



OPEN ACCESS

EDITED BY
Kürşad Turgay,
Max Planck Society,
Germany

REVIEWED BY
Axel Mogk,
Heidelberg University,
Germany
Marta Carroni,
Science for Life Laboratory (SciLifeLab),
Sweden

*CORRESPONDENCE
Karl R. Schmitz
✉ schmitzk@udel.edu

SPECIALTY SECTION
This article was submitted to
Evolutionary and Genomic Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 08 February 2023
ACCEPTED 24 March 2023
PUBLISHED 17 April 2023

CITATION
Jiang J and Schmitz KR (2023) Bioinformatic
identification of Clpl, a distinct class of Clp
unfoldases in Actinomycetota.
Front. Microbiol. 14:1161764.
doi: 10.3389/fmicb.2023.1161764

COPYRIGHT
© 2023 Jiang and Schmitz. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Bioinformatic identification of Clpl, a distinct class of Clp unfoldases in Actinomycetota

Jialiu Jiang¹ and Karl R. Schmitz^{1,2*}

¹Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, United States,

²Department of Biological Sciences, University of Delaware, Newark, DE, United States

All clades of bacteria possess Hsp100/Clp family unfoldase enzymes that contribute to aspects of protein quality control. In Actinomycetota, these include ClpB, which functions as an independent chaperone and disaggregase, and ClpC, which cooperates with the ClpP1P2 peptidase to carry out regulated proteolysis of client proteins. We initially sought to algorithmically catalog Clp unfoldase orthologs from Actinomycetota into ClpB and ClpC categories. In the process, we uncovered a phylogenetically distinct third group of double-ringed Clp enzymes, which we term Clpl. Clpl enzymes are architecturally similar to ClpB and ClpC, with intact ATPase modules and motifs associated with substrate unfolding and translation. While Clpl possess an M-domain similar in length to that of ClpC, its N-terminal domain is more variable than the strongly conserved N-terminal domain of ClpC. Surprisingly, Clpl sequences are divisible into sub-classes that either possess or lack the LGF-motifs required for stable assembly with ClpP1P2, suggesting distinct cellular roles. The presence of Clpl enzymes likely provides bacteria with expanded complexity and regulatory control over protein quality control programs, supplementing the conserved roles of ClpB and ClpC.

KEYWORDS

Actinomycetota, Clpl, unfoldase, protein quality control, proteolysis, Clp protease

Introduction

All bacteria possess double-ringed Hsp100/Clp family AAA+ (ATPases Associated with various cellular Activities) enzymes that harness chemical energy from ATP to unfold client proteins. In Actinomycetota (synonym Actinobacteria), these include the enzymes ClpB and ClpC, which participate in distinct aspects of protein quality control. ClpB operates as a chaperone and disaggregase, remodeling and releasing client proteins to refold in the cytosol (Weibezahn et al., 2004; Rosenzweig et al., 2013; Alam et al., 2021). ClpC functions as the unfoldase component of the ClpCP1P2 protease (Akopian et al., 2012; Schmitz and Sauer, 2014; Lunge et al., 2020). Target proteins unfolded by ClpC are translocated into the degradation chamber of the associated ClpP1P2 peptidase for degradation (Leodolter et al., 2015; Weinhäupl et al., 2022). Although the specific roles of ClpCP1P2 in Actinomycetota are poorly defined, this protease has emerged as a promising antibacterial target in *Mycobacterium tuberculosis*, and multiple compounds are known to kill *M. tuberculosis* by disrupting ClpCP1P2 activity (Schmitt et al., 2011; Ollinger et al., 2012; Gavrish et al., 2014; Gao et al., 2015; Famulla et al., 2016; Choules et al., 2019; Schmitz et al., 2020; Li et al., 2021).

ClpB and ClpC share a common architecture and operating principles. Both enzymes consist of a family-specific N-domain followed by two AAA+ ATPase modules, and both function as homomeric hexamers with a central axial channel (Lee et al., 2003; Kojetin et al., 2009; Lopez et al., 2020; Kim et al., 2022). Protein substrates are engaged by axial loops within the channel (Schlieker et al., 2004; Rizo et al., 2019). ATP hydrolysis events in individual ATP

modules drive conformational changes in the ring that apply an unfolding force to gripped substrates (Yu et al., 2018; Rizo et al., 2019). The interaction between ClpC and the ClpP1P2 peptidase is stabilized by flexible LGF-loops present on the underside of the unfoldase, which dock into hydrophobic pockets on the surface of the peptidase (Figure 1A; Kim et al., 2001; Weinhäupl et al., 2022). ClpB lacks LGF-loops and thus cannot collaborate with ClpP1P2 to carry out ATP-dependent proteolysis (Weibezahn et al., 2004; Duran et al., 2017). Indeed, the presence or absence of LGF-loops serves as a useful sequence marker for discriminating ClpB and ClpC sequences.

We initially sought to discriminate Actinomycetota ClpB and ClpC enzymes based on characteristic sequence features, as a prerequisite for analyzing the unique patterns of sequence conservation among ClpC orthologs. In the process, we uncovered a group of ClpC/B paralogs with intermediate characteristics, which we term ClpI enzymes (Figures 1B,C). Bioinformatic analyses reveals that ClpI sequences possess conserved features associated with ATP hydrolysis and unfoldase activity, but are evolutionarily distinct from both ClpC and ClpB. Notably, ClpI sequences occur with and without LGF-loops, suggesting that individual lineages of ClpI enzymes have evolved to work either as independent unfoldases or to proteolyze client proteins in cooperation with ClpP1P2. Our findings expand on the known diversity of AAA+ unfoldases, and reveal new points of regulation by which species within Actinomycetota can modulate protein quality control.

Methods

Sequence analysis of Actinomycetota ClpC paralogs

Ortholog sequences were collected using NCBI BLAST (Altschul et al., 1990) and the HMMER search algorithm (Finn et al., 2011). The BLAST bit-score was used to assess the similarity of hits to reference

sequences. 2D scatterplots illustrating the similarity of dataset members to reference sequences were constructed using the Python 3 matplotlib package (Hunter, 2007; Van Rossum and Drake, 2009). Each data point in the resulting plots corresponds to the sequence of a single ClpC paralog.

As an independent assessment of similarity, principal component analysis of datasets was used as a dimension reducing method (Jolliffe, 2014). Alignment scores of each sequence were determined and expressed as principal components with components equal to 10.

Multiple sequence alignments and phylogenetic analysis

Multiple sequence alignments of ClpC, ClpB, or ClpI orthologs were created using ClustalW (Thompson et al., 1994). Sequence logos were generated from alignments using WebLogo (Crooks et al., 2004). To compare the ClpC/B/I sequence occurrence among phylogenetic groups, the taxonomic order associated with each hit was extracted from the UniProt database using Biopython (Cock et al., 2009; Pundir et al., 2016). For phylogenetic analysis, 5 sequences were randomly selected from each taxonomic order (for orders with fewer than 5 representatives all sequences were included), sequences were aligned using ClustalW, and the resulting alignment was used to guide construction of a phylogenetic tree using ETE TOOLKIT (Huerta-Cepas et al., 2010, 2016). Clusters with bootstrap values greater than 50% were defined as confirmed subgroups (Tripathi and Sowdhamini, 2008). Genomic contexts of representative sequences were assessed using the MicrobesOnline resource (Dehal et al., 2010).

