterotetramer has a solvent accessible area of 36780 $Å^2$, the formation of the tetramer (in the presence of TPP) buries 17270 $Å^2$ or around 30% of the monomer solvent accessible area. However, in the absence of ChTK-C and TPP molecules, the interaction area between the two ChTK-N monomers is limited to around 800 $Å^2$ per monomer (6% of the ChTK-N monomer solvent accessible area) and the free ChTK-N in solution is likely to be in equilibrium between a monomer and a dimer (observed at high concentration of protein in size-exclusion chromatography). The ChTK-C dimer formation buries 1700 $Å^2$ per monomer (13% of its solvent accessible area), which suggests a dimer is a likely oligomeric

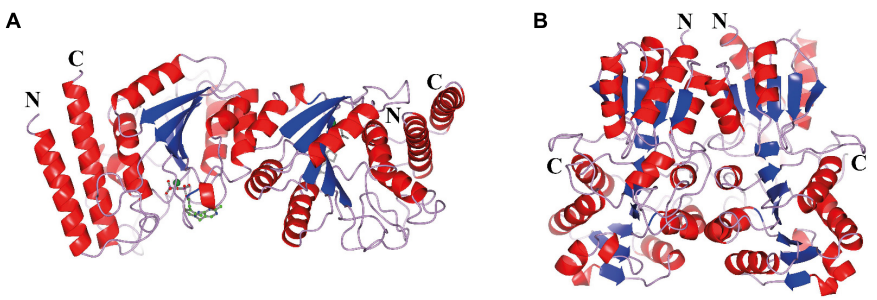


FIGURE 4 | The oligomeric state of the different ChTK polypeptides. **(A)** The modeled ChTK-N dimer is shown as a cartoon model with the N- and C-termini indicated. The catalytic Ca^{2+} is shown as a green sphere and the TPP cofactor as a stick model. **(B)** The ChTK-C homotetramer as determined by X-ray crystallography is shown as a cartoon model with N- and C termini indicated. The helices are shown in red, the strands are shown in blue and the loops in lilac. The **Figures 4–7B,C**, **8** were prepared using CCP4mg (McNicholas et al., 2011).

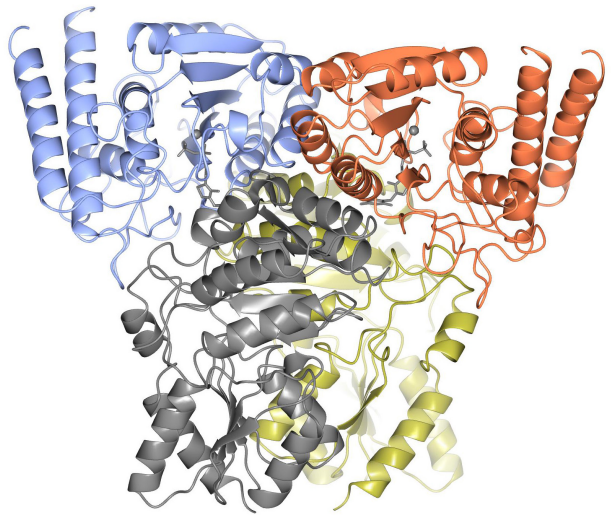


FIGURE 6 | The overall structure of ChTK-F heterotetramer is shown as a cartoon model with the ChTK-N monomers shown in blue and red and the ChTK-C monomers shown in gray and gold with the Ca^{2+} and TPP cofactor shown as stick models in gray.

form of ChTK-C in solution. The ChTK-C dimer has exposed surface hydrophobic patches on the interface with TPP and ChTK-N. Formation of the ChTK-C homotetramer observed in two crystal forms buries these hydrophobic patches, however only around 5% of each ChTK-C surface accessible area is buried. Such a tetramer is observed in the two crystal forms and in size exclusion chromatography, however it is unlikely to be stable at the lower concentrations of the protein in solution.

Active Site

The enzyme contains two identical active sites that contain the cofactor TPP and a Ca^{2+} ion bound between a subunit of ChTK-C and its corresponding ChTK-N. The TPP molecule has three moieties, a diphosphate group, a thiazolium ring and an amino-pyrimidine ring and adopts the higher energy V-conformation present in most TPP-dependant protein structures (**Figure 7A**) (Leeper et al., 2005). The diphosphate group is held in place by a number of hydrogen bonds formed with residues of the CHTK-N monomer (Lys68, His70, Gly149, Glu150, Asn178 and Lys240). The Ca^{2+} ion is held in place by interactions with residues Asp148, Asn178, Leu180 and the oxygen atoms on the diphosphate group. The thiazolium ring is held in place by hydrophobic contacts from both the ChTK-C and ChTK-N monomers (Leu31C, Ile53C, Leu119N and Ile182N). The amino-pyrimidine ring is held in place by another series of hydrogen bonds from both ChTK-C and ChTK-N (Glu55C, Gly117N and Leu119N) as well as a π-π stacking interaction between the pyrimidine ring and Phe80 from ChTK-C (**Figures 7B,C**).

Previous site-directed mutagenesis studies of the EcTK have shown it is possible to change the substrate specificity of this enzyme to accept bulky aromatic substrates (Payongsri et al., 2012; Payongsri et al., 2015). These studies have shown that mutations at R358, S385, D469, and R520 (EcTK numbering) improve the enzyme activity and yield toward aromatic aldehyde substrates (**Supplementary Figure S2**). A study by Saravanan et al. (2017b) also attempted to improve the GsTK enzymes activity toward aryl substrates and found another residue L382 (EcTK numbering) was also important for aromatic substrate binding. As none of the amino acids substitutions are present in the ChTK-F enzyme they cannot explain the substrate specificity. A computational docking approach was used to investigate potential the amino acid residues involved in substrate binding in the ChTK-F reconstituted enzyme. Docking studies were performed using Autodock (Kumar et al., 2018) to further rationalize the substrate specificity of the ChTK-F toward bulkier aldehyde substrates. The docking studies revealed a binding orientation of the phenylacetaldehyde substrate in the enzyme active site that places the aromatic ring of this substrate in a position where it could be stabilized by a cation- π interaction with a positively charged lysine residue (K313 ChTK-F). This interaction could not occur in the EcTK as there is a proline residue at this position. The docking suggested that the other residues in ChTK-F that interact with phenylacetaldehyde in the model obtained are in a similar position as in EcTK (**Supplementary Figure S3**).

Like other TK enzymes ChTK-F has a broad substrate specificity but was found not to accept pyruvate as a ketol donor. This is probably because hydroxypyruvate requires a lower energy of activation for the two-carbon unit transfer in comparison to pyruvate. We know that other TK-like enzymes (such as DXPS) are capable of activating the pyruvate molecule as a two-carbon unit donor but are only able to catalyze a reaction with a single sugar acceptor.

Structural Basis for Thermostability

The structural determination of the 'split-gene' ChTK-F has allowed further insight into the thermal stability of this enzyme when compared to the mesophilic EcTK. ChTK-F has been shown in this study to retain up to 50% activity when heated at 80°C for 1 hour while the EcTK enzyme loses most of its activity at 60°C (Jahromi et al., 2011) and the GsTK enzyme loses activity after 10 min at 75°C (Abdoul-Zabar et al., 2013). A number of structural features are thought to confer thermostability to proteins and these include higher number and clustering of salt bridges, the shortening of surface loops and an increase in hydrophobicity at domain and monomer interfaces (Littlechild et al., 2007, 2013). Previous studies have shown that the TPP cofactor binding plays an important role in preventing deactivation and aggregation of the EcTK at extreme pH, temperature and in the presence of organic solvents (Dalby et al., 2007; Martinez-Torres et al., 2007; Jahromi et al., 2011). This TPP binding is controlled by the formation of two cofactor binding loops containing residues 185–192 and 382–392 (EcTK numbering which correspond to 178–185 and 311–321 in ChTK-F numbering). Attempts have been made to make the EcTK more thermostable by engineering these loops to resemble their equivalent loops of the TK from *T. thermophilus* (TtTK, PDB: 2E6K, 32% identity). In the first cofactor loop Morris et al. (1996) made two mutations (G186R and H192P). The

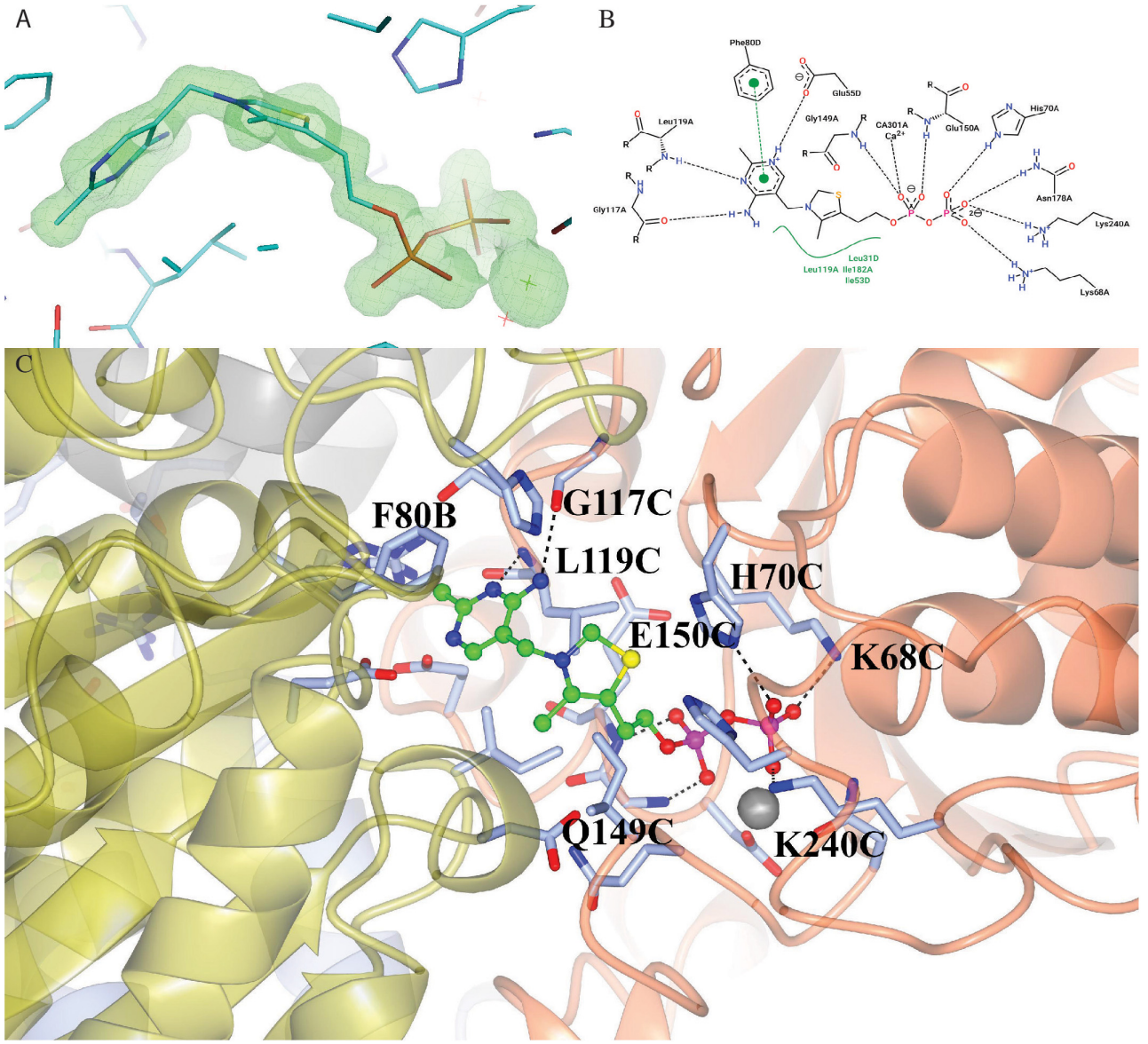


FIGURE 7 | (A) Electron- density map showing the TPP and Ca^{2+} ligands bound in the active site of ChTK-F. The weighted F_0-F_c omit map, where ligands where not used in the phase calculations, is shown in green and is contoured at 5.0 sigma. The neighboring residues are shown as stick models and the Ca^{2+} ion is shown as a green star. The figure was prepared using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. **(B)** A 2-D representation of the TPP binding site between ChTK-N (Chain-A) and ChTK-C (Chain-D). Hydrogen, ion and metal interactions between protein and ligands are drawn as dashed lines. Hydrophobic contacts are represented indirectly by spline section highlighting of the hydrophobic part of the ligand and the label of the contacting amino acid. **(C)** A cartoon model of the TPP and Ca^{2+} binding site between ChTK-N (ice-blue) and ChTK-C (gray) subunits. The TPP molecule is shown as a ball and stick model with carbon atoms colored in green. Amino acid side chains of residues implicated in ligand binding are shown stick models with the hydrogen bonds as black dashes. The Ca^{2+} ion is shown as a gray sphere with its interactions with the TPP molecule and amino acid side chains as red dashes.

most beneficial of these was H192P which not only increased the activity at 25°C but also increased the Topt to 60°C and retained more activity after heating at 60°C for 1 hour (Morris et al., 2016). This proline residue is conserved in the thermophilic ChTK-F enzyme and is located structurally in the same position as the TtTK enzyme (**Supplementary Figure S2**). Many thermophilic organisms, usually those with a high GC content in their DNA such as *T. thermophilus* contain more proline residues in the loop regions and these can contribute to high thermostability of the proteins (Suzuki et al., 1987). A comparison reveals that the EcTK and ChTK-F have the same number of prolines (30) compared to the higher number in TtTK (50). While this proline residue clearly plays a role in the enzyme performance at higher temperatures it is a combination of a number of factors that leads to ChTK-F stability and activity at extreme conditions.

Comparison to Other TPP Containing Enzymes

The ChTK-F enzyme was compared to the DXPS protein sequences from various organisms in an attempt to identify mutations that will allow the enzyme to use pyruvate as a substrate instead of the more expensive hydroxypyruvate. The sequence alignment of various TK and DXPS sequences show a high sequence homology in the residues shown to be involved in donor substrate binding (H66, H100, G114 and H473- EcTK numbering (**Supplementary Figure S2**). The exception to this is the TK His100 (EcTK numbering) residue which is consistently replaced by phenylalanine in DXPS (**Supplementary Figure S4**). While the mutation H100F in GsTK increased its activity toward pyruvate (1/10th that of DXPS) the mutation H100L increased it even further (1/3rd that of DXPS) (Saravanan et al., 2017a). As for other TKs the reconstituted ChTK-F has a histidine residue at this position.

There are also some differences in the way that the TKs and DXPSs bind the cofactor TPP. While the residues and domains (PP- and Pyr-) that are used to bind TPP remain highly similar between the two enzymes the TK enzyme binds TPP between domains on two different monomers whereas DXPS binds TPP between domains on the same monomer.

When comparing the structure of the 'full-length' TK enzymes with the 'split-gene', PDH and DXPS enzymes there is a linker (~ 70 amino acids) that bridges the PP and Pyr domains that is not present in the 'split-gene' TKs, PDH and DXPS (**Figure 8**). One benefit of the 'split-gene' is that absence of the linker allows the enzyme to have 'space' between the two domains that could account for its ability to use the bulkier phenylacetaldehyde substrate demonstrated in this study.

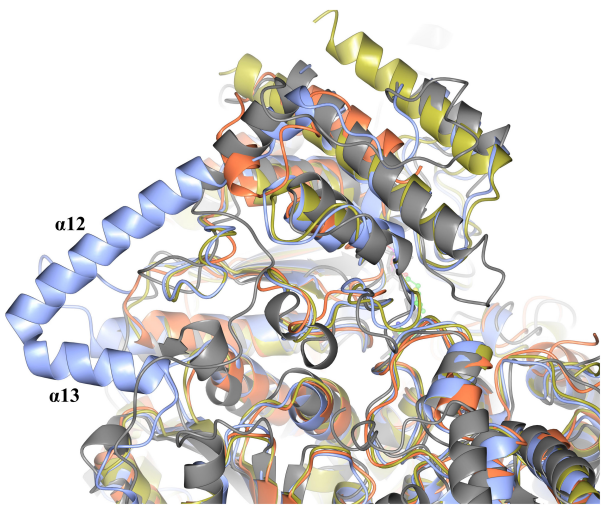


FIGURE 8 | Structural superposition showing the cartoon representation of two helices (α12 and α13) that are located on the linker region between the PP and Pyr domains of EcTK (cyan) which are not in ChTK-F (purple) or the other TK related enzymes DXPS (gray) and PDH (yellow).

TABLE 1 | Summary of crystallographic statistics.

Data collection statistics		
Crystal	**ChTK β4**	**ChTK α2β2**
Beamline	I04 DIAMOND	I04-1 DIAMOND
Wavelength (Å)	0.9795	0.9200
Space group	P43	I222
Unit Cell Parameters (Å)	$a = b = 101.9$, $c = 164.6$	$a = 123.0$, $b = 130.0$, $c = 165.9$
Total reflections[a]	687,440 (33,906)	1,762,697 (37,398)
Unique reflections[a]	131,289 (6,524)	288,149 (11,686)
Completeness (%)[a]	99.8 (99.8)	97.7 (81.3)
Rmeas (%)[a,b]	5.3 (424.5)	5.6 (161.9)
Mean I/σ[a]	13.1 (0.3)	16.4 (0.8)
Resolution range (Å)[a]	86.62 – 1.90 (1.93 –1.90)	60.80 – 1.34 (1.36 – 1.34)
CC1/2[a,c]	0.999 (0.339)	1.000 (0.266)
Wilson B-factor[d] (Å2)	53.2	23.4
Refinement Statistics		
Rwork	0.178	0.141
Rfree	0.207	0.159
Number of non-hydrogen atoms	9,872	11,492
Macromolecules	9,436	9,864
Solvent	436	1,628
Protein residues	1,213	1,194
RMS bond lengths (Å)	0.008	0.007
RMS bond angles (°)	1.60	1.47
Ramachandran favored (%)[e]	96.48	97.37
Ramachandran outliers (%)[e]	0.42	0.0
Clashscore[e]	5.98	5.04
Average B-factor protein (Å2)	59.4	20.5
Average B-factor solvent (Å2)	67.2	38.2
PDB code	6YAJ	6YAK

[a] *Values for the highest resolution shell are given in parentheses.*
[b] *Rmeas* = $\Sigma h[m/(m-1)]1/2\Sigma|ilh, i- < Ih >|/\Sigma h\Sigma ilh, i.$
[c] *CC1/2 is defined in Karplus and Diederichs (2012).*
[d] *Wilson B-factor was estimated by SFCHECK (Vaguine et al., 1999).*
[e] *The Ramachandran statistics and clashscore statistics were calculated using MOLPROBITY (Williams et al., 2018).*

CONCLUSION

The individual 'split-gene' products of this novel TK enzyme have been shown to have no activity in either a tetrazolium red based colorimetric assay or the more accurate HPLC assay. This is not unexpected since the crystal structure of ChTK-F shows that the PP and Pyr domains lie on separate parts of the gene and both are required to bind the cofactor TPP which is essential for the reaction to take place. The 'split-gene' TKs are only present in archaea and some thermophilic bacteria (although then a full-length TK gene is also present) so *C. hydrogenoformans* is unique in being the only bacteria to possess only a 'split-gene' TK.

The 'split-gene' products of ChTK can be separately expressed and purified prior to being incubated together and further purified by size exclusion chromatography to yield an active

transketolase enzyme (ChTK-F). This reconstituted enzyme has been shown to be active toward a broad range of bulkier aldehyde acceptors when used in the industrially important reaction using hydroxypyruvate as the ketol donor. The reconstituted enzyme has been biochemically characterized and has high thermal stability, is active at a range of pHs (pH 6–10, pH 8 optimum) and stable in 50% of a range of commonly used organic solvents.

The structure of the reconstituted enzyme reveals a heterotetrameric oligomeric state with two ChTK-N and two ChTK-C subunits forming the ChTK-F. The overall structure is similar to other full-length TK enzymes with TPP bound in the active site between the PP and Pyr domains. The major difference observed when the structures are compared is the absence of the linker between the PP and Pyr domains which could explain the ability of ChTK-F to turn over the bulkier substrates phenylacetaldehyde and cyclohexanecarboxaldehyde. The extensively studied EcTK is unable to use the same range of bulkier substrates which limits its synthetic range. Docking studies with ChTK-F involving phenylacetaldehyde have revealed a lysine (K33) residue that has the potential to help this substrate to be positioned during the reaction via a cation- π interaction with the positively charged lysine residue and the phenyl ring on the substrate (**Supplementary Figure S3**).

The high thermal stability of the ChTK-F enzyme can be explained by a combination of factors including the presence of the proline residue observed in other thermophilic TKs which when introduced into EcTK resulted in increased thermal stability of this mutant enzyme (Morris et al., 2016). The structure of the thermostable ChTK-F could help to further our understanding of thermostability of TK enzymes by predicting further mutations that could further increase the thermostability of mesophilic TKs. Further structural comparisons with the DXPS enzymes and ChTK-F could be used to predict mutations that will allow this enzyme to use a cheap substrate pyruvate as a ketol donor while keeping a broad acceptor substrate range to create high value chemicals.

The reconstituted active 'split-gene' TK enzyme ChTK-F has been shown in this study to have potential applications for industrial biocatalysis due to its ability to use bulkier aldehyde substrates and its overall stability to higher temperatures, a wider pH range and its tolerance to commonly used organic solvents.

MATERIALS AND METHODS

Materials

All reagents were obtained from Sigma-Aldrich (Buchs, Switzerland) unless otherwise stated. The chromatography columns were obtained from GE Healthcare (Little Chalfont, United Kingdom). The expression vector (pLATE51) was obtained from Thermofisher Scientific, (Rochford, United Kingdom).

Cloning, Expression, and Purification of ChTK-N and ChTK-C and Reconstitution of ChTK-F

The 'split-genes' were both cloned into the LIC site of pLATE51 vector and the expression carried out in *E. coli* BL21 DE3* cells. Cells were grown in 500 ml ZYM 5052 medium at 20°C and grown for 48 h. Cells were harvested by centrifugation (4700 x *g* at 18°C) and re-suspended in 50 mM Tris-HCl, pH 7.2, 2.4 mM TPP, 9 mM $CaCl_2$, 0.5 mM NaCl, and 20 mM imidazole. The cells were disrupted by sonication at 10 μm (Soniprep150; MSE, London, United Kingdom) on ice for 4 min and the cell debris was removed by centrifugation at 20 000 x *g* at 4°C for 30 min. The protein was expressed in the soluble fraction and the clarified cell lysate was then heat-treated at 50°C for 30 min before being centrifuged at 20 000 *g* at 4°C for 30 min to remove any denatured proteins. The protein was purified using a 1 ml HisTrap FF crude column (GE Healthcare, Little Chalfont, United Kingdom) using a gradient from 20 to 500 mM imidazole in 50 mM Tris-HCl pH 7.2, 2.4 mM TPP, 9 mM $CaCl_2$, 0.5 mM NaCl. The enzyme was then applied to a calibrated Superdex 200 HiLoad 16/60 gel filtration (GF) column (GE Healthcare, Little Chalfont, United Kingdom) and eluted with one column volume of 10 mM HEPES, 0.1 M NaCl, pH 7.2 at 1.0 ml min^{-1} with a yield of purified protein of 3.5 mg per liter of cell culture. Both individual enzymes (CHTK-N and CHTK-C) after GF were mixed in equimolar concentrations and left overnight at 4°C before being re-run on the calibrated Superdex 200 GF column.