Cross-alignment sequence comparison

Pre-calculated ClustalW alignments of individual ClpC/B/I paralogs were aligned to one another using the profile-profile method

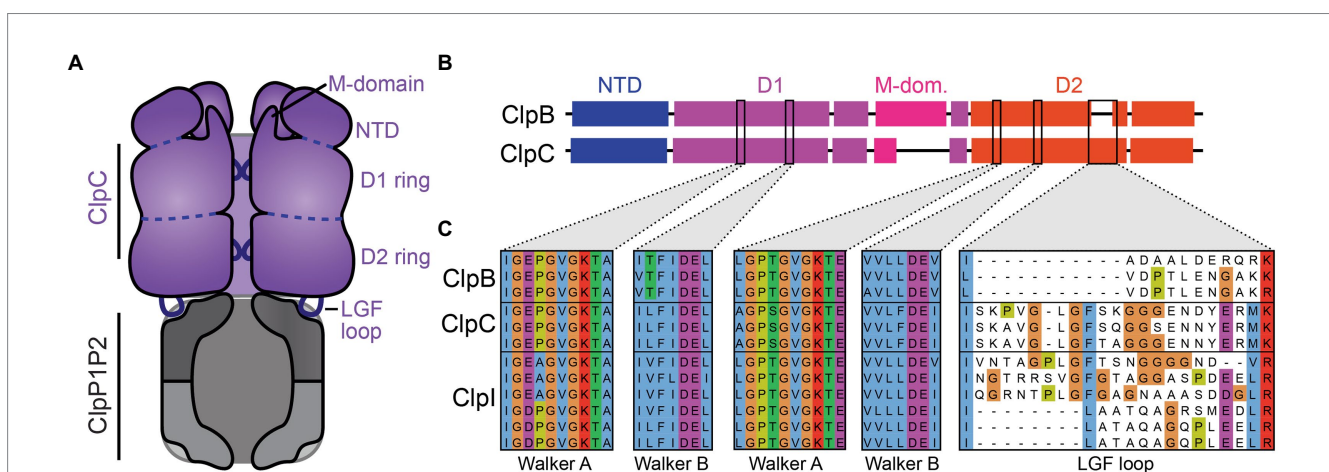


FIGURE 1 Clp enzyme architecture and domain organization. (A) A cartoon of ClpC illustrates notable features of Clp enzymes. In the context of proteolysis by ClpCP1P2, client substrates are recognized by ClpC, unfolded, and translocated through the axial pore into the peptidase degradation chamber. (B) ClpB and ClpC proteins contain a family-specific N-terminal domain (NTD), followed by two AAA+ ATPase modules (D1 and D2) with conserved Walker A and B motifs. The D1 ATPase module is interrupted by a helical M-domain that projects upward toward the NTD. ClpC enzymes additionally possess an LGF loop that makes stabilizing contacts with the ClpP2 ring of the peptidase. (C) Sequence alignment of select ClpC orthologs from Actinomycetota shows strong conservation of Walker A and B motifs. Notably, a subset of sequences possesses a short M-domain and distinct LGF loop regions, with or without identifiable LGF-like motifs.

in MUSCLE (Edgar, 2004). To illustrate positional conservation between top and bottom alignments, a 20-member profile array P of amino acid frequency f was constructed for each position in each alignment: $P = (f_1, \dots, f_{20})$ (Gribnikov et al., 1987). For each residue x in an alignment, a BLOSUM62-derived similarity score was calculated by summing BLOSUM62 substitution scores for that residue (Henikoff and Henikoff, 1992), weighted by the profile frequency of substituted residues y .

$$S_x = \sum_y \text{BLOSUM}(x,y) P(y).$$

Alignments were represented as color-coded bitmaps, with each position colored according to its score. A cross-alignment similarity score S was calculated for each position by iterating over all residue combinations (x, y) and summing BLOSUM62 scores weighted by the profile frequency in both the top (P_t) and bottom (P_b) frequency profile arrays.

$$S_{\text{cross-alignment}} = \sum_{x,y} \text{BLOSUM}(x,y) P_t(x) P_b(y).$$

Positional cross-comparison scores were represented as a one-dimensional heatmap between alignments.

Results

Actinomycetota possess a group of unusual Clp unfoldases

As an outgrowth of our interest in mycobacterial ClpC, we sought to assess the pattern of amino acid conservation across ClpC orthologs within Actinomycetota. We used the search algorithm HMMER (Finn et al., 2011) to collect Actinomycetota homologs of *Mycobacterium smegmatis* ClpC1 and aligned these using ClustalW (Thompson et al., 1994). ClpC bears substantial sequence and structural homology to ClpB, and our search results consequently included numerous hits annotated as ClpB or ambiguously as Clp enzymes. We attempted to use known sequence features to definitively sort homologs into separate ClpC and ClpB categories. Both groups of enzymes contain conserved motifs associated with ATP hydrolysis and unfolding/translocation (Figure 1C), but two major sequence features distinguish them: (i) the M-domain that projects upward from the D1 ATPase module is shorter in ClpC1 (~20 aa) than in ClpB (~90 aa) and (ii) ClpC, but not ClpB, contains an LGF-loop with a Leu-Gly-Phe or similar motif that stabilizes binding to the surface of the ClpP1P2 peptidase (Kim et al., 2001; Leodolter et al., 2015; Amor et al., 2016). Based on these features, we were able to categorize the majority of sequences as either ClpC or ClpB orthologs. However, a subset of sequences possessed short ClpC-like M-domains and either lacked LGF-loops or lacked identifiable LGF-like motifs. These unusual sequences had higher homology to ClpC and ClpB than to other AAA+ enzymes. From these sequence features alone, it was unclear whether intermediate sequences were atypical ClpC enzymes, atypical ClpB enzymes, or an entirely separate category.

To understand the relationship between atypical sequences and ClpC/ClpB groups, we used NCBI BLAST to assess the similarity of

each entry in our Actinomycetota dataset to *M. smegmatis* ClpC1 and ClpB reference sequences. BLAST bit-scores to each reference were plotted as a two-dimensional scatter plot (Figure 2A). The majority of sequences lie in two off-diagonal clusters: one with higher similarity to ClpC (Figure 2A purple box; 1,294 sequences) and one with higher similarity to ClpB (Figure 2A pink box; 1,162 sequences). Inspection of individual hits in these clusters confirmed that they comprise canonical ClpC and ClpB sequences, based on sequence features and annotations. However, a subset of sequences occupied a distinct third cluster, with moderate but approximately equal similarity to ClpC and ClpB references (Figure 2A blue box; 451 sequences). Importantly, sequences clustered into this third group based on overall sequence similarity to ClpB and ClpC references, rather than on specific sequence features. The intermediate cluster included the outlier sequences noted above that lack LGF-motifs, but also included a number of sequences with identifiable LGF-motifs (Figure 1C). Given their intermediate similarity to ClpC and ClpB references, we termed the proteins in this third group ClpI enzymes. ClpI subtypes with LGF-motifs were termed ClpIa, and those without LGF-motifs were labeled ClpIb.

For comparison, we performed an equivalent analysis of proteobacterial Clp enzymes. HMMER was used to compile proteobacterial homologs of *E. coli* ClpA, which were plotted based on BLAST similarity to *E. coli* ClpA and ClpB references (Figure 2B). The majority of hits appeared in either ClpA (orange box) or ClpB (green box) groups, with only a handful of sequences lying on the diagonal in two sparse clusters. A minor cluster located below the main ClpB group contained sequences of the Type VI secretion system ATPase ClpV/TssH (cyan box), which possess an M-domain but generally lack LGF-loops (Schlieker et al., 2005). Most other on-diagonal sequences were shorter than ClpA or ClpB, and thus possibly fragmentary genes or the result of mis-annotated translational start sites. The overall clustering pattern is in agreement with the expectation that ClpA and ClpB are the predominant groups of double-ringed Clp enzymes in Proteobacteria (Zolkiewski, 2006). Moreover, the clustering of ClpV sequences validates this approach for identifying distinct ortholog subtypes, and makes it unlikely that the ClpI cluster is an artifact of the analysis method.

As an independent assessment of sequence clustering, we combined sets of sequences and subjected them to principal component analysis (PCA; Jolliffe, 2014). PCA plots consistently sorted ClpI sequences into groups separate from ClpB and ClpC (Supplementary Figure S1). ClpIa and ClpIb subtypes were more similar, and only partially resolved into separate clusters.

ClpI orthologs occur in a subset of orders in Actinomycetota

ClpC, ClpB, and ClpI sequences were not equally abundant within the set of ClpC homologs identified by HMMER. ClpC and ClpB made up the majority of the dataset with similar individual abundance (~40% and ~44% of total, respectively), while ClpI sequences comprised only ~16% of total. The lower prevalence of ClpI suggests that these are present only in a subset of clades.