Colorimetric Assay

A reaction mixture (90 μl) containing propanal (50 mM), lithium hydroxypyruvate (LiHPA) (50 mM), TPP (2.4 mM), $CaCl_2$ (9 mM) and TK sample (50% total volume – CHTK-C, CHTK-N and CHTK-F) in glycylglycine (50 mM, pH7.5) was incubated at 20°C for 17 hrs. 10 μl of the reaction mixture was then transferred to a micro-well containing MP-carbonate resin (Biotage AB) (10 mg) and glycylglycine buffer (90 μl, 50 mM, pH 7.0) and the mixture was incubated at 20°C for 3 hrs. 50 μl of this mixture (without the beads) was then diluted with further glycylglycine buffer (50 μl, 50 mM, pH 7.0), then tetrazolium red solution (20 μl, 0.2% 2,3,5 triphenyltetrazolium chloride in methanol) and finally 3 M NaOH (10 μl) with mixing.

HPLC Assay

The enzyme kinetics was carried out using the HPLC assay that can be followed at 210 nm by the production of erythrulose or consumption of the ketol donor β-hydroxypyruvate (HPA). The cofactor solution (170 μl) (2 mM TPP, 9 mM $CaCl_2$, 10 mM HEPES pH 7.2, 0.1 M NaCl) was added to the purified CHTK (N, C or F) sample (30 μl) and incubated for 20 mins at room temperature. The substrate solution (100 μl) (0–100 mM glycolaldehyde (GA), 0–100 mM LiHPA, 10 mM HEPES pH 7.2, 0.1 M NaCl) was added to start the reaction. Samples (20 μl) were taken at regular intervals, and the reaction quenched with 0.1% trifluoroacetic acid (TFA) in water (180 μl). Samples were applied to a Rezex RHM-Monosaccharide H^+ (8%) (Phenomenex), using

0.1% TFA (mobile phase) 60°C and analyzed at 210 nm for LiHPA reduction and L/D-erythrulose production.

Thermal Stability of ChTK-F

The thermostability of ChTK-F was investigated by incubating enzyme samples at a range of temperatures (20–90°C) for one hour using the gradient function in a SensOQuest LabCycler (Geneflow) before samples are cooled to 4°C and assayed for activity using the HPLC method described above.

pH Stability of ChTK-F

The pH stability of ChTK-F was investigated by incubating the enzyme at room temperature for one hour in buffer solutions in the range of pH 5–12. The buffers used were 100 mM sodium acetate pH 5.0, 100 mM sodium phosphate pH 6.0, 100 mM Tris-HCl pH 7.0 – 9.0, 100 mM glycine-NaOH pH 10.0, 100 mM sodium dihydrogen orthophosphate-NaOH pH 11.0 – 12.0.

Solvent Stability of ChTK

The residual activity of the ChTK-F was tested after incubation in a range of common organic solvents. The enzyme was incubated for 1 hour in buffer containing 25 mM Tris-HCl pH 7.5, 100 mM NaCl, and either 10%, 25% or 50% of methanol, ethanol, isopropanol, DMSO, acetonitrile and acetone. Samples were then assayed for activity using the HPLC method described above.

Crystallization, Data Collection, and Structural Determination

The ChTK-F was concentrated to ∼10 mg ml^{-1} using a 10 kDa Vivaspin membrane (Vivaproducts, Littleton, Massachusetts, United States) and microbatch crystallization trials were set up using an Oryx 6 crystallization robot (Douglas Instruments, Hungerford, United Kingdom) using the JCSG+, PACT premier, MIDAS and Morpheus screens (Molecular Dimensions, Newmarket, United Kingdom; Newman et al., 2005). The droplet consisted of a 50:50 ratio of protein solution to screen solution and was covered with Al's oil (a 50:50 mixture of silicone oil and paraffin) before being stored at 18°C and was regularly checked for growth of crystals using a light microscope.

The first crystals appeared in MIDAS-plus C10 (35% w/v polyacrylate 2100 sodium salt, 0.2 M ammonium sulfate, 0.1 M HEPES-NaOH pH7.5). The crystals were frozen in cryoprotectant containing 30% PEG 400 and indexed in two space groups $P4_3$ and $P4_32_12$ ($a = b = 92.4$, $c = 170.2$ Å). Both crystal forms diffracted to 1.9Å, however the MR search for the ChTK-F components using MORDA molecular replacement pipeline (Vagin and Lebedev, 2015) could not position any ChTK-N in the crystal. Instead, four copies of the ChTK-C component were located in the $P4_3$ crystal form (and 2 copies in the $P4_32_12$ crystal) and these crystals only contain the ChTK-C component. Both crystal forms were subject to preliminary refinement. The ChTK-C homo-tetramers formed were similar in both crystal forms and the $P4_3$ form was chosen for further rebuilding on the basis of its better refinement statistics. The presence of sulfate ions in the crystallization conditions affected binding of the cofactor TPP resulting in the breakdown of ChTK-F. The crystallization of the protein in PEG-based conditions (0.1 MMT Buffer (malic acid/MES/Tris-HCl) pH 7.0, 25% w/v PEG 1500) resulted in crystals which diffracted to 2.1 Å and contained the intact CHTK-F heterotetramer with partial occupancy of TPP in the active site. An increase in the concentration of TPP in the crystallization conditions was found to prevent crystal growth.

However, when crystallized in the presence of DL-malic acid pH 7.0 and increased concentration of TPP [JCSG-plus F8 (2.1 M DL-malic acid pH 7.0)] better crystals were produced. The crystal was cryocooled in liquid N_2 straight from the droplet. Data were collected on beamline I04-1 at the Diamond Synchrotron light source (Didcot, United Kingdom) at 100 K in a stream of gaseous nitrogen using a PILATUS detector. Data were processed in space group I222 (**Table 1**) and scaled using XDS (Kabsch, 2010) and AIMLESS (Evans and Murshudov, 2013) in the xia2 (Winter et al., 2013) pipeline. All further data and model manipulation were carried out using the CCP4 suite of programs (Winn et al., 2011). Phases for the CHTK-F crystal were determined using the molecular-replacement method as implemented in MOLREP (Vagin and Teplyakov, 2010) using preliminary refined CHTK-C and the N-terminal part of the EcTK model (PDB: 1QGD).

Electron-density maps were calculated and the structure was positioned to give the best fit to both the $2F_o - F_c$ and $F_o - F_c$ maps. Maximum-likelihood refinement was performed using REFMAC 5 (Murshudov et al., 2011) after each session of model building performed in Coot (Emsley et al., 2010). Statistics of the data processing and the parameters of the final refined models are given in **Table 1**. The quality of the refined model was checked using PROCHECK (Laskowski et al., 1993) and MOLPROBITY (Williams et al., 2018). Images were created using the molecular-graphics programs PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC and CCP4mg (McNicholas et al., 2011).

Coordinates

The atomic coordinates and structure factors have been deposited in the Protein Data Bank as 6YAK ($\alpha_2\beta_2$) and 6YAJ (β_4).

Computational Docking of Phenylacetaldehyde in the ChTK Active-Site

Phenylacetaldehyde was docked into the active site of TPP bound ChTK-F structure elucidated in this study (PDB: 6YAK) using Autodock 4.2 (Morris et al., 2009). The ligand phenylacetaldehyde was obtained from structure data files in the PDB (Phenylacetaldehyde: HY1) and the explorable space for docking was defined as a cube 10 Å in length centered

at the carbanion on the thiazolium ring of TPP. Resulting docking solutions were studied using PyMOL.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.wwpdb.org/, 6YAK and 6YAJ.

AUTHOR CONTRIBUTIONS

PJ identified, cloned, over-expressed, and crystallized the protein. The enzyme was biochemically characterized by PJ, IC, and SD. The structure was determined by PJ, CS, and MI. JL coordinated the project and wrote the manuscript with MI and PJ with additional contributions from all authors.

FUNDING

This work was supported by the BBSRC (BB/L002035/1) as part of the THERMOGENE ERA-IB project.

ACKNOWLEDGMENTS

We would like to thank the Diamond Synchrotron Light Source for access to beamlines I04 and I04-1 (proposal 8889) and the beamline scientists for assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.592353/full#supplementary-material

REFERENCES

Abdoul-Zabar, J., Sorel, I., Hélaine, V., Charmantray, F., Devamani, T., Yi, D., et al. (2013). Thermostable transketolase from *Geobacillus stearothermophilus*: characterization and catalytic properties. *Adv. Synth. Catal.* 355, 116–128. doi: 10.1002/adsc.201200590

Arjunan, P., Nemeria, N., Brunskill, A., Chandrasekhar, K., Sax, M., Yan, Y., et al. (2002). Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85 Å resolution. *Biochemistry* 41, 5213–5221. doi: 10.1021/bi0118557

Bawn, M., Subrizi, F., Lye, G. J., Sheppard, T. D., Hailes, H. C., and Ward, J. M. (2018). One-pot, two-step transaminase and transketolase synthesis of l-glucoheptulose from l-arabinose. *Enzyme Microb. Technol.* 116, 16–22. doi: 10.1016/j.enzmictec.2018.05.006

Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* 78, 89–175. doi: 10.1128/MMBR.00041-13

Costelloe, S. J., Ward, J. M., and Dalby, P. A. (2008). Evolutionary analysis of the TPP-dependent enzyme family. *J. Mol. Evol.* 66, 36–49. doi: 10.1007/s00239-007-9056-2

Dalby, P. A., Aucamp, J. P., George, R., and Martinez-Torres, R. J. (2007). Structural stability of an enzyme biocatalyst. *Biochem. Soc. Trans.* 35, 1606–1609. doi: 10.1042/BST0351606

Egan, K., Kelleher, P., Field, D., Rea, M. C., Ross, R. P., Cotter, P. D., et al. (2017). Genome Sequence of *Geobacillus stearothermophilus* DSM 458, an antimicrobial-producing thermophilic bacterium, isolated from a sugar beet factory. *Genome Announc.* 5:3809. doi: 10.1128/genomeA.01172-17

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493

Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* 69, 1204–1214. doi: 10.1107/S0907444913000061

Fullam, E., Pojer, F., Bergfors, T., Jones, T. A., and Cole, S. T. (2012). Structure and function of the transketolase from *Mycobacterium tuberculosis* and comparison with the human enzyme. *Open Biol.* 2:110026. doi: 10.1098/rsob.110026

Gerhardt, S., Echt, S., Busch, M., Freigang, J., Auerbach, G., Bader, G., et al. (2003). Structure and properties of an engineered transketolase from maize. *Plant Physiol.* 132, 1941–1949. doi: 10.1104/pp.103.020982

Hibbert, E. G., Senussi, T., Smith, M. E., Costelloe, S. J., Ward, J. M., Hailes, H. C., et al. (2008). Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J. Biotechnol.* 134, 240–245. doi: 10.1016/j.jbiotec.2008.01.018

Hobbs, G. R., Lilly, M. D., Turner, N. J., Ward, J. M., Willets, A. J., and Woodley, J. M. (1993). Enzyme-catalysed carbon–carbon bond formation: use of transketolase from *Escherichia coli*. *J. Chem. Soc. Perkin Trans.* 1, 165–166. doi: 10.1039/p19930000165

Jahromi, R. R., Morris, P., Martinez-Torres, R. J., and Dalby, P. A. (2011). Structural stability of *E. coli* transketolase to temperature and pH denaturation. *J. Biotechnol.* 155, 209–216. doi: 10.1016/j.jbiotec.2011.06.023

Kabsch, W. (2010). Xds. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 125–132. doi: 10.1107/S0907444909047337

Karplus, P. A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030–1033. doi: 10.1126/science.1218231

Kato, M., Wynn, R. M., Chuang, J. L., Tso, S. C., Machius, M., Li, J., et al. (2008). Structural basis for inactivation of the human pyruvate dehydrogenase complex by phosphorylation: role of disordered phosphorylation loops. *Structure* 16, 1849–1859. doi: 10.1016/j.str.2008.10.010

Koonin, E. V., and Galperin, M. (2013). *Sequence — Evolution — Function*. Berlin: Springer Science & Business Media.

Kumar, K., Woo, S. M., Siu, T., Cortopassi, W. A., Duarte, F., and Paton, R. S. (2018). Cation-pi interactions in protein-ligand binding: theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* 9, 2655–2665. doi: 10.1039/c7sc04905f

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi: 10.1107/s0021889892009944

Leeper, F. J., Hawksley, D., Mann, S., Perez Melero, C., and Wood, M. D. (2005). Studies on thiamine diphosphate-dependent enzymes. *Biochem. Soc. Trans.* 33, 772–775. doi: 10.1042/BST0330772

Lilly, M. D., Chauhan, R., French, C., Gyamerah, M., Hobbs, G. R., Humphrey, A., et al. (1996). Carbon-carbon bond synthesis: the impact of rDNA technology on the production and use of *E. coli* transketolase. *Ann. N.Y. Acad. Sci.* 782, 513–525. doi: 10.1111/j.1749-6632.1996.tb40589.x

Lindqvist, Y., Schneider, G., Ermler, U., and Sundström, M. (1992). Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 A resolution. *EMBO J.* 11, 2373–2379. doi: 10.1002/j.1460-2075.1992.tb05301.x

Littlechild, J., Novak, H., James, P., and Sayer, C. (2013). "Mechanisms of thermal stability adopted by thermophilic proteins and their use in white biotechnology," in *Thermophilic Microbes in Environmental and Industrial Biotechnology: Biotechnology of Thermophiles*, eds T. Satyanarayana, J. Littlechild, and Y. Kawarabayasi (Dordrecht: Springer Netherlands), 481–507. doi: 10.1007/978-94-007-5899-5_19

Littlechild, J., Turner, N., Hobbs, G., Lilly, M., Rawas, A., and Watson, H. (1995). Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta Crystallogr. D. Biol. Crystallogr.* 51, 1074–1076. doi: 10.1107/S0907444995005415

Littlechild, J. A., Guy, J., Connelly, S., Mallett, L., Waddell, S., Rye, C. A., et al. (2007). Natural methods of protein stabilization: thermostable biocatalysts. *Biochem. Soc. Trans.* 35, 1558–1563. doi: 10.1042/BST0351558

Ludtke, S., Neumann, P., Erixon, K. M., Leeper, F., Kluger, R., Ficner, R., et al. (2013). Sub-angstrom-resolution crystallography reveals physical distortions that enhance reactivity of a covalent enzymatic intermediate. *Nat. Chem.* 5, 762–767. doi: 10.1038/nchem.1728

Marsden, S. R., Gjonaj, L., Eustace, S. J., and Hanefeld, U. (2017). Separating thermodynamics from Kinetics-A new understanding of the transketolase reaction. *ChemCatChem* 9, 1808–1814. doi: 10.1002/cctc.201601649

Martin, M. (2008). *Mechanistic Studies of Escherichia coli Transketolase*. Ph.D. thesis, University of Exeter, Exeter.

Martinez-Torres, R. J., Aucamp, J. P., George, R., and Dalby, P. A. (2007). Structural stability of *E. coli* transketolase to urea denaturation. *Enzyme Microb. Technol.* 41, 653–662. doi: 10.1016/j.enzmictec.2007.05.019

McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 386–394. doi: 10.1107/S0907444911007281

Mitschke, L., Parthier, C., Schroder-Tittmann, K., Coy, J., Ludtke, S., and Tittmann, K. (2010). The crystal structure of human transketolase and new insights into its mode of action. *J. Biol. Chem.* 285, 31559–31570. doi: 10.1074/jbc.M110.149955

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Morris, K. G., Smith, M. E. B., Turner, N. J., Lilly, M. D., Mitra, R. K., and Woodley, J. M. (1996). Transketolase from *Escherichia coli*: a practical procedure for using the biocatalyst for asymmetric carbon-carbon bond synthesis. *Tetrahedron Asymmetr.* 7, 2185–2188. doi: 10.1016/0957-4166(96)00266-2

Morris, P., Rios-Solis, L., Garcia-Arrazola, R., Lye, G. J., and Dalby, P. A. (2016). Impact of cofactor-binding loop mutations on thermotolerance and activity of *E. coli* transketolase. *Enzyme Microb. Technol.* 89, 85–91. doi: 10.1016/j.enzmictec.2016.04.003

Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G. E., Jordan, F., and Schneider, G. (1993). A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure* 1, 95–103. doi: 10.1016/0969-2126(93)90025-c

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., et al. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 355–367. doi: 10.1107/S0907444911001314

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., et al. (2005). Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr. D. Biol. Crystallogr.* 61, 1426–1431. doi: 10.1107/S0907444905024984

Nikkola, M., Lindqvist, Y., and Schneider, G. (1994). Refined structure of transketolase from *Saccharomyces cerevisiae* at 2.0 A resolution. *J. Mol. Biol.* 238, 387–404. doi: 10.1006/jmbi.1994.1299

Payongsri, P., Steadman, D., Hailes, H. C., and Dalby, P. A. (2015). Second generation engineering of transketolase for polar aromatic aldehyde substrates. *Enzyme Microb. Technol.* 71, 45–52. doi: 10.1016/j.enzmictec.2015.01.008

Payongsri, P., Steadman, D., Strafford, J., MacMurray, A., Hailes, H. C., and Dalby, P. A. (2012). Rational substrate and enzyme engineering of transketolase for aromatics. *Org. Biomol. Chem.* 10, 9021–9029. doi: 10.1039/c2ob25751c

Racker, E., De La Haba, G., and Leder, I. G. (1954). Transketolase-catalyzed utilization of fructose 6-phosphate and its significance in a glucose 6-phosphate oxidation cycle. *Arch. Biochem. Biophys.* 48, 238–240. doi: 10.1016/0003-9861(54)90331-8

Rodionova, I. A., Yang, C., Li, X., Kurnasov, O. V., Best, A. A., Osterman, A. L., et al. (2012). Diversity and versatility of the *Thermotoga maritima* sugar kinome. *J. Bacteriol.* 194, 5552–5563. doi: 10.1128/JB.01136-12

Saravanan, T., Junker, S., Kickstein, M., Hein, S., Link, M. K., Ranglack, J., et al. (2017a). Donor promiscuity of a thermostable transketolase by directed evolution: efficient complementation of 1-Deoxy-d-xylulose-5-phosphate synthase activity. *Angew. Chem. Int. Ed. Engl.* 56, 5358–5362. doi: 10.1002/anie.201701169

Saravanan, T., Reif, M.-L., Yi, D., Lorillière, M., Charmantray, F., Hecquet, L., et al. (2017b). Engineering a thermostable transketolase for arylated substrates. *Green Chemistry* 19, 481–489. doi: 10.1039/c6gc02017h

Schenk, G., Duggleby, R. G., and Nixon, P. F. (1998). Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int. J. Biochem. Cell Biol.* 30, 1297–1318. doi: 10.1016/s1357-2725(98)00095-8

Selivanov, V., Kovina, M., Kochevova, V., Meshalkina, L., and Kochetov, A. (2003). Studies of thiamin diphosphate binding to the yeast apotransketolase. *J. Mol. Catal. B Enzym* 26, 33–40. doi: 10.1016/S1381-1177(03)00115-2

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840. doi: 10.1073/pnas.141222098

Smith, M. E., Kaulmann, U., Ward, J. M., and Hailes, H. C. (2006). A colorimetric assay for screening transketolase activity. *Bioorg. Med. Chem.* 14, 7062–7065. doi: 10.1016/j.bmc.2006.06.008

Soderberg, T. (2005). Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes. *Archaea* 1, 347–352. doi: 10.1155/2005/314760

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234. doi: 10.1016/j.pep.2005.01.016

Suzuki, Y., Oishi, K., Nakano, H., and Nagayama, T. (1987). A strong correlation between the increase in number of proline residues and the rise in thermostability of five Bacillus oligo-1,6-glucosidases. *Appl. Microbiol. Biot.* 26, 546–551. doi: 10.1007/bf00253030

Svetlichny, V. A., Sokolova, T. G., Gerhardt, M., Ringpfeil, M., Kostrikina, N. A., and Zavarzin, G. A. (1991). *Carboxydothermus hydrogenoformans* gen. nov., sp. nov., a CO-utilizing Thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. *Syst. Appl. Microbiol.* 14, 254–260. doi: 10.1016/S0723-2020(11)80377-2

Vagin, A., and Lebedev, A. (2015). MoRDa, an automatic molecular replacement pipeline. *Acta Crystallogr. Section A Foundations Adv.* 71:s19. doi: 10.1107/s2053273315099672

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 22–25. doi: 10.1107/S0907444909042589

Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D. Biol. Crystallogr.* 55, 191–205. doi: 10.1107/S0907444998006684

Veitch, N. J., Maugeri, D. A., Cazzulo, J. J., Lindqvist, Y., and Barrett, M. P. (2004). Transketolase from *Leishmania mexicana* has a dual subcellular localization. *Biochem. J.* 382, 759–767. doi: 10.1042/BJ20040459

Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12984–12988. doi: 10.1073/pnas.1735403100

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* 27, 293–315. doi: 10.1002/pro.3330

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 235–242. doi: 10.1107/S0907444910045749

Winter, G., Lobley, C. M., and Prince, S. M. (2013). Decision making in xia2. *Acta Crystallogr. D. Biol. Crystallogr.* 69, 1260–1273. doi: 10.1107/S0907444913015308

Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: the complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. *PLoS Genet.* 1:e65. doi: 10.1371/journal.pgen.0010065

Xiang, S., Usunow, G., Lange, G., Busch, M., and Tong, L. (2007). Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* 282, 2676–2682. doi: 10.1074/jbc.M610235200

Yi, D., Devamani, T., Abdoul-Zabar, J., Charmantray, F., Helaine, V., Hecquet, L., et al. (2012). A pH-based high-throughput assay for transketolase: fingerprinting of substrate tolerance and quantitative kinetics. *Chembiochem* 13, 2290–2300. doi: 10.1002/cbic.201200364

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 December 2020
doi: 10.3389/fmicb.2020.605771

Extracellular Production, Characterization, and Engineering of a Polyextremotolerant Subtilisin-Like Protease From Feather-Degrading *Thermoactinomyces vulgaris* Strain CDF

Yidi Ding[1], Yong Yang[1], Yuxia Ren[1], Jingying Xia[1], Feng Liu[1], Yu Li[1], Xiao-Feng Tang[1,2] and Bing Tang[1,2]**

[1] State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, China, [2] Hubei Provincial Cooperative Innovation Center of Industrial Fermentation, Wuhan, China

OPEN ACCESS

Edited by:
Junpei Zhou,
Yunnan Normal University, China

Reviewed by:
Bassem Jaouadi,
Centre of Biotechnology of Sfax, Tunisia
Jaeho Cha,
Pusan National University, South Korea
Cheng Zhou,
Institute of Microbiology, Chinese Academy of Sciences, China

***Correspondence:**
Bing Tang
tangb@whu.edu.cn
Xiao-Feng Tang
tangxf@whu.edu.cn

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: 13 September 2020
Accepted: 26 November 2020
Published: 21 December 2020

Citation:
Ding Y, Yang Y, Ren Y, Xia J, Liu F, Li Y, Tang X-F and Tang B (2020) Extracellular Production, Characterization, and Engineering of a Polyextremotolerant Subtilisin-Like Protease From Feather-Degrading Thermoactinomyces vulgaris Strain CDF. Front. Microbiol. 11:605771. doi: 10.3389/fmicb.2020.605771

Here, the gene encoding a subtilisin-like protease (protease AIs) was cloned from *Thermoactinomyces vulgaris* strain CDF and expressed in *Escherichia coli*. The recombinant enzyme was released into the culture medium of *E. coli* as a mature form (mAIs). Purified mAIs displayed optimal activity at 60–70°C and pH 10.0 using azo-casein as the substrate, and showed a half-life of 13.8 h at 70°C. Moreover, the activity of thermostable mAIs was comparable to or higher than those of mesophilic subtilisin Carlsberg and proteinase K at low temperatures (10–30°C). Protease AIs was also stable in several organic solvents and showed high compatibility with commercial laundry detergents. Notably, mAIs exhibited approximately 100% of its activity at 3 M NaCl, and showed enhanced thermostability with the increase of NaCl concentration up to 3 M. Protease AIs possesses an excess of solvent-accessible acidic amino acid residues, which may account for the high halotolerance of the enzyme. Compared with homologous protease C2 from the same strain, protease AIs exhibits substantially lower activity toward insoluble keratin substrates but efficiently hydrolyzes soluble keratin released from chicken feathers. Additionally, direct substitution of the substrate-binding site of protease AIs with that of protease C2 improves its activity against insoluble keratin substrates. By virtue of its polyextremotolerant attribute and kerationolytic capacity, protease AIs may find broad applications in various industries such as laundry detergents, food processing, non-aqueous biocatalysis, and feather processing.