To establish the distribution of ClpI enzymes across phylogenetic lineages, we binned ClpB, ClpC, and ClpI sequences from our dataset by taxonomic order (Table 1). ClpB and ClpC sequences were found

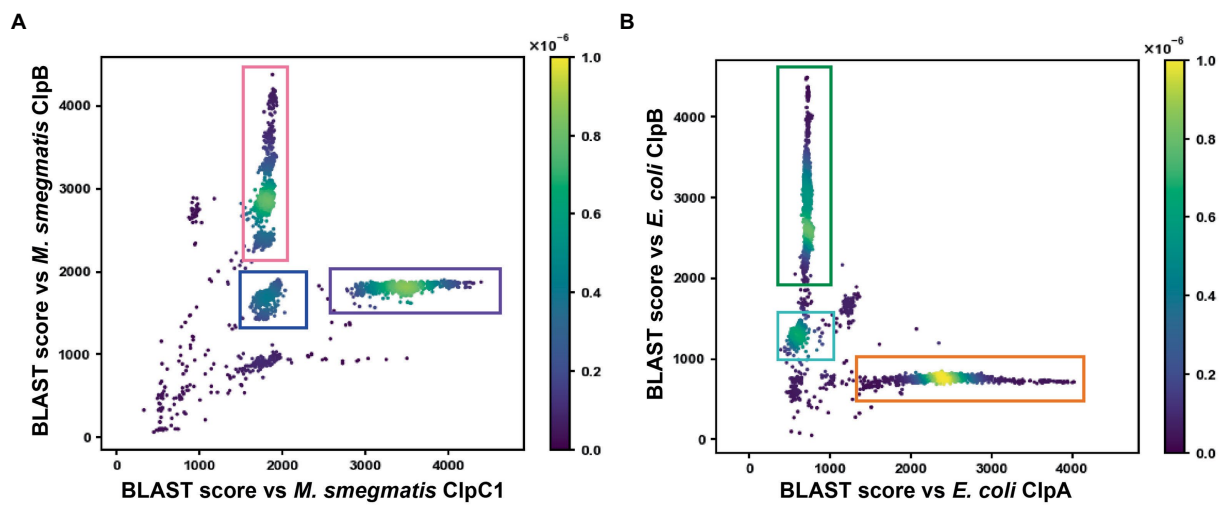


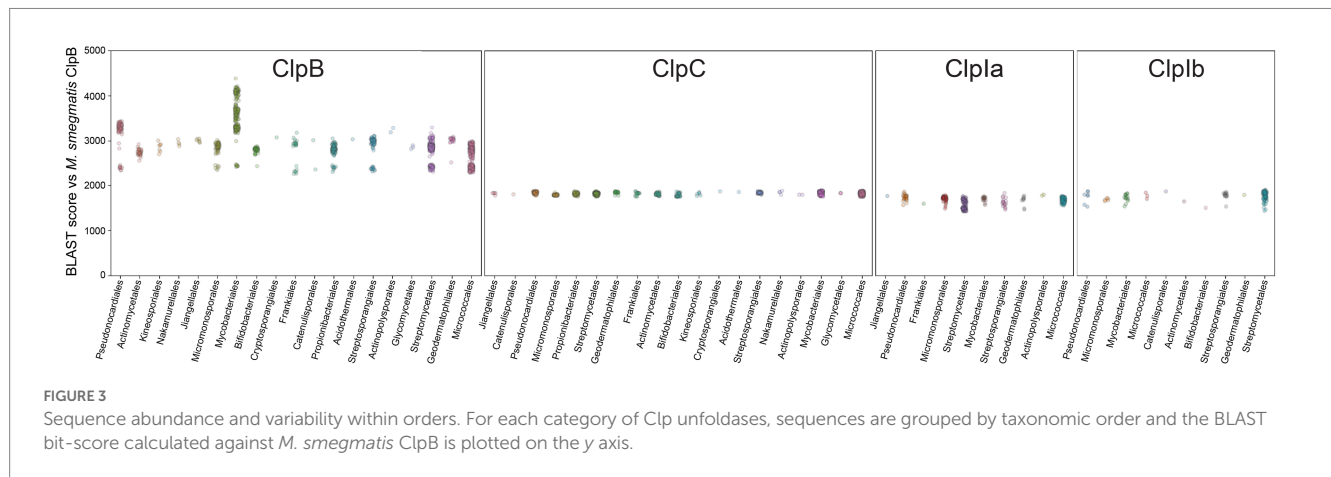
FIGURE 2
 BLAST analysis of ClpC/B/A orthologs. **(A)** Actinomycetota ClpC orthologs were compared to *Mycobacterium smegmatis* ClpB (UniProt ID: A0QQF0_MYCS2) and ClpC1 (CLPC1_MYCS2) reference sequences. The x axis represents the BLAST bit-score resulting from alignment of dataset sequences with the ClpC1 reference, while the y axis represents the BLAST bit-score associated with alignment to the ClpB reference. The density of hits in a region of the plot is shown as a heatmap. Typical ClpC orthologs cluster within the purple box, typical ClpB orthologs within the pink box, and atypical sequences with intermediate features cluster within the blue box. **(B)** Proteobacterial ClpA orthologs were compared to *E. coli* ClpA (CLPA_ECOLI) and ClpB (CLPB_ECOLI) references as in A. Conventional ClpA sequences lie within the orange box, conventional ClpB sequences lie within the green box, and orthologs of ClpV, a translocase associated with Type IV secretion systems, occupy the cyan box.

TABLE 1 Occurrence of ClpB, ClpC, and ClpI across Actinomycetota orders.

Order	ClpB	ClpC	ClpIa	ClpIb
Pseudonocardiales	83	73	25	7
Actinomycetales	36	39	0	1
Kineosporiales	7	6	0	0
Nakamurellales	4	4	0	0
Jiangellales	6	6	1	0
Micromonosporales	64	58	50	5
Bifidobacteriales	46	36	0	1
Cryptosporangiales	1	1	0	0
Actinopolysporales	2	2	2	0
Frankiales	23	16	1	0
Catenulesporales	2	1	0	1
Propionibacteriales	113	96	0	0
Mycobacteriales	204	185	15	14
Acidothermales	1	1	0	0
Streptosporangiales	75	50	18	14
Glycomycetales	3	3	0	0
Streptomycetales	296	230	117	78
Coriobacteriales	1	0	0	0
Geodermatophilales	22	21	9	1
Micrococcales	305	334	88	3
Total	1,294	1,162	326	125
% Total	44.5	40.0	11.2	4.3

in all 20 observed orders, suggesting that both enzymes play a critical role in cellular function and are evolutionary conserved across taxa. Interestingly, many individual orders contained unequal numbers of ClpB and ClpC sequences. For example, in Streptosporangiales we found 75 ClpB orthologs but only 50 ClpC orthologs. Conversely, Micrococcales contained 305 ClpB and 334 ClpC orthologs. As expected from the lower overall abundance, ClpI sequences were present in only 13 orders. In orders that possess ClpI orthologs, the number of ClpI sequences was generally lower than either ClpB or ClpC, suggesting that ancestral ClpI enzymes were lost in some lineages, and thus play a less critical role in cellular physiology. The prevalence of ClpIa and ClpIb subtypes within orders also varied. In most cases, (e.g., Pseudonocardiales, Micromonosporales, and Micrococcales) ClpIa sequences outnumbered ClpIb. Only in Mycobacteriales and Streptosporangiales were the two subtypes similarly abundant, although still less prevalent than ClpB or ClpC. Several orders (e.g., Actinomycetales) possessed only a single ClpI sequence, which may have arrived by horizontal gene transfer. ClpI enzymes do not necessarily replace ClpB or ClpC enzymes in a given species. We note species in which all four enzymes are clearly present—for example, *Rhodococcus* sp. *ABRD24* (ClpB Uniprot ID: ERC79_08130; ClpC: A0A4P6UCA9_9NOCA; ClpIa: A0A4P6UA74_9NOCA; ClpIb: A0A4P6UBT3_9NOCA).

To assess the similarity of Clp enzyme sequences within taxa, we grouped sequences by taxonomic order and plotted BLAST bit-scores generated against *M. smegmatis* ClpB (Figure 3). ClpB orthologs showed a wide range of scores between 2,000 and 4,500. The diversity of ClpB scores, particularly within Mycobacteriales, reflects the fact that some ClpB sequences are highly similar to the reference, while more distantly related ClpB orthologs have more divergent sequence composition. However, there also appears to be consistent



diversity in ClpB sequence composition across orders, which manifests as separate clusters spaced along the y-axis at scores of ~2,300 and ~2,600. Interestingly, all ClpC group members produced similar scores of ~2000, suggesting less variation of ClpC sequences across clades. ClpI sequences produced scores similar to ClpC, although both ClpIa and ClpIb subgroups exhibited greater variability in score than ClpC, suggesting lower sequence conservation.