Keywords: serine protease, thermostability, halotolerance, organic solvent, detergent, keratinase

INTRODUCTION

The members of subtilisin-like serine proteases (subtilases) superfamily are widely distributed in bacteria, archaea, and eukaryotes, and contribute to important biological processes such as protein metabolism, nutrition, protein processing, and pathogen invasion (Siezen and Leunissen, 1997; Rao et al., 1998). Subtilases have been extensively studied not only to provide insight into the mechanism

of enzyme catalysis and the structure-function relationship of proteins, but also because of their significant applications in detergents, leather processing, food and medicine fields (Gupta et al., 2002). In addition, many subtilases are able to hydrolyze keratin, which has a highly rigid structure rendered by extensive cross-linkages of disulfide bonds and is resistant to hydrolysis by commonly known proteases like trypsin and pepsin. Microbial/enzymatic degradation is regarded as an environmentally friendly approach to recycle keratin-containing wastes from poultry and leather industries (Gupta and Ramnani, 2006), and some keratinolytic microorganisms and proteases have been patented (Shih and William, 1992; Burtt and Ichida, 1999).

Subtilases from extremophiles have attracted increasing attention as promising materials for understanding the molecular basis of protein adaptation to harsh environments including extreme temperature, pH, and salt concentration etc., and studying these enzymes also greatly expands the reaction conditions of biocatalysis (Atomi et al., 2011; Elleuche et al., 2015; Salwan and Sharma, 2019). Thermostability is one of the main requirements for industrial enzymes since thermal inactivation represents a common problem in the application of biocatalysts. Meanwhile, protein substrates generally tend to be disordered at high temperatures and under highly alkaline conditions, rendering them more sensitive to proteolysis. Thermophile-derived thermostable subtilases with highly alkaline pH optima are highly desired in detergent and leather industries, and show great potential in the degradation of insoluble and hard-to-degrade animal proteins such as collagen, keratin, and prion proteins (Jang et al., 2002; Suzuki et al., 2006; Salwan and Sharma, 2019). In addition, subtilases from halophilic/halotolerant microorganisms usually function at high salt concentrations and could occasionally display polyextremotolerant attributes like tolerance to alkaline pH, elevated temperature, and organic solvent etc. (Mokashe et al., 2018); nevertheless, they are less thermostable than their counterparts from thermophiles. The exploration of more robust subtilases with polyextremotolerant attributes is not only scientifically significant for further understanding the mechanism of enzyme adaptation, but it is also of great practical importance in developing proteolytic biocatalysts with a wider versatility to multiple extreme conditions commonly encountered in industrial applications.

Thermoactinomyces species generally flourish in decaying hay, compost, and other high-temperature habitats, and have been used to produce a variety of thermostable proteases including metalloproteinase (Georgieva et al., 2000; Zabolotskaya et al., 2004; Majumder et al., 2013), carboxypeptidase (Akparov et al., 2015), glutamyl endopeptidase (Demidyuk et al., 1997), collagenase (Petrova et al., 2006), keratinase (Ignatova et al., 1999; Verma et al., 2016; Wang et al., 2019), and alkaline serine protease (Teplyakov et al., 1990; Gros et al., 1991; Tsuchiya et al., 1992, 1997; Lee et al., 1996). Among these enzymes, thermitase, the well-known alkaline serine protease from *Thermoactinomyces vulgaris*, serves as a model for studying enzyme structure–function relationship (Siezen and Leunissen, 1997). The genes encoding extracellular alkaline serine proteases from *Thermoactinomyces* sp. E79 (Lee et al., 1996) and *Thermoactinomyces* sp. YT06 (Wang et al., 2019) have been determined, showing that they belong to the thermitase family of subtilases. *T. vulgaris* strain CDF is capable of degrading chicken feathers at high temperatures (Cheng et al., 2009; Wang et al., 2015). A spore-associated subtilase (protease CDF) (Cheng et al., 2009), an extracellular subtilase (protease C2) (Wang et al., 2015), and a glutamyl endopeptidase (TS-GSE) (Liu et al., 2016) of the strain CDF have been characterized. The amino acid sequence of protease C2 is identical to that of protease E79 from *Thermoactinomyces* sp. E79 (Lee et al., 1996), albeit significant differences are seen in the upstream flanking regions of each gene. Protease C2 is able to efficiently hydrolyze chicken feathers at high temperatures and under alkaline conditions, representing a promising candidate for enzymatic processing keratinous wastes (Wang et al., 2015). By analyzing the complete genome sequence of the strain CDF (Li et al., 2019), a second gene encoding an extracellular subtilase (named protease Als; GenBank No. QBK13760) with an unusually low isoelectric point (pI) value of 4.26 was identified. In this study, the gene of protease Als was expressed in *Escherichia coli*, and enzymatic properties of the recombinant enzyme, including its tolerances to high temperature, high pH, high salinity, organic solvent, and detergent, were studied. The roles of charged amino acid residues on the enzyme surface in the polyextremotolerant behavior of protease Als were discussed. The capacity of protease Als to hydrolyze chicken feathers at high temperatures was also investigated, and keratinolytic activity of the enzyme was improved by modifying its substrate-binding site.

MATERIALS AND METHODS

Strains and Growth Conditions

Thermoactinomyces vulgaris strain CDF was isolated from the campus soil of Wuhan University, China (Cheng et al., 2009), and has been deposited in the China Center for Type Culture Collection (CCTCC) under the accession number AB206328. The strain CDF was grown at 55°C in Luria–Bertani (LB) medium and used for extraction of genomic DNA as described previously (Cheng et al., 2009). *E. coli* DH5α and *E. coli* BL21 (DE3) were used as hosts for cloning and protein expression, respectively, and were grown at 37°C in LB medium containing kanamycin (30 μg/ml) as needed.

Plasmid Construction and Mutagenesis

The plasmid pET26b (Novagen) was used as the vector for expressing recombinant proteins in *E. coli* BL21 (DE3). The primer sequences and the primer pairs used for PCR were listed in **Supplementary Tables 1, 2**, respectively. The DNA sequences encoding the protease Als precursor (pre-Als), the signal peptide-lacking proform (pro-Als), and the mature domain (mat-Als) were amplified from the genomic DNA of the strain CDF, and inserted into the *Nde*I-*Eco*RI site of pET26b to construct the expression plasmids pET26b-*pre-Als*, pET26b-*pro-Als*, and pET26b-*mat-Als* for the target proteins, each with a C-terminal

6 × His-tag. Using pET26b-*pre-Als* and pET26b-*mat-Als* as the templates, the expression plasmids for active-site variants of pre-Als (pre-S225A) and mat-Als (mat-S225A) were constructed by replacing the catalytic residue Ser225 with Ala via the QuikChange site-directed mutagenesis method (Papworth et al., 1996). The substrate-binding site variant of protease Als (AS14C) was constructed using the overlapping extension PCR method, as described previously (Bian et al., 2006). All recombinant plasmids were confirmed by DNA sequencing.

Expression and Purification

Escherichia coli BL21(DE3) cells harboring recombinant plasmids were cultured in LB medium, and the expression of recombinant proteins were carried out as described previously (Cheng et al., 2009). After induction with 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 6 h at 30°C, the *E. coli* cultures were centrifuged at 6,500 × *g* for 10 min at 4°C to separate the culture supernatants and the cells. After washing with buffer A (50 mM Tris–HCl, 10 mM $CaCl_2$, pH 8.0), *E. coli* cells were suspended in buffer A, followed by sonication on ice. Soluble and insoluble cellular fractions were separated by centrifugation at 13,400 × *g* for 10 min at 4°C. The recombinant proteins with a C-terminal 6× His-tag in the culture supernatants and the soluble cellular fractions were purified using affinity chromatography on a Ni^{2+}-charged Chelating Sepharose™ Fast Flow resin column (GE Healthcare, Little Chalfont, United Kingdom) as described previously (Bian et al., 2006). The purified mature protease C2 was prepared as described previously (Wang et al., 2015). The concentrations of purified enzyme samples were determined using the Bradford assay method (Bradford, 1976) with bovine serum albumin (BSA) as the standard.

Enzyme Activity Assays

The standard assay for azo-caseinolytic activity of the enzyme was carried out at 60°C for 15 min in 200 μl of reaction mixture containing 20 μl of the enzyme sample and 0.25% (w/v) azo-casein (Sigma, St. Louis, MO, United States) in buffer A. The reaction was terminated by adding 200 μl of 40% (w/v) trichloroacetic acid (TCA). After standing at room temperature (∼25°C) for 15 min, the mixture was centrifuged at 13,400 × *g* for 10 min, and the absorbance of the supernatant at 335 nm was measured in a 1-cm light-path cell. One unit (U) of activity was defined as the amount of enzyme required to increase the corresponding absorbance value by 0.01 unit per minute under the conditions described above.

The proteolytic activity of enzymes on the substrate *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (suc-AAPF-pNA) (Sigma) was measured at the temperatures indicated in buffer A containing 0.1 mM suc-AAPF-pNA. The activity was recorded by monitoring the initial velocity of suc-AAPF-pNA hydrolysis at 410 nm in a thermostated spectrophotometer (SP752; Shanghai Spectrum Instruments Co. Ltd, China). This velocity was calculated on the basis of an extinction coefficient for *p*-nitroaniline (pNA) of 8,480 M/cm at 410 nm (DelMar et al., 1979). One unit (U) of enzyme activity was defined as the amount of enzyme that produced 1 μmol of pNA per minute under assay conditions.

To determine the keratin/casein ratio of the enzyme, the proteolytic activity of the enzyme against keratin (TCI, Japan) or casein (Sigma) was measured as described previously (Jaouadi et al., 2010, 2014) with some modifications. The activity assay was carried out at 60°C for 30 min in 300 μl of reaction mixture containing 50 μl of suitably diluted enzyme solution and 1% (w/v) casein or keratin in buffer A. The reaction was terminated by adding 250 μl of 20% (w/v) TCA. After standing at room temperature (∼25°C) for 30 min, the mixture was centrifuged at 10,000 × *g* for 20 min. Thereafter, 100 μl of the supernatant was mixed with 500 μl of 0.5 M Na_2CO_3 and 100 μl of Folin-Ciocalteu's phenol reagent, followed by standing at room temperature for 30 min. After centrifugation at 10,000 × *g* for 20 min, the absorbance of the resulting supernatant at 660 nm was measured in a 1-cm light-path cell. One unit (U) of activity was defined as the amount of enzyme required to produce 1 μg of amino acid equivalent to tyrosine per minute under the conditions described above.

The proteolytic activity of the enzyme against insoluble substrate [keratin azure or collagen (Sigma)] was determined as described previously (Wang et al., 2015) with some modifications. Insoluble substrates were washed three times using buffer A. A reaction mixture (800 μl) containing 50 μl of the enzyme sample and 3 mg of the substrate in buffer A was incubated at 60°C for 40 min (collagen) or 2 h (keratin azure). The reaction was terminated by cooling on ice. After centrifugation at 13,400 × *g* for 10 min, the absorbance of the supernatant was measured in a 1-cm light-path cell at 595 nm for keratin azure or 280 nm for collagen. One unit (U) of activity was defined as the amount of enzyme required to increase the absorbance at 595 nm (keratin azure) or at 280 nm (collagen) by 0.01 unit per minute under the conditions described above.

Degradation of Chicken Feather and Soluble Feather Keratin

Enzymatic degradation of chicken feathers was conducted as described by Liang et al. (2010) with some modifications. Briefly, sterilized chicken feather (3 mg) was incubated with purified enzyme (100 μg/ml) at 60°C in buffer A containing 2% β-mercaptoethanol (β-ME). At different time intervals, the disintegration of feather was recorded photographically. The hydrolyzing activity of the enzyme toward soluble feather keratin was conducted as follows. Sterilized chicken feathers were cut into pieces ∼2–3 mm in length and incubated at 60°C for 2 h in buffer A containing 2% β-ME, followed by centrifugation at 13,400 × *g* for 10 min to collect the supernatant containing soluble feather keratin. After the addition of the enzyme (1 μg/ml) into the feather keratin solution, the reaction mixture (150 μl) was incubated at 60°C for 2 h. The reaction was terminated by adding 150 μl of 40% (w/v) TCA, and the precipitated degrading products of feather keratin were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

Effects of Surfactants, Oxidizing Agent, Organic Solvents and Detergents on Enzyme Activity and Stability

The enzyme (5 μg/ml) was pre-incubated at 40°C for 1 h in buffer A containing different concentrations of surfactants [sodium dodecyl sulfate (SDS), sodium dodecylbenzene sulfonate (SDBS), Triton X-100, Tween 20, and Tween 80], or in buffer B (50 mM Borate-NaOH, 10 mM $CaCl_2$, pH 8.0) containing different concentrations of H_2O_2. The pre-incubation of the enzyme with different concentrations of organic solvents [methanol, ethanol, isopropanol, acetone, dimethyl sulfoxide (DMSO)] was carried out in buffer A at 40°C for 1 h with shaking (Ibrahim et al., 2015). After incubation, the residual activity of the enzyme was determined by a standard assay for azo-caseinolytic activity at 60°C.

The effect of commercial laundry detergents on enzyme stability was investigated according to the method described by Ibrahim et al. (2015) with minor modification. The commercial laundry detergents include Tide (Procter & Gamble Co.), OMO (Unilever), Liby (Guangzhou Liby Enterprise Group Co.), and Whitecat (Shanghai Hutchison Whitecat Co., Ltd.). The detergent solutions (in tap water) were heat-treated at 90°C for 1 h to inactivate endogenous proteases. The enzyme (5 μg/ml) was pre-incubated at 40°C for 1 h or at 30°C for 24 h in buffer B containing 1.0% of the detergent. Thereafter, the residual activity of the enzyme against *N,N*-dimethylated casein (Sigma) was determined according to the method of Mechri et al. (2019), except that the reaction was carried out at 60°C for 30 min in buffer B.

Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis and Immunoblot Analyses

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was performed using the glycine-Tris buffer (King and Laemmli, 1971) or Tricine-Tris buffer systems (Schägger and von Jagow, 1987). To prevent self-degradation of the protease during sample preparation (boiling) or electrophoresis, the sample was precipitated by the addition of 20% TCA and then washed with acetone before being subjected to SDS-PAGE. The anti-His-tag monoclonal antibody (Novagen) was used for immunoblot analysis, as described previously (Cheng et al., 2009).

RESULTS

Extracellular Production of Protease Als in *E. coli*

Sequence alignment analysis shows that protease Als gene encodes a precursor comprising a predicted Sec-type signal peptide of 31 residues, an N-terminal propeptide of 79 residues, and a mature domain of 279 residues (**Figure 1**). The mature domain of protease Als shares 98.2, 66.8, and 65.6% amino acid sequence identities to protease KERTYT from *Thermoactinomyces* sp. YT06 (Wang et al., 2019), thermitase from *T. vulgaris* (Teplyakov et al., 1990), and protease C2 from *T. vulgaris* strain CDF (Wang et al., 2015), respectively.

The gene encoding protease Als precursor with a C-terminal 6× His-tag (pre-Als) was cloned and expressed in *E. coli*. For comparison purpose, a signal peptide-lacking proform (pro-Als) and active-site variants of the precursor (pre-S225A) and its mature domain (mat-S225A) were also constructed (**Figure 2A**). A 34-kDa product was detected in both total cellular protein (TCP) and culture supernatant of *E. coli* expressing pre-Als (**Figure 2B**). The 34-kDa product displayed the same apparent molecular mass as mat-S225A, and both could be detected using anti-His-tag monoclonal antibody (**Figure 2C**). These results demonstrate that recombinant pre-Als could be released into the culture medium and converted into its mature form (named mAls) by processing of the N-terminal propeptide. In contrast, pro-Als was detected only in TCP as the 34-kDa mAls (**Figure 2B**), suggesting that signal peptide is necessary for the release of the enzyme into the culture supernatant and does not prevent the maturation of the enzyme within the cell. Similar to the case of *E. coli* harboring a blank vector (control), no additional host cell proteins were detected in the culture supernatant of *E. coli* expressing pro-Als (**Figure 2B**), implying that intracellular accumulation of mAls does not cause detectable cell lysis. When pre-S225A was produced in *E. coli*, the recombinant protein with an apparent molecular mass of 45 kDa was also found in both TCP and culture supernatant, but the amount of the active-site variant in the culture supernatant was much lower than that of mAls derived from pre-Als (**Figure 2B**). These data suggest that the proteolytic activity of protease Als not only mediates the autoprocessing of the N-terminal propeptide but also contributes to the extracellular production of the enzyme in *E. coli*.

Protease Als Is a Thermostable Alkaline Enzyme With High Low-Temperature activity

The mature protease Als (mAls) with a His-tag at the C terminus was purified by affinity chromatography using a Ni^{2+}-charged column (**Figure 2B**). Using azo-casein or suc-AAPF-pNA as a substrate, the optimum temperature of purified mAls was determined to be 60–70°C at pH 8.0 (**Figure 3A**). The effect of pH on the enzyme activity was measured at 60°C over a pH range of 5.5–11.0, showing that mAls has an optimum pH of 10.0, with approximately 93% of this activity retained at pH 11.0 (**Figure 3B**). It was noticed that among the buffers used the Tris–HCl buffer is not supplemented with Na^+. When 100 mM NaCl was added in Tris–HCl buffers (pH 7.0–9.0), mAls exhibited higher activities (**Figure 3B**), implying that Na^+ could promote the activity of the enzyme. In the presence of 10 mM $CaCl_2$ at pH 8.0, mAls retained more than 90% of the original activity after incubation at 60°C for 12 h (**Supplementary Figure 1**), and showed half-lives of 13.8 h, 3.4 h, and 37 min at 70, 75, and 80°C, respectively (**Figure 3C**). At pH 10.0, mAls retained 20 and 50% of its initial activity after 1 h incubation at 60°C in the absence and presence of 10 mM $CaCl_2$, respectively (**Figure 3D**). These results

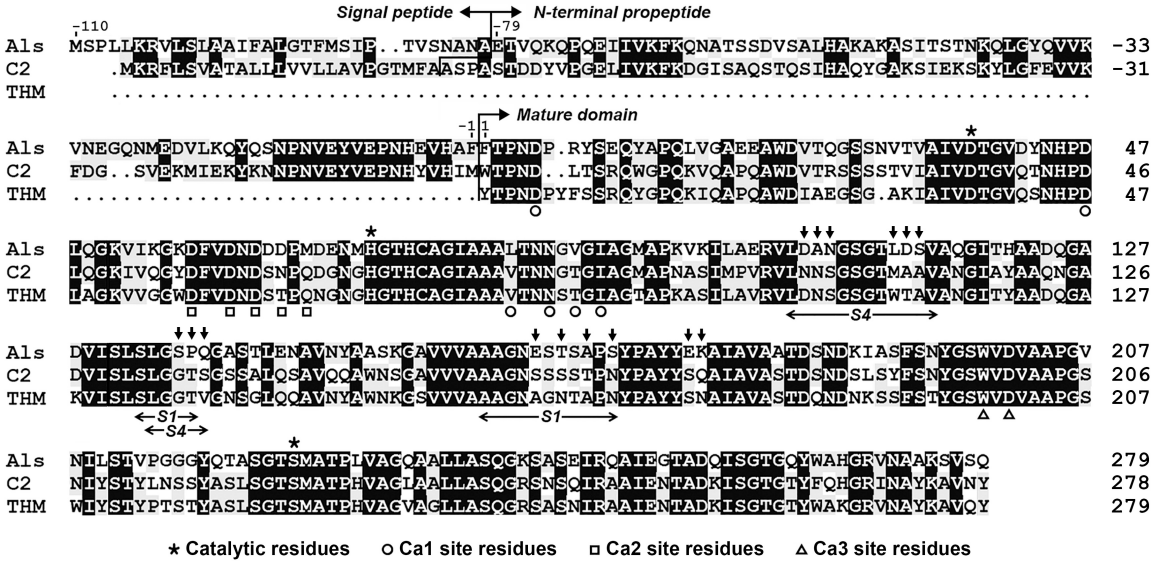


FIGURE 1 | Amino acid sequence alignment of protease Als (QBK13760) with protease C2 (ADD51544) and thermitase (THM, 1105242A). The signal peptide, N-terminal propeptide, mature domain, and the regions of the S1 and S4 pockets of the substrate-binding site are indicated. The catalytic residues and the residues of three Ca^{2+}-binding sites (Ca1, Ca2, and Ca3) in thermitase are marked. The residues indicated by vertical arrows above the sequence of protease Als represent those that were replaced by corresponding residues of protease C2 to construct the substrate-binding site variant AS14C. The amino acid residues are numbered starting from the N-terminus of the mature domain.