ClpIa and ClpIb subtypes are phylogenetically distinct

One curious characteristic of ClpI sequences is the presence of homologs with (ClpIa) and without (ClpIb) identifiable LGF-motifs. This heterogeneity stands in contrast to ClpC and ClpB groups, which uniformly possess or lack LGF-motifs, respectively. The existence of ClpI enzymes with and without LGF-motifs suggests either that ClpIa and ClpIb are evolutionarily distinct subgroups, or that LGF-motifs have been independently gained or lost over time in individual lineages.

To understand whether ClpIa and ClpIb subgroups represent separate evolutionary lineages, we performed phylogenetic analysis on a subset of ClpC, ClpB, and ClpI enzymes from each order within Actinomycetota (Figure 4). ClpB sequences form a distinct phylogenetic group, as do most ClpC and ClpI enzymes, in agreement with NCBI BLAST analysis (Figure 2A; Bifidobacteriales ClpC and ClpI are an exception, clustering together in a branch between the major ClpC and ClpI divisions.). Interestingly, we observe that ClpIa and ClpIb subtypes sort into separate divisions within the main ClpI branch. This pattern suggests that duplication of an ancestral ClpI-encoding gene (or horizontal acquisition of an ancestral ClpI) occurred early in Actinomycetota evolution, allowing differentiation into ClpIa and ClpIb paralogs with distinct cellular roles.

We also examined the genomic contexts of *clpI* genes. In some cases, these loci are found near genes associated with stress or protein quality control (Supplementary Figures S2A,B). For example, *clpIb* in *Saccharopolyspora erythraea* is found near *rpoE*, which is associated with the exocyttoplasmic stress response (Raivio and Silhavy, 2001). The *clpIa* gene in *Saccharopolyspora erythraea* is found near a cluster of universal stress protein genes (*uspA*; Nachin et al., 2005). In *Brevibacterium linens*, *clpIa* is located upstream of a secondary *clpP1-clpP2* locus, suggesting a role in proteolysis.

However, these local associations are not conserved across taxa. For comparison, ClpB- and ClpC-encoding genes are more commonly associated with specific neighbors: *clpB* frequently occurs near a short-chain dehydrogenase; *clpC* occurs near genes encoding lysyl-tRNA synthetase and the DNA repair protein *mutY* (Supplementary Figures S2C,D). Some other unfoldase-encoding genes possess strongly conserved genomic organization, such as *clpX*, which usually occurs downstream of genes encoding trigger factor and ClpP1P2 (Supplementary Figures S2E). The lack of strongly conserved genomic neighbors suggests that ClpI orthologs play diverse or less essential roles in cellular physiology.

Comparison of sequence conservation patterns among Clp unfoldases

To examine the sequence conservation within and across enzyme groups, we compared multiple sequence alignments generated for each group (Figure 5). Comparing sequence conservation patterns across alignments, we found that the large and small subdomains of the D1 and D2 ATPase modules were similarly conserved across all enzyme groups. ClpB enzymes are distinguished by substantially longer M-domains than either ClpC or ClpI (Figures 5A,B). The ClpB M-domain contains two sites that tolerate insertions of variable length, whereas the length of the ClpC M-domain is strictly conserved. The equivalent region in ClpI is similar in length to that of ClpC and possesses a conserved Tyr at the tip in place of Phe, but otherwise shares similar sequence composition (Figure 5C; Supplementary Figure S3A). Interestingly, M-domain insertions appear to be tolerated in ClpIb but not ClpIa enzymes (Figure 5D), suggesting that a specific M-domain length is more important in the context of proteolysis than disaggregation.

One stark difference among these enzyme groups is the extent of sequence conservation within the NTD. In both ClpB and ClpI, the NTD is less conserved than the ATPase modules (Figure 5B), with strict conservation among only a subset of residues that form its folded core. About 10% of ClpB sequences lack NTDs altogether, although these may be due to mis-annotated start sites. Several sites in the ClpI NTD tolerate insertions, including the extreme N-terminus. By contrast, the ClpC NTD is a fixed length with high sequence conservation even outside of the folded core, comparable to the ATPase modules (Figure 5C).

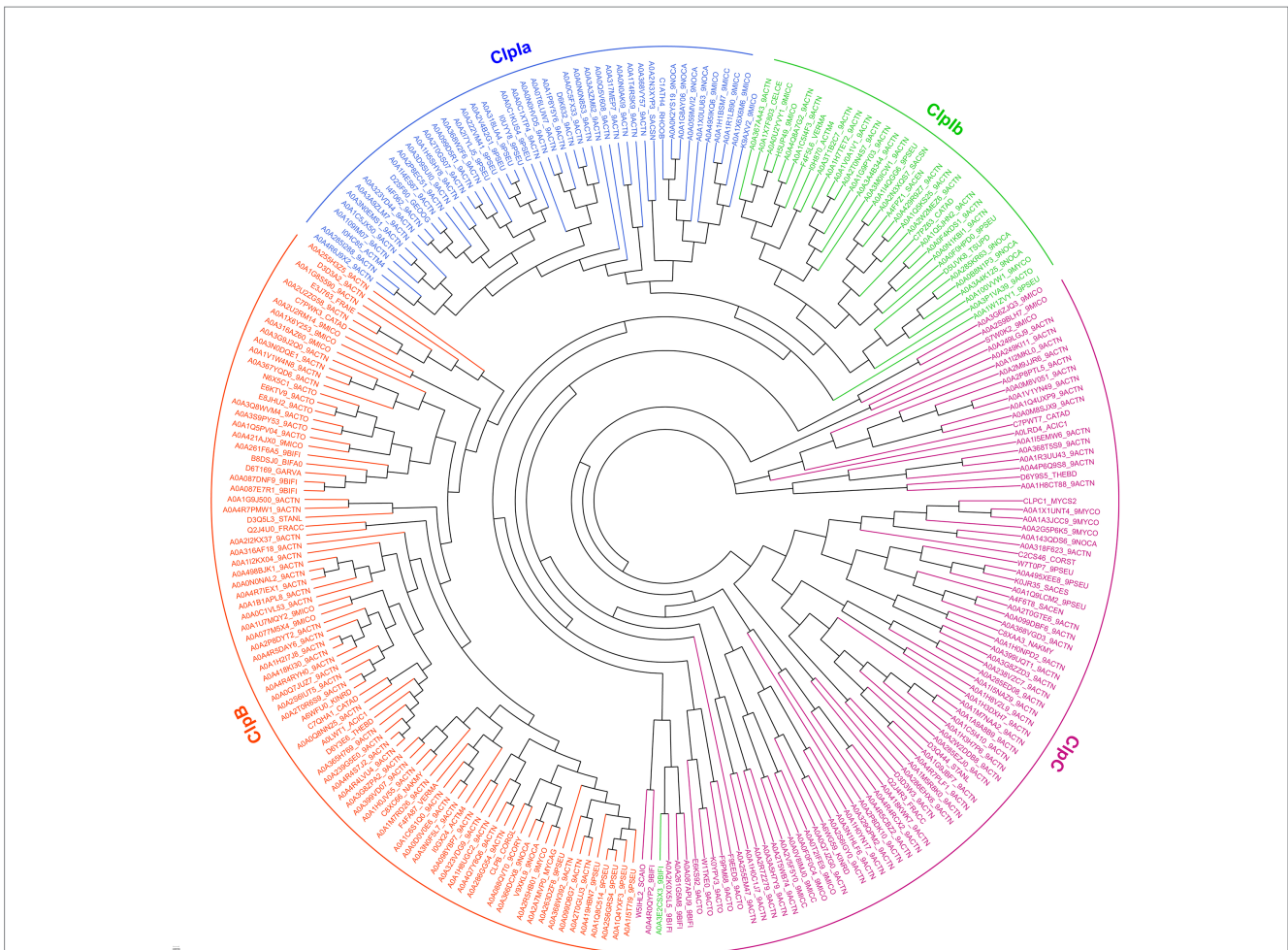
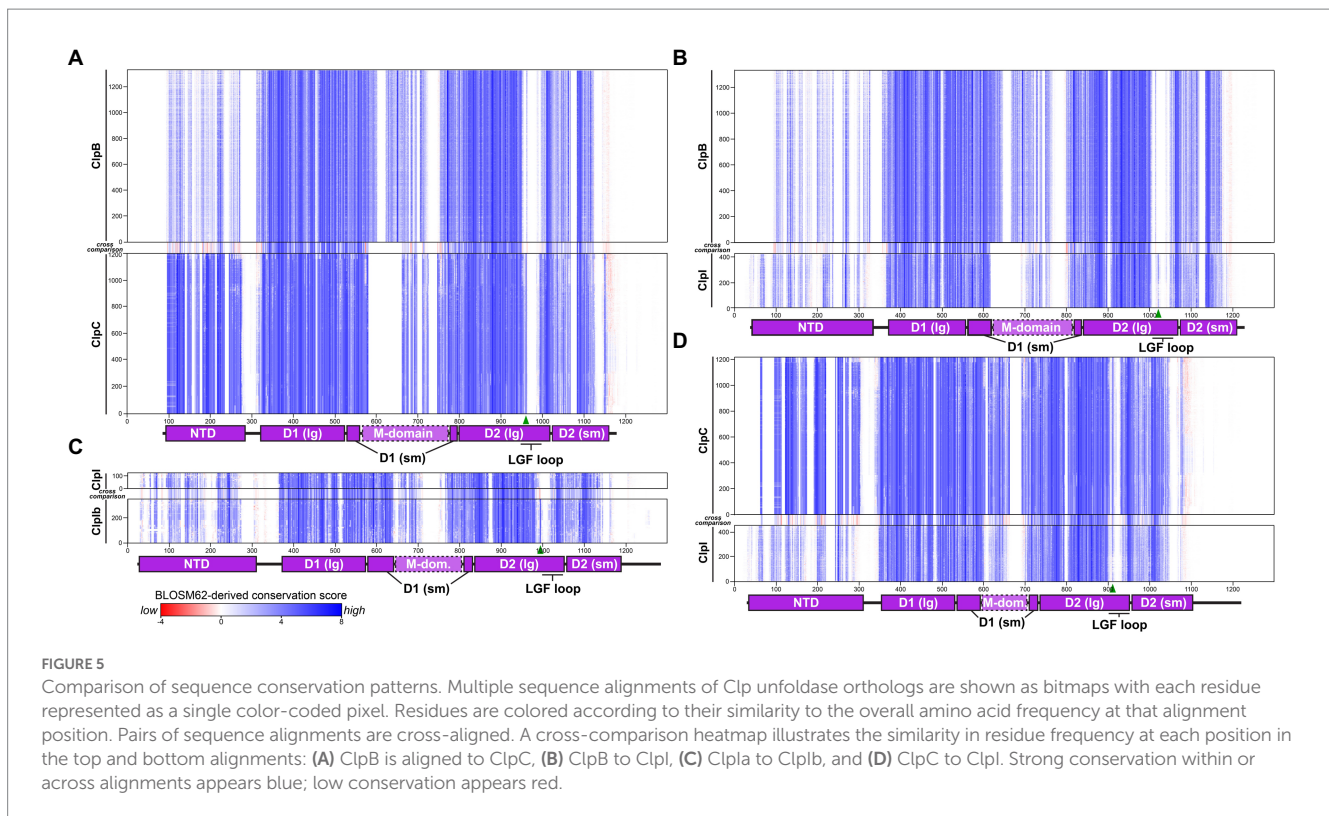