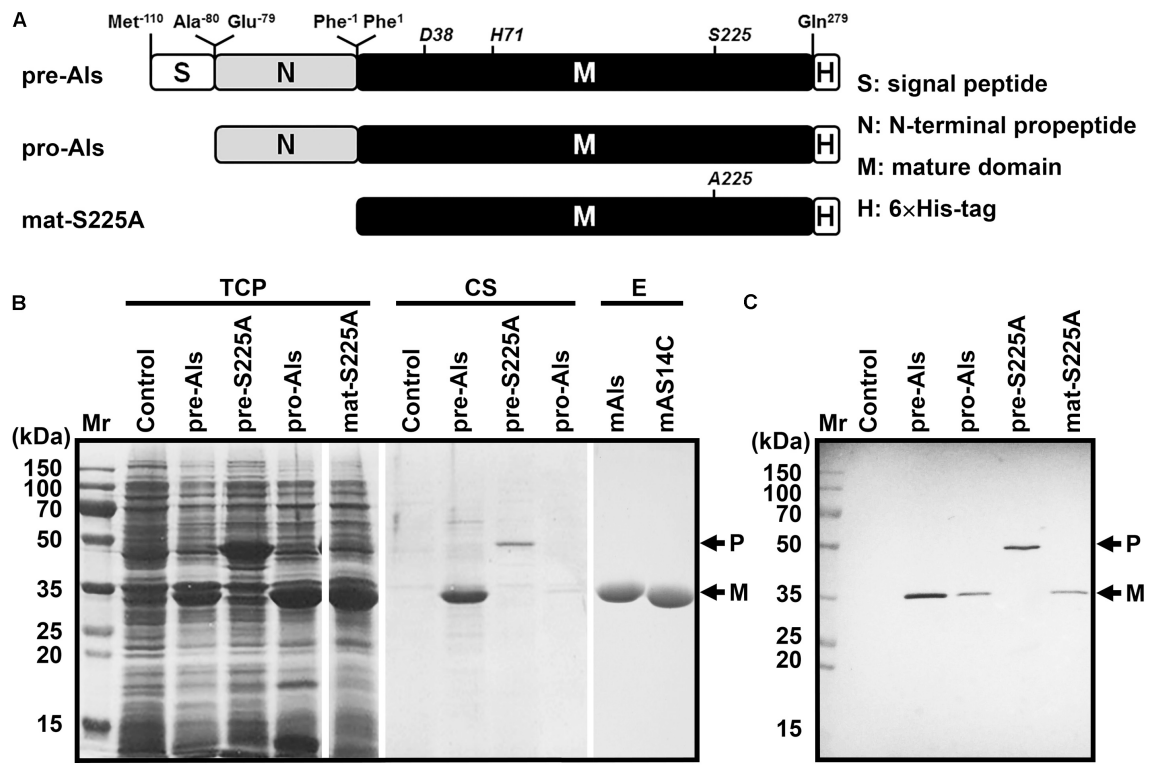


FIGURE 2 | Production and purification of recombinant protease Als and its variants expressed in *E. coli*. **(A)** Schematic representation of the primary structures of protease Als precursor and its derivatives. **(B)** SDS-PAGE analysis of total cellular protein (*TCP*) and culture supernatant (*CS*) of *E. coli* cells expressing recombinant proteins, as well as the purified enzymes (*E*). **(C)** Immunoblot analysis of recombinant proteins in TCP using anti-His-tag monoclonal antibody. The positions of the proform (*P*) and the mature form (*M*) are indicated.

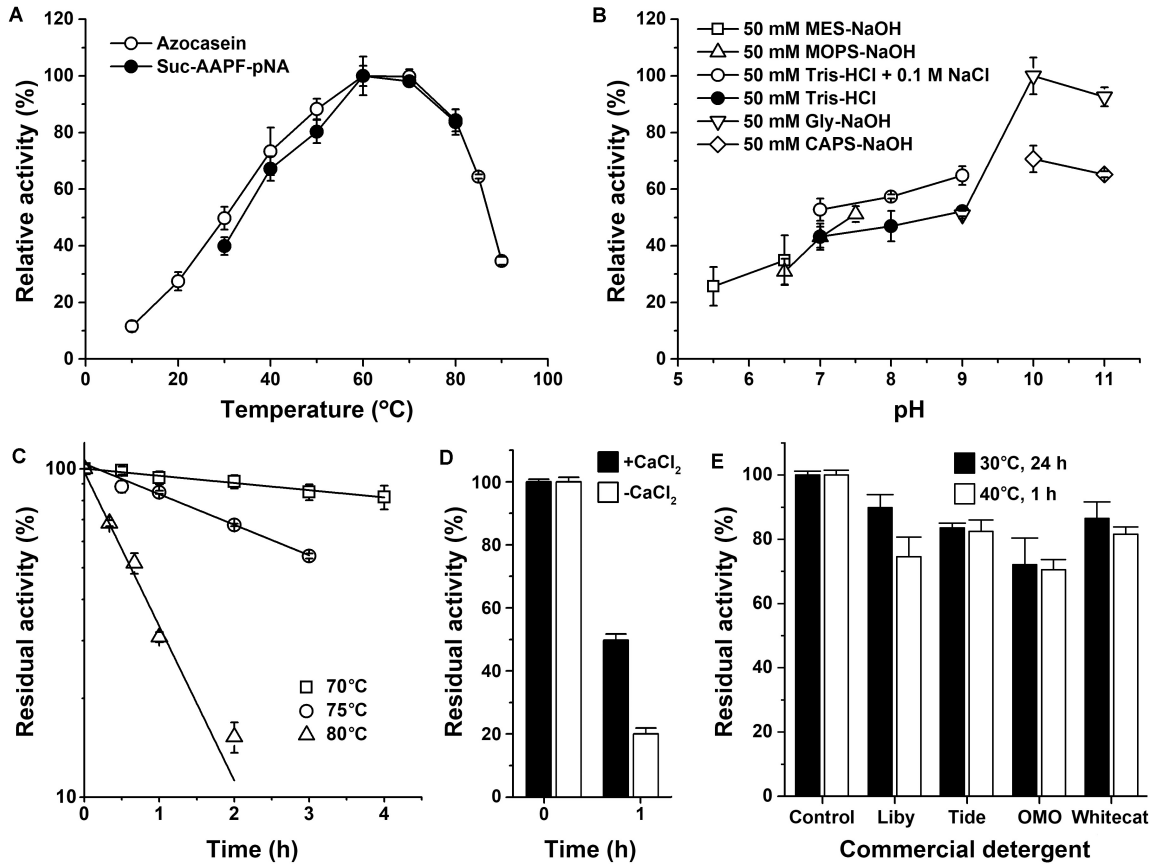


FIGURE 3 | Enzymatic properties of mAls. **(A)** Temperature dependence of proteolytic activity. Activity assays were performed in buffer A (pH 8.0) at the indicated temperatures using 0.25% azo-casein or 0.1 mM suc-AAPF-pNA as the substrates. Relative activity was calculated with the highest level of activity observed at 60°C defined as 100%. **(B)** pH dependence of enzyme activity. The azo-caseinolytic activity of mAls was determined at 60°C in the buffers with different pH values as indicated. The relative activity was calculated with the highest level of activity observed at pH 10 defined as 100%. **(C)** Thermostability of mAls at pH 8.0. The enzyme (1 μg/ml) was incubated in buffer A (50 mM Tris-HCl, 10 mM $CaCl_2$, pH 8.0) at different temperatures as indicated. At the time intervals indicated, aliquots were withdrawn and subjected to azo-caseinolytic activity assay at 60°C. The residual activity is expressed as a percentage of the initial activity. **(D)** Thermostability of mAls at pH 10.0. The enzyme (1 μg/ml) was incubated at 60°C for 1 h in 50 mM Glycine-NaOH (pH 10.0) in the absence (–) or presence (+) of 10 mM $CaCl_2$ and then subjected to azo-caseinolytic activity assay at 60°C. The residual activity is expressed as a percentage of the initial activity. **(E)** Stability of mAls in commercial laundry detergents. The enzyme (5 μg/ml) was incubated with 1% of each detergent at 40°C for 1 h or at 30°C for 24 h, and then subjected to activity assay at 60°C using N,N-dimethylated casein as the substrate. The residual activity is expressed as a percentage of the activity of the enzyme sample incubated under similar conditions in the absence of detergent. The values are expressed as means ± standard deviations (SDs) of two or three independent experiments performed in triplicate.

demonstrate that protease Als is a Ca^{2+}-dependent thermostable alkaline subtilase.

The thermostability and low-temperature activity of mAls were compared with those of thermophilic protease C2 and mesophilic subtilases such as subtilisin Carlsberg and proteinase K. After heat treatment at 70 or 80°C for 1 h, mAls retained 93 or 31% of its initial activity, respectively (**Figure 4A**). In contrast, subtilisin Carlsberg was completely inactivated by heat treatment at 70 or 80°C, while proteinase K retained 72% of its activity after 1 h-incubation at 70°C but was completely inactivated by heat treatment at 80°C (**Figure 4A**). These results suggest that protease Als is much more thermostable than subtilisin Carlsberg and proteinase K. Although mAls was less resistant to heat treatment at 80°C than protease C2 (**Figure 4A**), it exhibited a remarkably higher activity than the latter at 10–30°C (**Figure 4B**). Moreover, the activity of mAls was comparable to or higher than those of commercial subtilisin Carlsberg and proteinase K at the temperature range of 10–30°C (**Figure 4B**). These results suggest that the thermostable protease Als is a promising candidate for industrial application in a wide temperature range.

Protease Als Is Resistant to Organic Solvents and Compatible With Commercial Laundry Detergents

The effects of surfactants, oxidizing agent, and organic solvents on the stability of mAls were investigated by determination of the residual activity of the enzyme sample following pre-incubation with the additives at 40°C for 1 h (**Table 1**). The enzyme was highly stable in the presence of 1.0% non-ionic

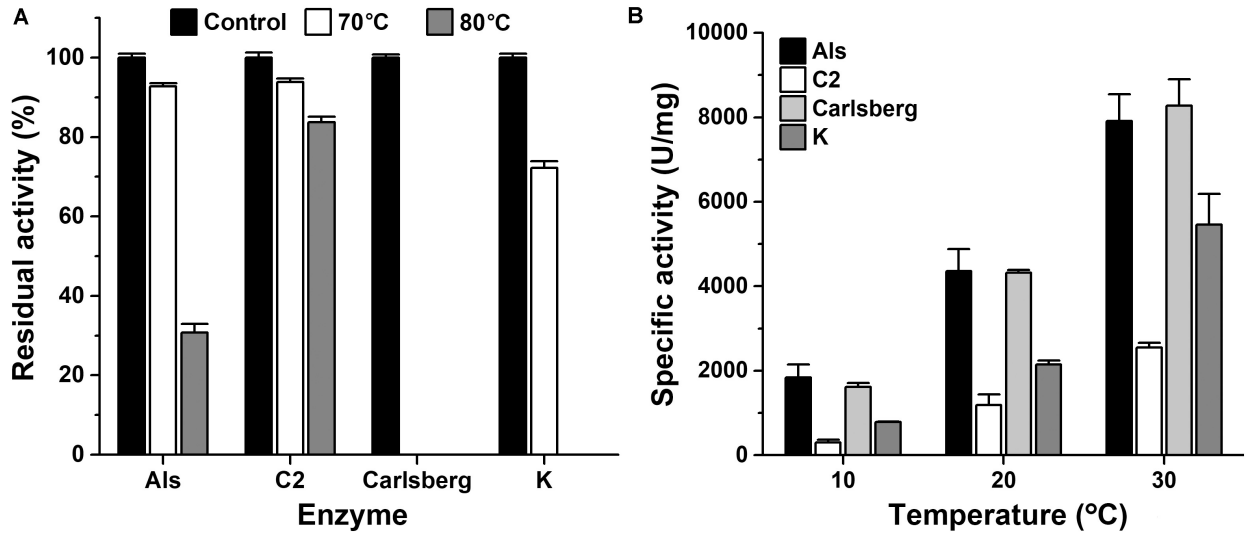


FIGURE 4 | Comparison of thermostabilities and low-temperature activities of mAls, protease C2, subtilisin Carlsberg, and proteinase K. **(A)** Heat inactivation of the enzymes. Each enzyme (1 μg/ml) was incubated at 70 or 80°C in buffer A for 1 h and then subjected to azo-caseinolytic activity assay at 60°C. The residual activity is expressed as a percentage of the initial activity. **(B)** Specific activities of the enzymes at low temperatures. Proteolytic activities of the enzymes were determined in buffer A at the indicated temperatures using 0.25% azo-casein as the substrate. The values are expressed as means ± SDs of three independent experiments performed in triplicate.

surfactants (Triton X-100, Tween 20, and Tween 80) or 0.1% anionic surfactants (SDS and SDBS), but it was less stable at 0.5–1.0% concentrations of SDS or SDBS. The enzyme retained 41.4% of its initial activity in the presence 1.0% H_2O_2, while a loss of approximately 90% of its activity was observed at 5.0% H_2O_2. In all organic solvents (20–40%) tested except isopropanol, mAls displayed approximately 82–134% of its initial activity, showing a high tolerance to organic solvents.

The compatibility of mAls with commercial laundry detergents was examined by pre-incubating the enzyme with detergents at 40°C for 1 h or at 30°C for 24 h, followed by activity assay. It was found that mAls retained 70–90% of its activity after incubation with the detergents tested (**Figure 3E**). The remarkable stability of protease Als in commercial laundry detergents suggests that protease Als has the potential to be used as a detergent additive.

Protease Als Is Highly Halotolerant

In comparison with its closely related homologous protease C2 and thermitase, a unique feature of protease Als is that it contains a larger number of acidic amino acid residues (19 Asp and 10 Glu), and thus a calculated pI value of 4.26 (**Supplementary Table 3**). A homology modeling of protease Als revealed that the acidic amino acid residues are distributed on the solvent-accessible surface area of the enzyme (**Supplementary Figure 2**). The possession of a high content of acidic amino acid residues accessible to solvent is rarely observed in common subtilases but is a typical feature of halophilic subtilases (halolysins) from haloarchaea (**Supplementary Table 3**; Kamekura et al., 1996; Shi et al., 2006). Based on these observations, and the evidence of increased activity of mAls by 0.1 M NaCl (**Figure 3B**), the effects of salt concentration on the activity and stability of protease Als were investigated.

It was found that the azo-caseinolytic activity of mAls increased by 47–61% in the presence of 0.5–2.0 M NaCl (**Figure 5A**). The azo-caseinolytic activities of protease C2, subtilisin Carlsberg, and proteinase K were also enhanced in the presence of 0.1–0.5 M NaCl, but to a lesser extent than that of mAls (**Figure 5A**). Notably, mAls exhibited approximately 100% of its activity at 3 M NaCl, while protease C2, subtilisin Carlsberg, and proteinase K retained 73, 56, and 65% of their activities, respectively, under the same salinity (**Figure 5A**). At either pH 8.0 or pH 10.0, the half-lives of mAls at 80 or 60°C in the presence of NaCl were longer than that in the absence of NaCl (**Figures 5B,C**). It is noticed that mAls is very stable at 80°C in the presence of 3 M NaCl, wherein the enzyme maintained more than 90% of its activity after 1-h incubation at pH 8.0 (**Figure 5B**). These results suggest that NaCl could not only promote the enzymatic activity but also enhance the thermostability of protease Als.

The Keratinolytic Activity of Protease Als Could Be Improved by Modifying Its Substrate-Binding Region

It was previously found that protease C2 from strain CDF possesses a high keratinolytic activity comparable to that of proteinase K and could hydrolyze collagen at high temperatures (Wang et al., 2015). Here, the keratinolytic and collagenolytic capacities of protease Als from the same strain were investigated. At 60°C, the activity of mAls against keratin azure (α-keratin) was only about 1/20 of that of protease C2, while the collagenolytic activities of the two enzymes were comparable (**Figure 6A**).

TABLE 1 | Effects of surfactants, oxidizing agent, and organic solvents on the stability of protease Als.

Agent	Concentration	Residual activity (%)[a]
Control		100.0 ± 1.5
SDS	0.1% (w/v)	104.4 ± 6.4
	0.5% (w/v)	57.7 ± 0.4
	1.0% (w/v)	45.6 ± 4.6
SDBS	0.1% (w/v)	98.4 ± 5.8
	0.5% (w/v)	57.4 ± 7.3
	1.0% (w/v)	22.1 ± 0.2
Triton X-100	1.0% (v/v)	112.8 ± 7.2
Tween 20	1.0% (v/v)	105.8 ± 3.8
Tween 80	1.0% (v/v)	116.2 ± 4.8
H_2O_2	1.0% (v/v)	41.4 ± 1.1
	5.0% (v/v)	10.3 ± 0.5
DMSO	20.0% (v/v)	115.2 ± 4.5
	40.0% (v/v)	133.6 ± 3.2
Acetone	20.0% (v/v)	93.7 ± 4.3
	40.0% (v/v)	94.0 ± 1.5
Methanol	20.0% (v/v)	101.0 ± 5.0
	40.0% (v/v)	86.9 ± 2.4
Ethanol	20.0% (v/v)	94.2 ± 2.7
	40.0% (v/v)	81.5 ± 1.1
Isopropanol	20.0% (v/v)	84.3 ± 1.2
	40.0% (v/v)	67.4 ± 1.3

[a]*After pre-incubation of mAls (5 μg/ml) with each agent at 40°C for 1 h, the azo-caseinolytic activity of the enzyme was determined at 60°C under standard assay conditions as described in section "Materials and Methods." The residual activity was calculated on the basis of the activity of the enzyme sample without pre-treatment (defined as 100%). The values are expressed as means ± standard deviations of three independent measurements.*

Meanwhile, protease C2 and proteinase K [well known for its high keratinolytic activity (Ebeling et al., 1974)] could completely degrade the barbules of chicken feather (β-keratin) within 4 h at 60°C, while mAls only partially disintegrated the barbules after 84 h (**Figure 6B**), showing that protease Als is a weak insoluble keratin-hydrolyzing enzyme compared to protease C2 and proteinase K. To determine the keratin/casein ratios of protease Als and protease C2, the proteolytic activities of the two enzymes against keratin and casein and were determined, showing that protease Als exhibited a slightly higher caseinolytic activity but a lower keratinolytic activity than protease C2 (**Figure 6A**). The keratin/casein ratios were calculated to be 0.31 for protease Als and 0.66 for protease C2 (**Figure 6C**). The hydrolytic activity of mAls toward soluble feather keratin, which was prepared by heat treatment (60°C) of chicken feathers under reducing conditions, was investigated further. The soluble feather keratins released from chicken feathers displayed a molecular mass of about 10 kDa (**Figure 6D**), in agreement with the theoretical molecular mass of intact chicken keratin (Fraser and Parry, 2011). It was found that both mAls and protease C2 were able to completely degrade the soluble feather keratin (**Figure 6D**).

When casein and collagen were used as the substrates, mAls and protease C2 exhibited comparable levels of hydrolyzing activities (**Figure 6A**). It was postulated that the remarkable difference between the two enzymes in their hydrolyzing activities toward insoluble keratin substrates may be due to the difference in substrate preference. To test this possibility, a substrate-binding site variant of protease Als (AS14C) was constructed by substituting its S1 and S4 pockets of the substrate-binding site with those of protease C2, and purified its mature form (mAS14C) (**Figure 2B**). The variant mAS14C not only showed enhanced activities against keratin substrates (keratin azure and keratin) (**Figure 6A**) and an increased keratin/casein ratio (**Figure 6C**), but also disintegrated the barbules of chicken feather more efficiently than protease Als (**Figure 6B**). These results confirm that the substrate-binding site of protease C2 has a stronger preference for insoluble keratin substrates than that of protease Als. Although mAS14C showed an improved keratinolytic activity, it is still less active than protease C2 in hydrolyzing keratin, keratin azure, and feathers (**Figures 6A,B**), indicating that, in addition to the substrate-binding site, other parts of enzyme molecule is also important for keratinolytic activity of the enzyme.

DISCUSSION

Protease Als belongs to the thermitase family of subtilases, and shares high amino acid sequence identity with thermitase (66.8%) and protease C2 (65.6%). Thermitase contains three Ca^{2+}-binding sites (Ca1, Ca2, and Ca3) that contribute to its structural stability (Gros et al., 1991). The ligand residues of the three Ca^{2+}-binding sites in thermitase are partially (Ca1 and Ca2) or fully (Ca3) conserved in proteases Als (**Figure 1**). The observed Ca^{2+}-dependent thermostability of protease Als confirms that the binding of Ca^{2+} is important for stabilizing the enzyme. In comparison with protease C2, protease Als shows a shorter half-life at high temperatures but is more active at low temperatures. In terms of stability-activity relationship, the behaviors of proteases Als and C2, two homologous enzymes coming from strain CDF, seem to follow the trade-off principle that enzymes can gain higher low-temperature activities by sacrificing their thermostability, and vice versa (Siddiqui and Cavicchioli, 2006). Despite being less stable than protease C2, protease Als is much more thermostable than mesophilic subtilisin Carlsberg and proteinase K. Moreover, protease Als exhibits proteolytic activity comparable to or higher than those of subtilisin Carlsberg and proteinase K at low temperatures. By virtue of its high thermostability and substantial low-temperature activity, protease Als is a promising candidate for practical application in a wide temperature range.

Besides the binding of Ca^{2+}, the formation of ion pairs between negatively and positively charged residues on protein surface also contributes to the thermostability of thermophilic proteins (Strickler et al., 2006; Karshikoff et al., 2015). It has been reported that thermitase possesses more surface ion pairs than mesophilic subtilisin BPN' (Voorhorst et al., 1997). At least eight of the ten surface ion pairs of thermitase are conserved in protease Als, involving six Asp residues, one Glu residue, three Arg residues, and two Lys residues (**Supplementary Table 4**). The

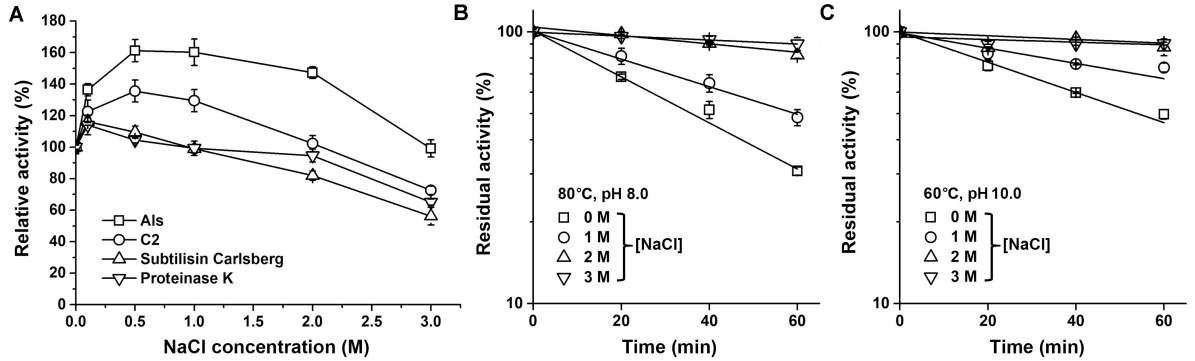


FIGURE 5 | Effects of salt concentration on enzyme activity and stability. **(A)** Salinity dependence of enzyme activity. Using 0.25% azo-casein as the substrate, proteolytic activities of the enzymes (1 μg/ml) were determined at 60°C in buffer A containing different concentrations of NaCl as indicated. Relative activity was calculated by defining the activity of the sample without NaCl as 100%. **(B,C)** Effect of NaCl on thermostability of mAls. The enzyme (1 μg/ml) was incubated at 80°C in buffer A (pH 8.0) **(B)** or at 60°C in 50 mM Glycine-NaOH (pH 10.0) containing 10 mM $CaCl_2$) **(C)** in the presence of different concentrations of NaCl. At the time intervals indicated, aliquots were withdrawn and subjected to azo-caseinolytic activity assay at 60°C. The residual activity is expressed as a percentage of the initial activity. The values are expressed as means ± SDs of three independent experiments performed in triplicate.

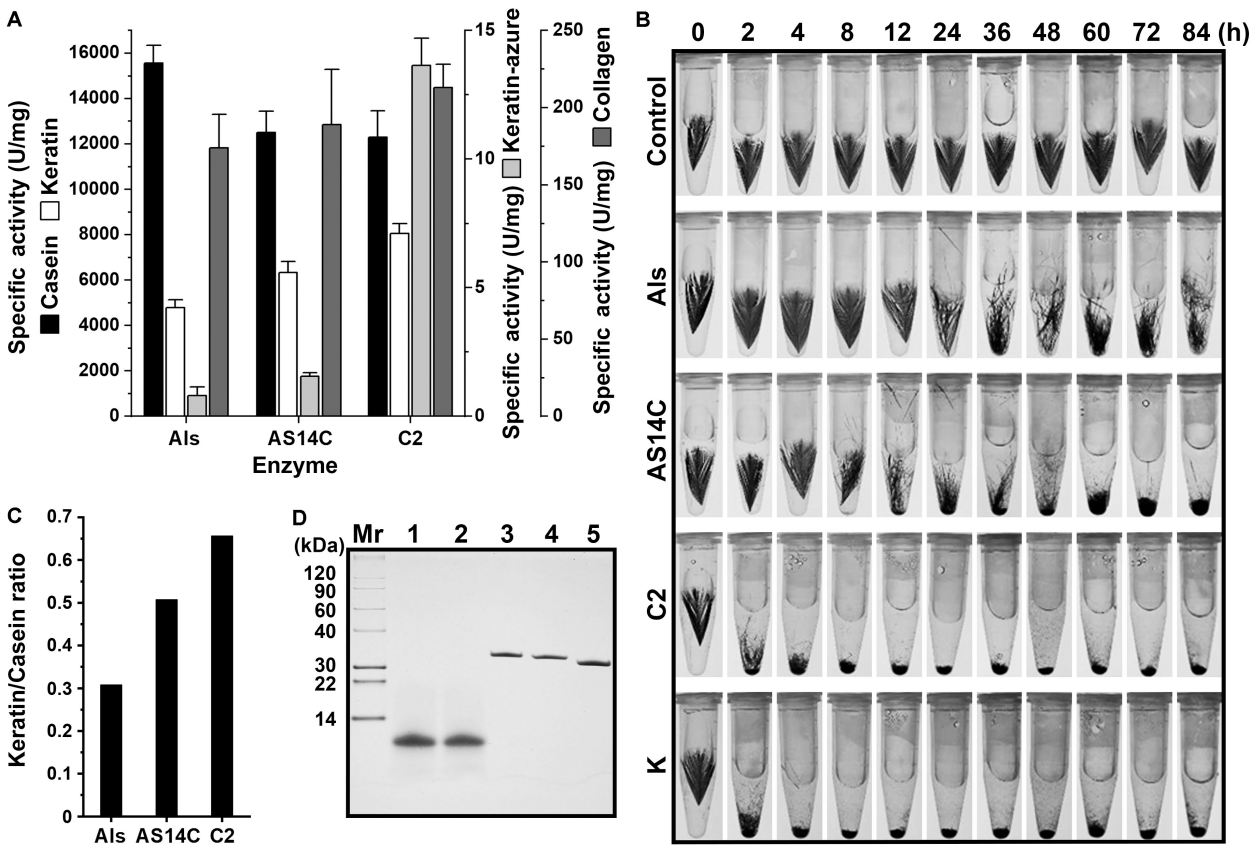


FIGURE 6 | Comparison of the keratinolytic and collagenolytic activities of mAls, mAS14C, and protease C2. **(A)** Specific activities of the enzymes toward casein keratin, collagen, and keratin azure at 60°C. The values are expressed as means ± SDs of three independent experiments performed in triplicate. **(B)** Degradation of chicken feather by the enzymes. Sterilized chicken feathers (3 mg each) were incubated at 60°C in buffer A containing 2% β-ME in the absence (*control*) or presence of each enzyme (100 μg/ml) for different time periods. Proteinase K was used as a reference enzyme. **(C)** Keratin/casein ratios of the enzymes. The keratin/casein ratios were calculated based on the data shown in **(A)**. **(D)** Hydrolysis of soluble feather keratin by the enzymes. Soluble feather keratin (*lane 1*) in buffer A containing 2% β-ME was incubated at 60°C for 2 h in the absence (*lane 2*) or presence of mAls (*lane 3*), AS14C (*lane 4*), or protease C2 (*lane 5*), followed by Tricine-Tris SDS-PAGE analysis.

data presented here showed that protease Als is less thermostable but more active at pH 10.0 than at pH 8.0. A possible explanation for this is that the basic residues forming ion pairs tend to be deprotonated at high pH values, and thus the ion-pair interactions important for thermostability would be weakened. Meanwhile, the attenuation of the ion-pair interactions at high pH may confer certain flexibility favorable for enzyme catalysis, reflecting a trade-off between stability and activity (Siddiqui and Cavicchioli, 2006).

Protease Als possesses an excess of solvent-accessible acidic amino acid residues and displays high halotolerance, as evidenced by the findings that the enzyme exhibited improved activity at 0.5-2.0 M NaCl, maintained approximately 100% of its activity at 3 M NaCl, and showed enhanced thermostability with increasing salt concentration. The possession of an excess of acidic amino acid residues located on protein surface is regarded as an adaptive mechanism of halophilic enzymes, wherein the solvent-accessible acidic residues are involved in the formation of a strong hydration shell to maintain structural stability of halophilic enzymes at high salt concentrations (Mevarech et al., 2000; DasSarma and DasSarma, 2015; Mokashe et al., 2018). Meanwhile, halophilic enzymes could maintain their structural flexibility necessary for catalytic activity at high salinity via the electrostatic repulsion of negatively charged residues, while the hydrophobic core in non-halophilic enzymes would be rigidified in the presence of a high concentration of salts, leading to a decrease of the flexibility required for efficient catalysis (Mevarech et al., 2000). Protease Als has such features as halophilic enzymes; nevertheless, it is a halotolerant rather than halophilic enzyme. In contrast to halophilic enzymes, which tend to be destabilized and inactivated at low-salt concentration due to strong repulsion force exerted by the acidic residues (Chakravorty et al., 2017; Mokashe et al., 2018), protease Als is highly stable and active in the absence of NaCl. It was noticed that protease Als possesses a higher content of basic lysine residues compared with homologous halolysins (**Supplementary Table 3**) which commonly contain few lysine residue along with the excess of acidic residues (Mokashe et al., 2018). The abovementioned surface ion pairs, including those involving lysine residues, may contribute to the stabilization of protease Als in the absence of NaCl. At low salt concentrations, the destabilizing effect of strong electrostatic repulsion by negatively charged residues could be compensated by the stabilizing effect of the ion pairs. At high salt concentrations, both electrostatic repulsion and attraction would be weakened due to increased ionic strength of the solvent, and the enzyme could be stabilized via the formation of a strong hydration shell, mediated by an excess of negatively charged surface residues. Such subtle balance between electrostatic repulsion and attraction plays an important role in the halotolerant behavior of protease Als, allowing it to be stable and active in a wide salinity range.

Some proteases from halophilic/halotolerant microorganisms were reported to be both halotolerant and thermotolerant, but they are generally less stable at high temperatures than their homologs from thermophiles, especially in the absence of salt (Mokashe et al., 2018). Additionally, the information about gene/protein sequences of the reported halo-thermotolerant proteases from halophilic/halotolerant microbes is very limited (Mokashe et al., 2018), thus the structural basis for their polyextremtolerant attributes remains to be elucidated. On the other hand, a common feature of thermophile-derived proteases are their high thermostability, and some of them (e.g., those from marine thermophiles and hyperthermophiles) are resistant to salinity (Barzkar et al., 2018). However, to the best of our knowledge, by far there is no literature showing that a (hyper)thermophile-derived thermostable protease could exhibit 100% of its activity at extremely high salt concentrations (e.g., 3 M NaCl). In the case of protease Als, it showed a half-life of 13.8 h at 70°C in the absence of NaCl, retained more than 90% of its activity after 1-h incubation with 3 M NaCl at 80°C, and exhibited the same level of activity at 3 M NaCl as that in the absence NaCl, thereby representing one of the most robust proteases coupling high thermostability and high halotolerance reported so far. Moreover, the comparative analysis of protease Als with its homologs (e.g., protease C2, thermitase, and halolysins) provides important clues about the roles of surface ion pairs and negatively charged residues in the tolerance of the enzyme to both high temperature and high salinity.

The salt-stable characteristics of halophilic and halotolerant enzymes generally allow these enzymes to be stable and functional in low-water media such as organic solvents (Mokashe et al., 2018). Meanwhile, enzymes from thermophiles are usually reported to be organic solvent-resistant due to their intrinsically stable structures (Sellek and Chaudhuri, 1999). It is not astonishing that protease Als, a thermostable and halotolerant enzyme, is highly resistant to organic solvents. Both thermostable and halophilic/halotolerant proteases have been successfully used for peptide synthesis in the presence of organic solvents (Nagayasu et al., 1994; Ryu et al., 1994; Toplak et al., 2015). Accordingly, the organic solvent-tolerance of protease Als makes the enzyme a potential candidate for use in non-aqueous biocatalysis.

In comparison with protease C2, protease Als exhibited substantially lower hydrolyzing activity toward insoluble keratin substrates such as keratin azure and chicken feathers. Meanwhile, protease Als showed a low keratin/casein ratio (0.31), suggesting that this enzyme is not a true keratinase (Jaouadi et al., 2014) but a subtilase with low keratinolytic activity. Nevertheless, protease Als efficiently hydrolyzes the soluble products released from feathers. Similarly, the recently reported protease KERTYT, which shares 98.2% amino acid sequence identity to protease Als, has the ability to hydrolyze soluble keratin, although it is unclear whether it can degrade insoluble keratin substrates or not (Wang et al., 2019). *T. vulgaris* strain CDF can degrade and grow on chicken feathers not only by secreting keratinolytic proteases but also by providing a supply of reducing power necessary for the cleavage of keratin disulfide bonds (Wang et al., 2015). It seems that both of protease C2 and protease Als are important for strain CDF to use chicken feathers as a nutrient. Protease C2 plays a major role in degrading insoluble keratin substrates, and both protease C2 and protease Als hydrolyze soluble keratins, which may be released from feathers by the action of reducing power of the cells, into smaller peptides available for the assimilation by strain CDF. Notably, substitution of the substrate-binding

site of protease Als with that of protease C2 improves the hydrolyzing activity against insoluble keratin substrates and the keratin/casein ratio. Jaouadi et al. (2014) and Fang et al. (2015) reported that a rational design of the substrate-binding region of keratinases could improve the keratinolytic activity of the enzyme. The results presented here suggest that direct replacement of the substrate-binding site of a protease with that of a highly keratinolytic protease is an alternative strategy for engineering of protease with improved keratinolytic activity.

Recombinant protease Als could be released into the culture supernatant of *E. coli* as a mature form. The extracellular production of recombinant proteins in *E. coli* greatly facilitates downstream processing and protein preparation, and is highly desired in industrial applications (Burdette et al., 2018). Proteins released into the culture supernatant of *E. coli* have to pass through both the cytoplasmic and outer membranes. The predicted Sec-type signal peptide of protease Als precursor is necessary for the extracellular release of the enzyme, indicating that recombinant protease Als is translocated across the cytoplasmic membrane via the Sec pathway and secreted into the periplasm. Furthermore, proteolytic activity of protease Als promotes the release of the enzyme into the culture media, implying that the outer membrane permeability of *E. coli* could be improved through the action of the enzyme. It is possible that active mature protease Als in the periplasm could hydrolyze outer membrane proteins, leading to an increase of the outer membrane permeability. Nevertheless, the proteolytic activity is not essential for extracellular release of protease Als, as its active site variant could also be detected in the culture supernatant. The detailed mechanism for the release of recombinant protease Als from the periplasm into the extracellular milieu remains to be further elucidated.

In summary, protease Als from *T. vulgaris* strain CDF is stable and functional under polyextreme conditions. Besides possessing high thermostability and being active at high pH values, it is highly tolerant to high salinity, organic solvents, and detergents. Moreover, the low-temperature activity of protease Als is comparable to or higher than that of some commercial proteases, and the recombinant enzyme could be extracellularly produced by *E. coli*. With these merits, protease Als may find broad applications in laundry detergents, food processing, and non-aqueous biocatalysis, etc. Although protease Als only weakly degrade insoluble keratin substrates, it could efficiently hydrolyze soluble keratin and thus contribute to the utilization of feathers as a nutrient by strain CDF for growth. In addition, the direct replacement of the substrate-binding site by that of a keratinase has been proved to be a valuable method to improve the keratinolytic activity of a protease here. Finally, the excess of acidic amino acid residues and ion pairs on enzyme surface most likely contribute to the high halotolerance and thermostability of protease Als. Future mutational analysis of the surface charged residues of protease Als is warranted to probe the molecular basis for the polyextremotolerant property of the enzyme.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, QBK13760.

AUTHOR CONTRIBUTIONS

YD, YY, YR, JX, FL, and YL conducted the experiments. YD, X-FT, and BT analyzed and interpreted the results and contributed to writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (grant numbers 31470185 and 31770072) and the National Infrastructure of Natural Resources for Science and Technology Program of China (grant number NIMR-2014-8).

ACKNOWLEDGMENTS

We thank Dr. Fei Gan for helpful discussion and critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.605771/full#supplementary-material

REFERENCES

Akparov, V., Timofeev, V. I., Khaliullin, I. G., Švedas, V., Chestukhina, G. G., and Kuranova, I. P. (2015). Structural insights into the broad substrate specificity of carboxypeptidase T from *Thermoactinomyces vulgaris*. *FEBS J.* 282, 1214–1224. doi: 10.1111/febs.13210

Atomi, H., Sato, T., and Kanai, T. (2011). Application of hyperthermophiles and their enzymes. *Curr. Opin. Biotech.* 22, 618–626. doi: 10.1016/j.copbio.2011.06.010

Barzkar, N., Homael, A., Hemmati, R., and Patel, S. (2018). Thermostable marine microbial proteases for industrial applications: scopes and risks. *Extremophiles.* 22, 335–346. doi: 10.1007/s00792-018-1009-8

Bian, Y., Liang, X., Fang, N., Tang, X. F., Tang, B., Shen, P., et al. (2006). The roles of surface loop insertions and disulfide bond in the stabilization of thermophilic WF146 protease. *FEBS Lett.* 580, 6007–6014. doi: 10.1016/j.febslet.2006.09.068

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1006/abio.1976.9999

Burdette, L. A., Leach, S. A., Wong, H. T., and Tullman-Ercek, D. (2018). Developing Gram-negative bacteria for the secretion of heterologous proteins. *Microb. Cell Fact.* 17:196. doi: 10.1186/s12934-018-1041-5

Burtt, E. H., and Ichida, J. M. (1999). *Bacteria Useful for Degrading Keratin*. US Patent 6214676. Oxford: Oxford University Press.

Chakravorty, D., Khan, M. F., and Patra, S. (2017). Multifactorial level of extremostability of proteins: can they be exploited for protein engineering? *Extremophiles*. 21, 419–444. doi: 10.1007/s00792-016-0908-9

Cheng, G., Zhao, P., Tang, X. F., and Tang, B. (2009). Identification and characterization of a novel spore-associated subtilase from *Thermoactinomyces* sp. CDF. *Microbiology* 155, 3661–3672. doi: 10.1099/mic.0.031336-0

DasSarma, S., and DasSarma, P. (2015). Halophiles and their enzymes: negativity put to good use. *Curr. Opin. Microbiol.* 25, 120–126. doi: 10.1016/j.mib.2015.05.009

DelMar, E. G., Largman, C., Brodrick, J. W., and Geokas, M. C. (1979). A sensitive new substrate for chymotrypsin. *Anal. Biochem.* 99, 316–320. doi: 10.1016/s0003-2697(79)80013-5

Demidyuk, I. V., Nosovskaya, E. A., Tsaplina, I. A., Karavaiko, G. I., and Kostrov, S. V. (1997). Purification and characterization of serine proteinase of the Glu, Asp-specific enzyme family from *Thermoactinomyces* species. *Biochem. Moscow*. 62, 171–175.

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., and Lang, H. (1974). Proteinase K from *Tritirachium album* limber. *Eur. J. Biochem.* 47, 91–97. doi: 10.1111/j.1432-1033.1974.tb03671.x

Elleuche, S., Schäfers, C., Blank, S., Schröder, C., and Antranikian, G. (2015). Exploration of extremophiles for high temperature biotechnological processes. *Curr. Opin. Microbiol.* 25, 113–119. doi: 10.1016/j.mib.2015.05.011

Fang, Z., Zhang, J., Liu, B. H., Du, G. C., and Chen, J. (2015). Insight into the substrate specificity of keratinase KerSMD from *Stenotrophomonas maltophilia* by site-directed mutagenesis studies in the S1 pocket. *RSC Adv.* 5, 74953–74960. doi: 10.1039/c5ra12598g

Fraser, R. D., and Parry, D. A. (2011). The structural basis of the filament-matrix texture in the avian/reptilian group of hard β-keratins. *J. Struct. Biol.* 173, 391–405. doi: 10.1016/j.jsb.2010.09.020

Georgieva, D. N., Stoeva, S., Ivanova, V., Gusterova, A., and Voelter, W. (2000). Specificity of a neutral Zn-dependent proteinase from *Thermoactinomyces sacchari* toward the oxidized insulin B chain. *Curr. Microbiol.* 41, 70–72. doi: 10.1007/s002840010094

Gros, P., Kalk, K. H., and Hol, W. G. (1991). Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium. *J. Biol. Chem.* 266, 2953–2961. doi: 10.2210/pdb3tec/pdb

Gupta, R., Beg, Q. K., and Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biot.* 59, 15–32. doi: 10.1007/s00253-002-0975-y

Gupta, R., and Ramnani, P. (2006). Microbial keratinases and their prospective applications: an overview. *Appl. Microbiol. Biot.* 70, 21–33. doi: 10.1007/s00253-005-0239-8

Ibrahim, A. S., Al-Salamah, A. A., El-Badawi, Y. B., El-Tayeb, M. A., and Antranikian, G. (2015). Detergent-, solvent- and salt-compatible thermoactive alkaline serine protease from halotolerant alkaliphilic *Bacillus* sp. NPST-AK15: purification and characterization. *Extremophiles* 19, 961–971. doi: 10.1007/s00792-015-0771-0

Ignatova, Z., Gousterova, A., Spassov, G., and Nedkov, P. (1999). Isolation and partial characterisation of extracellular keratinase from a wool degrading thermophilic actinomycete strain *Thermoactinomyces candidus*. *Can. J. Microbiol.* 45, 217–222. doi: 10.1139/w98-230

Jang, H. J., Kim, B. C., Pyun, Y. R., and Kim, Y. S. (2002). A novel subtilisin-like serine protease from *Thermoanaerobacter yonseiensis* KB-1: its cloning, expression, and biochemical properties. *Extremophiles* 6, 233–243. doi: 10.1007/s00792-001-0248-1

Jaouadi, B., Aghajari, N., Haser, R., and Bejar, S. (2010). Enhancement of the thermostability and the catalytic efficiency of *Bacillus pumilus* CBS protease by site-directed mutagenesis. *Biochimie* 92, 360–369. doi: 10.1016/j.biochi.2010.01.008

Jaouadi, N. Z., Jaouadi, B., Hlima, H. B., Rekik, H., Belhoul, M., Hmidi, M., et al. (2014). Probing the crucial role of Leu31 and Thr33 of the *Bacillus pumilus* CBS alkaline protease in substrate recognition and enzymatic depilation of animal hide. *PLoS One* 9:e108367. doi: 10.1371/journal.pone.0108367

Kamekura, M., Seno, Y., and Dyall-Smith, M. (1996). Halolysin R4, a serine proteinase from the halophilic archaeon *Haloferax mediterranei*; gene cloning, expression and structural studies. *BBA Protein Struct. Mol. Enzymol.* 1294, 159–167. doi: 10.1016/0167-4838(96)00016-7

Karshikoff, A., Nilsson, L., and Ladenstein, R. (2015). Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J.* 282, 3899–3917. doi: 10.1111/febs.13343

King, J., and Laemmli, U. K. (1971). Polypeptides of the tail fibres of bacteriophage T4. *J. Mol. Biol.* 62, 465–477. doi: 10.1016/0022-2836(71)90148-3

Lee, J. K., Kim, Y. O., Kim, H. K., Park, Y. S., and Oh, T. K. (1996). Purification and characterization of a thermostable alkaline protease from *Thermoactinomyces* sp. E79 and the DNA sequence of the encoding gene. *Biosci. Biotechol. Biochem.* 60, 840–846. doi: 10.1271/bbb.60.840

Li, B., Liu, F., Ren, Y., Ding, Y., Li, Y., Tang, X. F., et al. (2019). Complete genome sequence of *Thermoactinomyces vulgaris* strain CDF, a thermophilic bacterium capable of degrading chicken feathers. *Microbiol. Resour. Announc.* 8:e00530-19. doi: 10.1128/mra.00530-19

Liang, X., Bian, Y., Tang, X. F., Xiao, G., and Tang, B. (2010). Enhancement of keratinolytic activity of a thermophilic subtilase by improving its autolysis resistance and thermostability under reducing conditions. *Appl. Microbiol. Biot.* 87, 999–1006. doi: 10.1007/s00253-010-2534-2

Liu, F., Zhao, Z. S., Ren, Y., Cheng, G., Tang, X. F., and Tang, B. (2016). Autocatalytic activation of a thermostable glutamyl endopeptidase capable of hydrolyzing proteins at high temperatures. *Appl. Microbiol. Biot.* 100, 10429–10441. doi: 10.1007/s00253-016-7697-z

Majumder, D. R., Kanekar, P. P., and Gaikwad, S. M. (2013). Purification and characterization of a thermolysin like protease from *Thermoactinomyces thalpophilus* MCMB-380. *Protein Peptide Lett.* 20, 918–925. doi: 10.2174/0929866511320080009

Mechri, S., Bouacem, K., Jaouadi, N. Z., Rekik, H., Elhoul, M. B., Benmrad, M. O., et al. (2019). Identification of a novel protease from the thermophilic *Anoxybacillus kamchatkensis* M1V and its application as laundry detergent additive. *Extremophiles* 23, 687–706. doi: 10.1007/s00792-019-01123-6

Mevarech, M., Frolow, F., and Gloss, L. M. (2000). Halophilic enzymes: proteins with a grain of salt. *Biophys. Chem.* 86, 155–164. doi: 10.1016/s0301-4622(00)00126-5

Mokashe, N., Chaudhari, B., and Patil, U. (2018). Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. *Int. J. Biol. Macromol.* 117, 493–522. doi: 10.1016/j.ijbiomac.2018.05.217

Nagayasu, T., Miyanaga, M., Tanaka, T., Sakiyama, T., and Nakanishi, K. (1994). Synthesis of aspartame precursor with an immobilized thermolysin in tert-amyl alcohol. *Biotechnol. Bioeng.* 43, 1118–1123. doi: 10.1002/bit.260431116

Papworth, D., Bauer, J. C., Braman, J., and Wright, D. A. (1996). Site-directed mutagenesis in one day with greater than 80% efficiency. *Strategies* 9, 3–4. doi: 10.1080/08924562.1996.11000299

Petrova, D. H., Shishkov, S. A., and Vlahov, S. S. (2006). Novel thermostable serine collagenase from *Thermoactinomyces* sp. 21E: purification and some properties. *J. Basic Microbiol.* 46, 275–285. doi: 10.1002/jobm.200510063

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. R.* 62, 597–635. doi: 10.1128/mmbr.62.3.597-635.1998

Ryu, K., Kim, J., and Dordick, J. S. (1994). Catalytic properties and potential of an extracellular protease from an extreme halophile. *Enzyme Microb. Technol.* 16, 266–275. doi: 10.1016/0141-0229(94)90165-1

Salwan, R., and Sharma, V. (2019). Trends in extracellular serine proteases of bacteria as detergent bioadditive: alternate and environmental friendly tool for detergent industry. *Arch. Microbiol.* 201, 863–877. doi: 10.1007/s00203-019-01662-8

Schägger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379. doi: 10.1016/0003-2697(87)90587-2

Sellek, G. A., and Chaudhuri, J. B. (1999). Biocatalysis in organic media using enzymes from extremophiles. *Enzyme Microb. Tech.* 25, 471–482. doi: 10.1016/s0141-0229(99)00075-7

Shi, W., Tang, X. F., Huang, Y., Gan, F., Tang, B., and Shen, P. (2006). An extracellular halophilic protease SptA from a halophilic archaeon *Natrinema* sp. J7: gene cloning, expression and characterization. *Extremophiles* 10, 599–606. doi: 10.1007/s00792-006-0003-8

Shih, J. C. H., and William, C. M. (1992). *Purified Bacillus licheniformis PWD-1 Keratinase*. US Patent US5171682. Washington, DC: U.S. Patent and Trademark Office.