FIGURE 4 Phylogenetic analysis of ClpC paralogs. Distance methods were used to generate a phylogenetic tree of ~250 representative sequences of Actinomycetota ClpB (orange), ClpC (violet), ClpIa (blue), and ClpIb (green). Leaves are labeled with UniProt accession IDs.

This suggests that interactions made by the ClpC NTD, such as those with substrates, adaptors, or the M-domain, are both important for function and conserved across Actinomycetota. Conversely, sequence-specific functions of ClpB and ClpI NTDs are either less important or more diverse. Whereas M-domain length serves as a strong marker for ClpB sequences, the conserved NTD sequence is the strongest diagnostic feature distinguishing ClpC orthologs. Given the marked difference in NTDs between enzyme groups, we were curious whether the NTD alone drives bioinformatic partitioning of sequences into distinct ClpB, ClpC, and ClpI groups. To test this, we truncated NTDs from our dataset and performed BLAST analysis against *M. smegmatis* ClpB and ClpC1, as in Figure 2A. Truncated sequences similarly clustered into ClpB, ClpC, and ClpI groups (Supplementary Figure S4), demonstrating that characteristics beyond the NTD support the existence of an independent ClpI cluster.

Several functionally significant interaction sites are structurally defined in the NTDs of ClpB, ClpC, and ClpA. Both ClpB and ClpC NTDs possess a hydrophobic groove that bind hydrophobic regions in protein substrates (Supplementary Figure S5B; Li and Sha, 2003; Rosenzweig et al., 2015; Rizo et al., 2019). Hydrophobic residues that

line the groove are partially conserved in all enzymes (Supplementary Figure S3B). However, compared to ClpB and ClpC, most ClpI sequences have additional material at the N-terminus of the NTD (Figures 5B,C). AlphaFold2 models (Jumper et al., 2021) predict that these extensions form a folded cap that blocks access to the ClpI groove (Supplementary Figure S5C), which may prevent substrate binding to this site. Firmicutes and Actinomycetota ClpC enzymes possess two conserved phosphoarginine (pArg) binding sites (Fuhrmann et al., 2009; Weinhäupl et al., 2018; Ogonna et al., 2022). One of the two sites is partially conserved in ClpB, while neither pArg binding site is present in ClpI (Supplementary Figure S3C). Additionally, we examined residues involved in binding of *B. subtilis* MecA to ClpC (Wang et al., 2011) and *E. coli* ClpS to ClpA (Guo et al., 2002; Zeth et al., 2002). These sites, which overlap on the surface of the NTD, are conserved in ClpC and partially conserved in ClpI (Supplementary Figure S3C). Interactions between *M. tuberculosis* ClpS and ClpC1 have been observed by bacterial two-hybrid assays, and ClpS-dependent degradation of substrates by *M. tuberculosis* ClpC1P1P2 has been demonstrated (Li et al., 2020; Ziemski et al., 2020). ClpI sequence conservation patterns suggest that ClpS and other adaptors may similarly deliver substrates to some ClpI orthologs.



ClpC and ClpI NTDs are followed by a region of low conservation and variable length (Figures 5B,C), consistent with a flexible linker leading to the large subunit of the D1 ring. Indeed, the NTD was not resolved in recent cryo-EM structures of *M. tuberculosis* ClpC1 (Weinhäupl et al., 2022), presumably due to its conformational flexibility with respect to the enzyme core. Conversely, the linkage between NTD and D1 is shorter in ClpB and lacks insertions (Figure 5A), suggesting lower conformational flexibility between the ClpB NTD and the body of the unfoldase. We compared the state of this linker region in AlphaFold2 models of ClpB, ClpC, and ClpI (Supplementary Figure S5A). The short ClpB linker is predicted to pack against the NTD, tethering it close to the D1 AAA+ module. Most ClpC and ClpI predictions place the NTD against the D1 ring, but all incorporate a flexible segment in the linker, suggesting that domains can adopt flexibly tethered conformations under some circumstances. [Notably, the N-terminal NTD extension that caps the hydrophobic groove in ClpI also is predicted to form a loop that binds between the D1 large and small subdomains, well positioned to regulate D1 ATPase activity in the docked state (Supplementary Figure S5C)].

We examined sequence alignments to identify positions that were strongly conserved within individual enzyme groups but differed in residue identity between groups. These sites of contrasting conservation generally correlate with red lines in the cross-comparison strips shown in Figure 5. Positions with notable enzyme-specific amino acid identities are listed in Supplementary Table S1, and are useful in sorting uncharacterized Actinomycetota Clp enzymes into ClpB/C/I groups. A subset of these sites were previously identified as markers for Clp enzyme classes in Proteobacteria (Miller et al., 2018).

Differences in NTD

Our original motivation for classifying Clp unfoldases was to understand sequence conservation patterns among ClpC orthologs. Because the degree of overall NTD conservation varied among ClpB/C/I enzymes, we used ConSurf (Glaser et al., 2003; Ashkenazy et al., 2010, 2016) to map sequence conservation onto models of the NTD from each group and compared the results (Figure 6). In agreement with our sequence analysis (Figure 5), strong conservation was observed over most of the ClpC NTD surface (Figure 6B), including around two putative phosphoarginine binding sites (Wolf et al., 2020; Ogbonna et al., 2022). By contrast, surface conservation was lower and more localized on models of ClpB and ClpI NTDs (Figures 6A,C). The high surface conservation on the ClpC NTD reinforces the idea that this module has multiple conserved interaction partners that contribute to the cellular function and regulation of ClpC enzymes.