Siddiqui, K. S., and Cavicchioli, R. (2006). Cold-adapted enzymes. *Annu. Rev. Biochem.* 75, 403–433. doi: 10.1146/annurev.biochem.75.103004.142723

Siezen, R. J., and Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* 6, 501–523. doi: 10.1002/pro.5560060301

Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., et al. (2006). Protein stability and surface electrostatics: a charged relationship. *Biochem. US* 45, 2761–2766. doi: 10.1021/bi0600143

Suzuki, Y., Tsujimoto, Y., Matsui, H., and Watanabe, K. (2006). Decomposition of extremely hard-to-degrade animal proteins by thermophilic bacteria. *J. Biosci. Bioeng.* 102, 73–81. doi: 10.1263/jbb.102.73

Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frömmel, C., Höhne, W. E., et al. (1990). Crystal structure of thermitase at 1.4 Å resolution. *J. Mol. Biol.* 214, 261–279. doi: 10.1016/0022-2836(90)90160-n

Toplak, A., Nuijens, T., Quaedflieg, P. J., Wu, B., and Janssen, D. B. (2015). Peptide synthesis in neat organic solvents with novel thermostable proteases. *Enzyme Microb. Technol.* 73-74, 20–28. doi: 10.1016/j.enzmictec.2015.03.003

Tsuchiya, K., Ikeda, I., Tsuchiya, T., and Kimura, T. (1997). Cloning and expression of an intracellular alkaline protease gene from alkalophilic *Thermoactinomyces* sp. HS682. *Biosci. Biotechnol. Biochem.* 61, 298–303. doi: 10.1271/bbb.61.298

Tsuchiya, K., Nakamura, Y., Sakashita, H., and Kimura, T. (1992). Purification and characterization of a thermostable alkaline protease from alkalophilic *Thermoactinomyces* sp. HS682. *Biosci. Biotechnol. Biochem.* 56, 246–250. doi: 10.1271/bbb.56.246

Verma, A., Singh, H., Anwar, M. S., Kumar, S., Ansari, M. W., and Agrawal, S. (2016). Production of thermostable organic solvent tolerant keratinolytic protease from *Thermoactinomyces* sp. RM4: IAA production and plant growth promotion. *Front. Microbiol.* 7:1189. doi: 10.3389/fmicb.2016.01189

Voorhorst, W. G., Warner, A., De Vos, W. M., and Siezen, R. J. (1997). Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. *Protein Eng.* 10, 905–914. doi: 10.1093/protein/10.8.905

Wang, L., Cheng, G., Ren, Y., Dai, Z., Zhao, Z. S., Liu, F., et al. (2015). Degradation of intact chicken feathers by *Thermoactinomyces* sp. CDF and characterization of its keratinolytic protease. *Appl. Microbiol. Biot.* 99, 3949–3959. doi: 10.1007/s00253-014-6207-4

Wang, L., Zhou, Y., Huang, Y., Wei, Q., Huang, H., and Guo, C. (2019). Cloning and expression of a thermostable keratinase gene from *Thermoactinomyces* sp. YT06 in *Escherichia coli* and characterization of purified recombinant enzymes. *World J. Microb. Biot.* 35:135. doi: 10.1007/s11274-019-2710-1

Zabolotskaya, M. V., Demidyuk, I. V., Akimkina, T. V., and Kostrov, S. V. (2004). A novel neutral protease from *Thermoactinomyces* species 27a: sequencing of the gene, purification, and characterization of the enzyme. *Protein J.* 23, 483–492. doi: 10.1007/s10930-004-5225-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 March 2021
doi: 10.3389/fmicb.2021.577001

Thermoacidophilic *Alicyclobacillus* Superoxide Dismutase: Good Candidate as Additives in Food and Medicine

*Xueqian Dong[1,2], Wei Wang[2], Shannan Li[1], Hongyu Han[2], Peiwen Lv[1] and Chunyu Yang[1]**

[1] State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao, China, [2] Shandong Food Ferment Industry Research & Design Institute, QiLu University of Technology (Shandong Academy of Sciences), Jinan, China

OPEN ACCESS

Edited by:
Massimiliano Fenice, University of Tuscia, Italy

Reviewed by:
Maria Bogomilova Angelova, The Stephan Angeloff Institute of Microbiology, Bulgarian Academy of Sciences, Bulgaria
Anna Maria Garzillo, University of Tuscia, Italy
Khelifa Bouacem, Mouloud Mammeri University of Tizi Ouzou, Algeria

***Correspondence:**
Chunyu Yang
ycy21th@sdu.edu.cn

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *28 June 2020*
Accepted: *26 February 2021*
Published: *18 March 2021*

Citation:
Dong X, Wang W, Li S, Han H, Lv P and Yang C (2021) Thermoacidophilic Alicyclobacillus Superoxide Dismutase: Good Candidate as Additives in Food and Medicine. Front. Microbiol. 12:577001. doi: 10.3389/fmicb.2021.577001

Thermoacidophilic *Alicyclobacillus* strains attract great interests as the resource of thermostable or acidic enzymes. In this study, a putative gene encoding superoxide dismutase (*Aa*SOD) was identified in a thermoacidophilic *Alicyclobacillus* strain. With a 16-fold activity observed, the *Aa*SOD activity expressing in the medium of manganese enrichment was much higher than that in the iron medium. In addition, the purified *Aa*SOD can be reconstituted exclusively with either Fe^{2+} or Mn^{2+}, with its Mn-bound protein showing 25-fold activity than that of Fe-bound form. The optimal temperature for *Aa*SOD reaction was 35°C, and was highly stable at any certain temperature up to 80°C. Of particular interest, the enzyme is found to be very stable across a wide pH range spanning from 2.0 to 10.0, which confers its robust stability in the acidic stomach environment and implies striking potentials as food additive and for medical use.

Keywords: thermoacidophilic *Alicyclobacillus* strain, superoxide dismutase, acid tolerant, thermostability, cambialistic Fe/Mn type

INTRODUCTION

Superoxide dismutases (SODs, EC 1.15.1.1) are one type of antioxidant enzymes derived from living organisms which can protect themselves against oxidative stress, by achieving disproportionation of superoxide anion radical (O_2^-) to hydrogen peroxide (H_2O_2) and dioxygen (O_2) through a redox cycle of metal ions (Abreu and Cabelli, 2010). SODs have been confirmed to be closely related to various physiological processes in living bodies, since they are related not only to the prevention and treatment of various diseases, but also to anti-aging therapy and prevention of skin pigmentation. SODs, also known as metalloenzymes, are divided into four groups depending on their metal preferences: manganese SOD (Mn-SOD), iron SOD (Fe-SOD), copper/zinc SOD (Cu/Zn-SOD), and nickel SOD (Ni-SOD) (Perry et al., 2010; Bafana et al., 2011). The resolved crystal structures of different types of SODs revealed that these proteins differ not only with regard to the coordinated metal ion, but also to the protein folding. Mn-SOD and Fe-SOD are closely related when compared by their amino acids and tertiary structures, and assigned into one family (Miller, 2012). In this family, some use both Fe^{2+} and Mn^{2+} as their cofactors, and thus are referred to as the cambialistic Fe/Mn type SODs (Schmidt et al., 1996; Yamano et al., 1999). In spite of metal ion specificity,

these Fe-SODs, Mn-SODs, and the cambialistic SODs were found to be phylogenetically related, which is proved by the high sequence identity, as well as many common features in their tertiary structures (Jackson and Cooper, 1998). In strain *Escherichia coli*, the Mn-SOD and Fe-SOD are extremely similar in sequence (45% identity) and crystal structure (91% homology), but are strictly specific to their cognate metal ions (Hunter et al., 2018). The cambialistic SODs can accommodate both Fe^{2+} and Mn^{2+} as their cofactors for catalysis but display different ions preference. ApeSOD from *Aeropyrum pernix* is less active in its Fe-bound form while the cambialistic SOD from *Propionibacterium shermanii* exhibits similar enzymatic activity in the presence of Fe^{2+} and Mn^{2+} (Meier et al., 1982; Nakamura et al., 2011). Comparison of the crystal structures of both SODs shows that these two cambialistic SODs have different binding forms with Fe^{2+} and Mn^{2+}, which was proposed as a structural explanation for their ions preference (Schmidt et al., 1996; Nakamura et al., 2011).

Supplementation of exogenous SOD has been reported to boost the antioxidant defense of host (Manolov et al., 2017). Considerable evidences from clinical or animal models have been accumulated and revealed that SOD is beneficial in a wide variety of applications, including reduction fibrosis following radiation treatment, preventions of aging, diabetes, tumor formation, and hepatitis C related fibrosis (Matès and Sánchez-Jimènez, 2000; Emerit et al., 2006; Bafana et al., 2011; Naso et al., 2011), or reduction of the cytotoxic and cardiotoxic effects of anticancer drugs (Trotti, 1997). Due to its excellent antioxidant and therapeutic properties, various SOD products have been applied in the industries of medicine, health care products, food additives, and cosmetics (Johnson and Giulivi, 2005). However, the thermal denaturation is an important factor leading to enzyme inactivation because the high temperature treatment is often processed in practical applications of the enzyme (Song et al., 2009). Therefore, it has stimulated a widespread interest in exploring thermotolerant enzymes from thermophilies (Morozkina et al., 2010). A plenty of previous research confirmed that the thermostable enzymes have unique structural and functional characteristics that give them good adaptability in high temperature environment. To date, SODs have been identified from various thermophilic archaea and bacteria, including *Bacillus* (Boyadzhieva et al., 2010), thermoacidophilic crenarchaeon (Slutskaya et al., 2012), *Pyrobaculum* (Amo et al., 2003), *Chloroflexus* (Lancaster et al., 2004), *Thermus* species (Mandelli et al., 2013) and so on. In addition, SODs from thermophilic fungi have also been reported by a few studies, such as *Thermomyces lanuginosus* Mn-SOD (Li et al., 2005) and *Thermoascus aurantiacus* var. *levisporus* Cu/Zn-SOD, which also display remarkable tolerance to the extreme temperature (Song et al., 2009). Besides these wild type thermophilic SODs, thermostable SODs were also constructed by engineering protocols. The heat-resistance and stress-tolerance of thermophilic cambialistic Fe/Mn-SOD from *A. pernix* K1 was successfully developed through the fusion with the N-terminal domain of *Geobacillus thermodenitrificans* SOD (Li et al., 2017).

Comparing with its thermostability, the acidic tolerance of SOD is far less addressed, since few species of SOD are enzymatically stable under extremely acidic conditions (Ken et al., 2005). *Alicyclobacillus* strains are heterotrophic organisms that mostly inhabit acidic geothermal environments such as hot springs and acid mine waters (Zhang et al., 2015; López et al., 2018). As an extraordinary resource for exploring unique enzymes with acidophilic and (or) thermophilic properties, the *Alicyclobacillus* isolates are of great interests in recent years because of their double physiological characteristics of acidophilic and thermotolerant (McKnight et al., 2010). Some thermostable and acidic tolerant glycoside hydrolases, including α-amylase (Zhang et al., 2018), glucanase (Bai et al., 2010), xylanase (Lee et al., 2018) etc., have been identified from *Alicyclobacillus* species. Correa-Llantén et al. (2014) have purified SOD from *Alicyclobacillus* isolate CC2 that was most active at 55°C, pH 7.4. However, there are no further studies performed on *Alicyclobacillus*-derived SOD. In the present study, we engaged in exploring a novel SOD from a strain *Alicyclobacillus* sp. HJ previously isolated from the hot-spring (Zhang et al., 2018). As a result, the Fe/Mn-SOD with remarkable acid and thermal tolerance was identified and catalytically characterized in this study.

MATERIALS AND METHODS

Strains and Plasmids

Strain *Alicyclobacillus* sp. HJ was isolated from Tengchong hot-spring, Yunnan, China and deposited in the Marine Culture Collection of China (MCCC 1K03506). The 16S rDNA sequence analysis revealed 100% similarity with gene of *Alicyclobacillus acidocaldarius* DSM451. In the M63 medium, the strain grows optimally at conditions of 65°C and pH 4.0 (Zhang et al., 2018). The clone vector pEASY-blunt and *E. coli* DH5α was used for gene cloning and fidelity confirmation. The recombinant expression vector pET-24a(+) with *Aa*SOD encoding gene was transformed into *E. coli* BL21 CodonPlus for protein expression, and Luria-Bertani (LB) medium was used for the cultivation of recombinants.

Phylogenetic Analysis of *Aa*SOD and Modeling

The amino acid sequence of *Aa*SOD were submitted to National Center for Biotechnology Information (NCBI) for blast analysis. A phylogenetic tree was constructed to establish the evolutionary relationship among *Aa*SOD and corresponding SOD sequences retrieved from the NCBI protein database. MEGA 7.0 software was used to build the phylogenetic tree, by using Neighbor-Joining method with bootstrap replications of 1,000 (Kumar et al., 2016). Amino acid sequences multiple alignment was performed by ClustalX 2.0 and DNAMAN programs. The protein sequence of *Aa*SOD was submitted to SWISS-MODEL server for homology modeling (Waterhouse et al., 2018).

Protein Expression and Purification

Genomic DNA of *Alicyclobacillus* sp. HJ was isolated with ChargeSwitch® gDNA Mini Bacteria Kit (Life Technologies) and

used as template for *Aa*SOD gene amplification. Primers for its PCR were designed as *Aa*SOD-F (5′ –CGGGATCCATGCCACATCAACTCCCAC–3′) and *Aa*SOD-R (5′ –CCCAAGCTTGCCGTTCAGCGCGGCCTCGT–3′), which incorporate restriction sites *Bam*HI and *Hin*dIII (underlined), respectively. The PCR products was ligated into pET-24a(+) and transformed in the competent cells *E. coli* BL21 CodonPlus. The correct clones were picked and further verified by PCR sequencing.

For *Aa*SOD expression, cells of *E. coli* harboring pET-24a(+)-*Aa*SOD were cultured in 5 ml LB medium at 37°C for 12 h. Subsequently, the culture was diluted at the ratio of 1:100 into 100 ml fresh LB media containing 50 mg l^{-1} ampicillin and grown at 37°C, 180 rpm. When OD_{600} of the culture reached to 0.6, isopropyl-β-D-thiogalactoside (IPTG) was added up to the final concentration of 0.5 mM for protein induction. After cultivation at 18°C, 100 rpm for 12 h, cells were harvested by centrifugation at 8,000 × *g* for 10 min at 4°C. The cell debris was resuspended in 10 ml buffer A (20 mM Tris–HCl, 0.5 M sodium chloride, pH 8.0) with a protease inhibitor cocktail and disrupted using a high-pressure homogenizer (ATS Engineering Inc., Canada). The cell lysis was then centrifuged at 17,000 × *g* for 30 min at 4°C, and the supernatant was applied to HisTrap™ FF crude column (GE Healthcare) for protein purification at 4°C. The target protein was eluted with buffer A that contains 250 mM imidazole. For further analysis, the elution was concentrated and dialyzed overnight with 50 mM Tris–HCl buffer (pH 8.0) at 4°C. The protein concentration was determined by Brad-Ford method with bovine serum albumin (BSA) as standard (Bradford, 1976). The expression level and purity of target protein were verified by SDS-PAGE (12% w/v), with Bio-Rad Mini-PROTEAN TETRA electrophores system (United States).

SOD Activity Assay

The superoxide dismutase activity was measured by pyrogallic acid spontaneous oxidation assay (Marklund and Marklund, 1974). Briefly, 50 mM pyrogallic acid was added to 50 mM Tris–HCl buffer (pH 8.0) at the final volume of 3 ml reaction mixture. The absorbance of the mixture was then measured at 325 nm at every 30 s, during a 4-min reaction process at room temperature. The rate of the self-oxidation was approximately kept at 0.07 OD min^{-1} and a certain amount *Aa*SOD was added to the reaction mixture at an inhibition ratio of 50%. One unit of superoxide dismutase activity was defined by the amount of enzyme that inhibited the rate of pyrogallic acid spontaneous oxidation by 50%. The activities of each sample were performed in three replicates.

Influence of Fe^{2+} and Mn^{2+} on the Native Protein Expression

The *E. coli* strain carrying pET-24a(+)-*Aa*SOD was prepared as above described and then inoculated into 100 ml LB medium, or LB media containing 1 mM $FeCl_2$, $MnCl_2$, or both ions. After IPTG induction for 12 h, cells were collected by centrifugation and lysed in high-pressure homogenizer. After centrifugation, the supernatant flowed through the HisTrap crude column for protein purification and then the activity of SOD was measured by standard procedures described above.

Reconstitution of Metals Into SOD

The apo-enzyme was prepared as previous described (He et al., 2007). In brief, the enzyme was dialyzed with 50 mM acetate buffer (pH 3.8) containing 8 M urea and 10 mM EDTA for 16 h at room temperature. Then the enzyme was dialyzed at 4°C sequentially with 50 mM acetate buffer (pH 3.8) containing 8 M urea, 50 mM Tris–HCl buffer (pH 7.0) containing 8 M urea, and 50 mM Tris–HCl buffer (pH 8.0). To prepare Mn- or Fe-reconstituted *Aa*SOD, the apo-protein was subsequently dialyzed for 4 h at 4°C with the following buffers: 8 M urea, 50 mM acetate buffer, 10 mM $MnSO_4$ and (or) $FeSO_4$, pH 3.8; 8 M urea, 50 mM phosphate buffer, 10 mM $MnSO_4$ and (or) $FeSO_4$, pH 7.0; 4 M urea, 50 mM phosphate buffer, 10 mM $MnSO_4$ and (or) $FeSO_4$, pH 7.0; 2 M urea, 50 mM phosphate buffer, 10 mM $MnSO_4$ and (or) $FeSO_4$, pH 7.0; 50 mM phosphate buffer, 1 mM $MnSO_4$ and (or) $FeSO_4$, pH 7.0; and 50 mM Tris–HCl buffer/0.5 mM EDTA, pH 8.0, respectively. The reconstituted enzymes obtained were analyzed by SDS-PAGE and the specific activity was determined by pyrogallic acid spontaneous oxidation assay.

Metal Analysis

Inductively Coupled Plasma Mass Spectrometry (ICP-MS) was used to measure the contents of Fe^{2+} and Mn^{2+} Briefly, the mixture of 500 μl sample (1 mg ml^{-1}) and 8 ml 65% nitric acid was digested by microwave digestion system. ddH_2O was added to the mixture at a total volume of 10 ml subsequently and further analyzed by ICP8000 ICP-OES. The Fe^{2+} and Mn^{2+} concentrations were calculated based on their corresponding standard curves. Under experimental conditions, the standard recovery rate is between 96.23–102.65% for Fe^{2+}, and 98.31–103.05% for Mn^{2+}.

Effect of Temperature on *Aa*SOD

The optimum temperature was determined by measuring the activity of SOD at different temperatures ranging from 25 to 80°C. For thermal stability measurement, 1 mg ml^{-1} protein was incubated in 50 mM Tris–HCl buffer (pH 8.0) at room temperature (RT), 60, 70, 80, 90, and 100°C for 1 h, respectively. Subsequently, the residual activities were determined by the standard condition (pH 8.0, 25°C) as described above. Each temperature was performed in triplicate. The relative activity was calculated by the percentage of the maximum activity at 35°C.

After dialyzed overnight with 200 mM HEPES (pH 8.0), the purified *Aa*SOD was tested for its thermal denaturation in the MicroCal VP-DSC (Malven). Protein sample (1 mg ml^{-1}) was stirred to degassed under vacuum, cooled down to 20°C, and then gradually heated to 110°C. Using the dialysis buffer as a baseline, *Aa*SOD sample was scanned from a temperature range of 70 to 110°C at a rate of 1.5°C min^{-1}. By subtracting the baseline, the transition curve was fitted by non-2-state model and the T_m value was calculated (MN2state) in the Origin software (MicroCal Software, Inc., Northampton, MA, United States).

pH Tolerance of *Aa*SOD

The pH stability of *Aa*SOD was determined by incubating the enzyme in various buffers with different pH values at 25°C for 1 h, and then measuring the residual activity under standard assay conditions described above. The buffer systems include 200 mM KCl-HCl buffer (pH 1.0–2.0), acetate buffer (pH 2.0–5.0), sodium phosphate buffer (pH 5.0–7.0), Tris–HCl buffer (pH 7.0–9.0), and glycine–NaOH buffer (pH 9.0–10.0). Each experiment was performed in triplicate. The maximal activity of *Aa*SOD was defined as 100% for relative activity calculation.

Effects of Divalent Metal Ions on SOD Activity

Effects of metal ions on *Aa*SOD activity were determined by adding various divalent metal salts ($MgCl_2$, $ZnSO_4$, $BaCl_2$, $CaCl_2$, $CuSO_4$, $NiSO_4$, $CoCl_2$, $MnSO_4$, and $FeSO_4$), at a final concentration of 1 mM in 50 mM Tris–HCl buffer (pH 8.0). After incubating the purified *Aa*SOD with each ion at 25°C for 30 min, the activities were measured under the standard conditions. The activity of the enzyme without addition of metal ions was defined as 100% and used for relative activity calculation.

Statistical Method

The data was analyzed using origin 9.0 (OriginLab corp., United States), the means and standard deviations were calculated by descriptive statistics.

Nucleotide Sequence Accession Number

The nucleotide sequence encoding *Aa*SOD was deposited in the GenBank database under accession number of MT559075.

RESULTS AND DISCUSSION

Sequence Analysis and Structural Analyses of *Aa*SOD

Based on the genomic analysis of *Alicyclobacillus* sp. HJ, a putative 204 amino acid protein encoded by *orf0188* was predicted to be involved in detoxifying superoxide radicals, converting them to hydrogen peroxide and oxygen. Blast analysis indicated that *Aa*SOD exhibits 72.5% identity with a Mn-SOD in *Geobacillus stearothermophilus* (Adams et al., 2019) and 69.7% identity with the Mn-SOD in *Bacillus subtilis* 168 (Liu et al., 2007). As shown in **Figure 1**, *Aa*SOD was phylogenetically clustered with some other identified Mn-SODs and the Fe/Mn SODs from *Staphylococcus* species, while relatively distant with the Fe-SOD proteins. As both Mn- and Fe-SODs from *Pseudomonas aeruginosa* PAO1 or *E. coli* K12 were blasted as homologs of *Aa*SOD, the amino sequence of *Aa*SOD showed 56.8% identity with that of Mn-SOD (Steinman, 1978) from *E. coli* K12, and also 48.3% with its Fe-dependent SOD (Carlioz et al., 1988). This firmly verified the conclusion that Fe- and Mn-SODs are highly conserved in sequences and three dimension (3D), while possessing different dependence to metal ions (Lah et al., 1995). Based on the multiple sequence alignment, the

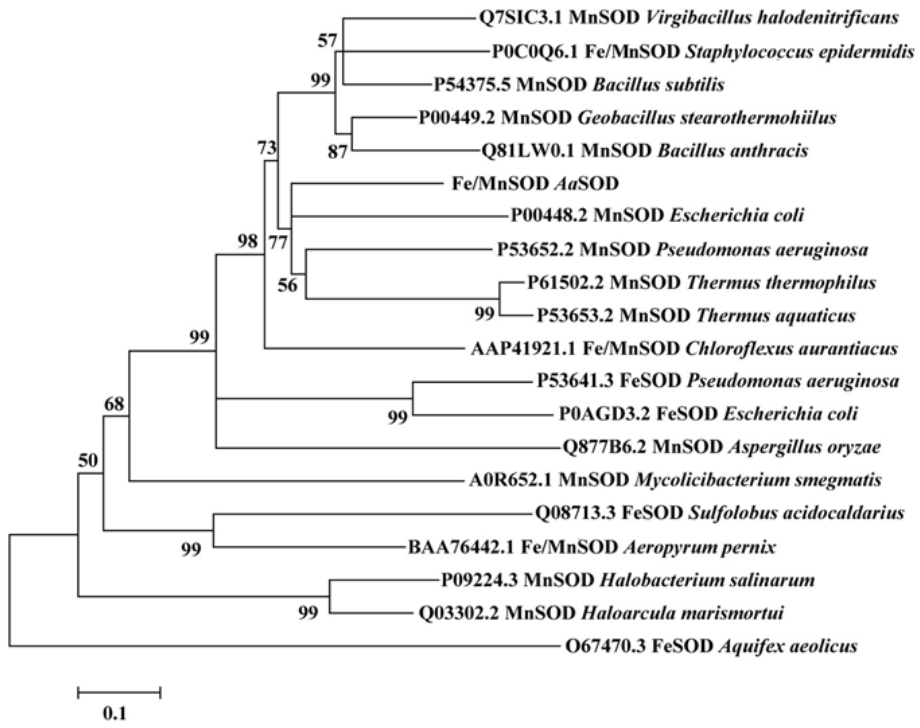


FIGURE 1 | Neighbor-joining phylogenetic tree based on the amino sequences of *Aa*SOD and its closest relatives retrieved from NCBI database. The tree was built by the program MEGA version 7.0. Bootstrap percentages >50% (based on 1000 replications) are shown at branch points.

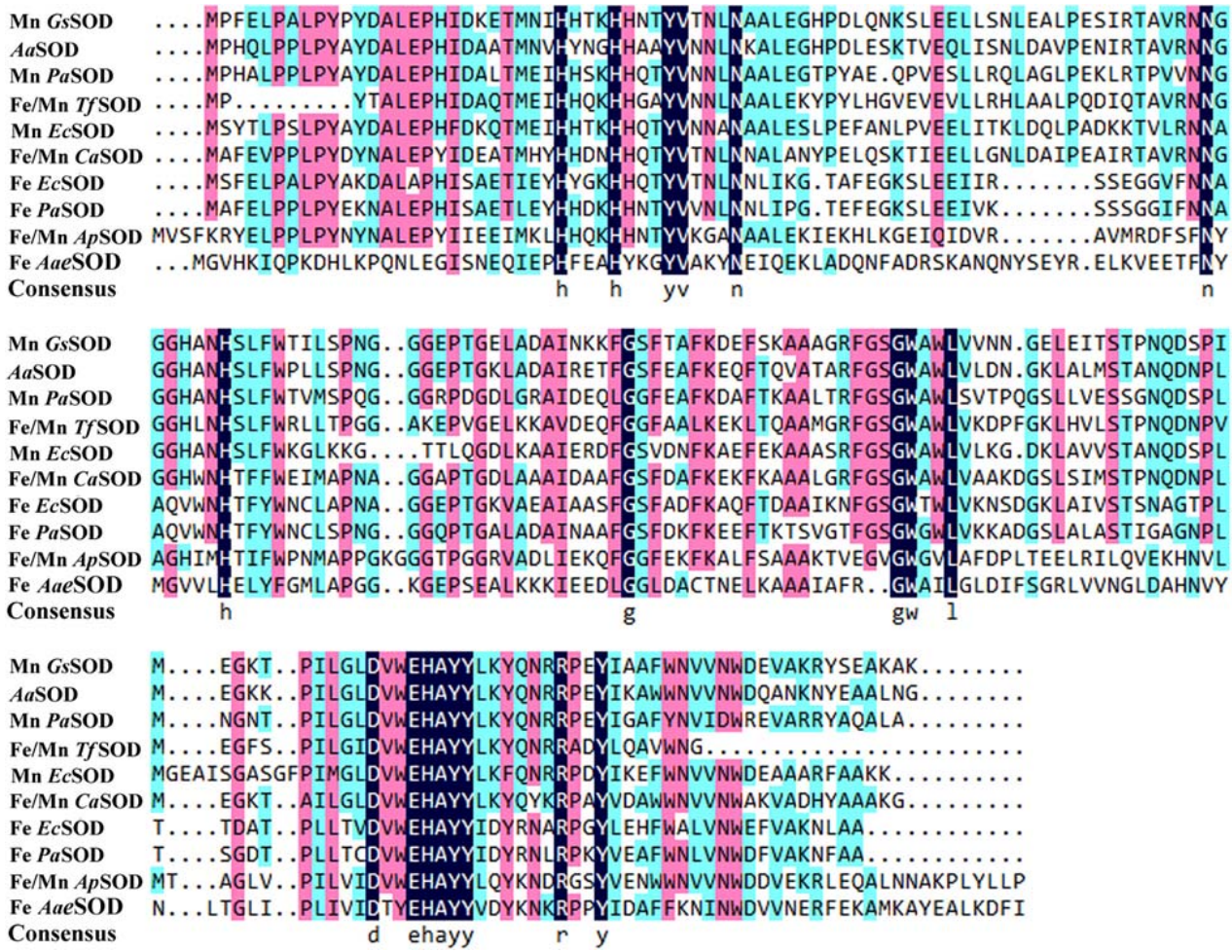


FIGURE 2 | Multiple alignment of *Aa*SOD with other identified Fe, Mn, or Fe/Mn type SODs that retrieved from NCBI database. *Gs*SOD: *Geobacillus stearothermophilus*, *Aa*SOD: this study, *Pa*SOD: *Pseudomonas aeruginosa*, *Tf*SOD: *Thermus filiformis*, *Ec*SOD: *Escherichia coli*, *Ca*SOD: *Chloroflexus aurantiacus*, *Ap*SOD: *Aeropyrum pernix*, *Aae*SOD: *Aquifex aeolicus*.

Mn- and Fe/Mn SODs share more conserved sequences as shown in **Figure 2**, while possessing more variations from our Fe SODs. Consequently, we assumed that the catalytic process of *Aa*SOD may use Mn or Fe/Mn as its cofactors.

A homology model of *Aa*SOD was generated by the auto modeling system in the SWISS-MODEL server, based on the recently resolved structure of *Gs*-MnSOD (PDB: 6pro) as the template (Adams et al., 2019), which is from strain *G. stearothermophilus* and aligns 72.5% with the query sequence. As shown in **Figure 3A**, the superimposition of the predicted 3D structure of *Aa*SOD significantly conformed to the tertiary structure of *Gs*-MnSOD. The four residues for manganese binding, His 26, His 81, Asp 163, and His 167, are all well conserved in the *Aa*SOD sequence (**Figure 2**) and occupy identical spatial locations in its homologs structure (**Figure 3B**).

Protein Purification

To avoid the formation of inclusion body during protein expression process and gain high expression level, *Aa*SOD was heterologous expressed in the *E. coli* host of BL21 Codonplus, which contains extra copies of the argU, ileY, and leuW tRNA genes. In addition, the heterologous expression of *Aa*SOD was induced at a low temperature of 18°C. As a result, a large amount of SOD was produced after 10-h induction by shaking at 100 rpm. As shown in **Figure 4**, a tense band of approximately 25 kDa was detected in the SDS-PAGE, which is in accordance with the theoretical molecular weight of 22.82 kDa. After one-step purification on the His-Trap affinity column by 250 mM imidazole elution, a high yield of 99.8 mg l^{-1} protein with high purity of 95% were obtained. We assumed that the high purification level and recovery rate would be benefit from the high expression levels of the protein. This confers *Aa*SOD great potentials for scaled-up production and industrial applications. However, extracellular expression protocol could be needed for preparing *Aa*SOD products considering its purification efficiency. To choose host strains that are suitable for extracellular protein expression, such as *Bacillus subtilis*, would be addressed in the *Aa*SOD production.

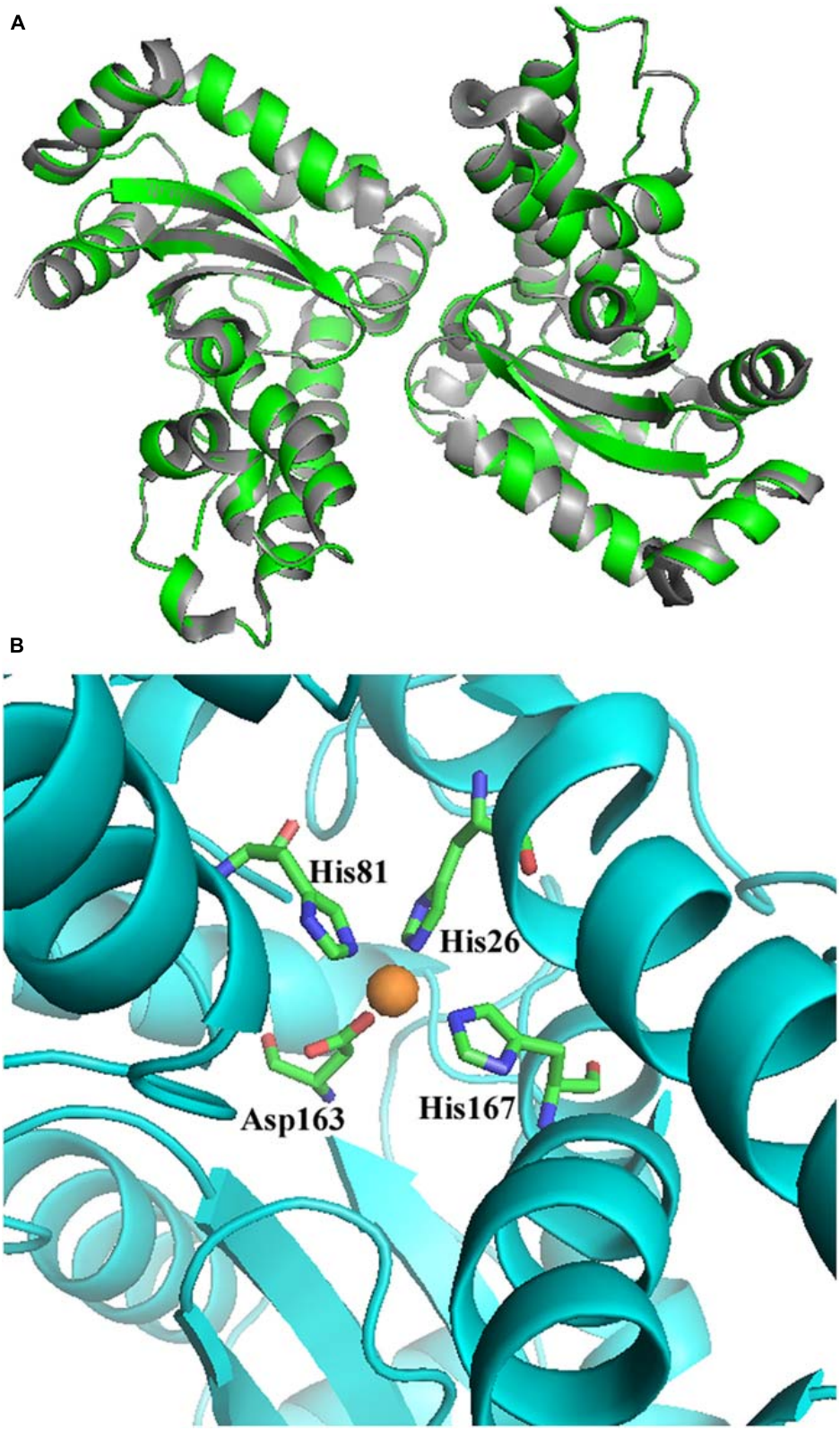


FIGURE 3 | Superimposition of the predicted structure of *Aa*SOD (Green) over its template of *Geobacillus stearothermophilus Gs*-MnSOD (gray) **(A)** and four predicted residues for manganese binding **(B)**.

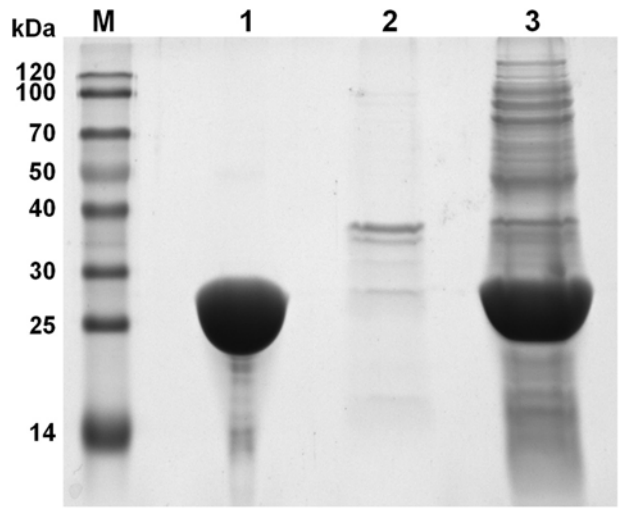


FIGURE 4 | SDS-PAGE spectrum of *Aa*SOD. Lane M, Marker; lane 1, purified protein after His-Trap affinity column; lane 2, cell debris of the *E. coli* recombinant harboring *Aa*SOD; lane 3, supernatant of the *E. coli* recombinant harboring *Aa*SOD induced by IPTG for 12 h at 18°C.

Metal Incorporation of *Aa*SOD

To explore the metal incorporation of *Aa*SOD in its native form, 1 mM $MnSO_4$ and (or) $FeSO_4$ was added into LB medium during the expression process. As shown in **Table 1**, the proteins purified from the LB medium without ion supplementation (native protein) possessed a specific activity of 90.94 U mg^{-1}, which is obviously higher than the apo-enzyme (5.5 U mg^{-1}, **Table 1**). The fact is that the natural LB medium contains trace metal ions, and thus provides Fe^{2+} or Mn^{2+} to the *Aa*SOD for cooperation. It was further confirmed by the presence of coordinated Fe^{2+} and Mn^{2+} in the ICP-MS analysis. In the natural *Aa*SOD, the Fe^{2+} and Mn^{2+} contents per one molar proteins were 0.051 and 6.85×10^{-3} mol, respectively. However, only 8.18×10^{-4} mol of Fe^{2+} and 5.20×10^{-5} mol of Mn^{2+} were detected in the apo-SOD. Different from the lower activity of proteins that expressed in the natural or Fe^{2+}-rich LB medium, there was significantly higher *Aa*SOD activity of 1198.55 U mg^{-1} purified from the *E. coli* culture enriched with 1 mM $MnSO_4$. And *Aa*SOD from the medium containing these two ions, displayed a modest activity of 534.56 U mg^{-1}. Differently, the Fe^{2+}-enriched cultures induced similar amount of *Aa*SOD but with very low activity (74.65 U mg^{-1}), indicating that *Aa*SOD is a cambialistic enzyme because of Mn^{2+} binding being much more effective than Fe^{2+} binding, and therefore should be assigned into Fe/Mn type SOD. The activity bias between Fe- and Mn-bound SOD was also identified in some thermostable SODs, in which their Fe-reconstituted SOD activity is much lower than that of Mn-reconstituted SOD (Li et al., 2016). For examples, the SOD from thermophilic *C. aurantiacus* was most efficient with manganese incorporated, up to 30% of the activity was retained with iron (Lancaster et al., 2004). What's more, the $MnSOD_{cd}$ from *Clostridium difficile* exhibited the highest activity while Fe-sub-$MnSOD_{cd}$ showed only 1/10 activity of $MnSOD_{cd}$ (Li et al., 2014). In the ApeSOD of *A. pernix* K1, it was hypothesized that the Fe-bound mode could mimic the product-inhibited form by residue Tyr39, and was proposed as an explanation for the lower activity of Fe/SOD than Mn/SOD (Nakamura et al., 2011). Therefore, we suspected that the binding mode of metal cofactors in *Aa*SOD was similar as those SODs, in which Fe^{2+} competitively binds to the catalytic sites and thus restrains the activities of these cambialistic SODs. Therefore, it would be significantly beneficial to the SOD activity when we created an Mn^{2+}-enriched environment for the expression of these Fe/Mn SODs, as our observations of high SOD activity in the Mn^{2+}-enriched medium of this study.

Metal fidelity of *Aa*SOD was further investigated by an *in vitro* reconstitution experiment with 1 mM $MnSO_4$ and(or) $FeSO_4$. As a result, the metal specificities were highly consistent with corresponding proteins expressed in the $MnSO_4$ or $FeSO_4$ enriched medium. As shown in **Table 1**, the specific activities of all treatments displayed higher activities than the apo-enzyme. The *Aa*SOD sample that was reconstituted exclusively in the presence of Fe^{2+} had lower specific activity of 21.05 U mg^{-1}, which was higher than the apo-enzyme while being lower than native *Aa*SOD (90.94 U mg^{-1}) expressed in the crude LB medium. Similarly, the Mn^{2+}-bound form of *Aa*SOD also exhibited a remarkably higher specific activity of 517.63 U mg^{-1}, while the reconstituted sample upon both ions obtained a decreased activity of 270.27 U mg^{-1}.

Thermo- and pH-Stabilities of *Aa*SOD

For industrial purpose, many enzymes are preferred to be processed into solid state, which could be realized by freeze drying or spray drying technology. In the case of freeze drying, the enzymes can preserve high activity under lower temperature. However, the large-scaled production is limited by its high energy consumption and processing cost. Thus, the spray drying process is more favorable and widely used. In addition, thermostable enzymes are often associated with robust resistance to harsh environments, including detergent, strong acid and alkali, or chemical and denaturants (Vieille and Zeikus, 2001). Therefore, thermostable SODs are of great interest and screened as good candidates for industrial production. As we previously described, to explore excellent SODs for industrial

TABLE 1 | Superoxide dismutase (SOD) activity incorporating various ions.

Sample	Specific activity (U·mg^{-1})
Natural medium	90.94 ± 0.47
*Aa*SOD from Fe^{2+} enriched medium	74.65 ± 2.92
*Aa*SOD from Mn^{2+} enriched medium	1198.55 ± 20.36
*Aa*SOD from Fe^{2+} and Mn^{2+} enriched medium	534.56 ± 27.33
Apo-enzyme	5.50 ± 0.40
Fe^{2+}-reconstitution	21.05 ± 1.10
Mn^{2+}- reconstitution	517.63 ± 48.35
Fe^{2+}/Mn^{2+}- reconstitution	270.27 ± 127.83

The apo-enzyme was prepared by dialyzing the purified AaSOD with 50 mM acetate buffer (pH 3.8) containing 8 M urea and 10 mM EDTA.

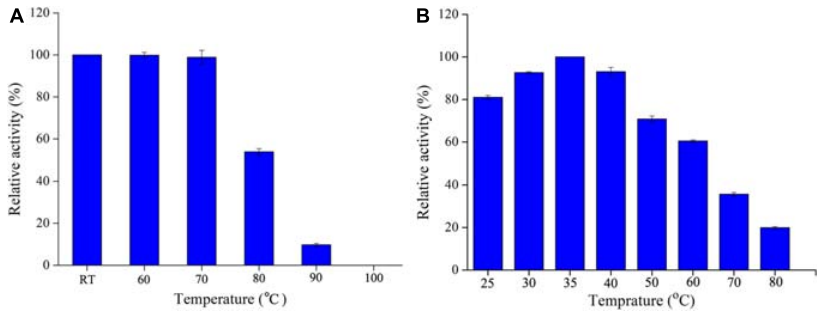


FIGURE 5 | Thermostability **(A)** and optimal temperature **(B)** of *Aa*SOD. For thermostability measurement, the purified *Aa*SOD was incubated at different temperatures (RT, 60, 70, 80, 90, and 100°C) for 1 h, and subsequently measured the residual activities. The relative activity was calculated as the percentage of the maximum activity at 35°C. To achieve the optimal catalytic temperature, the activity of purified *Aa*SOD was tested at different temperatures (25–80°C) and the highest activity was set as 100% for relative activity calculation.

application, some thermostable SODs had been identified from thermophilic strains. *Alicyclobacillus* sp. HJ is a strain from hot-spring and possesses thermoacidophilic properties, i.e., grows optimally at pH 4.0 and 65°C. Therefore, it was regarded as an ideal source of thermostable or acid stable proteins, as our previous observation on a thermoacidophilic α-amylase (Zhang et al., 2018). Expectedly, *Aa*SOD exhibited high stability under high temperatures, with whole activity retaining after incubation at 60–70°C for 1 h (**Figure 5A**). In addition, 52.7% activity was still detected after incubation at 80°C for 1 h. Despite of robust thermostability, the optimal catalytic temperature of *Aa*SOD is mesophilic. During a range of temperatures of 20–70°C, the highest activity was observed at 35°C (**Figure 5B**). This is different from many SODs from thermopiles, which generally possess an optimum activity at high temperature (Seatovic et al., 2004; Zhu et al., 2013).