Discussion

Double-ringed Hsp100/Clp family unfoldases are ubiquitous in Actinomycetota. Based on sequence annotations and prior studies, Actinomycetota were known to possess two groups with distinct roles in protein homeostasis: ClpB enzymes that function as independent unfoldases and disaggregases (Yu et al., 2018), and ClpC enzymes that cooperate with the ClpP1P2 peptidase to carry out regulated proteolysis of folded proteins (Kar et al., 2008; Schmitt et al., 2011; Schmitz and Sauer, 2014; Leodolter et al., 2015; Lunge et al., 2020; Ogbonna et al., 2022; Weinhäupl et al., 2022). An ability to confidently

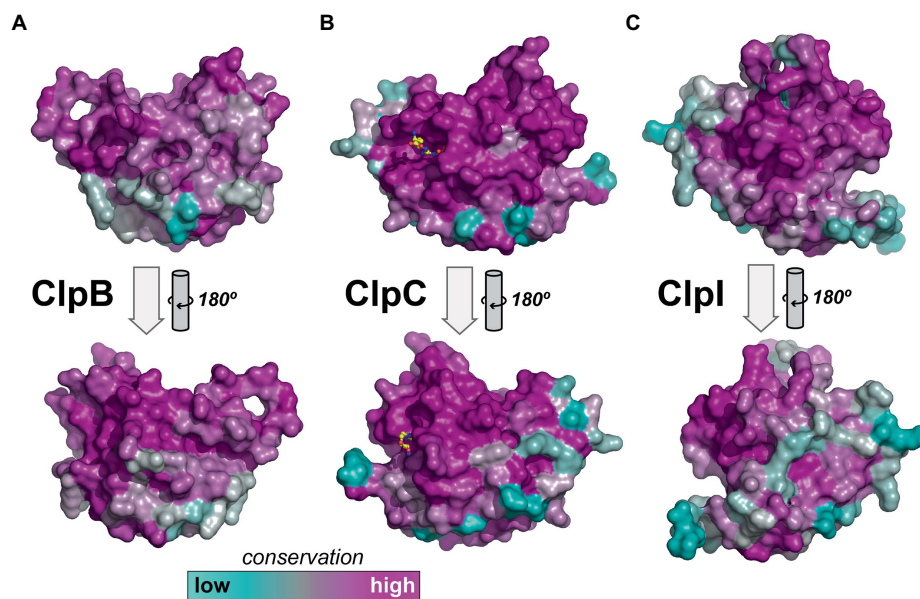


FIGURE 6

Surface conservation of Clp unfoldase NTDs. (A) ClpB sequence conservation was projected onto the NTD surface from the AlphaFold model AF-P9WPD1-F1 of *M. tuberculosis* ClpB. (B) ClpC sequence conservation was projected onto the surface of the crystal structure *M. tuberculosis* ClpC1 NTD (PDB ID: 3WDB). Phosphoarginine, shown as ball-and-sticks, was modeled into conserved binding sites (Ogbonna et al., 2022) as observed in the *B. subtilis* NTD, PDB ID: 5HBN. (C) ClpI sequence conservation was projected onto the NTD from the AlphaFold model AF-A0A4P6UA74-F1 of *Rhodococcus* sp. ABRD24 ClpI.

sort homologs into discrete enzyme families is critical for unraveling their individual contributions to cellular physiology. However, our initial sequence analysis of Actinomycetota ClpC homologs revealed numerous enzymes with similarity to ClpC and ClpB, but possessing sequence features that defy simple classification into these established groups. Our analysis here reveals the existence of a third evolutionarily distinct category of double-ringed Clp enzymes in Actinomycetota, which we term ClpI.

ClpI orthologs are less abundant than ClpB and ClpC enzymes, and appear in only a subset of Actinomycetota orders. Indeed, ClpI enzymes are absent in some of the more extensively studied species in Actinomycetota, including *Mycobacterium tuberculosis* and *Corynebacterium glutamicum* which may explain why they have not been previously described. Where they do occur, ClpI enzymes appear to exist alongside ClpB and ClpC. Accordingly, ClpI orthologs likely expand the complexity of protein quality control programs in these cells, providing additional points of regulation.

Interestingly, ClpI enzymes are divisible into subgroups, which likely have distinct cellular roles. The ClpIa subtype possess an LGF-loop similar in length to that of ClpC with an identifiable LGF-like motif, which is critical for docking of Clp unfoldases to the Clp peptidase (Kim et al., 2001; Amor et al., 2016, 2019; Fei et al., 2020). ClpIa enzymes thus presumably serve as alternative unfoldase partners for ClpP1P2, and function in regulated proteolysis of client proteins. The region equivalent to the LGF-loop in the ClpIb subtype is short, ~5 amino acids, and lacks a conserved LGF-like motif. Thus, ClpIb enzymes likely operate as independent unfoldases, although we cannot rule out the possibility that they interact in an LGF-independent manner with ClpP2; with the ClpP1 ring, which has no known interaction partners to its outer face; or even with the 20S

peptidase complex—although terminal GQYL-like motifs that canonically mediate 20S binding are not present in these enzymes (Pearce et al., 2006; Ziemski et al., 2018).

While ClpP-associated unfoldases have been shown to operate independently of ClpP in some contexts (Wickner et al., 1994; Pak and Wickner, 1997; Burton and Baker, 2005; Baytshtok et al., 2015; Kardon et al., 2015), to our knowledge, ClpI enzymes are the only group of Clp unfoldases that exist with and without LGF-motifs. This raises questions about whether these two subtypes possess differences in substrate preference or unfolding ability. An obligate unfoldase that works to rescue diverse misfolded proteins would likely benefit from more promiscuous substrate recognition than a ClpP-associated unfoldase, for which substrate selection must be tightly regulated to avoid harmful off-target proteolysis. Differences might also be expected in the degree of processivity. Successful proteolysis of large and recalcitrant substrates is thought to require strong processivity, permitting stepwise unfolding and translocation over the course of many minutes (Martin et al., 2005; Glynn et al., 2012; Fei et al., 2020; Ripstein et al., 2020; Kim et al., 2022). By contrast, the obligate unfoldase/disaggregase ClpB functions in a non-processive tug-and-release fashion under some circumstances, which may aid refolding of targets (Li et al., 2015; Rizo et al., 2019). *In vitro* biochemical studies of representative ClpI enzymes will help to answer these questions, and clarify their individual roles.

One other important finding, which arises from our improved ability to categorize ClpC orthologs, is the remarkable sequence conservation specifically among ClpC NTDs. The NTD is not known to possess enzymatic activity, nor to directly participate in substrate unfolding. Instead, it is thought to interact with substrates [e.g.,

through phosphoarginine modifications (Trentini et al., 2016; Weinhäupl et al., 2018; Ogbonna et al., 2022)], with proteolytic adaptors [e.g., the N-end rule adaptor ClpS (Ziemski et al., 2020)], and with the D1 ring and M-domain to influence the functional and oligomeric state of the entire unfoldase (Weinhäupl et al., 2018; Maurer et al., 2019; Taylor et al., 2022). Its conservation highlights the importance of these interactions across Actinomycetota. Moreover, it helps explain why multiple naturally occurring antibiotics have evolved to bind the ClpC NTD, and how antibiotics with overlapping binding sites can have different mechanistic effects on ClpC function (Choules et al., 2019; Maurer et al., 2019; Wolf et al., 2020; Hong et al., 2022).

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://doi.org/10.6084/m9.figshare.22013591.v1>, <https://doi.org/10.6084/m9.figshare.22013519.v1>, <https://doi.org/10.6084/m9.figshare.22013180.v1>, <https://doi.org/10.6084/m9.figshare.22013162.v1>.

Author contributions

JJ and KS conceived of the project, carried out analyses, and prepared the manuscript. All authors contributed to the article and approved the submitted version.

References

- Akopian, T., Kandror, O., Raju, R. M., Unnikrishnan, M., Rubin, E. J., and Goldberg, A. L. (2012). The active ClpP protease from *M. tuberculosis* is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J.* 31, 1529–1541. doi: 10.1038/emboj.2012.5
- Alam, A., Bröms, J. E., Kumar, R., and Sjöstedt, A. (2021). The role of ClpB in bacterial stress responses and virulence. *Front. Mol. Biosci.* 8:668910. doi: 10.3389/fmolb.2021.668910
- 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
- Amor, A. J., Schmitz, K. R., Baker, T. A., and Sauer, R. T. (2019). Roles of the ClpX IGF loops in ClpP association, dissociation, and protein degradation. *Protein Sci.* 28, 756–765. doi: 10.1002/pro.3590
- Amor, A. J., Schmitz, K. R., Sello, J. K., Baker, T. A., and Sauer, R. T. (2016). Highly dynamic interactions maintain kinetic stability of the ClpXP protease during the ATP-fueled mechanical cycle. *ACS Chem. Biol.* 11, 1552–1560. doi: 10.1021/acscchembio.6b00083
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi: 10.1093/nar/gkw408
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38, W529–W533. doi: 10.1093/nar/gkq399
- Baytshok, V., Baker, T. A., and Sauer, R. T. (2015). Assaying the kinetics of protein denaturation catalyzed by AAA+ unfolding machines and proteases. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5377–5382. doi: 10.1073/pnas.1505881112
- Burton, B. M., and Baker, T. A. (2005). Remodeling protein complexes: Insights from the AAA+ unfoldase ClpX and mu transposase. *Protein Sci.* 14, 1945–1954. doi: 10.1110/ps.051417505
- Choules, M. P., Wolf, N. M., Lee, H., Anderson, J. R., Grzelak, E. M., Wang, Y., et al. (2019). Rufomycin targets ClpC1 proteolysis in mycobacterium tuberculosis and *M. abscessus*. *Antimicrob. Agents Chemother.* 63, e02204–e02218. doi: 10.1128/aac.02204-18
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Duran, E. C., Weaver, C. L., and Lucius, A. L. (2017). Comparative analysis of the structure and function of AAA+ motors ClpA, ClpB, and Hsp104: Common threads and disparate functions. *Front. Mol. Biosci.* 4:54. doi: 10.3389/fmolb.2017.00054
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Famulla, K., Sass, P., Malik, I., Akopian, T., Kandror, O., Alber, M., et al. (2016). Acyldepsipeptide antibiotics kill mycobacteria by preventing the physiological functions of the ClpP1P2 protease. *Mol. Microbiol.* 101, 194–209. doi: 10.1111/mmi.13362
- Fei, X., Bell, T. A., Jenni, S., Stinson, B. M., Baker, T. A., Harrison, S. C., et al. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *elife* 9:e52774. doi: 10.7554/eLife.52774
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Fuhrmann, J., Schmidt, A., Spiess, S., Lehner, A., Turgay, K., Mechtler, K., et al. (2009). MscB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science* 324, 1323–1327. doi: 10.1126/science.1170088
- Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., et al. (2015). The cyclic peptide ecumicin targeting ClpC1 is active against mycobacterium tuberculosis in vivo. *Antimicrob. Agents Chemother.* 59, 880–889. doi: 10.1128/AAC.04054-14
- Gavriš, E., Sit, C. S., Cao, S., Kandror, O., Spoering, A., Peoples, A., et al. (2014). Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* 21, 509–518. doi: 10.1016/j.chembiol.2014.01.014

Funding

KS was supported by NIH NIGMS award P20GM104316 and by startup funds from the University of Delaware.

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.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1161764/full#supplementary-material>

- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., et al. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19, 163–164. doi: 10.1093/bioinformatics/19.1.163
- Glynn, S. E., Nager, A. R., Baker, T. A., and Sauer, R. T. (2012). Dynamic and static components power unfolding in topologically closed rings of a AAA+ proteolytic machine. *Nat. Struct. Mol. Biol.* 19, 616–622. doi: 10.1038/nsmb.2288
- Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. U. S. A.* 84, 4355–4358. doi: 10.1073/pnas.84.13.4355
- Guo, F., Esser, L., Singh, S. K., Maurizi, M. R., and Xia, D. (2002). Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *J. Biol. Chem.* 277, 46753–46762. doi: 10.1074/jbc.M208104200
- 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
- Hong, J., Duc, N. M., Jeong, B. C., Cho, S., Shetye, G., Cao, J., et al. (2022). Identification of the inhibitory mechanism of ecumicin and rufomycin 4-7 on the proteolytic activity of mycobacterium tuberculosis ClpC1/ClpP1/ClpP2 complex. *Tuberculosis (Edinb.)* 138:102298. doi: 10.1016/j.tube.2022.102298
- Huerta-Cepas, J., Dopazo, J., and Gabaldón, T. (2010). ETE: A python environment for tree exploration. *BMC Bioinformatics* 11:24. doi: 10.1186/1471-2105-11-24
- Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of Phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Jolliffe, I. (2014). “Principal component analysis” in *Wiley StatsRef: Statistics Reference Online*. eds. T. C. N. Balakrishnan, B. Everitt, W. Piegorisch, F. Ruggeri and J. L. Teugels (New York, NY: Wiley).
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kar, N. P., Sikriwal, D., Rath, P., Choudhary, R. K., and Batra, J. K. (2008). Mycobacterium tuberculosis ClpC1: Characterization and role of the N-terminal domain in its function. *FEBS J.* 275, 6149–6158. doi: 10.1111/j.1742-4658.2008.06738.x
- Kardon, J. R., Yien, Y. Y., Huston, N. C., Branco, D. S., Hildick-Smith, G. J., Rhee, K. Y., et al. (2015). Mitochondrial ClpX activates a key enzyme for Heme biosynthesis and erythropoiesis. *Cells* 161, 858–867. doi: 10.1016/j.cell.2015.04.017
- Kim, S., Fei, X., Sauer, R. T., and Baker, T. A. (2022). AAA+ protease-adaptor structures reveal altered conformations and ring specialization. *Nat. Struct. Mol. Biol.* 29, 1068–1079. doi: 10.1038/s41594-022-00850-3
- Kim, Y. I., Levchenko, I., Fraczkowska, K., Woodruff, R. V., Sauer, R. T., and Baker, T. A. (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat. Struct. Biol.* 8, 230–233. doi: 10.1038/84967
- Kojetin, D. J., McLaughlin, P. D., Thompson, R. J., Dubnau, D., Prepiak, P., Rance, M., et al. (2009). Structural and motional contributions of the Bacillus subtilis ClpC N-domain to adaptor protein interactions. *J. Mol. Biol.* 387, 639–652. doi: 10.1016/j.jmb.2009.01.046
- Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., et al. (2003). The structure of ClpB: A molecular chaperone that rescues proteins from an aggregated state. *Cells* 115, 229–240. doi: 10.1016/s0092-8674(03)00807-9
- Leodolter, J., Warweg, J., and Weber-Ban, E. (2015). The mycobacterium tuberculosis ClpP1P2 protease interacts asymmetrically with its ATPase partners ClpX and ClpC1. *PLoS One* 10:e0125345. doi: 10.1371/journal.pone.0125345
- Li, Y., Corro, J. H., Palmer, C. D., and Ojha, A. K. (2020). Progression from remodeling to hibernation of ribosomes in zinc-starved mycobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 117, 19528–19537. doi: 10.1073/pnas.2013409117
- Li, L., MacIntyre, L. W., Ali, T., Russo, R., Koirala, B., Hernandez, Y., et al. (2021). Biosynthetic interrogation of soil Metagenomes reveals Metamarin, an uncommon Cyclomarin congener with activity against mycobacterium tuberculosis. *J. Nat. Prod.* 84, 1056–1066. doi: 10.1021/acs.jnatprod.0c01104
- Li, J., and Sha, B. (2003). Crystal structure of the E. coli Hsp100 ClpB N-terminal domain. *Structure* 11, 323–328. doi: 10.1016/s0969-2126(03)00030-3
- Li, T., Weaver, C. L., Lin, J., Duran, E. C., Miller, J. M., and Lucius, A. L. (2015). Escherichia coli ClpB is a non-processive polypeptide translocase. *Biochem. J.* 470, 39–52. doi: 10.1042/bj20141457
- Lopez, K. E., Rizo, A. N., Tse, E., Lin, J., Scull, N. W., Thwin, A. C., et al. (2020). Conformational plasticity of the ClpAP AAA+ protease couples protein unfolding and proteolysis. *Nat. Struct. Mol. Biol.* 27, 406–416. doi: 10.1038/s41594-020-0409-5
- Lunge, A., Gupta, R., Choudhary, E., and Agarwal, N. (2020). The unfoldase ClpC1 of mycobacterium tuberculosis regulates the expression of a distinct subset of proteins having intrinsically disordered termini. *J. Biol. Chem.* 295, 9455–9473. doi: 10.1074/jbc.RA120.013456
- Martin, A., Baker, T. A., and Sauer, R. T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* 437, 1115–1120. doi: 10.1038/nature04031
- Maurer, M., Linder, D., Franke, K. B., Jäger, J., Taylor, G., Gloge, F., et al. (2019). Toxic activation of an AAA+ protease by the antibacterial drug Cyclomarin. *Cell Chem. Biol.* 26, 1169–1179.e4. doi: 10.1016/j.chembiol.2019.05.008
- Miller, J. M., Chaudhary, H., and Marsee, J. D. (2018). Phylogenetic analysis predicts structural divergence for proteobacterial ClpC proteins. *J. Struct. Biol.* 201, 52–62. doi: 10.1016/j.jsb.2017.11.003
- Nachin, L., Nannmark, U., and Nyström, T. (2005). Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.* 187, 6265–6272. doi: 10.1128/jb.187.18.6265-6272.2005
- Ogbonna, E. C., Anderson, H. R., and Schmitz, K. R. (2022). Identification of arginine phosphorylation in *Mycobacterium smegmatis*. *Microbiol. Spectr.* 10:e0204222. doi: 10.1128/spectrum.02042-22
- Ollinger, J., O'Malley, T., Kesicki, E. A., Odingo, J., and Parish, T. (2012). Validation of the essential ClpP protease in *Mycobacterium tuberculosis* as a novel drug target. *J. Bacteriol.* 194, 663–668. doi: 10.1128/JB.06142-11
- Pak, M., and Wickner, S. (1997). Mechanism of protein remodeling by ClpA chaperone. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4901–4906. doi: 10.1073/pnas.94.10.4901
- Pearce, M. J., Arora, P., Festa, R. A., Butler-Wu, S. M., Gokhale, R. S., and Darwin, K. H. (2006). Identification of substrates of the *Mycobacterium tuberculosis* proteasome. *EMBO J.* 25, 5423–5432. doi: 10.1038/sj.emboj.7601405
- Pundir, S., Martin, M. J., and O'Donovan, C. (2016). UniProt tools. *Curr. Protoc. Bioinformatics* 53, 1.29.1–1.29.15. doi: 10.1002/0471250953.bi0129s53
- Raivio, T. L., and Silhavy, T. J. (2001). Periplasmic stress and ECF sigma factors. *Annu. Rev. Microbiol.* 55, 591–624. doi: 10.1146/annurev.micro.55.1.591
- Ripstein, Z. A., Vahidi, S., Houry, W. A., Rubinstein, J. L., and Kay, L. E. (2020). A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *elife* 9:e52158. doi: 10.7554/eLife.52158
- Rizo, A. N., Lin, J., Gates, S. N., Tse, E., Bart, S. M., Castellano, L. M., et al. (2019). Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. *Nat. Commun.* 10:2393. doi: 10.1038/s41467-019-10150-y
- Rosenzweig, R., Farber, P., Velyvis, A., Rennella, E., Latham, M. P., and Kay, L. E. (2015). ClpB N-terminal domain plays a regulatory role in protein disaggregation. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6872–E6881. doi: 10.1073/pnas.1512783112
- Rosenzweig, R., Moradi, S., Zarrine-Afsar, A., Glover, J. R., and Kay, L. E. (2013). Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. *Science* 339, 1080–1083. doi: 10.1126/science.1233066
- Schlieker, C., Weibezahn, J., Patzelt, H., Tessarz, P., Strub, C., Zeth, K., et al. (2004). Substrate recognition by the AAA+ chaperone ClpB. *Nat. Struct. Mol. Biol.* 11, 607–615. doi: 10.1038/nsmb787
- Schlieker, C., Zentgraf, H., Dersch, P., and Mogk, A. (2005). ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol. Chem.* 386, 1115–1127. doi: 10.1515/bc.2005.128
- Schmitt, E. K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., et al. (2011). The natural product cyclomarin kills mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angew. Chem. Int. Ed. Engl.* 50, 5889–5891. doi: 10.1002/anie.201101740
- Schmitz, K. R., Handy, E. L., Compton, C. L., Gupta, S., Bishai, W. R., Sauer, R. T., et al. (2020). Acyldepsipeptide antibiotics and a bioactive fragment thereof differentially perturb *Mycobacterium tuberculosis* ClpXP1P2 activity in vitro. *ACS Chem. Biol.* doi: 10.1021/acscchembio.9b00454
- Schmitz, K. R., and Sauer, R. T. (2014). Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Mol. Microbiol.* 93, 617–628. doi: 10.1111/mmi.12694
- Taylor, G., Frommherz, Y., Katikaridis, P., Layer, D., Sinning, I., Carroni, M., et al. (2022). Antibacterial peptide CyclomarinA creates toxicity by deregulating the *Mycobacterium tuberculosis* ClpC1/ClpP1P2 protease. *J. Biol. Chem.* 298:102202. doi: 10.1016/j.jbc.2022.102202
- 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
- Trentini, D. B., Suskiewicz, M. J., Heuck, A., Kurzbauer, R., Deszcz, L., Mechtler, K., et al. (2016). Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature* 539, 48–53. doi: 10.1038/nature20122
- Tripathi, L. P., and Sowdhamini, R. (2008). Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics* 9:549. doi: 10.1186/1471-2164-9-549
- Van Rossum, G., and Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Wang, F., Mei, Z., Qi, Y., Yan, C., Hu, Q., Wang, J., et al. (2011). Structure and mechanism of the hexameric MecA-ClpC molecular machine. *Nature* 471, 331–335. doi: 10.1038/nature09780
- Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., et al. (2004). Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cells* 119, 653–665. doi: 10.1016/j.cell.2004.11.027

- Weinhäupl, K., Brennich, M., Kazmaier, U., Lelievre, J., Ballell, L., Goldberg, A., et al. (2018). The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond dynamics in the N-terminal domain of ClpC1 from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 293, 8379–8393. doi: 10.1074/jbc.RA118.002251
- Weinhäupl, K., Gragera, M., Bueno-Carrasco, M. T., Arranz, R., Krändor, O., Akopian, T., et al. (2022). Structure of the drug target ClpC1 unfoldase in action provides insights on antibiotic mechanism of action. *J. Biol. Chem.* 298:102553. doi: 10.1016/j.jbc.2022.102553
- Wickner, S., Gottesman, S., Skowrya, D., Hoskins, J., McKenney, K., and Maurizi, M. R. (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci. U. S. A.* 91, 12218–12222. doi: 10.1073/pnas.91.25.12218
- Wolf, N. M., Lee, H., Zagal, D., Nam, J. W., Oh, D. C., Lee, H., et al. (2020). Structure of the N-terminal domain of ClpC1 in complex with the antituberculosis natural product ecumicin reveals unique binding interactions. *Acta Crystallogr. D Struct. Biol.* 76, 458–471. doi: 10.1107/s2059798320004027
- Yu, H., Lupoli, T. J., Kovach, A., Meng, X., Zhao, G., Nathan, C. F., et al. (2018). ATP hydrolysis-coupled peptide translocation mechanism of *Mycobacterium tuberculosis* ClpB. *Proc. Natl. Acad. Sci. U. S. A.* 115, e9560–e9569. doi: 10.1073/pnas.1810648115
- Zeth, K., Ravelli, R. B., Paal, K., Cusack, S., Bukau, B., and Dougan, D. A. (2002). Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat. Struct. Biol.* 9, 906–911. doi: 10.1038/nsb869
- Ziemski, M., Jomaa, A., Mayer, D., Rutz, S., Giese, C., Veprintsev, D., et al. (2018). Cdc48-like protein of actinobacteria (Cpa) is a novel proteasome interactor in mycobacteria and related organisms. *elife* 7:e34055. doi: 10.7554/eLife.34055
- Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A., and Weber-Ban, E. (2020). Genome-wide interaction screen for mycobacterium tuberculosis ClpCP protease reveals toxin-antitoxin systems as a major substrate class. *FEBS J.* 288, 111–126. doi: 10.1111/febs.15335
- Zolkiewski, M. (2006). A camel passes through the eye of a needle: Protein unfolding activity of Clp ATPases. *Mol. Microbiol.* 61, 1094–1100. doi: 10.1111/j.1365-2958.2006.05309.x