By directly measuring the forces stabilizing the conformational structure, DSC analysis is ideally suitable to evaluate protein thermal denaturation (Singh and Singh, 2004). Hence, the thermal denaturation of *Aa*SOD was further measured with DSC analysis, by scanning from 70 to 120°C based on the purified *Aa*SOD. After processes of buffer correction, normalization, and baseline subtraction, the DSC curve was fitted by the Gauss function (**Figure 6**). In the meanwhile, the melting temperature (T_m) was calculated as 89.5°C and the ΔH value was -2097 cal mol^{-1}. This denaturation temperature is somewhat higher than the incubation temperature, of which the maximal activity was observed (**Figure 5B**). To our best knowledge, the thermophilic SODs are generally active at high temperatures, e.g., SOD_ASAC from *Acidilobus saccharovorans* has an optimal activity at 70°C and denaturation temperature of 107.3°C (Slutskaya et al., 2012). Due to their vast bioavailability of SODs, they are widely used in cosmetics and food additives, as well as in pharmaceuticals (Bafana et al., 2011). Compared to the recorded thermostable or themophilic SODs, the thermostable but mesophilic properties of *Aa*SOD not only meet the requirement for the processing technology, but also endow its advantage to keep full functions after being taken into the organisms.

To date, available methods for measuring SOD activity are almost based on inhibitive assay, which are seriously restricted

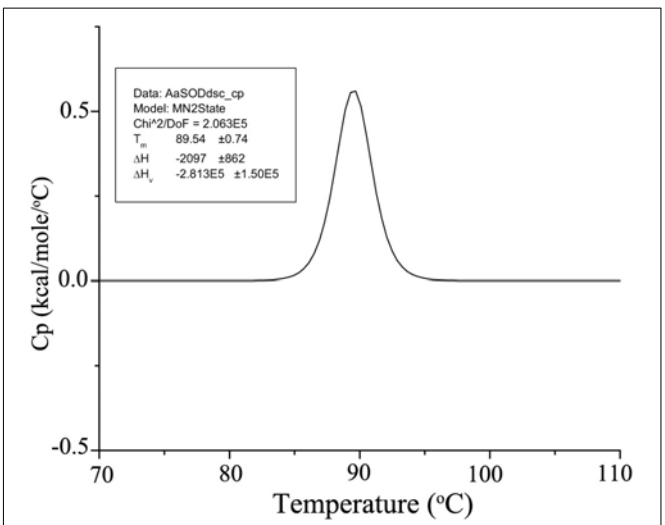


FIGURE 6 | DSC spectrum of *Aa*SOD. The purified *Aa*SOD was scanned from a temperature range of 70–110°C at a rate of 1.5°C min^{-1}. The transition curve was fitted by non-2-state model and the T_m value was calculated (MN2state) in the Origin software.

to the reaction pH. Not unexpectedly, we failed to measure the acidic activities of *Aa*SOD, even at pH 6.0 conditions (data not shown). Therefore, we only tested the pH tolerance profiles by incubating the enzyme in various pH buffers for 1 h. Using this maximum activity as 100%, the relative activities of those proteins incubating in other pH buffers were calculated and shown in **Figure 7A**. Inspiringly, the enzyme showed good stability under a broad range of pH conditions with the highest stability at pH 4.0. In particular, it retained attractively high activities at extremely acidic conditions, with around 70 and 60% activities retained at pH 3.0 and 2.0, respectively. This provides compelling evidence that *Aa*SOD is an acidophilic enzyme, which is in accordance to our previously observations, that the α-amylase from strain *Alicyclobacillus* sp. HJ is also highly active at pH 4.0 (Zhang et al., 2018). To our best knowledge, no recorded SOD showing such extremely acidic

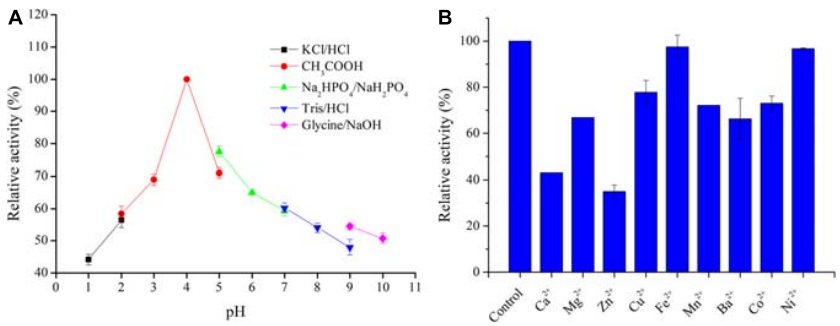


FIGURE 7 | Stabilities of *Aa*SOD in buffers with different pH values **(A)** and effect of divalent ions on the *Aa*SOD activity **(B)**. The purified enzyme was incubated in the buffer systems of 200 mM KCl–HCl buffer (pH 1.0–2.0), acetate buffer (pH 2.0–5.0), sodium phosphate buffer (pH 5.0–7.0), Tris–HCl buffer (pH 7.0–9.0), and glycine–NaOH buffer (pH 9.0–10.0) for 1 h, 25°C. The maximal activity was taken as 100% for relative activity calculation. For metal ions effects, the purified enzyme was incubated with 1 mM of divalent metal ions in 50 mM Tris–HCl buffer (pH 8.0), at 25°C for 30 min. The mixture with no ions supplementation was used as a control for relative activity calculation.

tolerance so far, and thus *Aa*SOD would be served as an ideal model protein for exploring the mechanism of acidic tolerance. Moreover, abundant studies concluded that exogenous SOD supplementation has pharmacological and therapeutic benefits to mammals, including reduce liver oxidative stress in diabetic animals, resist inflammatory diseases, boost antioxidant defense and so on (Stephenie et al., 2020). Consequently, being an oral medicine or dietary supplement would be a favorable form for SOD administration. Regarding the advantage of excellent acidic tolerance, *Aa*SOD will be stable in the stomach after being digested by the organisms, and thus assure its activity as food additive and for medicine use.

Influence of Divalent Ions on *Aa*SOD Activity

As shown in **Figure 7B**, various divalent metal ions were supplemented into the reaction mixture and tested for their influence on the catalytic activity of purified *Aa*SOD. As a result, the SOD activities were partly suppressed by some tested metal ions. However, still more than 60% activities were reserved upon tested metal ions, with exceptions of Zn^{2+} and Ca^{2+}. As shown in **Figure 7B**, only around 40% activity was preserved in the presence of 1 mM of Zn^{2+}. Differently, Fe^{2+} and Ni^{2+} showed no influence on the enzymatic activity, with the activity retained completely upon Fe^{2+} and 96.7% in the presence of 1 mM Ni^{2+}. These results are mostly in agreement with the Mn-SOD from *Geobacillus* sp. EPT3, in which most metal ions except Mn^{2+} partially inhibited the enzyme activities but not lethal (Zhu et al., 2014). This guarantee high activities of these SODs when exposing to trace metal ions in applications.

CONCLUSION

This manuscript firstly describes one thermostable and cambialistic SOD from a thermoacidophilic *Alicyclobacillus* strain. There are ample potentials in various applications with its advantages: it is significantly active in the mesophilic environment but highly stable under high temperatures, and it is an acidophilic protein with high stability at extremely acidic environments. Its large amount of heterologous expression and superior activity in the presence of Mn^{2+} create great feasibility for its industrial scale producing. Future studies would focus on optimizing its large-scaled expression and purifications protocols, achieving exogenous expression in the biosafety strain like *B. subtilis*, as well as experimentally testing its bioactivity and bioavailability in the animal models.

DATA AVAILABILITY STATEMENT

The nucleotide sequence encoding AaSOD was deposited in the GenBank database, accession number LC573735. The Strain Alicyclobacillus sp. HJ was deposited in the Marine Culture Collection of China, accession number MCCC 1K03506.

AUTHOR CONTRIBUTIONS

CY and XD conceptualized and defined the experimental design. XD, WW, HH, and SL performed the experiments. CY, HH, and PL carried out the data analysis and manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the financial supports from the National Natural Science Foundation of China (31870094) and Collaborative Innovation Project of Shandong Academy of Science (2019-CXY9).

ACKNOWLEDGMENTS

The authors thank Jingyao Qu and Zhifeng Li for the technical help in DSC analysis.

REFERENCES

Abreu, I. A., and Cabelli, D. E. (2010). Superoxide dismutases-a review of the metal-associated mechanistic variations. *Biochim. Biophys. Acta* 1804, 263–274. doi: 10.1016/j.bbapap.2009.11.005

Adams, J. J., Morton, C. J., and Parker, M. W. (2019). The crystal structure of the manganese superoxide dismutase from *Geobacillus stearothermophilus*: Parker and Blake (1988) revisited. *Aust. J. Chem.* 73, 145–150. doi: 10.1071/ch19346

Amo, T., Atomi, H., and Imanaka, T. (2003). Biochemical properties and regulated gene expression of the superoxide dismutase from the facultatively aerobic hyperthermophile *Pyrobaculum calidifontis*. *J. Bacteriol.* 185, 6340–6347. doi: 10.1128/jb.185.21.6340-6347.2003

Bafana, A., Dutt, S., Kumar, S., and Ahuja, P. S. (2011). Superoxide dismutase: an industrial perspective. *Crit. Rev. Biotechnol.* 31, 65–76. doi: 10.3109/07388551.2010.490937

Bai, Y., Wang, J., Zhang, Z., Shi, P., Luo, H., Huang, H., et al. (2010). A novel family 9 β-1,3(4)-glucanase from thermoacidophilic *Alicyclobacillus* sp. A4 with potential applications in the brewing industry. *Appl. Microbiol. Biotechnol.* 87, 251–259. doi: 10.1007/s00253-010-2452-3

Boyadzhieva, I. P., Atanasova, M., and Emanuilova, E. (2010). A novel, thermostable manganese-containing superoxide dismutase from *Bacillus licheniformis*. *Biotechnol. Lett.* 32, 1893–1896. doi: 10.1007/s10529-010-0368-8

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Carlioz, A., Ludwig, M. L., Stallings, W. C., Fee, J. A., and Touati, D. (1988). Iron superoxide dismutase nucleotide sequence of the gene from *Escherichia coli* K12 and correlations with crystal structures. *J. Biol. Chem.* 263, 1555–1562. doi: 10.1016/s0021-9258(19)57340-9

Correa-Llantén, D. N., Amenábar, M. J., Muñoz, P. A., Monsalves, M. T., Castro, M. E., and Blamey, J. M. (2014). *Alicyclobacillus* sp. strain CC2, a thermo-acidophilic bacterium isolated from Deception Island (Antarctica) containing a thermostable superoxide dismutase enzyme. *Adv. Polar Sci.* 25, 92–96.

Emerit, J., Samuel, D., and Pavio, N. (2006). Cu-Zn super oxide dismutase as a potential antifibrotic drug for hepatitis C related fibrosis. *Biomed. Pharmacother.* 60, 1–4. doi: 10.1016/j.biopha.2005.09.002

He, Y. Z., Fan, K. Q., Jia, C. J., Wang, Z. J., Pan, W. B., Huang, L. Y., et al. (2007). Characterization of a hyperthermostable Fe superoxide dismutase from hot spring. *Appl. Microbiol. Biotechnol.* 75, 367–376. doi: 10.1007/s00253-006-0834-3

Hunter, T., Bonetta, R., Sacco, A., Vella, M., Sultana, P. M., Trinh, C. H., et al. (2018). A single mutation is sufficient to modify the metal selectivity and specificity of a eukaryotic manganese superoxide dismutase to encompass iron. *Chem. Eur. J.* 24, 5303–5308. doi: 10.1002/chem.201704655

Jackson, S. M. J., and Cooper, J. B. (1998). An analysis of structural similarity in the iron and manganese superoxide dismutases based on known structures and sequences. *Biometals* 11, 159–173.

Johnson, F., and Giulivi, C. (2005). Superoxide dismutases and their impact upon human health. *Mol. Aspects Med.* 26, 340–352. doi: 10.1016/j.mam.2005.07.006

Ken, C. F., Hsiung, T. M., Huang, Z. X., Juang, R. H., and Lin, C. T. (2005). Characterization of Fe/Mn-superoxide dismutase from diatom *Thallassiosira weissflogii*: cloning, expression, and property. *J. Agric. Food Chem.* 53, 1470–1474. doi: 10.1021/jf048269f

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lah, M. S., Dixon, M. M., Pattridge, K. A., Stallings, W. C., and Ludwig, M. L. (1995). Structure-function in *Escherichia coli* iron superoxide dismutase: comparisons with the manganese enzyme from *Thermus thermophilus*. *Biochemistry* 34, 1646–1660. doi: 10.1021/bi00005a021

Lancaster, V. L., LoBrutto, R., Selvaraj, F. M., and Blankenship, R. E. (2004). A cambialistic superoxide dismutase in the thermophilic photosynthetic bacterium *Chloroflexus aurantiacus*. *J. Bacteriol.* 186, 3408–3414. doi: 10.1128/jb.186.11.3408-3414.2004

Lee, B. D., Apel, W. A., Sheridan, P. P., and Deveaux, L. C. (2018). Glycoside hydrolase gene transcription by *Alicyclobacillus acidocaldarius* during growth on wheat arabinoxylan and monosaccharides: a proposed xylan hydrolysis mechanism. *Biotechnol. Biofuels* 11:110.

Li, D. C., Gao, J., Li, Y. L., and Lu, J. A. (2005). Thermostable manganese-containing superoxide dismutase from the thermophilic fungus *Thermomyces lanuginosus*. *Extremophiles* 9, 1–6. doi: 10.1007/s00792-004-0413-4

Li, H., Feng, Z. M., Sun, Y. J., Ning, S. J., Zhou, W. L., Liu, A., et al. (2016). Engineering a thermostable iron superoxide dismutase based on manganese superoxide dismutase from *Thermus thermophilus*. *Process Biochem.* 51, 39–47. doi: 10.1016/j.procbio.2015.11.001

Li, M., Guo, S., Li, X., Wang, Q., and Wang, W. (2017). Engineering a highly thermostable and stress tolerant superoxide dismutase by N-terminal modification and metal incorporation. *Biotechnol. Bioprocess Eng.* 22, 725–733. doi: 10.1007/s12257-017-0243-8

Li, W., Wang, H., Chen, Z., Ye, Q., Tian, Y., Xu, X., et al. (2014). Probing the metal specificity mechanism of superoxide dismutase from human pathogen *Clostridium difficile*. *Chem. Commun.* 50, 584–586. doi: 10.1039/c3cc47859a

Liu, P., Ewis, H. E., Huang, Y. J., Lu, C. D., Tai, P. C., and Weber, I. T. (2007). Structure of *Bacillus subtilis* superoxide dismutase. *Acta Crystallogr. F. Struct. Biol. Cryst. Commun.* 63, 1003–1007.

López, G., Díaz-Cárdenas, C., Alzate, J. D., Gonzalez, L. N., and Baena, S. (2018). Description of *Alicyclobacillus montanus* sp. nov. a mixotrophic bacterium isolated from acidic hot springs. *Int. J. Syst. Evol. Microbiol.* 68, 1608–1615. doi: 10.1099/ijsem.0.002718

Mandelli, F., Franco Cairo, J. P. L., Citadini, A. P. S., Mandelli, F., Alvarez, T. M., Oliveira, R. J., et al. (2013). The characterization of a thermostable and cambialistic superoxide dismutase from *Thermus filiformis*. *Lett. Appl. Microbiol.* 57, 40–46. doi: 10.1111/lam.12071

Manolov, V., Yonova, D., Bogov, B., Petrova, J., Vasilev, V., and Vazelov, E. (2017). Hepcidin, selenium and superoxide dismutase in oxidative stress and in dialysis patients. *J. Urol. Nephrol.* 2:000116.

Marklund, S. L., and Marklund, G. (1974). Involvement of the superoxide anion radical in the autoxidation of pyrogallol and a convenient assay for superoxide dismutase. *Eur. J. Biochem.* 47, 469–474. doi: 10.1111/j.1432-1033.1974.tb03714.x

Matès, J. M., and Sánchez-Jimènez, F. (2000). Role of active oxygen species in apoptosis: implications for cancer therapy. *Int. J. Biochem. Cell. Biol.* 32, 157–170. doi: 10.1016/s1357-2725(99)00088-6

McKnight, I. C., Eiroa, M. N., Sant, A. S., and Massaguer, P. R. (2010). *Alicyclobacillus acidoterrestris* in pasteurized exotic Brazilian fruit juices: isolation, genotypic characterization and heat resistance. *Food Microbiol.* 27, 1016–1022. doi: 10.1016/j.fm.2010.06.010

Meier, B., Barra, D., Bossa, F., Calabrese, L., and Rotilio, G. (1982). Synthesis of either Fe- or Mn-superoxide dismutase with an apparently identical protein moiety by an anaerobic bacterium dependent on the metal supplied. *J. Biol. Chem.* 257, 13977–13980. doi: 10.1016/s0021-9258(19)45329-5

Miller, A. F. (2012). Superoxide dismutases: ancient enzymes and new insights. *FEBS Lett.* 586, 585–595. doi: 10.1016/j.febslet.2011.10.048

Morozkina, E. V., Slutskaya, E. S., Fedorova, T. V., Tugay, T. I., Golubeva, L. I., and Korolev, O. V. (2010). Extremophilic microorganisms: biochemical adaptation and biotechnological application (Review). *Appl. Biochem. Microbiol.* 46, 1–14. doi: 10.1134/s0003683810010011

Nakamura, T., Nakamura, T., Torikai, K., Uegaki, K., Morita, J., Machida, K., et al. (2011). Crystal structure of the cambialistic superoxide dismutase from *Aeropyrum pernix* K1–insights into the enzyme mechanism and stability. *FEBS J.* 278, 598–609. doi: 10.1111/j.1742-4658.2010.07977.x

Naso, F. C. D., Dias, A. S., Porawski, M., and Marroni, N. A. P. (2011). Exogenous superoxide dismutase: action on liver oxidative stress in animals with streptozotocin-induced diabetes. *Exp. Diabetes Res.* 2011:754132.

Perry, J. J. P., Shin, D. S., Getzoff, E. D., and Tainer, J. A. (2010). The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta* 1804, 245–262.

Schmidt, M., Meier, B., and Parak, F. (1996). X-ray structure of the cambialistic superoxide dismutase from *Propionibacterium shermanii* active with Fe or Mn. *J. Biol. Inorg. Chem.* 1, 532–541. doi: 10.1007/s007750050089

Seatovic, S., Gligic, L., Radulovic, Z., and Jankov, R. M. (2004). Purification and partial characterization of superoxide dismutase from the thermophilic bacteria *Thermothrix* sp. *J. Serb. Chem. Soc.* 69, 9–16. doi: 10.2298/jsc0401009s

Singh, S., and Singh, J. (2004). Controlled release of a model protein lysozyme from phase sensitive smart polymer systems. *Inter. J. Pharm.* 271, 189–196. doi: 10.1016/j.ijpharm.2003.11.010

Slutskaya, E. S., Bezsudnova, E. Y., Mardanov, A. V., Safenkova, I. V., Kleimenov, S. Y., Chebotareva, N. A., et al. (2012). Iron-dependent superoxide dismutase from novel thermoacidophilic crenarchaeon *Acidilobus saccharovorans*: from gene to active enzyme. *Biochemistry* 77, 1368–1376. doi: 10.1134/s0006297912120048

Song, N. N., Zheng, Y., E, S. J., and Li, D. C. (2009). Cloning, expression, and characterization of thermostable manganese superoxide dismutase from *Thermoascus aurantiacus* var. *levisporus*. *J. Microbiol.* 47, 123–130. doi: 10.1007/s12275-008-0217-9

Steinman, H. M. (1978). The amino acid sequence of Mangano superoxide dismutase from *Escherichia coli* B. *J. Biol. Chem.* 253, 8708–8720. doi: 10.1016/s0021-9258(17)34235-7

Stephenie, S., Chang, Y. P., Gnanasekaran, A., Esa, N. M., and Gnanaraj, C. (2020). An insight on superoxide dismutase (SOD) from plants for mammalian health enhancement. *J. Funct. Foods* 68, 103917. doi: 10.1016/j.jff.2020.103917

Trotti, A. (1997). Toxicity antagonists in cancer therapy. *Curr. Opin. Oncol.* 9, 569–578. doi: 10.1097/00001622-199711000-00013

Vieille, C., and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* 65, 1–43. doi: 10.1128/mmbr.65.1.1-43.2001

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-model: homology modelling of protein structures and complexes. *Nucleic. Acids. Res.* 46, 296–303.

Yamano, S., Sako, Y., Nomura, N., and Maruyama, T. (1999). A cambialistic SOD in a strictly aerobic hyperthermophilic archaeon, *Aeropyrum pernix*. *J. Biochem.* 126, 218–225. doi: 10.1093/oxfordjournals.jbchem.a022426

Zhang, B., Wu, Y. F., Song, J. L., Huang, Z. S., and Jiang, C. Y. (2015). *Alicyclobacillus fodiniaquatilis* sp. nov, isolated from acid mine water. *Int. J. Syst. Evol. Microbiol.* 65, 4915–4920. doi: 10.1099/ijsem.0.000695

Zhang, L., Yin, H., Zhao, Q., Yang, C., and Wang, Y. (2018). High alkaline activity of a thermostable α-amylase (cyclomaltodextrinase) from thermoacidophilic *Alicyclobacillus* isolate. *Ann. Microbiol.* 68, 811–888.

Zhu, H., Liu, J., Qu, J., Gao, X., and Lu, J. R. (2013). Stress fermentation strategies for the production of hyperthermostable superoxide dismutase from *Thermus thermophilus* HB27: effects of ions. *Extremophiles* 17, 995–1002. doi: 10.1007/s00792-013-0581-1

Zhu, Y., Wang, G., Ni, H., Xiao, A., and Cai, H. (2014). Cloning and characterization of a new manganese superoxide dismutase from deep-sea thermophile *Geobacillus* sp. EPT3. *World J. Microbiol. Biotechnol.* 30, 1347–1357. doi: 10.1007/s11274-013-1536-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership