frontiers | Research Topics

Cell-free synthetic biology, volume II

Edited by
Jian Li, Yong-Chan Kwon, Yuan Lu and Simon J. Moore

Published in
Frontiers in Bioengineering and Biotechnology

ISSN 1664-8714
ISBN 978-2-8325-2683-5
DOI 10.3389/978-2-8325-2683-5

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

Cell-free synthetic biology, volume II

Topic editors

Jian Li — ShanghaiTech University, China

Yong-Chan Kwon — Louisiana State University, United States

Yuan Lu — Tsinghua University, China

Simon J. Moore — Queen Mary University of London, United Kingdom

Citation

Li, J., Kwon, Y.-C., Lu, Y., Moore, S. J., eds. (2023). *Cell-free synthetic biology, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2683-5

Table of contents

frontiers in Bioengineering and Biotechnology

REVIEW
published: 30 November 2021
doi: 10.3389/fbioe.2021.787881

Circular RNA: Biosynthesis *in vitro*

*Xinjie Chen and Yuan Lu**

Key Laboratory of Industrial Biocatalysis, Ministry of Education, Department of Chemical Engineering, Tsinghua University, Beijing, China

Circular RNA (circRNA) is a unique type of noncoding RNA molecule. Compared with traditional linear RNA, circRNA is a covalently closed circle produced by a process called backsplicing. CircRNA is abundant in many cells and has rich functions in cells, such as acting as miRNA sponge, protein sponge, protein scaffold, and mRNA regulator. With the continuous development of circRNA study, circRNA has also played an important role in medical applications, including circRNA vaccines and gene therapy. In this review, we illustrate the synthesis of circRNAs *in vitro*. We focus on biological ligation methods, such as enzymatic ligation from the bacteriophage T4 and ribozyme method. In addition, we summarize the current challenges in the design, synthesis, application, and production of circRNAs, and propose possible solutions in the future. CircRNA is expected to play an essential role in basic research and medical applications.

OPEN ACCESS

Edited by:
Dawei Zhang,
Tianjin Institute of Industrial Biotechnology (CAS), China

Reviewed by:
Marcos De La Peña,
Polytechnic University of Valencia, Spain
Subha Ranjan Das,
Carnegie Mellon University, United States

***Correspondence:**
Yuan Lu
yuanlu@tsinghua.edu.cn

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *01 October 2021*
Accepted: *16 November 2021*
Published: *30 November 2021*

Citation:
Chen X and Lu Y (2021) Circular RNA: Biosynthesis in vitro. *Front. Bioeng. Biotechnol. 9:787881. doi: 10.3389/fbioe.2021.787881*

Keywords: circular RNA, RNA synthesis, *in vitro* transcription, enzymatic ligation, ligase, permuted intron-exon

INTRODUCTION

A newly described RNA, called circular RNA (circRNA) (Sanger et al., 1976; Arnberg et al., 1980; Kos et al., 1986; Flores et al., 2011; Salzman et al., 2012; Jeck et al., 2013; Memczak et al., 2013; Guo et al., 2014; Jeck and Sharpless, 2014; Wang et al., 2014), has got much attention because of the development of high-throughput RNA-sequencing technology in recent years. Compared with traditional linear RNA, circRNA is a 3′-5′ covalently closed ring (Holdt et al., 2018) and does not need 5′-cap or 3′-poly(A) tails to keep it stable (Petkovic and Müller, 2015). These circRNAs have been found in a wide range of cells (Sanger et al., 1976; Arnberg et al., 1980; Kos et al., 1986; Salzman et al., 2012; Memczak et al., 2013; Wang et al., 2014), and most of them are noncoding RNAs (ncRNAs) (Santer et al., 2019). According to the components of circRNAs, circRNAs are classified as exonic (ecircRNA), exon-intron (EIcircRNA), or intronic (ciRNA) (He et al., 2021; Tao et al., 2021). Among them, ecircRNAs are the major circRNAs and are mainly produced by a process called backsplicing *in vivo* (Schindewolf et al., 1996; Salzman et al., 2012; Jeck et al., 2013; Memczak et al., 2013; Guo et al., 2014; Barrett et al., 2015; Starke et al., 2015; Szabo et al., 2015). EcircRNAs are mainly located in the cytoplasm and have various functions. The most well-known function of circRNAs is the microRNA (miRNA) sponge (Li et al., 2020; He et al., 2021), such as circRNA *CDR1as* and *Sry* (Hansen et al., 2013a; Memczak et al., 2013). Besides, circRNAs are also proved to serve as protein sponge (Ashwal-Fluss et al., 2014) and scaffold for protein complexes (Du et al., 2017a; Du et al., 2020). In addition, circRNAs also play a role in regulating the translation and stability of mRNA levels (Chen et al., 2020b; Huang et al., 2020) and the activity of proteins (Jeck et al., 2013; Ashwal-Fluss et al., 2014; Abdelmohsen et al., 2017; Du et al., 2017b). As biomarkers, circRNAs are also proved to be relative to the age-dependent neural accumulation in *drosophila* (Westholm et al., 2014). Moreover, circRNAs can also be involved in cancer or other diseases (Hansen et al., 2013b; Li et al., 2015; Qu et al., 2015).

With the development of RNA vaccines, the improvement of the stability of RNAs has been a great challenge. Fortunately, circRNAs show great potential ahead of this challenge. The structural

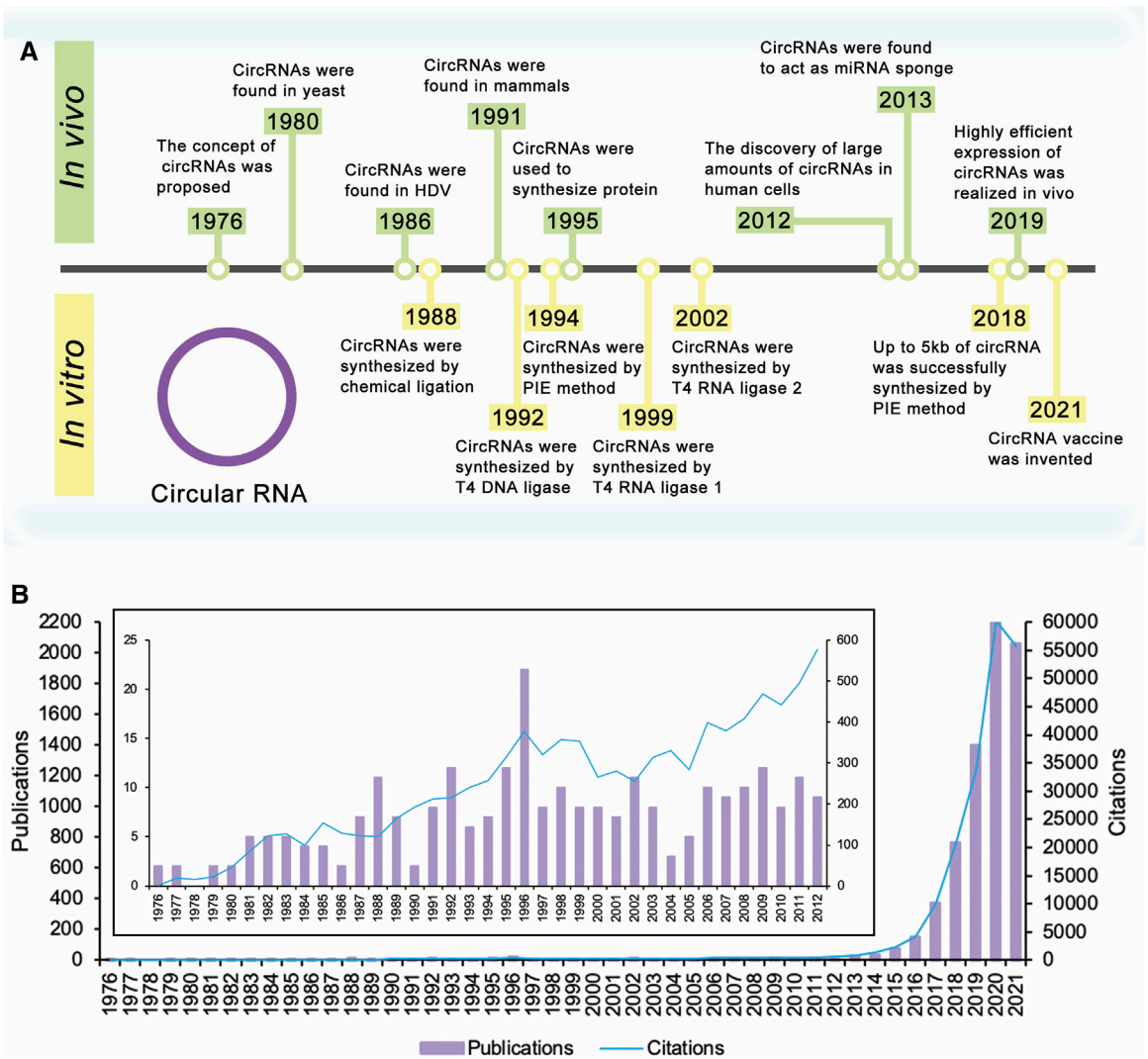


FIGURE 1 | Timeline of circular RNA. **(A)** The concept of circRNAs was proposed in 1976. Extensive research on circRNAs began in 2012 with the discovery of large amounts of circRNAs in human cells. These researches have improved yields of circRNAs, realized the synthesis of large circRNA molecules, and expanded the applications of circRNAs. **(B)** Times cited and publications over time. There were a few publications about circRNAs before 2012. Since the discovery of large amounts of circRNAs in human cells in 2012, the number of publications and citations has increased year by year and is still rising.

advantage of circRNAs brings it higher stability, especially against the degradation of exonucleases (Wesselhoeft et al., 2018). Endogenously produced circRNAs are 2–5 times more stable than linear RNAs (Enuka et al., 2016; Holdt et al., 2018). Therefore, the possibility that circRNAs serve as vaccines has been the main concern for researchers. Although circRNAs lack the essential elements for cap-dependent translation, the discovery of internal ribosome entry sites (IRES) makes it possible for circRNAs to act as translation templates (Legnini et al., 2017; Pamudurti et al., 2017; Wesselhoeft et al., 2018). N^6-methyladenosine (m^6A) is also proved to promote the extensive translation of circRNAs (Yang et al., 2017). Based on these studies, circRNA vaccines are likely to become the dominant form of the next generation of vaccines.

In this review, we introduced the synthesis of circRNAs *in vitro* and focused on the challenges and opportunities of circRNA biosynthesis. Since circRNAs have unique structures and special functions different from linear RNAs, further study of circRNAs will help understand biogenesis and expand the application of circRNAs in biological therapy and vaccine research.

A BRIEF HISTORY OF CIRCRNAS

The concept of circRNAs was first proposed by Sanger et al., in 1976 (Sanger et al., 1976) (**Figure 1A**). They studied four different highly purified viroids and found that viroids were covalently closed circRNA molecules. For the next 20 years, circRNAs were found in several kinds of life, such as yeast (Arnberg et al., 1980) and hepatitis delta virus (HDV) (Kos

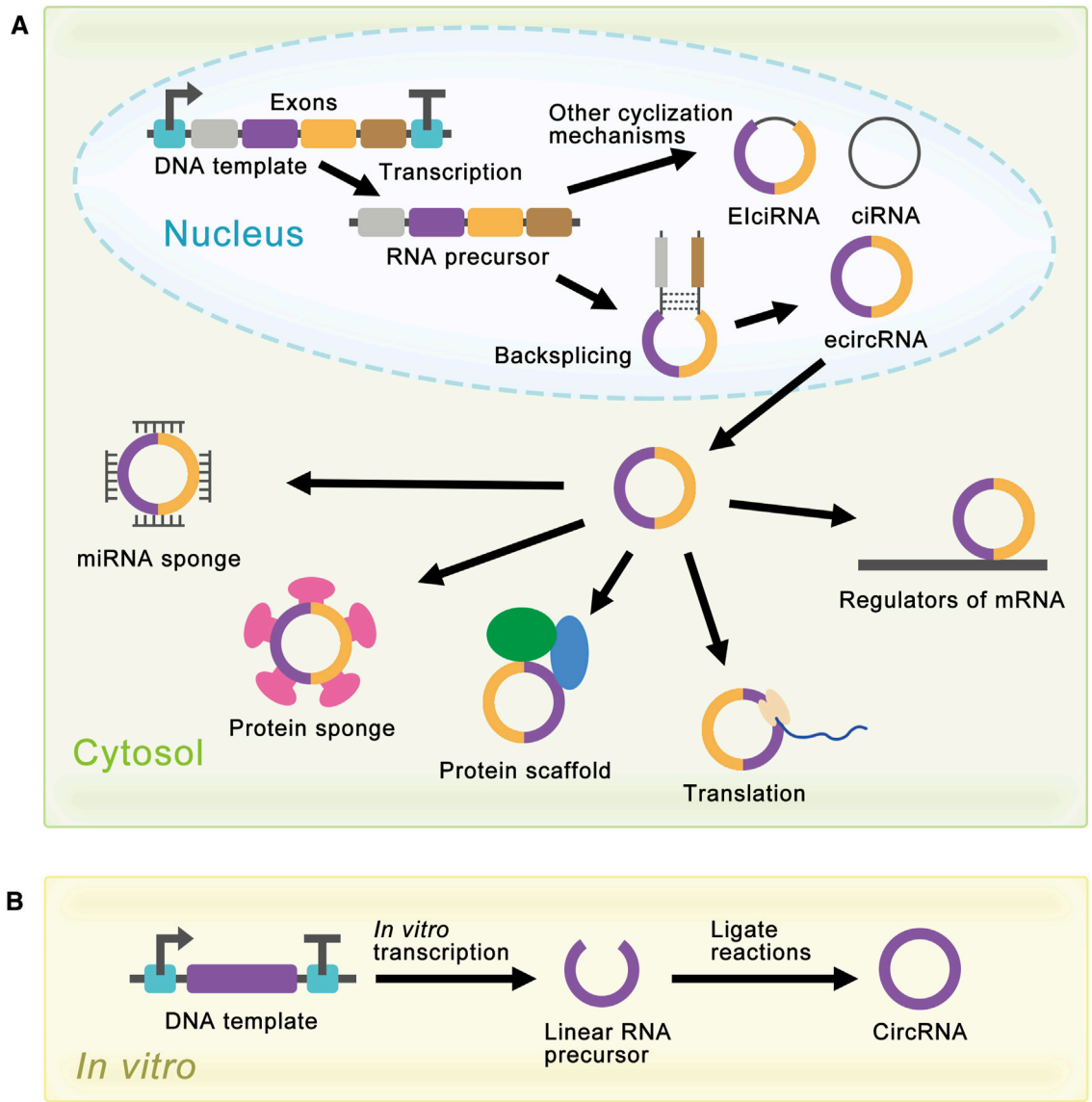


FIGURE 2 | Schematic diagram of circRNA synthesis *in vivo* and *in vitro*. **(A)** EcircRNAs are the major circRNAs and are mainly produced by a process called backsplicing *in vivo*. EIciRNAs and ciRNAs are produced by other cyclization reactions and located in the nucleus. EcircRNAs are mainly located in the cytoplasm and have various functions. In the cytosol, circRNAs can act as miRNA sponges, protein sponges, protein scaffolds, translation templates, and regulators of mRNA. **(B)** Linear RNA precursor is produced from *in vitro* transcription, and researchers have developed several methods for linear RNA precursor ligation to synthesize circRNA *in vitro*.

et al., 1986). CircRNAs were first discovered in mammals in 1991. Nigro et al. identified several abnormally spliced transcripts, in which exons from a candidate tumor suppressor gene (DCC) were scrambled during the splicing process *in vivo* (Nigro et al., 1991). After that, more circRNAs were found in mammalian cells (Cocquerelle et al., 1992; Capel et al., 1993; Zaphiropoulos, 1996; Zaphiropoulos, 1997; Surono et al., 1999). However, since the expression level of these circRNAs observed was so low, researchers believed that these circRNAs were the products of aberrant RNA splicing (Nigro et al., 1991; Cocquerelle et al., 1993). Therefore, researchers only knew the existence of circRNAs, but had no further understanding of theirs functions or impacts, and circRNAs did not receive much attention.

Extensive research on circRNAs began in 2012 with the discovery of large amounts of circRNAs in human cells (**Figure 1B**). With the development of high-throughput sequencing technology and computational analysis, Salzman et al. found that a substantial fraction of the spliced transcripts from hundreds of genes were circRNAs by deep sequencing of RNA from a variety of normal and malignant human cells (Salzman et al., 2012). This result proved that the circRNA was not a mistake of aberrant RNA splicing, but a general feature of the gene expression program in human cells.

The discovery revived great interest in circRNA research, and relative research was growing exponentially. In 2013, the function of circRNAs as the miRNA sponge was reported (Hansen et al., 2013a; Memczak et al., 2013). Besides, circRNAs were proved to be more stable than associated linear mRNAs *in vivo* (Jeck et al., 2013). The confirmation of the abundance, function, and stability of circRNAs laid the foundation of circRNAs research. Since then, more functions of circRNAs have been explored by researchers (**Figure 2A**).

Although natural circRNAs were ncRNAs and were considered incapable of translation, research had proved that manufactured circRNAs with IRES could be translated *in vivo* and *in vitro* (Chen and Sarnow, 1995; Perriman and Ares, 1998; Wang and Wang, 2015; Barrett and Salzman, 2016; Szabo and Salzman, 2016), which opened the way for the synthesis of proteins with circRNAs. More and more research was devoted to artificially controlling the expression level of circRNAs to realize the enrichment of circRNAs functions (Liang and Wilusz, 2014; Zhang et al., 2014; Kramer et al., 2015; Zhang et al., 2016; Garikipati et al., 2019). For the past few years, researchers have achieved highly efficient expression of circRNAs in cells using autocatalytic transcripts (Litke and Jaffrey, 2019) or viroid scaffolds (Daròs, 2021). At the same time, researchers had also developed several methods for the synthesis of circRNAs *in vitro*, such as chemical method (Sokolova et al., 1988), enzymatic method (Moore, 1999), and ribozyme method (Puttaraju and Been, 1992). These methods produced circRNAs by ligating the ends of linear RNA precursor, which would be covered later (**Figure 2B**).

SYNTHESIS OF CIRCRNAS *IN VITRO*

Synthesis of Linear RNA Precursor *in vitro*

At present, the main method of circRNA synthesis *in vitro* is ligating the ends of linear RNA precursor to produce a covalently closed circle. Linear RNA can be produced by chemical synthesis (Müller and Appel, 2017; Obi and Chen, 2021) or enzymatic strategy (Obi and Chen, 2021). The advantage of chemical synthesis is that the 5′ monophosphate can be directly introduced during the synthesis process for future cyclization. However, limited by the high cost of purification and low yield, chemical synthesis can only produce RNA of less than 50 to 70 nucleotides in length. Therefore, enzymatic strategy is the primary linear RNA synthesis method at present. Enzymatic strategy is usually realized through an *in vitro* transcription (IVT) reaction (Beckert and Masquida, 2011), which includes DNA template, reaction buffer, and phage RNA polymerase. The phage RNA polymerase usually derives from the T7, SP6, or T3 bacteriophages, and T7 RNA polymerase (Rio, 2013) is the most common phage RNA polymerase. IVT reaction allows for longer RNA synthesis at a lower cost. However, the run-off nature of phage polymerases may result in incomplete RNA. Some studies have improved transcription quality and reduced side reactions by mutating wild-type phage RNA polymerase.

Chemosynthesis or Biosynthesis?

Researchers have developed several methods for linear RNA precursor ligation *in vitro*. These ligation methods include chemical ligation, enzymatic ligation, and ribozyme method. Chemical ligation is realized by using cyanogen bromide (BrCN) or 1-ethyl-3-(3′-dimethylaminopropyl) carbodiimide to link DNA-RNA hybrids (Sokolova et al., 1988). However, this method suffers from low ligating efficiency (Dolinnaya et al., 1991) and biosafety concerns. In addition, chemical ligation forms 2′, 5′-phosphodiester bonds instead of the natural 3′, 5′-phosphodiester bonds (**Table 1**). Therefore, chemical ligation is not a common ligation method. Researchers are more interested in the biosynthesis of circRNAs, which includes enzymatic ligation and ribozyme method.

Enzymes From Bacteriophage T4

Enzymatic ligations are realized by catalytic reactions of several enzymes from the bacteriophage T4, including T4 DNA ligase (T4 Dnl), T4 RNA ligase 1 (T4 Rnl 1), and T4 RNA ligase 2 (T4 Rnl 2). It is worth noting that linear RNA precursor needs a 3′-OH on the acceptor substrate and a 5′ monophosphate on the donor substrate for enzymatic ligation (Moore, 1999). If the linear RNA precursor is produced by chemical synthesis, a 5′ monophosphate can be incorporated during the synthesis or added after the synthesis using ATP and T4 polynucleotide kinase. However, if the linear RNA precursor is synthesized by IVT reaction, it usually starts with 5′-pppG. Therefore, the 5′-terminus has to be dephosphorylated prior using calf intestinal (CIP) enzyme or other phosphatases (Petkovic and Müller, 2018). Then, 5′ monophosphate can be added using ATP and T4 polynucleotide kinase. In addition, the addition of GMP to the IVT reaction mixture is also proved to be useful for phosphorylating the transcript at its 5′ end (Abe et al., 2018). However, this method must ensure that the first base in the transcript is G, so there are some limitations to this method.

T4 Dnl ligation reaction includes inactivated kinase reaction and hybridized reaction. T4 Dnl can help ligate double-stranded duplexes, such as DNA/RNA hybrids (Moore, 1999) (**Figure 3A**). Therefore, this method needs a complementary DNA (cDNA) template or bridge to achieve RNA ligation. Generally, the cDNA bridge needs at least 10 nucleotides on either side of the ligation junction to guarantee high-quality ligation. The advantage of this method is that the accuracy of the linkage sites is greatly improved due to the addition of cDNA bridge. However, the ends of linear RNA precursor should be free of significant RNA secondary structure, and there should not be a high percentage of Us in the duplex region. Besides, T4 Dnl is less efficient in DNA/RNA hybrids linkage than double-stranded DNA (dsDNA) linkage. Due to these characteristics, only a few studies use this method for RNA ligation (Chen et al., 2017).

T4 Rnl 1 is a more common ligase for RNA ligation. T4 Rnl 1 catalyzes the nucleophilic attack of the 3′-OH terminus onto the activated 5′-terminus to form a covalent 5′, 3′-phosphodiester bond (Petkovic and Müller, 2018) and produces circRNAs (**Figure 3B**). Some studies have used cDNA bridge to prevent the linear RNA precursor from folding into an unsuitable structure (Wang and Ruffner, 1998). It is worth noting that

TABLE 1 | Advantages and disadvantages of different ligation method.

Ligation method		Advantages	Disadvantages
Chemical ligation		• Only chemical reagents • No biological components	• Low ligating efficiency • Biosafety concern • 2′, 5′-phosphodiester bonds
Enzymatic ligation	T4 DNA ligase	• Accurate	• Affected by significant RNA secondary structure • Affected by high percentage of Us • Low efficiency • Intermolecular end joining side reactions
	T4 RNA ligase 1	• High efficiency	• Low ligating efficiency for large RNA molecules
		• Synthesize as little as 6 to 8 nucleotides of circRNAs	• Affected by significant RNA secondary structure • Intermolecular end joining side reactions
	T4 RNA ligase 2	• More efficient for linear RNA precursor folding into a secondary structure with the ligation junction in a double-stranded region	• Low ligating efficiency for large RNA molecules • Intermolecular end joining side reactions
Ribozyme method	Group I intron self-splicing system	• Simple reaction condition and purification method • Can be used for RNA cyclization *in vitro* and *in vivo* • Can synthesize large circRNAs	• Affected by significant RNA secondary structure
	Group II intron self-splicing system	• Accurate ligation	• 2′, 5′-phosphodiester bonds • The mechanism remains unclear *in vitro*
	Hairpin ribozyme method	• High efficiency for small circRNAs	• Unstable • Exogenous HPR sequences

T4 Rnl 1 has different preferences for the nucleotides of the 5′-terminus and 3′-terminus: A > G ≥ C > U for the 3′-terminal nucleotide acceptor, and pC > pU > pA > pG for the 5′-terminal nucleotide donor (England and Uhlenbeck, 1978; McLaughlin et al., 1982; Petkovic and Müller, 2015; Müller and Appel, 2017). In this way, as little as 6 to 8 nucleotides of circRNAs can be synthesized (Kaufmann et al., 1974; Petkovic and Müller, 2015). This method can achieve high-efficiency single-stranded RNA (ssRNA) linkage. However, the RNA ligation efficiency reduces with large RNA molecules (Costello et al., 2020). Similar to the T4 Dnl ligation reaction, significant RNA secondary structure at the ends of linear RNA precursor can greatly reduce the ligation efficiency of T4 Rnl 1. Besides, intermolecular end joining (oligomerization) is also a serious side reaction, which cannot be completely suppressed (Petkovic and Müller, 2018). This side reaction will increase with the increase of the concentration of linear RNA precursor, which limits the amount of RNA cyclization.

T4 Rnl 2 can also be used for RNA ligation (Ho and Shuman, 2002; Nandakumar and Shuman, 2004; Yin et al., 2004). Similar to T4 Rnl 1, T4 Rnl 2 also catalyzes the nucleophilic attack of the 3′-OH terminus onto the activated 5′-terminus to form a covalent 5′, 3′-phosphodiester bond. However, T4 Rnl 2 is much more active at joining nicks in double-stranded RNA (dsRNA) than at ligating the ends of ssRNA (Nandakumar et al., 2004; Bullard and Bowater, 2006) (**Figure 3C**). Based on this feature, when the linear RNA precursor folds into a secondary structure with the ligation junction in a double-stranded region, the efficiency of T4 Rnl 2 is much higher than that of T4 Rnl 1 (Petkovic and Müller, 2018). Besides, with the help of RNA splint, T4 Rnl 2 can also realize the ligation of ends of ssRNA. However, RNA splints cannot be used to ligate short RNA precursors, because the splint-precursor complex is sterically unstable when the length of linear RNA precursor is shorter than 30 nucleotides. Same as T4 Rnl 1, T4 Rnl 2 also suffers from low efficiency for large RNA molecules and side reactions. To overcome these challenges, there are studies using software to simulate the secondary structure of the target circRNA and hypothetically cut at one site so that a few intramolecular base pairs are formed at the terminal (Chen et al., 2020a). This method can achieve efficient RNA cyclization at high concentrations with T4 Rnl 2, but it is still affected by different RNA sequences.

In a word, T4 Dnl and T4 Rnl 1 are suitable for RNA ligations without complex secondary structures. T4 Rnl 2 is more suitable for linear RNA precursor with the ligation junction in a double-stranded region. Therefore, different T4 ligases need to be selected according to the secondary structure of the linear RNA precursor. However, all these enzymatic ligation methods cannot realize large RNA molecules ligation, and cannot totally avoid intermolecular end-joining side reactions. These problems still need to be solved (**Table 1**).

Ribozyme Method

Modified group I intron self-splicing system is the most common ribozyme method, which is also called permuted introns and exons (PIE) method (Puttaraju and Been, 1992; Ford and Ares,

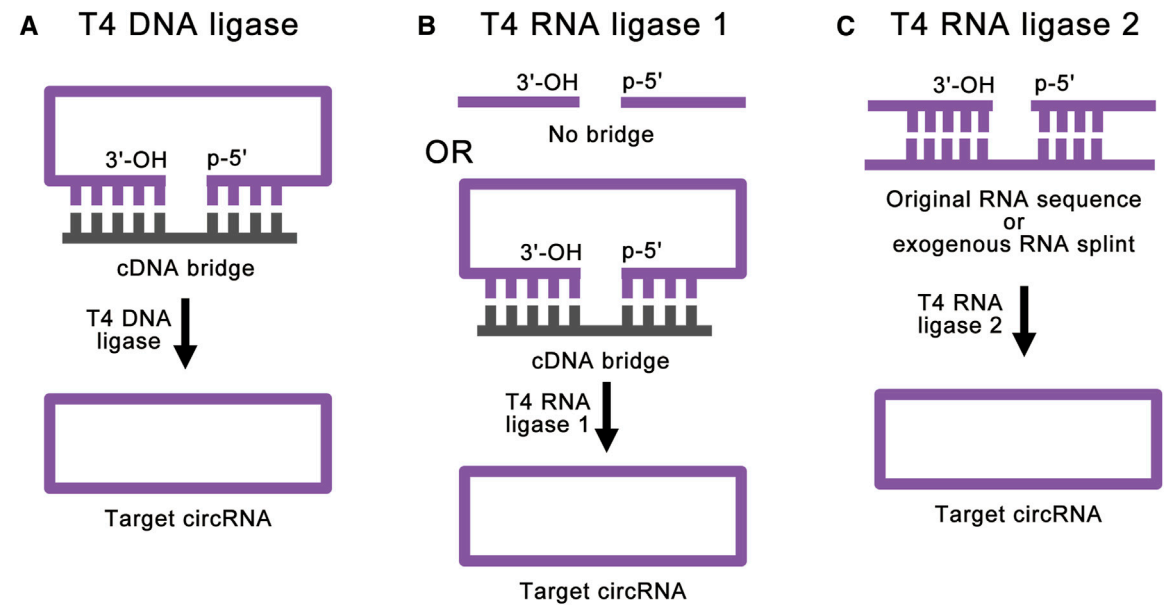


FIGURE 3 | Strategies for enzymatic ligations. **(A)** T4 DNA ligase can help ligate double-stranded duplexes, such as DNA/RNA hybrids. With the help of cDNA bridge, T4 DNA ligase can achieve accurate RNA ligation. **(B)** T4 RNA ligase 1 catalyzes the nucleophilic attack of the 3′-OH terminus onto the activated 5′-terminus to form a covalent 5′, 3′-phosphodiester bond. The cDNA bridge can prevent the linear RNA precursor from folding into an unsuitable structure. **(C)** T4 RNA ligase 2 is suitable for linear RNA precursor with the ligation junction in a double-stranded region. With the help of RNA splint, T4 RNA ligase 2 can also realize the ligation of ends of ssRNA.

1994; Umekage and Kikuchi, 2009; Wesselhoeft et al., 2018; Rausch et al., 2021). PIE method requires only the addition of GTP and Mg^{2+} as cofactors and shows great potential for protein synthesis. This method realized RNA ligation through a regular group I intron self-splicing reaction, including two transesterifications at defined splice sites (**Figure 4A**). PIE method can be used for RNA cyclization *in vitro* and *in vivo* (Meganck et al., 2021), which broadens its application. Compared to chemical ligation and enzymatic ligation, PIE method could be applied for the cyclization of larger linear RNA precursor, and the reaction condition and purification method of PIE method is simpler. A recent study has realized accurate RNA ligation by designing custom-tailored PIE transcription templates from which synthetic circRNAs of almost any sequence may be efficiently synthesized without exogenous exon sequences (Rausch et al., 2021). Based on these advantages, PIE method is currently the most studied and most widely used RNA ligation method. However, PIE method still has its disadvantages. Due to the complexity of RNA secondary structure, different RNA sequences will lead to a great difference in the final circRNA field, which limits the application of PIE method. Due to the introduction of exogenous exon sequences, the final circRNA sequence will be different from the original linear RNA precursor sequence, which may negatively affect the validation of some circRNA functions.

Group II introns can also be used for circRNA synthesis, which involves an inverse splicing reaction (Jarrell, 1993; Mikheeva et al., 1997). This splicing reaction involves the joining of the 5′ splice site at the end of an exon to the 3′ splice site at the beginning of the same exon (**Figure 4B**). Compared to the group I introns, all exon sequences are dispensable for group II intron-catalyzed inverse splicing. Therefore, this method can enable more accurate linear RNA precursor ligation. However, this method forms 2′, 5′-phosphodiester bonds at the ligation site instead of the natural 3′, 5′-phosphodiester bonds, and the mechanism remains unclear *in vitro* (Petkovic and Müller, 2015; Müller and Appel, 2017; Obi and Chen, 2021). At present, there are few studies concentrating on group II introns.

Hairpin ribozyme (HPR) can produce circRNAs through rolling circle reaction and the self-splicing reaction from circular single-strand DNA template (Diegelman and Kool, 1998; Kazakov et al., 2006; Dallas et al., 2008; Petkovic and Müller, 2013). The linear RNA precursor with HPR will fold into two alternative cleavage-active conformations to remove the 3′-end and the 5′-end. As a result, the intermediate will contain a 5′-OH and a 2′, 3′-cyclic phosphate to produce the target circRNA (**Figure 4C**). In this method, small circRNAs can be produced from long repeating RNAs transcribed by RNA polymerase through a rolling circle mechanism *in vitro* (Diegelman and Kool, 1998). This method is mainly used for efficient production of small circRNAs (Hieronymus and Müller, 2019). Freezing stimulation, ionic conditions, and additional cofactors are proved to affect the activity of HPR (Nesbitt et al., 1999; Kazakov et al., 2006; Strohbach et al., 2006; Dallas et al., 2008). The disadvantage of this method is that the circRNA is not stable due to the dynamic equilibrium of HPR-catalyzed cleavage and ligation (Müller and Appel, 2017). Besides, the circRNA contains HPR sequences, which may negatively affect the function of some circRNAs.

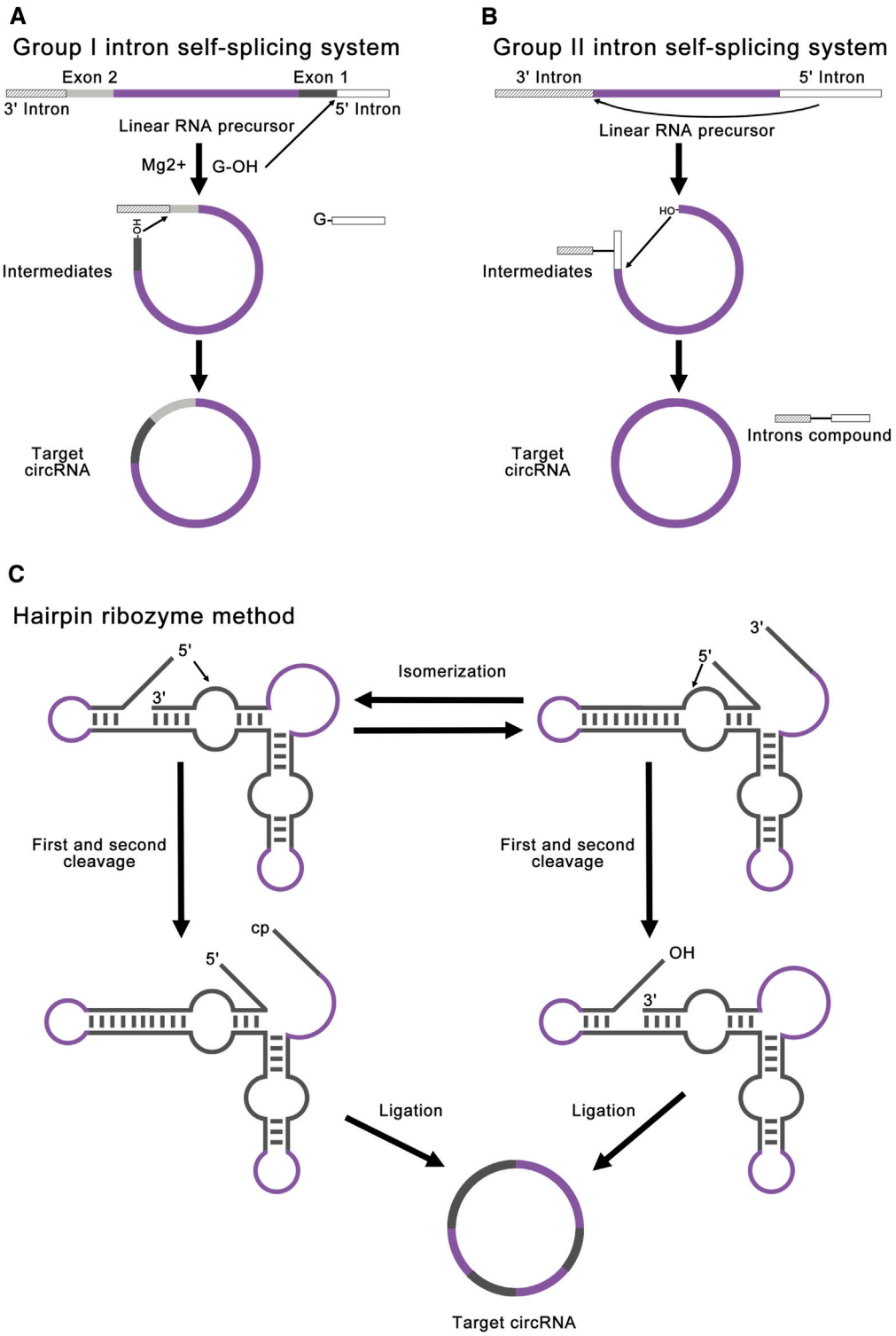


FIGURE 4 | Strategies for ribozyme methods. **(A)** Group I intron self-splicing system requires only the addition of GTP and Mg^{2+} as cofactors and shows great potential for protein synthesis. This method realized RNA ligation through a normal group I intron self-splicing reaction, including two transesterifications at defined splice sites. The final circRNA will contain exogenous exon sequences. **(B)** Group II intron self-splicing system involves the joining of the 5′ splice site at the end of an exon to the 3′ splice site at the beginning of the same exon. All exon sequences are dispensable for group II intron catalyzed inverse splicing. This method can enable more accurate linear RNA precursor ligation. **(C)** Hairpin ribozyme method can produce circRNA through the rolling circle reaction and the self-splicing reaction. The linear RNA precursor with HPR will fold into two alternative cleavage-active conformations to remove the 3′-end and the 5′-end. As a result, the intermediate will contain a 5′-OH and a 2′, 3′-cyclic phosphate to produce target circRNA.

TABLE 2 | Challenges and potential solutions for circRNAs.

Challenges	Potential solutions
Secondary structure of linear RNA precursor	o Unnatural nucleotides o RNA-binding proteins
Cyclization efficiency, especially for large RNA molecules	o Mutating the wild enzyme from the bacteriophage T4 o Rational design
Side reactions, especially for intermolecular end joining reaction	o Optimizing the cyclization reaction conditions and controlling the linear RNA precursor concentration o Immobilizing the ligase with hydrogels or other materials
Production of modified circRNAs	o Using unnatural nucleotides for the synthesis of linear RNA precursor o Incorporating chemically modified groups or unnatural nucleotides during the ligation reaction
Untapped potential	o Transferring scientific research achievements to commercial application o Further exploring the function and mechanism of circRNAs
The yield of circRNAs	o Optimizing current reaction components and conditions o New type reactors
The cost of raw materials	o Using cells to synthesize nucleotides or directly synthesize linear RNA precursors

In a word, PIE method using group I introns is suitable for most circRNAs production at present, which is still limited by the secondary structure of linear RNA precursor. Group II introns method can enable more accurate linear RNA precursor ligation, but the mechanism remains unclear *in vitro*. HPR method can efficiently produce small circRNAs but suffers from unstable product and exogenous HPR sequences. Among these ribozyme methods, PIE method shows the greatest potential for circRNAs production. At present, PIE method has been widely used in basic research and industrial production (**Table 1**).

CHALLENGES AND OUTLOOKS

Although there are many ways to synthesize circRNAs *in vitro*, there are still many challenges and opportunities. CircRNA is still a hot area of research, and there are many difficulties to be solved (**Table 2**).

Although PIE method has realized accurate linear RNA precursor ligation (Rausch et al., 2021), the structure of the linear RNA precursor still has a marked effect on circRNA synthesis yield. To overcome this problem, embedding unnatural nucleotides into linear RNA precursor may help change its secondary structure and improve the efficiency of ligation. Besides, RNA-binding proteins (RBPs) may also help the cyclization of RNA.

In terms of enzyme design, although existing natural enzymes from the bacteriophage T4 are capable of RNA cyclization, there is still much room for improvement. It is necessary to improve the cyclization efficiency of the ligases by mutation or rational design, especially for large RNA molecules. The critical point of mutation and rational design is to improve the binding efficiency of ligase to substrate RNA and the selectivity of cyclic products.

Reducing side reactions during circRNAs synthesis is also a big challenge. At present, the main side reaction is the intermolecular end-joining reaction, which exists in most circRNA synthesis methods and cannot be completely avoided. Therefore, it is necessary to optimize the cyclization reaction conditions and control the linear RNA precursor concentration to reduce the occurrence of such side reactions. In addition, immobilizing the ligase with hydrogels or other materials may realize the compartmentalization of the reaction and the cyclization of a single RNA substrate.

With the growing application of circRNAs, the synthesis of circRNAs with chemically modified or unnatural nucleotides is also difficult to be solved. Since circRNA does not have a terminal structure, the modification of circRNA is more difficult than linear RNA. One solution is to use unnatural nucleotides for the synthesis of linear RNA precursors, which can embed unnatural nucleotides into the final circRNA to improve the stability of circRNA and diversify its functions. In addition, the incorporation of chemically modified groups or unnatural nucleotides during the ligation reaction is also a possible solution.

Although circRNAs have been proved to hold the potential to become a novel vaccine (Qu et al., 2021), the commercial potential of circRNAs has yet to be fully exploited. Because of the unique structure of circRNAs, circRNAs can be used in the synthesis of some specific proteins, such as protein materials with lots of repeated sequences. Besides, circRNA can also be used as an expression template in cell-free expression systems. At the same time, the function of circRNA as ncRNA is also important. CircRNAs can not only be used as biomarkers of some diseases but also be used for gene therapy (Holdt et al., 2018). Since the function and mechanism of circRNAs in cells have not been fully studied, applications for circRNAs are still being developed. Further efforts are still needed.

Faced with the growing demand for circRNAs, large quantities of circRNAs need to be produced. Generally, increasing the yield of circRNAs requires increasing the concentration of the ligation reaction components, such as linear RNA precursors. However, an increase in the concentration of linear RNA precursors will lead to an increase in side reactions. These side reactions will decrease the yield of target circRNAs and increase the difficulty of product purification. Therefore, further optimization of current reaction components and conditions may be required based on scaled-up production systems. This

problem may also be solved by new-type reactors, such as microreactors. At the same time, the cost of current circRNAs production is still a concern for researchers. The main reason is the cost of raw materials, such as expensive nucleotides. One potential solution is to use cells to synthesize nucleotides or directly synthesize linear RNA precursors. However, how to efficiently purify these raw materials produced by cells still needs further research.

In a word, circRNA is still a hot area of current research and holds great potentials. With the continuous efforts in this field, circRNAs will play an essential role in basic research and medical applications, and become an important part of human health in the near future.

AUTHOR CONTRIBUTIONS

XC and YL contributed to conception and design of the study. XC wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by Beijing Natural Science Foundation (2192023), National Natural Science Foundation of China (21878173), and National Key R&D Program of China (2018YFA0901700).

REFERENCES

Abdelmohsen, K., Panda, A. C., Munk, R., Grammatikakis, I., Dudekula, D. B., De, S., et al. (2017). Identification of HuR Target Circular RNAs Uncovers Suppression of PABPN1 Translation by CircPABPN1. *RNA Biol.* 14, 361–369. doi:10.1080/15476286.2017.1279788

Abe, N., Kodama, A., and Abe, H. (2018). Preparation of Circular RNA *In Vitro*. *Methods Mol. Biol.* 1724, 181–192. doi:10.1007/978-1-4939-7562-4_15

Arnberg, A. C., Van Ommen, G.-J. B., Grivell, L. A., Van Bruggen, E. F. J., and Borst, P. (1980). Some Yeast Mitochondrial RNAs Are Circular. *Cell* 19, 313–319. doi:10.1016/0092-8674(80)90505-X

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). CircRNA Biogenesis Competes with Pre-mRNA Splicing. *Mol. Cel* 56, 55–66. doi:10.1016/j.molcel.2014.08.019

Barrett, S. P., and Salzman, J. (2016). Circular RNAs: Analysis, Expression and Potential Functions. *Dev* 143, 1838–1847. doi:10.1242/dev.128074

Barrett, S. P., Wang, P. L., and Salzman, J. (2015). Circular RNA Biogenesis Can Proceed through an Exon-Containing Lariat Precursor. *Elife* 4, 1–18. doi:10.7554/eLife.07540

Beckert, B., and Masquida, B. (2011). Synthesis of RNA by *In Vitro* Transcription. *Methods Mol. Biol.* 703, 29–41. doi:10.1007/978-1-59745-248-9_3

Bullard, D. R., and Bowater, R. P. (2006). Direct Comparison of Nick-Joining Activity of the Nucleic Acid Ligases from Bacteriophage T4. *Biochem. J.* 398, 135–144. doi:10.1042/BJ20060313

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., et al. (1993). Circular Transcripts of the Testis-Determining Gene Sry in Adult Mouse Testis. *Cell* 73, 1019–1030. doi:10.1016/0092-8674(93)90279-Y

Chen, C.-y., and Sarnow, P. (1995). Initiation of Protein Synthesis by the Eukaryotic Translational Apparatus on Circular RNAs. *Science* 268, 415–417. doi:10.1126/science.7536344

Chen, Y. G., Kim, M. V., Chen, X., Batista, P. J., Aoyama, S., Wilusz, J. E., et al. (2017). Sensing Self and Foreign Circular RNAs by Intron Identity. *Mol. Cel* 67, 228–238. doi:10.1016/j.molcel.2017.05.022

Chen, H., Cheng, K., Liu, X., An, R., Komiyama, M., and Liang, X. (2020a). Preferential Production of RNA Rings by T4 RNA Ligase 2 without Any Splint through Rational Design of Precursor Strand. *Nucleic Acids Res.* 48, e54. doi:10.1093/nar/gkaa181

Chen, L., Kong, R., Wu, C., Wang, S., Liu, Z., Liu, S., et al. (2020b). Circ-MALAT1 Functions as Both an mRNA Translation Brake and a microRNA Sponge to Promote Self-Renewal of Hepatocellular Cancer Stem Cells. *Adv. Sci.* 7, 1900949. doi:10.1002/advs.201900949

Cocquerelle, C., Daubersies, P., Majérus, M. A., Kerckaert, J. P., and Bailleul, B. (1992). Splicing with Inverted Order of Exons Occurs Proximal to Large Introns. *EMBO J.* 11, 1095–1098. doi:10.1002/j.1460-2075.1992.tb05148.x

Cocquerelle, C., Mascrez, B., Hétuin, D., and Bailleul, B. (1993). Mis-splicing Yields Circular RNA Molecules. *FASEB J.* 7, 155–160. doi:10.1096/fasebj.7.1.7678559

Costello, A., Lao, N. T., Barron, N., and Clynes, M. (2020). Reinventing the Wheel: Synthetic Circular RNAs for Mammalian Cell Engineering. *Trends Biotechnol.* 38, 217–230. doi:10.1016/j.tibtech.2019.07.008

Dallas, A., Balatskaya, S. V., Kuo, T.-C., Ilves, H., Vlassov, A. V., Kaspar, R. L., et al. (2008). Hairpin Ribozyme-Antisense RNA Constructs Can Act as Molecular Lassos. *Nucleic Acids Res.* 36, 6752–6766. doi:10.1093/nar/gkn637

Daròs, J.-A. (2021). Production of Circular Recombinant RNA in *Escherichia coli* Using Viroid Scaffolds. *Methods Mol. Biol.* 99, 99–107. doi:10.1007/978-1-0716-1499-0_8

Diegelman, A., and Kool, E. T. (1998). Generation of Circular RNAs and Trans-cleaving Catalytic RNAs by Rolling Transcription of Circular DNA Oligonucleotides Encoding Hairpin Ribozymes. *Nucleic Acids Res.* 26, 3235–3241. doi:10.1093/nar/26.13.3235

Dolinnaya, N. G., Sokolova, N. I., Ashirbekova, D. T., and Shabarova, Z. A. (1991). The Use of BrCN for Assembling Modified DNA Duplexes and DNA-RNA Hybrids; Comparison with Water-Soluble Carbodiimide. *Nucl. Acids Res.* 19, 3067–3072. doi:10.1093/nar/19.11.3067

Du, W. W., Fang, L., Yang, W., Wu, N., Awan, F. M., Yang, Z., et al. (2017a). Induction of Tumor Apoptosis through a Circular RNA Enhancing Foxo3 Activity. *Cell Death Differ.* 24, 357–370. doi:10.1038/cdd.2016.133

Du, W. W., Yang, W., Chen, Y., Wu, Z.-K., Foster, F. S., Yang, Z., et al. (2017b). Foxo3 Circular RNA Promotes Cardiac Senescence by Modulating Multiple Factors Associated with Stress and Senescence Responses. *Eur. Heart J.* 38, ehw001–1412. doi:10.1093/eurheartj/ehw001

Du, W. W., Yang, W., Li, X., Fang, L., Wu, N., Li, F., et al. (2020). The Circular RNA circSKA3 Binds Integrin β1 to Induce Invadopodium Formation Enhancing Breast Cancer Invasion. *Mol. Ther.* 28, 1287–1298. doi:10.1016/j.ymthe.2020.03.002

England, T. E., and Uhlenbeck, O. C. (1978). Enzymic Oligoribonucleotide Synthesis with T4 RNA Ligase. *Biochemistry* 17, 2069–2076. doi:10.1021/bi00604a008

Enuka, Y., Lauriola, M., Feldman, M. E., Sas-Chen, A., Ulitsky, I., and Yarden, Y. (2016). Circular RNAs Are Long-Lived and Display Only Minimal Early Alterations in Response to a Growth Factor. *Nucleic Acids Res.* 44, 1370–1383. doi:10.1093/nar/gkv1367

Flores, R., Grubb, D., Elleuch, A., Nohales, M.-Á., Delgado, S., and Gago, S. (2011). Rolling-circle Replication of Viroids, Viroid-like Satellite RNAs and Hepatitis delta Virus: Variations on a Theme. *RNA Biol.* 8, 200–206. doi:10.4161/rna.8.2.14238

Ford, E., and Ares, M. (1994). Synthesis of Circular RNA in Bacteria and Yeast Using RNA Cyclase Ribozymes Derived from a Group I Intron of Phage T4. *Proc. Natl. Acad. Sci.* 91, 3117–3121. doi:10.1073/pnas.91.8.3117

Garikipati, V. N. S., Verma, S. K., Cheng, Z., Liang, D., Truongcao, M. M., Cimini, M., et al. (2019). Circular RNA CircFndc3b Modulates Cardiac Repair after Myocardial Infarction via FUS/VEGF-A axis. *Nat. Commun.* 10. doi:10.1038/s41467-019-11777-7

Guo, J. U., Agarwal, V., Guo, H., and Bartel, D. P. (2014). Expanded Identification and Characterization of Mammalian Circular RNAs. *Genome Biol.* 15, 409. doi:10.1186/s13059-014-0409-z

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013a). Natural RNA Circles Function as Efficient microRNA Sponges. *Nature* 495, 384–388. doi:10.1038/nature11993

Hansen, T. B., Kjems, J., and Damgaard, C. K. (2013b). Circular RNA and miR-7 in Cancer. *Cancer Res.* 73, 5609–5612. doi:10.1158/0008-5472.CAN-13-1568

He, A. T., Liu, J., Li, F., and Yang, B. B. (2021). Targeting Circular RNAs as a Therapeutic Approach: Current Strategies and Challenges. *Sig. Transduct. Target. Ther.* 6, 1–14. doi:10.1038/s41392-021-00569-5

Hieronymus, R., and Müller, S. (2019). Engineering of Hairpin Ribozyme Variants for RNA Recombination and Splicing. *Ann. N.Y. Acad. Sci.* 1447, 135–143. doi:10.1111/nyas.14052

Ho, C. K., and Shuman, S. (2002). Bacteriophage T4 RNA Ligase 2 (gp24.1) Exemplifies a Family of RNA Ligases Found in All Phylogenetic Domains. *Proc. Natl. Acad. Sci.* 99, 12709–12714. doi:10.1073/pnas.192184699

Holdt, L. M., Kohlmaier, A., and Teupser, D. (2018). Circular RNAs as Therapeutic Agents and Targets. *Front. Physiol.* 9, 1262. doi:10.3389/fphys.2018.01262

Huang, Q., Guo, H., Wang, S., Ma, Y., Chen, H., Li, H., et al. (2020). A Novel Circular RNA, circXPO1, Promotes Lung Adenocarcinoma Progression by Interacting with IGF2BP1. *Cell Death Dis.* 11, 1031. doi:10.1038/s41419-020-03237-8

Jarrell, K. A. (1993). Inverse Splicing of a Group II Intron. *Proc. Natl. Acad. Sci.* 90, 8624–8627. doi:10.1073/pnas.90.18.8624

Jeck, W. R., and Sharpless, N. E. (2014). Detecting and Characterizing Circular RNAs. *Nat. Biotechnol.* 32, 453–461. doi:10.1038/nbt.2890

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs Are Abundant, Conserved, and Associated with ALU Repeats. *RNA* 19, 141–157. doi:10.1261/rna.035667.112

Kaufmann, G., Klein, T., and Littauer, U. Z. (1974). T4 RNA Ligase: Substrate Chain Length Requirements. *FEBS Lett.* 46, 271–275. doi:10.1016/0014-5793(74)80385-6

Kazakov, S. A., Balatskaya, S. V., and Johnston, B. H. (2006). Ligation of the Hairpin Ribozyme in Cis Induced by Freezing and Dehydration. *RNA* 12, 446–456. doi:10.1261/rna.2123506

Kos, A., Dijkema, R., Arnberg, A. C., Van Der Meide, P. H., and Schellekens, H. (1986). The Hepatitis delta (δ) Virus Possesses a Circular RNA. *Nature* 323, 558–560. doi:10.1038/323558a0

Kramer, M. C., Liang, D., Tatomer, D. C., Gold, B., March, Z. M., Cherry, S., et al. (2015). Combinatorial Control ofDrosophilacircular RNA Expression by Intronic Repeats, hnRNPs, and SR Proteins. *Genes Dev.* 29, 2168–2182. doi:10.1101/gad.270421.115

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cel* 66, 22–37. doi:10.1016/j.molcel.2017.02.017

Li, F., Zhang, L., Li, W., Deng, J., Zheng, J., An, M., et al. (2015). Circular RNA ITCH Has Inhibitory Effect on ESCC by Suppressing the Wnt/β-Catenin Pathway. *Oncotarget* 6, 6001–6013. doi:10.18632/oncotarget.3469

Li, J., Mohammed-Elsabagh, M., Paczkowski, F., and Li, Y. (2020). Circular Nucleic Acids: Discovery, Functions and Applications. *ChemBioChem* 21, 1547–1566. doi:10.1002/cbic.202000003

Liang, D., and Wilusz, J. E. (2014). Short Intronic Repeat Sequences Facilitate Circular RNA Production. *Genes Dev.* 28, 2233–2247. doi:10.1101/gad.251926.114

Litke, J. L., and Jaffrey, S. R. (2019). Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts. *Nat. Biotechnol.* 37, 667–675. doi:10.1038/s41587-019-0090-6

McLaughlin, L. W., Romaniuk, E., Romaniuk, P. J., and Neilson, T. (1982). The Effect of Acceptor Oligoribonucleotide Sequence on the T4 RNA Ligase Reaction. *Eur. J. Biochem.* 125, 639–643. doi:10.1111/j.1432-1033.1982.tb06730.x

Meganck, R. M., Liu, J., Hale, A. E., Simon, K. E., Fanous, M. M., Vincent, H. A., et al. (2021). Engineering Highly Efficient Backsplicing and Translation of Synthetic circRNAs. *Mol. Ther. Nucleic Acids* 23, 821–834. doi:10.1016/j.omtn.2021.01.003

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency. *Nature* 495, 333–338. doi:10.1038/nature11928

Mikheeva, S., Hakim-Zargar, M., Carlson, D., and Jarrell, K. (1997). Use of an Engineered Ribozyme to Produce a Circular Human Exon. *Nucleic Acids Res.* 25, 5085–5094. doi:10.1093/nar/25.24.5085

Moore, M. J. (1999). Joining RNA Molecules with T4 DNA Ligase. *Methods Mol. Biol.* 118, 11–19. doi:10.1385/1-59259-676-2:11

Müller, S., and Appel, B. (2017). *In Vitro* circularization of RNA. *RNA Biol.* 14, 1018–1027. doi:10.1080/15476286.2016.1239009

Nandakumar, J., and Shuman, S. (2004). How an RNA Ligase Discriminates RNA versus DNA Damage. *Mol. Cel* 16, 211–221. doi:10.1016/j.molcel.2004.09.022

Nandakumar, J., Ho, C. K., Lima, C. D., and Shuman, S. (2004). RNA Substrate Specificity and Structure-Guided Mutational Analysis of Bacteriophage T4 RNA Ligase 2. *J. Biol. Chem.* 279, 31337–31347. doi:10.1074/jbc.M402394200

Nesbitt, S. M., Erlacher, H. A., and Fedor, M. J. (1999). The Internal Equilibrium of the Hairpin Ribozyme: Temperature, Ion and pH Effects. *J. Mol. Biol.* 286, 1009–1024. doi:10.1006/jmbi.1999.2543

Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., et al. (1991). Scrambled Exons. *Cell* 64, 607–613. doi:10.1016/0092-8674(91)90244-S

Obi, P., and Chen, Y. G. (2021). The Design and Synthesis of Circular RNAs. *Methods* S1046-2023, 00065–00067. doi:10.1016/j.ymeth.2021.02.020

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of CircRNAs. *Mol. Cel* 66, 9–21. doi:10.1016/j.molcel.2017.02.021

Perriman, R., and Ares, M. (1998). Circular mRNA Can Direct Translation of Extremely Long Repeating-Sequence Proteins *In Vivo*. *RNA* 4, 1047–1054. doi:10.1017/S135583829898061X

Petkovic, S., and Müller, S. (2013). RNA Self-Processing: Formation of Cyclic Species and Concatemers from a Small Engineered RNA. *FEBS Lett.* 587, 2435–2440. doi:10.1016/j.febslet.2013.06.013

Petkovic, S., and Müller, S. (2015). RNA Circularization Strategies *In Vivo* and *In Vitro*. *Nucleic Acids Res.* 43, 2454–2465. doi:10.1093/nar/gkv045

Petkovic, S., and Müller, S. (2018). Synthesis and Engineering of Circular RNAs. *Methods Mol. Biol.* 1724, 167–180. doi:10.1007/978-1-4939-7562-4_14

Puttaraju, M., and Been, M. (1992). Group I Permuted Intron-Exon (PIE) Sequences Self-Splice to Produce Circular Exons. *Nucl. Acids Res.* 20, 5357–5364. doi:10.1093/nar/20.20.5357

Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., et al. (2015). Circular RNA: A New star of Noncoding RNAs. *Cancer Lett.* 365, 141–148. doi:10.1016/j.canlet.2015.06.003

Qu, L., Yi, Z., Shen, Y., Xu, Y., Wu, Z., Tang, H., et al. (2021). Circular RNA Vaccines against SARS-CoV-2 and Emerging Variants. *bioRxiv*. doi:10.1101/2021.03.16.435594

Rausch, J. W., Heinz, W. F., Payea, M. J., Sherpa, C., Gorospe, M., and Le Grice, S. F. J. (2021). Characterizing and Circumventing Sequence Restrictions for Synthesis of Circular RNA *In Vitro*. *Nucleic Acids Res.* 49, E35. doi:10.1093/nar/gkaa1256

Rio, D. C. (2013). Expression and Purification of Active Recombinant T7 RNA Polymerase from *E. coli*. *Cold Spring Harb. Protoc.* 2013, pdb.prot078527. doi:10.1101/pdb.prot078527

Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS One* 7, e30733. doi:10.1371/journal.pone.0030733

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids Are Single-Stranded Covalently Closed Circular RNA Molecules Existing as Highly Base-Paired Rod-like Structures. *Proc. Natl. Acad. Sci.* 73, 3852–3856. doi:10.1073/pnas.73.11.3852

Santer, L., Bär, C., and Thum, T. (2019). Circular RNAs: A Novel Class of Functional RNA Molecules with a Therapeutic Perspective. *Mol. Ther.* 27, 1350–1363. doi:10.1016/j.ymthe.2019.07.001

Schindewolf, C., Braun, S., and Domdey, H. (1996). *In Vitro* generation of a Circular Exon from a Linear Pre-mRNA Transcript. *Nucleic Acids Res.* 24, 1260–1266. doi:10.1093/nar/24.7.1260

Sokolova, N. I., Ashirbekova, D. T., Dolinnaya, N. G., and Shabarova, Z. A. (1988). Chemical Reactions within DNA Duplexes Cyanogen Bromide as an Effective Oligodeoxyribonucleotide Coupling Agent. *FEBS Lett.* 232, 153–155. doi:10.1016/0014-5793(88)80406-X

Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L.-H., et al. (2015). Exon Circularization Requires Canonical Splice Signals. *Cel Rep.* 10, 103–111. doi:10.1016/j.celrep.2014.12.002

Strohbach, D., Novak, N., and Müller, S. (2006). Redox-active Riboswitching: Allosteric Regulation of Ribozyme Activity by Ligand-Shape Control. *Angew. Chem. Int. Ed.* 45, 2127–2129. doi:10.1002/anie.200503820

Surono, A., Takeshima, Y., Wibawa, T., Ikezawa, M., Nonaka, I., and Matsuo, M. (1999). Circular Dystrophin RNAs Consisting of Exons that Were Skipped by Alternative Splicing. *Hum. Mol. Genet.* 8, 493–500. doi:10.1093/hmg/8.3.493

Szabo, L., and Salzman, J. (2016). Detecting Circular RNAs: Bioinformatic and Experimental Challenges. *Nat. Rev. Genet.* 17, 679–692. doi:10.1038/nrg.2016.114

Szabo, L., Morey, R., Palpant, N. J., Wang, P. L., Afari, N., Jiang, C., et al. (2015). Statistically Based Splicing Detection Reveals Neural Enrichment and Tissue-specific Induction of Circular RNA during Human Fetal Development. *Genome Biol.* 16, 126. doi:10.1186/s13059-015-0690-5

Tao, M., Zheng, M., Xu, Y., Ma, S., Zhang, W., and Ju, S. (2021). CircRNAs and Their Regulatory Roles in Cancers. *Mol. Med.* 2794, 94. doi:10.1186/s10020-021-00359-3

Umekage, S., and Kikuchi, Y. (2009). *In Vitro* and *In Vivo* Production and Purification of Circular RNA Aptamer. *J. Biotechnol.* 139, 265–272. doi:10.1016/j.jbiotec.2008.12.012

Wang, L., and Ruffner, D. E. (1998). Oligoribonucleotide Circularization by 'template-Mediated' Ligation with T4 RNA Ligase: Synthesis of Circular Hammerhead Ribozymes. *Nucleic Acids Res.* 26, 2502–2504. doi:10.1093/nar/26.10.2502

Wang, Y., and Wang, Z. (2015). Efficient Backsplicing Produces Translatable Circular mRNAs. *RNA* 21, 172–179. doi:10.1261/rna.048272.114

Wang, P. L., Bao, Y., Yee, M.-C., Barrett, S. P., Hogan, G. J., Olsen, M. N., et al. (2014). Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS One* 9, e90859. doi:10.1371/journal.pone.0090859

Wesselhoeft, R. A., Kowalski, P. S., and Anderson, D. G. (2018). Engineering Circular RNA for Potent and Stable Translation in Eukaryotic Cells. *Nat. Commun.* 9, 2629. doi:10.1038/s41467-018-05096-6

Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., et al. (2014). Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-dependent Neural Accumulation. *Cel Rep.* 9, 1966–1980. doi:10.1016/j.celrep.2014.10.062

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive Translation of Circular RNAs Driven by N6-Methyladenosine. *Cell Res* 27, 626–641. doi:10.1038/cr.2017.31

Yin, S., Kiong Ho, C., Miller, E. S., and Shuman, S. (2004). Characterization of Bacteriophage KVP40 and T4 RNA Ligase 2. *Virology* 319, 141–151. doi:10.1016/j.virol.2003.10.037

Zaphiropoulos, P. G. (1996). Circular RNAs from Transcripts of the Rat Cytochrome P450 2C24 Gene: Correlation with Exon Skipping. *Proc. Natl. Acad. Sci.* 93, 6536–6541. doi:10.1073/pnas.93.13.6536

Zaphiropoulos, P. G. (1997). Exon Skipping and Circular RNA Formation in Transcripts of the Human Cytochrome P-450 2C18 Gene in Epidermis and of the Rat Androgen Binding Protein Gene in Testis. *Mol. Cel. Biol.* 17, 2985–2993. doi:10.1128/mcb.17.6.2985

Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary Sequence-Mediated Exon Circularization. *Cell* 159, 134–147. doi:10.1016/j.cell.2014.09.001

Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.-L., et al. (2016). The Biogenesis of Nascent Circular RNAs. *Cel Rep.* 15, 611–624. doi:10.1016/j.celrep.2016.03.058

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 | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 01 April 2022
doi: 10.3389/fbioe.2022.832176

HyperXpress: Rapid Single Vessel DNA Assembly and Protein Production in Microliterscale

Darius Leon Zibulski[1], *Niels Schlichting*[1] *and Johannes Kabisch*[1,2]*

[1]Computer-aided Synthetic Biology, Darmstadt, Germany, [2]Department of Biotechnology and Food Science, NTNU, Trondheim, Norway

Rapid prototyping of biological functions has the common aim of generating, screening, and selecting variant libraries as quickly as possible. This approach is now to be extended by the HyperXpress workflow, which connects ligase cycling reaction for DNA assembly, multiply-primed rolling circle amplification for signal amplification, and cell-free protein synthesis to a single vessel reaction in the lower µl scale. After substantial optimization of the method a proof-of-principle demonstrating the high flexibility of HyperXpress for semi-rational protein engineering by expanding, reducing, and replacing β-strands of three different green fluorescent proteins is described. These single-day experiments resulted in six functional, new-to-nature GFP prototypes.

Keywords: rapid prototyping, ligase cycling reaction, rolling circle amplification, cell-free protein synthesis, semi-rational protein engineering

OPEN ACCESS

Edited by:
Jian Li,
ShanghaiTech University, China

Reviewed by:
Cheemeng Tan,
University of California, Davis, United States
Yifan Liu,
ShanghaiTech University, China

***Correspondence:**
Johannes Kabisch
johannes.kabisch@ntnu.no

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *09 December 2021*
Accepted: *28 February 2022*
Published: *01 April 2022*

Citation:
Zibulski DL, Schlichting N and Kabisch J (2022) HyperXpress: Rapid Single Vessel DNA Assembly and Protein Production in Microliterscale. Front. Bioeng. Biotechnol. 10:832176. doi: 10.3389/fbioe.2022.832176

1 INTRODUCTION

Prototyping in bioengineering usually starts at the level of digital sequence information, which can be recombined to obtain novel biological functions of genetic systems (Duportet et al., 2014), proteins (Dopp et al., 2019a), or cells (Robinson et al., 2020). Proteins have a special place among them, as they are being used extensively in many areas of industry, chemistry, medicine, and agriculture, which necessitate the rapid and efficient identification of variants with new and improved properties (Cole and Gaucher, 2011). For this protein engineering process, numerous mutagenesis and recombination methods in combination with a wide variety of selection and screening processes are available (Nannemann et al., 2011), whose common endeavor is to select or screen a highly diverse, functional variant library as effectively as possible (Cole and Gaucher, 2011).

In order to reduce the necessary screening effort and still obtain the most functional and diverse library possible, semi-rational, and focused libraries offer an optimal basis. Various methods, such as whole-gene synthesis, site-saturation mutagenesis, and site-directed mutagenesis, are used to generate these libraries (Lutz, 2010). Of these, site-saturation mutagenesis is one of the most prominent methods, as it allows the targeted introduction of mutations through the use of degenerated primers in a PCR. However, it harbors the risk that if the randomized sequence is expanded too much, secondary and tertiary structural elements can be impaired, thus limiting the functionality of the library (Marcheschi et al., 2013). Apart from this, the associated screening campaigns for such focused libraries often include the use of cell lysates (Chen et al., 2010) or cellular protein synthesis and purification (Heinzelman et al., 2009) of the individual variants, which is quite time-consuming and labor-intensive (Chen et al., 2010). Established systems consisting of RCA and CFPS can be used to circumvent these problems in order to amplify the gene library by RCA and express it via CFPS. But the potential of these RCA-CFPS systems has not yet been fully exhausted, as

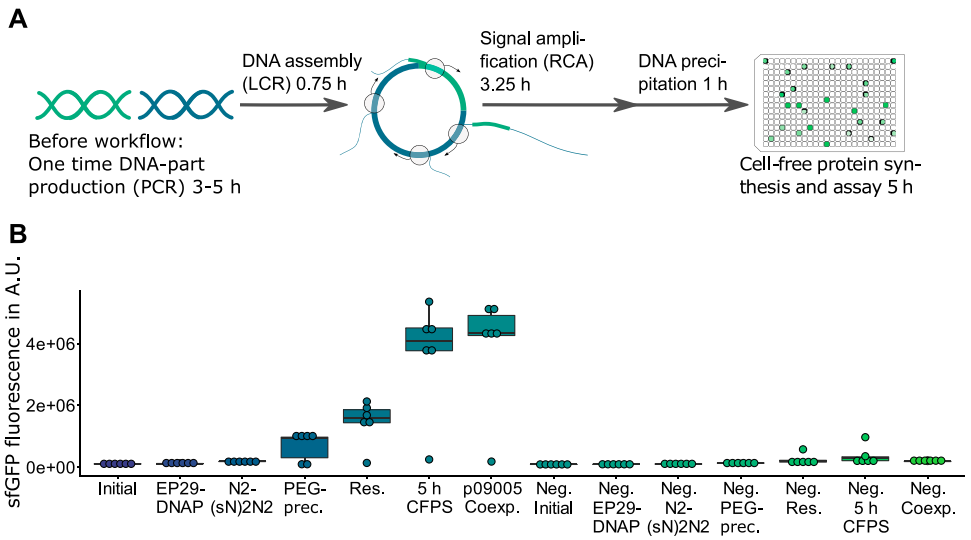


FIGURE 1 | HyperXpress workflow and effect on product formation of the most relevant optimizations. **(A)** The HyperXpress workflow starts with the assembly of DNA parts using LCR. This assembly includes a circularization of the DNA construct which upon addition of the components required for RCA results in an amplification of circular constructs. In order to provide optimal reaction conditions for CFPS a buffer exchange is achieved through DNA precipitation and resuspension. Upon addition of the components required for CFPS each DNA assembly is assayed for product formation. **(B)** Results of the sequential optimizations of the workflow compared to the corresponding 0 nM bridging oligo reaction as a negative control (Neg.). EP29: EquiPhi29™ DNA Polymerase; N2(sN)2N2: endonuclease protected random hexamer oligonucleotides; PEG prec.: in-well DNA precipitation using polyethylenglycol; res.: resuspension of the DNA precipitate in ultrapure water; p09005 coexp.: bicistronic operon plasmid with sfGFP and mKate2 genes used instead of p10024 template with monocistronic GFP reporter construct (n = 6).

their practical implementation is still accompanied by cumbersome work steps and increased material consumption (Dopp et al., 2019b). Thus previous RCA-CFPS systems suffer from high total reaction volumes of 15 µl, long execution times of at least 12 h, a high number of workflow steps of six or more and changing of reaction vessels (Dopp et al., 2019b; Hadi et al., 2020).

As a novel rapid approach to prototyping, the HyperXpress workflow (see **Figure 1**) represents a partially automated, successive single vessel reaction consisting of ligase cycling reaction (LCR) (Kok et al., 2014; Schlichting et al., 2019), multiply-primed rolling circle amplification (RCA) (Dean et al., 2001), and cell-free protein synthesis (CFPS) (Sun et al., 2013; Kwon and Jewett, 2015) within one 384-well plate on the lower microliter scale (Zibulski, 2019). At the beginning of the workflow, the LCR mediates a non-homologous assembly of selected DNA fragments to form a circular product (Kok et al., 2014), which is amplified in the subsequent RCA to form complex dsDNA concatamers (Dean et al., 2001), serving in the final CFPS as a template for the gene expression of the encoded proteins (Chong, 2014). After substantial optimization of this workflow in respect to achieving a low-volume, single vessel reaction of the RCA with LCR and CFPS, semi-rational protein engineering was carried out by LCR-based non-homologous recombination within the GFP protein family in order to work out the advantages and disadvantages of HyperXpress. As a result, deletion, substitution, and expansion of the GFP-β-barrel by β-strands between the GFP wild type superfolder GFP [sfGFP (Pédelacq et al., 2006)], mAvicFP1 [AF (Lambert et al., 2020)], and mNeonGreen [NG (Shaner et al., 2013)] could be demonstrated in a semi-rational fashion, resulting in 51 new-to-nature sequence variants, six of which were functional.

2 RESULTS AND DISCUSSION

2.1 Development of the HyperXpress Workflow

A traditional workflow for testing new biological functions includes assembling DNA and its transformation into a heterologous host, both for the purpose of selecting and amplifying a circularized DNA assembly as well as producing and analyzing the resulting product. This evaluation of assembled genetic information can be rapidly sped up by using cell-free gene-expression systems (Silverman et al., 2020). HyperXpress further speeds up the path from DNA part to prototype by allowing the combining of all steps in lower µl volume, single vessel reactions. LCR (Kok et al., 2014) is used for DNA assemblies as it allows scarless recombination of DNA parts, while RCA is used to selectively amplify successful and thus circularized assemblies, a function usually performed by a host organism. **Supplementary Figure S1** demonstrates the required RCA-amplification of circular constructs. **Supplementary Figure S2** with 1:1 vector-to-insert ratios demonstrates that the addition of bridging oligos resulting in circularizing provides a signal an order of magnitude higher than the corresponding non-assembled, linear fragments. This figure as well stresses the importance of later discussed control criteria: increasing ratios of functional expression fragments results in fluorescence signals comparable to the negative controls (ratios of 1:3 and 1:4).

Large batch-to-batch variations of CFPS with respect to its productivity (**Supplementary Figure S3**) can occur (Takahashi et al., 2015). In order to obtain batch-consistent data by using only one batch for a large prototyping campaign as well as to reduce material consumption, a protocol was developed that allows a 3.6 µl CFPS. Compared to the RCA-CFPS system with the previous CFPS volume minimum of 15 µl (Dopp et al., 2019b), this protocol reduces the required CFPS mix 4.2 fold with a lower necessary working concentration and increases the number of reactions available per batch to over 4,800 reactions for 4 ml of *E. coli* cell extract. The initially employed GenomiPhi™ V2 DNA Amplification Kit does not allow for an easy adjustment of volumes, as reaction volumes are pre-defined and while resulting in strong signals this solution is costly. The kit was thus replaced with a less costly, self-made RCA solution. As shown in **Supplementary Figure S4**, the phi29-based RCA mix described in this work provided a sufficient signal-to-background ratio and reduced the price by roughly 50%. Improvements to this custom kit came from using phi29 DNA polymerase with its supplied buffer in combination with an expanded RCA incubation time of 180 min as well as the utilization of a pyrophosphatase (PPase) for the removal of RCA-inhibiting pyrophosphate from the reaction (Dean et al., 2001).

2.2 Optimization of the HyperXpress Workflow

Next, the sfGFP-gene of the reporter-operon was split into seven parts and a workflow with a sufficient signal-to-background ratio for detecting low activity GFP variants was built up. Moreover, a genetic system producing both an internal reference mKate2 parallel to GFP and a higher sensitivity for GFP variants with low activity was realized. The relative fluorescence used to evaluate the experiments is defined as the $\frac{\text{mean fluorescence of the sample}}{\text{mean fluorescence of the corresponding negative control}}$ (Duportet et al., 2014). All optimizations were performed sequentially and not as a full factorial, so no conclusion can be made about potential combinatorial effects. The results of the optimizations of the workflow are summarized in **Figure 1** as well as additional figures in the supplement (**Supplementary Figures S5–S8, S10, S12**).

The rolling circle amplification step of the workflow was examined with respect to the optimal temperature as well as the utilization of random hexamers required for priming the polymerization reaction. Moreover, two phi-polymerases were tested to achieve a higher RCA efficiency: The wild type polymerase with a temperature optimum at 30°C (phi29 DNA Polymerase, NEB) and an engineered variant with an optimum at 40°C (EquiPhi29™, ThermoScientific). No significant difference in performance could be observed (**Supplementary Figure S5**, t test: $p = 0.158 > 0.05$). Future experiments will be performed with EquiPhi29™ because a higher stability and a lower error rate than the wild type DNA polymerase are reported (Povilaitis et al., 2016). For the optimization of the priming of the polymerization reaction, both different lengths and protection strategies for the random oligos added to the RCA reaction were tested. Oligos ranging in length from six to nine nucleotides were examined (**Supplementary Figure S6**). Since an increase in length did not result in a better signal-to-background ratio, less costly hexamers were chosen for the workflow. Both, the wild type phi29 as well as the EquiPhi29™, exhibit prominent 3′-5′-exonuclease activity as part of the proof-reading activity (Salas et al., 2008; Stockbridge and Wolfenden, 2011; Povilaitis et al., 2016), thus protecting the priming hexamers by incorporating 5′-phosphorothioate modifications in different positions was tested (**Supplementary Figures S7, S8**). Random hexamers with modifications of the two central oligonucleotides (N2sN2N2) showed the strongest increase in signal. It is likely that this central position of the phosphorothioate linkages allows, on the one hand, the degradation of non-annealed 3′-nucleotides of partially-annealed hexamers to generate an elongatable dsDNA-3′-terminus and, on the other hand, protection from complete exonucleolysis of non-annealed hexamers.

Variating the amount of RCA-reaction used in the subsequent CFPS indicated that components contained in the RCA reaction have an inhibiting effect on the CFPS (**Supplementary Figure S9**). Therefore a potentially automatable buffer exchange based on DNA precipitation and re-suspension in 384 microwell plates (MTP) using centrifugation was developed (see detailed protocol in the **Supplemental Material**). Besides being faster and not requiring cooling of the sample, PEG-precipitation outperformed ethanol precipitation in terms of signal strength when excluding the two failed reactions in the following CFPS (**Supplementary Figures S10, S11**). Efficient re-suspension was achieved by adding ultrapure water, sealing the MTP with a hydrophobic film and centrifuging the inverted plate at low rpm (**Supplementary Figure S12**). The precipitation and re-suspension steps frequently result in the failure of one out of every six replicate reactions performed (e.g. **Supplementary Figures S10, S12**), yet the small scale of the reactions and the high degree of parallelization can compensate for such events by allowing high numbers of replicates. While DNA precipitation strongly increases the signal, it could be omitted by using a CFPS-adapted RCA buffer which substitutes *in vitro* translation-inhibiting chloride anions (Weber et al., 1977) with L-glutamate (**Supplementary Figure S13**). Finally, extending the reaction time of the CFPS from 2.5 to 5 h doubled the signal strength (**Figure 1**).

By optimizing the workflow, HyperXpress now offers a number of advantages over existing RCA-CFPS systems. First of all, the sequence of all workflow steps, including DNA assembly, is performed in the same reaction vessel, while existing systems have to change the reaction vessel between each step. In addition, the CFPS volume has been reduced to 3.6 µl, which results in lower material consumption compared to other RCA-CFPS systems with a minimum final CFPS volume of 15 µl. This results in costs below 0.35 € per 3.6 µl reaction (price calculation see **Supplementary Data**) for the full workflow, compared to 0.20 € per 15 µl which includes only RCA-CFPS and excludes the price for three required DNA purification steps (Dopp et al., 2019a; Dopp et al., 2019b). Besides, HyperXpress only comprises a minimum of four steps, including DNA precipitation, in contrast to other systems requiring six or eight steps, including several necessary purification or dilution steps (see Supplement for

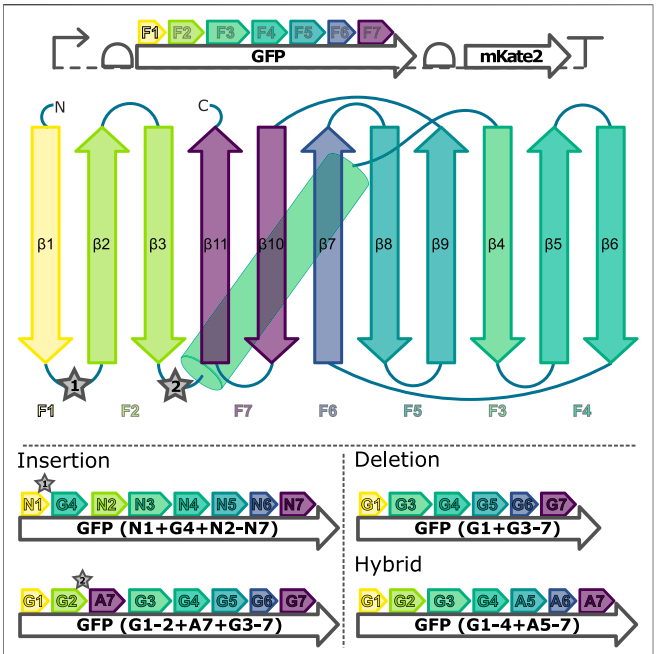


FIGURE 2 | Genetic construct and secondary structure indicating the fragments used for assemblies of the GFP proteins. *Top:* Bicistronic operon structure encoding the engineering target GFP and the assembly-control reporter mKate2. *Middle:* Schematic depicting the F1-F7 GFP DNA fragments used for the family shuffling. mKate2: red fluorescent protein; β1-11: β-sheets of the GFPs; ★ 1 and 2: insertion sites for additional β-sheets. *Bottom:* Examples for the three fragment recombination strategies performed including their nomenclature. G*n*: sfGFP fragment *n*; A*n*: mAvicFP1 fragment *n*; N*n*: mNeonGreen fragment *n*.

comparison). As a consequence, this workflow can be more readily scaled and, by omitting the precipitation step, which reduces the signal-to-background ratio, can be readily automated. Moreover, the LCR as the first step of the workflow directly enables DNA recombination and circularization to create its own focused DNA libraries and RCA templates, whereas other systems depend on pre-synthesized DNA libraries for starting. Finally, starting with a DNA-part library, HyperXpress can be executed in 10 h compared to at least 12 h (Kumar and Chernaya, 2009; Dopp et al., 2019a; Dopp et al., 2019b; Hadi et al., 2020).

To enable discrimination in engineering experiments between reactions without GFP fluorescence due to a failed LCR and reactions without GFP fluorescence due to assembling a genetic construct of non-functional GFP, the operon structure depicted at the top of **Figure 2** was designed and demonstrated (**Supplementary Figure S14**).

Testing the co-expression of sfGFP as the target to be assembled and mKate2 as the LCR-control resulted in a sfGFP signal comparable to that of a monocistronic operon with only a sfGFP gene (**Figure 1**, p10024 vs. p09005). This sfGFP-mKate2 co-expression without signal loss of sfGFP indicated that the resources in the CFPS (e.g., amino acids, NTPs) and the gene expression rate were sufficient to express mKate2 in parallel to sfGFP.

This operon was used to apply two control criteria to the proof-of-principle described below: At least 2 mKate2 fluorescence units were measured (see Equation 1), and thus the LCR was considered successful, thus the screening result was mKate2^{pos}. Otherwise, the LCR was deemed to have failed and the screening result was considered mKate2^{neg} (**Supplementary Figure S14**). This is the necessary criterion. A relative sfGFP fluorescence of at least 2 suggests an active GFP variant and thus a GFP^{pos} result. In this respect, a relative sfGFP fluorescence of less than 2 causes an inactive variant and thus a GFP^{neg} result. This is a sufficient criterion.

2.3 Proof-Of-Principle: GFP Domain Shuffling

Based on the optimized workflow, a LCR-based non-homologous recombination between the GFP wild types sfGFP, AF, and NG is carried out, involving the semi-rational insertion, deletion, and substitution of β-strands of the GFP-β-barrel. For this purpose, all three GFP wild types are divided into 7 fragments according to amino acid sequence homologies and, in the case of sfGFP and NG, additionally according to crystal structure similarities (Pedelacq et al., 2005; Clavel et al., 2016), so that all fragments always carry complete secondary structural elements (**Figure 2**).

On the one hand, among the 27 deletion variants, any deletion of fragments 2 (**Supplementary Figure S14**), 4 (**Supplementary Figure S15**), or 6 (**Supplementary Figure S16**) leads to a complete loss of GFP fluorescence activity, which is likely caused by excessive interference (Liu et al., 2015) of the β-barrel structure. On the other hand, among the 12 substitution (**Figure 3**; **Supplementary Figures S17, 18**) and 12 insertion variants (**Figure 4**; **Supplementary Figures S19, S20**), nine new-to-nature GFP variants with GFP fluorescence activity can be identified. Yet four of the nine positive variants need to be categorized as mKate2^{neg}-GFP^{pos} due to their failure to fuly meet the control criteria. Only the *in vivo* and *in vitro* validations (**Figure 5**; **Supplementary Figures S21 and S22**) confirm these four mKate2^{neg}-GFP^{pos} variants as active. Besides, among all 51 insertion, deletion, and substitution variants, there are 18 mKate2^{neg}-GFP^{neg} variants as a result of a non-functional LCR, which means that 35.3% of all variants are not even accessible for an evaluation of the fluorescence activity due to insufficient LCR efficiency.

There are also 24 mKate2^{pos}-GFP^{neg} variants, for which no activity can be detected. In spite of this, the workflow produces 9 mKate2^{pos}- and mKate2^{neg}-GFP^{pos} GFP variants, of which the *in vivo* and *in vitro* validations show four insertion and two substitution variants to be classified as active, while the remaining three insertion variants cannot be definitively evaluated because of the existence of point mutations at the time of the validations. Overall, the HyperXpress workflow yields 11.8% of validated active GFP variants from a focused library of 51 variants.

The three validated, active sfGFP insertion variants suggest that the sfGFP-β-barrel allows the homology-independent insertion of the AF-or NG-fragment 4 or 7 (see insertion examples in **Figure 2**) C-terminal to the sfGFP-fragment 1or 2

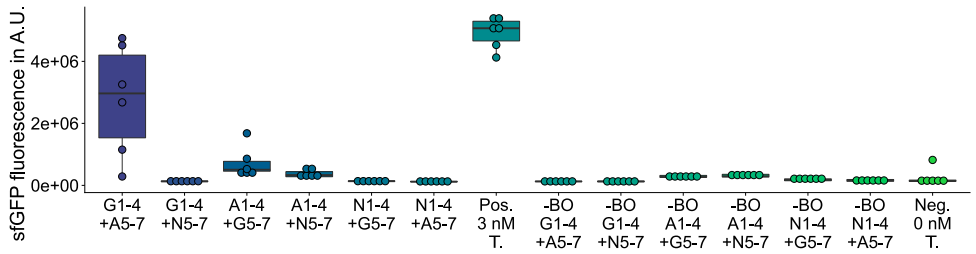


FIGURE 3 | GFP measurements of assemblies with hybrid sequences. G*n*: sfGFP fragment *n*; A*n*: mAvicFP1 fragment *n*; N*n*: mNeonGreen fragment *n*; Pos.: 3 nM of the template plasmid used to obtain the vector for assembly; -BO: no bridging oligos added; Neg.: No Template DNA added.

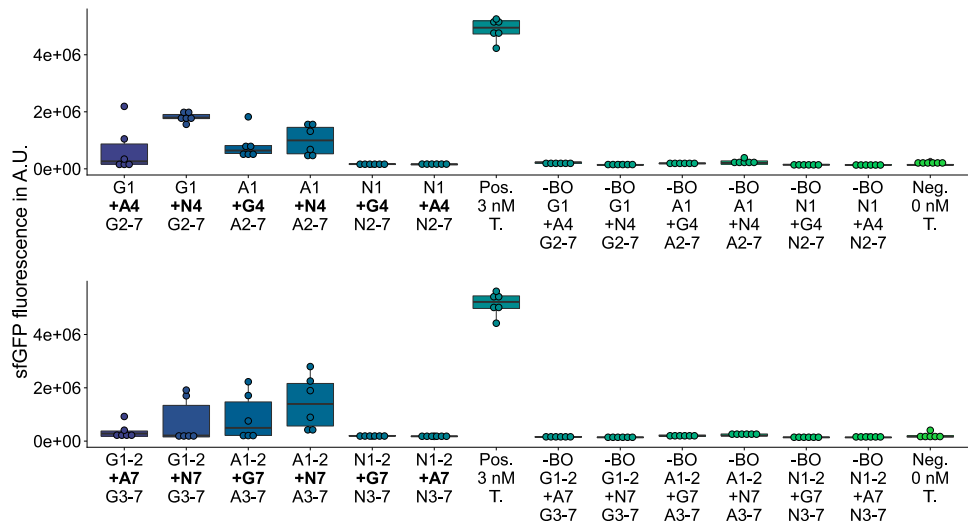


FIGURE 4 | GFP measurements of assemblies with additional sequences inserted. *Top:* Insertion of an additional fragment 4 between fragment 1 and 2. *Bottom:* Insertion of an additional fragment 7 between fragment 2 and 3. G*n*: sfGFP fragment *n*; A*n*: mAvicFP1 fragment *n*; N*n*: mNeonGreen fragment *n*; Pos.: 3 nM of the template plasmid used to obtain the vector for assembly; -BO: no bridging oligos added; Neg.: No Template DNA added.

without complete loss of activity. At the same time, there is an active AF insertion variant with an insertion of the sfGFP-fragment 4 C-terminal to the AF-fragment 1, which suggests a milieu-dependent loss of activity being only active *in vitro* in the context of CFPS of the HyperXpress workflow. Otherwise, there are two active substitution variants where the sfGFP- respectively AF-fragments 5 to 7 are substituted by the corresponding homologous AF-respectively sfGFP-β-strands. This manifests itself as an amino acid sequence homology dependency of this kind of substitution.

3 CONCLUSION

HyperXpress enables rapid prototyping as a single vessel *in vitro* workflow by enabling flexible, combinatorial DNA assembly using LCR, selection and amplification of correctly circularized DNA via RCA, and protein production using cell-free protein synthesis. The extensively optimized reaction steps and reagents allow down-scaling to low µL volumes, which allows a batch consistent and affordable throughput of thousands of samples. As a proof-of-concept, the HyperXpress workflow was used for the first time in semi-rational protein engineering and allows a fast, partially automated, material-saving screening of new GFP variants. Future work will include semi-rational protein engineering in other protein families, the engineering of multi-protein complexes, the construction of new metabolic pathways, and the design of genetic circuits (Gregorio et al., 2019).

4 MATERIALS AND METHODS

A detailed step-by-step protocol for the optimized HyperXpress workflow is provided in the **Supplementary Material**. A table of the utilized oligonucleotides can be found in **Supplementary Table S1** (bridging oligos) and **Supplementary Table S2** (primers). Strains are listed in **Supplementary Table S3**. Plasmid p09008 containing the hybrid-GFP (A1-4+G5-7), which showed a very high *in vitro* signal and can also be used

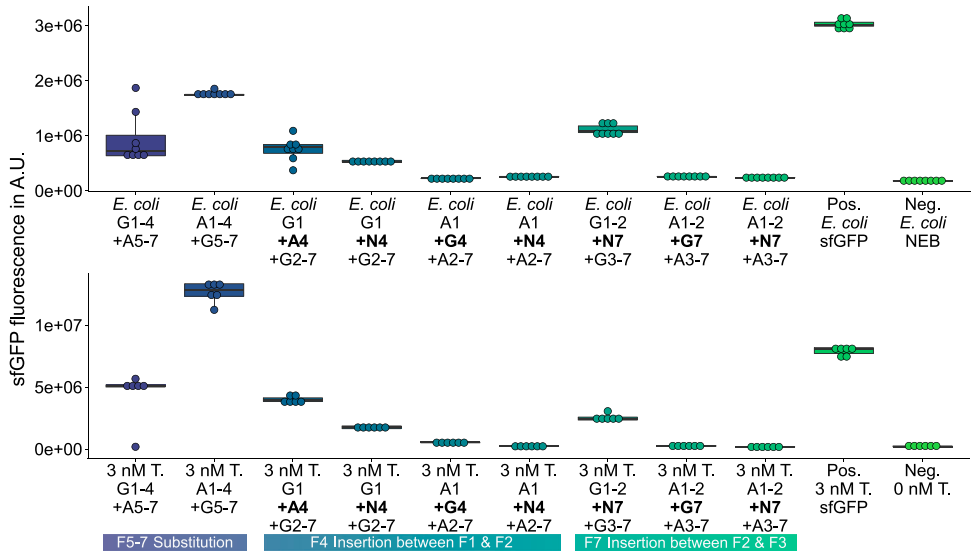


FIGURE 5 | GFP measurements validating constructs previously assembled and screened with HyperXpress. *Top: E. coli in vivo* validation of functional GFP-variants Pos.: *E. coli* expressing template plasmid; Neg.: *E. coli* carrying no plasmid (n = 8). *Bottom: in vitro* validation using 3 nM of isolated plasmid DNA (T.) instead of the LCR in the HyperXpress workflow; Neg.: No Template DNA added (n = 6).

to combine arbitrary LCR assemblies with mKate2 as a control, is deposited at Addgene (#173715).

4.1 5′-Phosphorylation and PCR Amplification

The 5′-phosphorylation of the DNA primers [modified after Schlichting et al. (2019), ordered lyophilized and desalted; Sigma-Aldrich] takes place in a 50 µl reaction mixture consisting of 10 µM DNA primers, 1x T4-PNK/T4-polynucleotide kinase buffer, 0.2 U/µl T4-PNK (New England BioLabs) and 1.7 mM ATP. This phosphorylation mixture is incubated for 70 min at 37°C, followed by inactivation for 20 min at 65°C. The 5'-phosphorylated primers are stored at −20°C and can be used in PCRs without further purification. Afterwards, all DNA fragments are amplified via 50 µl PCRs consisting of 0.2 mM of each dNTP, 1x Q5® Reaction Buffer Pack, 0.04 U/µl Q5® High-Fidelity DNA Polymerase (New England BioLabs), 250 nM of each 5'-phosphorylated forward and reverse primer, and 1 fM of template DNA (0.01 fM for the vector amplification). After the PCR, the PCR mixtures are digested with DpnI (1x CutSmart® Buffer, 0.4 U/µl DpnI; New England BioLabs) for 1 h at 37°C and denatured for 20 min at 80°C. Finally, the DpnI digested mixtures are purified via the innuPREP PCRpure Kit (Analytik Jena GmbH) and analyzed photometrically with regard to DNA concentration and quality (UV5 Nano; Mettler Toledo). The purified DNA fragments are stored at −20°C.

4.2 Ligase Cycling Reaction

In each well of a 384-well plate, 0.612 µl of LCR mixture is placed, which consists of 3 nM per DNA fragment, 30 nM per BO (ordered lyophilized and desalted; Sigma-Aldrich), 1x Ampligase® buffer [self-made, see Schlichting et al. (2019) or protocol in the **Supplementary Material**], 10 mM MgCl2, 0.5 mM NAD^+, and 0.3 U/µl Ampligase® (Lucigen) [modified after Schlichting et al. (2019)]. This LCR mixture is cycled through the following temperature program: 1. 92°C for 2 min; 2. 92°C for 5 s; 3. 66°C for 90 s; 4. repeat steps 2 and 3 another 24 times.

4.3 Rolling Circle Amplification

The multiply-primed RCA (modified after Dean et al., 2001) is divided into three sub-steps: annealing, isothermal amplification, and inactivation. For the annealing, a 0.468 µl solution of random N2(sN)2N2 DNA primers (ordered lyophilized and desalted; Sigma-Aldrich) and 10x phi29 DNA Polymerase Reaction Buffer/phi29 buffer (New England BioLabs) is dispensed to the 0.612 µl LCR via a contactless nanoliter-dispenser (I.DOT, DISPENDIX) so that the resulting 1.08 µl annealing mixture consists of 56.7% (v/v) LCR, 166.7 µM random hexamers, and 1.67x phi29 buffer. That annealing mixture is denatured for 3 min at 95°C and then cooled down to 4°C as quickly as possible. This is followed by the isothermal amplification, for which a 0.72 µl solution of dNTPs, BSA (New England BioLabs), EquiPhi29™ DNA polymerase (Thermo Scientific), inorganic *E. coli* pyrophosphatase (New England BioLabs), and DTT is dispensed to the 1.08 µl annealing mixture via the I.DOT. All of that results in the final 1.8 µl RCA mixture consisting of 34% (v/v) LCR, 100 µM N2(sN)2N2, 1 mM of each dNTP, 0.4 µg/µl BSA, 1x phi29 buffer, 0.36 U/µl EquiPhi29™ DNA polymerase, 0.004 U/µl inorganic *E. coli* pyrophosphatase, and 4 mM DTT. This 1.8 µl of RCA mixture is incubated for 3 h at 40°C, followed by inactivation for 10 min at 65°C.

4.4 384-Well DNA Precipitation and Re-Suspension

At the beginning of the PEG DNA precipitation [modified after Paithankar and Prasad (1991)] carried out at room temperature, a 1.8 µl solution of 26% (w/v) PEG-8000 and 20 mM MgCl2 is added via a dispenser pipette to the 1.8 µl RCA, resulting in a 3.6 µl precipitation mixture with 50% (v/v) RCA, 13% (w/v) PEG-8000 and 10 mM MgCl2. This mixture is centrifuged at 4,000 rpm at 20°C for 30 min. In order to be able to remove the supernatant from the DNA precipitate, the 384-well plate is covered with a piece of paper and briefly centrifuged in an inverted position in the table centrifuge for a few seconds so that the supernatant is centrifuged out of the wells onto the paper. Afterwards, a washing step is executed in which 3.6 µl of 70% (v/v) ethanol is dispensed into the wells and centrifuged for 5 min at 4,000 rpm and 20°C, followed by discarding the supernatant by centrifugation in the inverted position on the table centrifuge. That washing step is repeated once. The washed DNA precipitate is then dried for 10 min at 30°C; 1.8 µl ultrapure water is added via the nanoliter-dispenser and dissolved again at 50°C for 10 min. To resuspend the DNA solution, the 384-well plate is tightly closed with Parafilm (Pechiney Plastic Packaging). This sealed 384-well plate is centrifuged in an inverted position for a few seconds in a table-top centrifuge, resulting in the collection of the solution on the Parafilm lid of the wells. The plate is then centrifuged again in the non-inverted position for a few seconds to transfer the solution back into the wells.

4.5 Cell-Free Protein Synthesis

To carry out the 3.6 µl CFPS [modified after Sun et al. (2013)], a 1.8 µl solution of *E. coli* cell extract (CE, self-made, see protocol in the **Supplementary Material**) and cell extract buffer (CEB) is dispensed via the nanoliter-dispenser to the 1.8 µl DNA solution to get 50% (v/v) DNA solution, 22.73% (v/v) CE and 27.27% (v/v) CEB (under permanent cooling of the 384-well plate). In contrast to Sun et al. (2013), the CE is produced via modified cell cultivation and the cell lysis is done via sonification (see protocol in the **Supplementary Material**). This CFPS mixture is covered with an optical adhesive cover, incubated for 5 h at 29°C in a microplate reader (ClarioStar Plus, BMG LabTech) and the fluorescence is measured online via top optic with the following wavelength settings: sfGFP and AF (λex = 470 ± 7.5 nm/λem = 515 ± 10 nm; mKate2 (λex = 588 ± 7.5 nm/λem = 633 ± 10 nm); NG (λex = 491 ± 7 nm/λem = 533 ± 10 nm).

4.6 Sequence Verification

All DNA assemblies described have been verified by sequencing. For this, 50 µl electro-competent *E. coli* (*E. coli* NEB®-10β, New England Biolabs) were mixed with 5 µl LCR mixture, filled into an electroporation cuvette and incubated for 5 min on ice. These cells were electroporated for 4.6 ms at 2.5 kV, resuspended in 1 ml of LB media and incubated for 1 h at 37°C and 200 rpm. Finally, the cells were plated out on amp-containing LB agar plates so that they could grow over night at 37°C. Afterwards, selected colonies were cultivated in amp-containing LB media to create -80°C cryo-stocks for the *in vivo* validation and to obtain plasmids for sequencing and *in vitro* validation.

4.7 *In vivo* Validation

An aliquot of the *E. coli* −80°C cryo-stock is plated out on ampicillin (amp)-containing LB agar plates and incubated at 37°C overnight. The next day, the grown *E. coli* colonies are used to inoculate an overnight ampicillin culture, which is incubated at 37°C and 200 rpm. From this overnight culture, an OD_{600} of 0.5 is set in 1.5 ml of LB-amp medium (preheated to 30°C), representing the final culture to be measured. The final culture is distributed in 8-fold repetitions with 150 µl per well over a 96-well plate. These 150 µl final cultures are incubated for 5 h at 30°C and 200 rpm using 3D-printed plate holders (Bruder et al., 2019) so that after 5 h an end point measurement can be carried out in a microplate reader for the absorption at 600 nm and for the sfGFP, mKate2, and NG fluorescence. Before the first measurement, the plate is shaken for 15 s at 200 rpm.

4.8 Software

Plots were generated using R (R Development Core Team, 2008) and R-Studio (RStudio Team, 2018) as an IDE. *In silico* planning was done using Geneious R10.2 (https://www.geneious.com). Combinatorial DNA assemblies were planed using DIVA (https://public-diva.jbei.org/) with custom add-ons for automatic design of bridging oligos as well as automated generation of dispensing protocols (yet unpublished work).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

DZ—conceptualization, investigation, visualization, and writing; NS—conceptualization, investigation; JK—conceptualization, visualization, writing, resources, and supervision.

ACKNOWLEDGMENTS

The authors thank Woody Fessner for his valuable feedback, Nathan Hilson and Hector Plahar for giving us access to the DIVA code base, and Harald Gültig for software engineering of the LCR and dispensing add-ons for DIVA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.832176/full#supplementary-material

REFERENCES

Bruder, S., Moldenhauer, E. J., Lemke, R. D., Ledesma-Amaro, R., and Kabisch, J. (2019). Drop-in Biofuel Production Using Fatty Acid Photodecarboxylase from *Chlorella Variabilis* in the Oleaginous Yeast *Yarrowia Lipolytica*. *Biotechnol. Biofuels* 12, 202. doi:10.1186/s13068-019-1542-4

Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., et al. (2010). Reconstructed Evolutionary Adaptive Paths Give Polymerases Accepting Reversible Terminators for Sequencing and SNP Detection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1948–1953. doi:10.1073/pnas.0908463107

Chong, S. (2014). Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Curr. Protoc. Mol. Biol.* 108, 16. doi:10.1002/0471142727.mb1630s108

Clavel, D., Gotthard, G., and Royant, A. (2016). Structure of the Yellow-Green Fluorescent Protein mNeonGreen From Branchiostoma lanceolatum at the Near physiological pH 8.0. doi:10.2210/pdb5ltr/pdb

Cole, M. F., and Gaucher, E. A. (2011). Exploiting Models of Molecular Evolution to Efficiently Direct Protein Engineering. *J. Mol. Evol.* 72, 193–203. doi:10.1007/s00239-010-9415-2

Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome Res.* 11, 1095–1099. doi:10.1101/gr.180501

Dopp, J. L., Jo, Y. R., and Reuel, N. F. (2019a). Methods to Reduce Variability in *E. Coli*-Based Cell-free Protein Expression Experiments. *Synth. Syst. Biotechnol.* 4, 204–211. doi:10.1016/j.synbio.2019.10.003

Dopp, J. L., Rothstein, S. M., Mansell, T. J., and Reuel, N. F. (2019b). Rapid Prototyping of Proteins: Mail Order Gene Fragments to Assayable Proteins within 24 hours. *Biotechnol. Bioeng.* 116, 667–676. doi:10.1002/bit.26912

Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., et al. (2014). A Platform for Rapid Prototyping of Synthetic Gene Networks in Mammalian Cells. *Nucleic Acids Res.* 42, 13440–13451. doi:10.1093/nar/gku1082

Gregorio, N. E., Levine, M. Z., and Oza, J. P. (2019). A User's Guide to Cell-Free Protein Synthesis. *MPs* 2, 24. doi:10.3390/mps2010024

Hadi, T., Nozzi, N., Melby, J. O., Gao, W., Fuerst, D. E., and Kvam, E. (2020). Rolling circle Amplification of Synthetic DNA Accelerates Biocatalytic Determination of Enzyme Activity Relative to Conventional Methods. *Sci. Rep.* 10, 10279. doi:10.1038/s41598-020-67307-9

Heinzelman, P., Snow, C. D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., et al. (2009). A Family of Thermostable Fungal Cellulases Created by Structure-Guided Recombination. *PNAS* 106, 5610–5615. doi:10.1073/pnas.0901417106

Kok, S. d., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., et al. (2014). Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* 3, 97–106. doi:10.1021/sb4001992

Kumar, G., and Chernaya, G. (2009). Cell-free Protein Synthesis Using Multiply-Primed Rolling circle Amplification Products. *BioTechniques* 47, 637–639. doi:10.2144/000113171

Kwon, Y.-C., and Jewett, M. C. (2015). High-throughput Preparation Methods of Crude Extract for Robust Cell-free Protein Synthesis. *Sci. Rep.* 5, 8663. doi:10.1038/srep08663

Lambert, G. G., Depernet, H., Gotthard, G., Schultz, D. T., Navizet, I., Lambert, T., et al. (2020). Aequorea's Secrets Revealed: New Fluorescent Proteins with Unique Properties for Bioimaging and Biosensing. *Plos Biol.* 18, e3000936. doi:10.1371/journal.pbio.3000936

Liu, S.-s., Wei, X., Dong, X., Xu, L., Liu, J., and Jiang, B. (2015). Structural Plasticity of green Fluorescent Protein to Amino Acid Deletions and Fluorescence rescue by Folding-Enhancing Mutations. *BMC Biochem.* 16, 17. doi:10.1186/s12858-015-0046-5

Lutz, S. (2010). Beyond Directed Evolution-Semi-Rational Protein Engineering and Design. *Curr. Opin. Biotechnol.* 21, 734–743. doi:10.1016/j.copbio.2010.08.011

Marcheschi, R. J., Gronenberg, L. S., and Liao, J. C. (2013). Protein Engineering for Metabolic Engineering: Current and Next-Generation Tools. *Biotechnol. J.* 8, 545–555. doi:10.1002/biot.201200371

Nannemann, D. P., Birmingham, W. R., Scism, R. A., and Bachmann, B. O. (2011). Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis. *Future Med. Chem.* 3, 809–819. doi:10.4155/fmc.11.48

Paithankar, K. R., and Prasad, K. S. N. (1991). Precipitation of DNA by Polyethylene Glycol and Ethanol. *Nucl. Acids Res.* 19, 1346. doi:10.1093/nar/19.6.1346

Pedelacq, J. D., Cabantous, S., Tran, T. H., Terwilliger, T. C., and Waldo, G. S. (2005). Crystal Structure of a Superfolder Green Fluorescent Protein. doi:10.2210/pdb2b3p/pdb

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and Characterization of a Superfolder green Fluorescent Protein. *Nat. Biotechnol.* 24, 79–88. doi:10.1038/nbt1172

Povilaitis, T., Alzbutas, G., Sukackaite, R., Siurkus, J., and Skirgaila, R. (2016). In Vitroevolution of Phi29 DNA Polymerase Using Isothermal Compartmentalized Self Replication Technique. *Protein Eng. Des. Selection* 29, 617–628. doi:10.1093/protein/gzw052

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org

Robinson, C. J., Carbonell, P., Jervis, A. J., Yan, C., Hollywood, K. A., Dunstan, M. S., et al. (2020). Rapid Prototyping of Microbial Production Strains for the Biomanufacture of Potential Materials Monomers. *Metab. Eng.* 60, 168–182. doi:10.1016/j.ymben.2020.04.008

RStudio Team (2018). *RStudio*. Boston, MA: Integrated Development Environment for R; RStudio, Inc.

Salas, M., Blanco, L., Lázaro, J. M., and de Vega, M. (2008). The Bacteriophage Phi29 DNA Polymerase. *IUBMB Life* 60, 82–85. doi:10.1002/iub.19

Schlichting, N., Reinhardt, F., Jager, S., Schmidt, M., and Kabisch, J. (2019). Optimization of the Experimental Parameters of the Ligase Cycling Reaction. *Synth. Biol. (Oxf)* 4, ysz020. doi:10.1093/synbio/ysz020

Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., et al. (2013). A Bright Monomeric green Fluorescent Protein Derived from *Branchiostoma lanceolatum*. *Nat. Methods* 10, 407–409. doi:10.1038/nmeth.2413

Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020). Cell-free Gene Expression: an Expanded Repertoire of Applications. *Nat. Rev. Genet.* 21, 151–170. doi:10.1038/s41576-019-0186-3

Stockbridge, R. B., and Wolfenden, R. (2011). Enhancement of the Rate of Pyrophosphate Hydrolysis by Nonenzymatic Catalysts and by Inorganic Pyrophosphatase. *J. Biol. Chem.* 286, 18538–18546. doi:10.1074/jbc.m110.214510

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for Implementing an *Escherichia coli* Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp.* 79, 50762. doi:10.3791/50762

Takahashi, M. K., Hayes, C. A., Chappell, J., Sun, Z. Z., Murray, R. M., Noireaux, V., et al. (2015). Characterizing and Prototyping Genetic Networks with Cell-free Transcription-Translation Reactions. *Methods* 86, 60–72. doi:10.1016/j.ymeth.2015.05.020

Weber, L. A., Hickey, E. D., Maroney, P. A., and Baglioni, C. (1977). Inhibition of Protein Synthesis by Cl-. *J. Biol. Chem.* 252, 4007–4010. doi:10.1016/s0021-9258(17)40350-4

Zibulski, D. (2019). *Entwicklung einer Zellextrakt-basierten* in vitro *Ligase-Cycling-Reaction. Thesis*. Darmstadt: Technical University Darmstadt.

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. TU Darmstadt has applied for a patent in the name of the authors of this publication (EP 21155857.2, pending) covering aspects of the method described in the supplemental protocol.

frontiers in Bioengineering and Biotechnology

METHODS
published: 07 April 2022
doi: 10.3389/fbioe.2022.873854

Rapid and Facile Preparation of Giant Vesicles by the Droplet Transfer Method for Artificial Cell Construction

Yasuhiro Shimane[1,2] *and Yutetsu Kuruma*[1,3,4]*

[1]Institute for Extra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan, [2]Research Institute of Industrial Technology, Toyo University, Saitama, Japan, [3]PRESTO, Japan Science and Technology Agency (JST), Saitama, Japan, [4]Graduate School of Nanobioscience, Yokohama City University, Yokohama, Japan

Giant vesicles have been widely used for the bottom-up construction of artificial (or synthetic) cells and the physicochemical analysis of lipid membranes. Although methods for the formation of giant vesicles and the encapsulation of molecules within them have been established, a standardized protocol has not been shared among researchers including non-experts. Here we proposed a rapid and facile protocol that allows the formation of giant vesicles within 30 min. The quality of the giant vesicles encapsulating a cell-free protein expression system was comparable to that of the ones formed using a conventional method, in terms of the synthesis of both soluble and membrane proteins. We also performed protein synthesis in artificial cells using a lyophilized cell-free mixture and showed an equivalent level of protein synthesis. Our method could become a standard method for giant vesicle formation suited for artificial cell research.

Keywords: giant vesicle, artificial cells, protocells, droplet transfer, cell-free synthetic biology, cell-free protein synthesis, liposome, lyophilization

OPEN ACCESS

Edited by:
Jian Li,
ShanghaiTech University, China

Reviewed by:
Allen Liu,
University of Michigan, United States
Emiliano Altamura,
University of Bari Aldo Moro, Italy

***Correspondence:**
Yutetsu Kuruma
ykuruma@jamstec.go.jp

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *11 February 2022*
Accepted: *03 March 2022*
Published: *07 April 2022*

Citation:
Shimane Y and Kuruma Y (2022) Rapid and Facile Preparation of Giant Vesicles by the Droplet Transfer Method for Artificial Cell Construction. Front. Bioeng. Biotechnol. 10:873854. doi: 10.3389/fbioe.2022.873854

INTRODUCTION

The droplet transfer method, which is also known as the inverted emulsion method, has been widely used for the formation of giant vesicles (GVs) with a diameter of tens of micrometers (Pautot et al., 2003; Walde et al., 2010). GVs have been applied as a model cell membrane for the investigation of the physical properties of lipid membranes in the field of soft-matter physics (Jimbo et al., 2016; Oda et al., 2020; Lowe et al., 2022), and for the construction of artificial cells in the field of synthetic biology (Lu et al., 2021). Although several methods have been developed to form GVs, such as electro formation (Angelova and Dimitrov, 1986) or film hydration (Reeves and Dowben, 1969; Tsumoto et al., 2009), the droplet transfer method (Moga et al., 2019) has advantages in the generation of single lamellar membrane vesicles and its application to biochemical experiments requiring physiological conditions. The principle of this method is as follows: the water-in-oil (W/O) droplets, which are stabilized by a monolayer of phospholipids, pass through another monolayer formed at the interface of a lipid-oil layer and an aqueous layer; as a consequence, lipid bilayer vesicles are formed in the aqueous solution (Noireaux and Libchaber, 2004).

Although GVs are highly versatile materials, it is not always easy to obtain quality GVs with high reproducibility, which often fails even if following the same experimental steps. Moreover, beginners who have no experience in preparing GVs may be more likely to have an experimental failure. This tendency increases as the composition of the internal solution become more complex. To reduce these difficulties, some tailor-made devices have been created and applied in

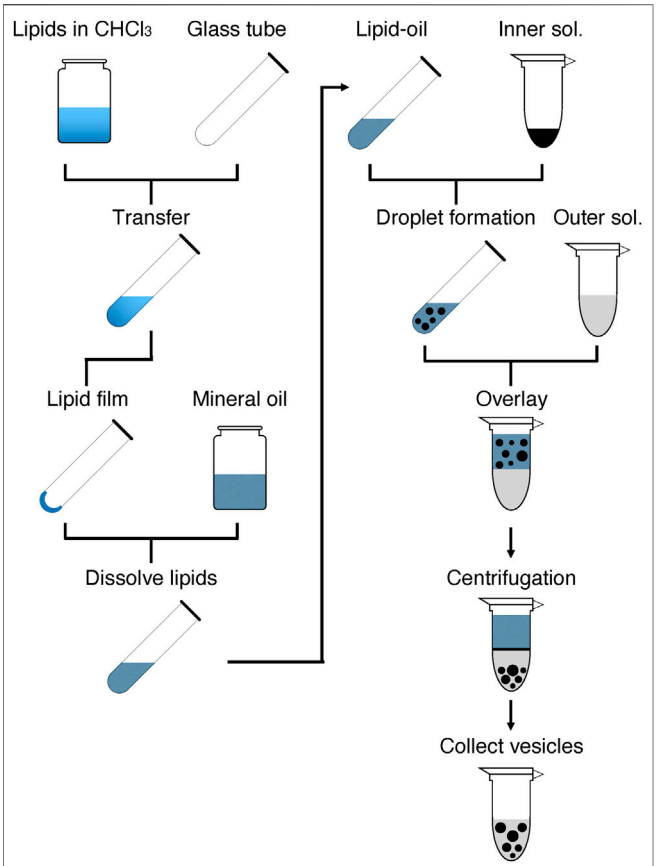


FIGURE 1 | Overview of the rapid and facile preparation of giant vesicles. The left half shows the steps for the preparation of lipid-oil mixture, and the right half shows the steps for the formation of GVs. The details are described in the Stepwise Procedures.

the droplet transfer method (Morita et al., 2018; Bashirzadeh et al., 2021). In addition to the droplet transfer method, more advanced methods using microfluidic devices have recently been developed and will become the main technology in this field (Robinson, 2019; Kamiya, 2020). However, designing the devices, creating the template mold, and the relatively high cost for the system set up hesitate us to choose this method. Therefore, establishing a concise protocol that allows the easy and rapid preparation of GVs with high reproducibility is important to reduce experimental errors.

In this report, we present a handy protocol that allows the completion of all experimental steps, from the lipid-oil preparation to the GV formation, within 30 min without the use of specific devices. The reproducibility of the method is high and a sufficient population of GVs can be produced. We show that we have successfully prepared artificial cells by this method using a cell-free protein expression system (cell-free system) that has been lyophilized and rehydrated. Because the technique of GV formation is the basis of artificial cell experiments, we aim to standardize our method for the development of artificial (synthetic) cell research.

Overview of the Experimental Approach Used for GV Formation

As shown in **Figure 1**, GV formation is initiated by preparing a lipid-oil mixture containing the desired lipid composition. The prepared lipid-oil mixture is used to form W/O droplets by mixing with the inner solution, which is composed of the desired reaction mixture, e.g., a cell-free system. The prepared droplets are layered on the outer solution. Giving a force by centrifugation induces the passage of the droplets through the interface between the oil and aqueous layers, thus forming lipid bilayer GVs in the aqueous solution. The formed GVs are collected and used for a subsequent experiment. If needed, the GVs can be washed with the outer solutions to remove the components that leaked from the broken droplets.

Lipid Composition of GV Membranes

Although the type of lipids used for GV formation depends on the purpose of each experiment, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) or 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) is generally used as a basic lipid, because of its high stability. When preparing GV using several types of lipids, the lipids dissolved in an organic solvent such as chloroform at the defined concentration are mixed to obtain the desired composition. When synthesizing membrane proteins inside GVs, an acidic phospholipid such as phosphatidylglycerol (PG) or phosphatidylserine (PS) is added to the lipid composition—for example, 10–30 mol% acidic phospholipids are used—because the negative charge on the membrane surface is important for maintaining the correct structure of membrane protein and affects to its function (Pöyry and Vattulainen, 2016). The charge is also important for the localization of peripheral membrane proteins onto the membrane surface (Furusato et al., 2018). The use of 30 mol% cholesterol makes the membrane rigid (Gracià et al., 2010; Chakraborty et al., 2020). However, it should be noted that a certain amount of cholesterol inhibits the spontaneous membrane insertion when integral membrane proteins were synthesized inside GVs or outside of liposomes (Nakamura et al., 2018; Berhanu et al., 2019). Besides cholesterol, a physiological concentration of diacylglycerol also inhibits the spontaneous insertion (Nakamura et al., 2018). Fluorescent lipids or hydrophobic dyes, such as rhodamine-phosphatidylethanolamine (PE), NBD-PE, or Nile Red, are used for labeling GV membranes. Moreover, polyethylene glycol (PEG)-lipids (2–5 KDa in size) are used to avoid the adhesion of GVs. Chemically modified unnatural lipids are also used for experimental purposes (Kurihara et al., 2011; Bhattacharya et al., 2019). Conversely, natural lipids extracted from the membranes of various organisms, such as soybean or egg yolk, are not suitable for the formation of stable GV by this method, whereas they are applicable for the electroformation

TABLE 1 | Lipid composition of GV used for artificial cell experiments.

Type of synthesized or encapsulated proteins	Lipid composition (mol%)	References
Integral membrane proteins	DOPC 50%, DOPE 36%, DOPG 12%, CL (18:1) 2%, DSPE-PEG-biotin 1 mass%, DHPE-Texas Red 0.5 mass%. Or, DOPC 75%, DOPG 25%, DSPE-PEG-biotin 1 mass%, DHPE-Texas Red 0.5 mass%	Blanken et al. (2020)
Soluble and integral membrane proteins	POPC and cholesterol with or without 1% DSPE-PEG(2000) biotin	Toparlak et al. (2020)
Soluble protein	DOPC and a chemically modified lysophospholipid	Bhattacharya et al. (2019)
Soluble protein	POPC 40%, POPE 20%, POPG 20%, cholesterol 20%, NBD-PE 0.25%	Lee et al. (2018)
Soluble and integral membrane proteins	POPC 57.5%, cholesterol 40%, PEG2000PE[a] 0.25%	Berhanu et al. (2019)
Integral and peripheral membrane proteins	POPC 80%, POPG 20%, rhodamine-DOPE 0.5%	Furusato et al. (2018)

[a]*1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[biotinyl(polyethyleneglycol)-2000].*

method (Méléard et al., 2009) or for preparation of small-size liposomes (<200 nm in diameter). Examples of lipid compositions reported in recent artificial cell studies are shown in **Table 1**.

Inner Solution

The inner solution corresponds to the internal aqueous phase of GVs. Regarding the inner solution, any component can be encapsulated in GVs, except hydrophobic molecules and detergents. For example, by adding purified proteins into the inner solution, various enzymatic reactions or structural formations in the GVs can be generated, e.g., polymerase chain reaction (Oberholzer et al., 1995), transcription (Altamura et al., 2021), actin polymerization (Lee et al., 2018; Litschel et al., 2021), Min oscillation (Litschel et al., 2018; Yoshida et al., 2019), etc. In more advanced applications, the encapsulation of a cell-free system reconstituted with purified multi-components involved in transcription and translation (Shimizu et al., 2001) or a cell extract from a certain organism such as *Escherichia coli* (Noireaux and Liu, 2020) allows the synthesis of desired proteins from the genes of interest. This technology has an advantage, particularly for the synthesis of membrane proteins (Furusato et al., 2018; Berhanu et al., 2019). In general, the purification of membrane proteins requires the use of a detergent that dissolves the cell membrane to isolate the proteins. The purified membrane protein sample contains a detergent, therefore it cannot be encapsulated inside GVs, although there are some exceptions (Yanagisawa et al., 2011; Altamura et al., 2017). To solve this problem, synthesizing membrane proteins inside GVs is a rational approach and allows the expression of biochemical functions on the GV membranes. This approach is depending on the phenomenon of spontaneous membrane insertion of nascent membrane proteins (Nishiyama et al., 2006). It should be noted that, when a certain amount of cholesterol or diacylglycerol is included in the composition of the GV membrane, this spontaneous insertion does not occur. The limitation of this approach is that not all membrane proteins can be integrated into the lipid membrane in a native form (Berhanu et al., 2019). For example, if a membrane protein has a large hydrophilic domain at the outside of the GV membrane (the opposite side of the translating ribosome), the spontaneous insertion with the correct membrane orientation of the protein does not occur.

TABLE 2 | Example of the inner solution of artificial cells using the PURE system (PURE*frex*2.0).

Reagent	Volume (μL)
Sol. I (Buffer, etc.)	10
Sol. II (Enzyme mix)	1
Sol. III (Ribosomes)	2
2 M Sucrose	2
DNA (plasmid or linear)	X (1–5 nM f.c.)
Water	5-X
Total	20

In addition to purified proteins and cell-free systems, small-sized liposomes with a diameter of <200 nm can also be encapsulated, presenting the vesicle-in-vesicle structure (Lee et al., 2018; Berhanu et al., 2019; Altamura et al., 2021). This technique can mimic intracellular organelles. Interestingly, by coupling with a cell-free system that synthesizes a membrane protein, it is possible to localize the synthesized membrane proteins onto the liposome membrane (Berhanu et al., 2019). The membrane localization of the protein can be oriented using cholesterol (Nakamura et al., 2018), i.e., when cholesterol is used in the GV membrane but not in the liposome membrane, a large part of the synthesized membrane proteins are localized to the internal liposomes. The encapsulated liposome organelle is also useful for generating a proton gradient between the GV lumen and the liposome inside (Berhanu et al., 2019).

To mimic intracellular molecular crowding, Ficoll PM70 (Furusato et al., 2018; Berhanu et al., 2019) or bovine serum albumin (Fujiwara and Yanagisawa, 2014) is often encapsulated within GVs. This molecular crowding effect is essential for various types of molecular assembly and, especially, for the membrane localization of peripheral membrane proteins in association with a negative charge on the membrane surface.

In any case, the droplet transfer method generally uses sucrose in the inner solution to make it heavier than the

TABLE 3 | Example of the outer solution of artificial cell encapsulating the PURE system.

Component	2× preparation	1× final conc
HEPES-KOH (pH 7.6)	40 mM	20 mM
Potassium glutamate	360 mM	180 mM
MgOAc	28 mM	14 mM
Spermidine	4 mM	2 mM
10-Formyl-tetrahydrofolate	20 μg/ml	10 μg/ml
Dithiothreitol	4 mM	2 mM
18-amino-acid mixture (w/o cysteine and tyrosine)	1 mM	0.5 mM
Cysteine	1 mM	0.5 mM
Tyrosine	1 mM	0.5 mM
Creatine phosphate	40 mM	20 mM

glucose-containing outer solution, while maintaining equal osmotic pressure. In some experiments, sucrose may have a side effect on the internal reaction. In such a case, OptiPrep, a density gradient medium, is often used instead of sucrose (Van de Cauter et al., 2021).

An example of the inner solution used for the encapsulation of a commercial cell-free kit (PURE system) in GVs is shown in **Table 2**.

Outer Solution

The outer solution should be made when preparing the inner solution or in advance. The composition of the outer solution should be same or very similar to that of the inner solution, except reactive molecules and genes, i.e., enzymes and DNAs. The osmolality of the outer solution should be adjusted by changing the concentration of glucose, or alternative material, to be the same as the inner osmotic pressure. For the use of cell-free systems, the basic buffer composition and amino acids should be maintained, whereas a tRNA mixture and NTPs are removed. Note that the outer solution should be prepared from a double-concentrated solution, as the final solution can be diluted with the glucose solution (see an example in **Table 3**). The prepared outer solution can also be used for washing GVs after their formation.

Formation of GVs

The prepared W/O droplets are transferred to the aqueous phase by giving a force from the top to the bottom of the microtube by centrifugation (**Figure 1**). Although a large number of the droplets are broken when they pass through the interface, successful droplets form a lipid bilayer membrane, resulting in the generation of GVs. Usually, after centrifugation, white debris appears at the interface. This has to be removed entirely together with the oil phase. The formed GVs appear at the bottom of the tube because those are containing sucrose, which is heavier than glucose in the outer solution. A portion of the bottom fraction is collected carefully and transferred into a fresh tube for further experimentation. GVs can be washed with the same outer solution. Additional reagents can be supplied to the outer solution as needed while paying attention to the change in osmotic pressure. For example, ATP can be supplied to the exterior of the GVs at the same concentration as in the internal cell-free system, considering the possible leakage of ATP from the inside.

The Storable Artificial Cell Mixture

To make the component solutions of artificial cells portable and easy to store, the mixed inner cell-free reaction solution (without DNA) and the outer solution were lyophilized. These are used to make a ready-to-use kit together with the dried lipid film and mineral oil. This enabled us to generate artificial cells within 30 min starting from the preparation of the lipid-oil to the formation of artificial cells. The successful protein synthesis using the lyophilized and rehydrated cell-free components has been reported previously (Hunt et al., 2017; Lee et al., 2020; Yang et al., 2021). Therefore, it is possible to ship the samples at ambient temperature while avoiding moisture. This has a great advantage for working outside the laboratory or exchanging materials between laboratories.

MATERIALS AND EQUIPMENT

Materials

- Glass tube: diameter, 10 mm; length, 50 mm (Maruemu Corp., cat. #0407-03)
- Microtube, 1.5 ml (Eppendorf, 3810X)
- PCR tube, 0.2 ml (Nippon Genetics, Co., Ltd., cat. # FG-021D)
- Parafilm (Bemis™, PM996)

Reagents

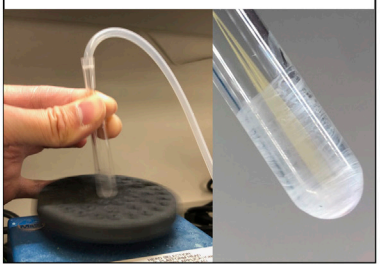
- Chloroform ($CHCl_3$) (Fujifilm, cat. #038-02606). CAUTION: the vapor and liquid forms of chloroform are toxic
- Mineral oil (MP Biomedicals, Inc., cat. #194836)
- 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (sodium salt), POPC (Avanti Polar, cat # 850457P)
- 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1′-rac-glycerol) (sodium salt), POPG (Avanti Polar, cat # 840457P)
- Sucrose (Fujifilm, cat. # 196-00015)
- d(+)-Glucose (Fujifilm, cat. # 049-31165)
- Cell-free system, e.g., PURE*frex*2.0 (GeneFrontier, cat. # PF201-0.25-5-EX)
- Genes of interest (under the control of the T7 promoter and ribosome-binding site)
- (Optional) Ficoll PM70 (Sigma-Aldrich, cat. #F2878)

1. Preparation of inner and outer solutions (for cell-free reaction inside)

Mix the components of the cell-free reaction mixture and sucrose on ice. Also, prepare the outer solution with adjusted osmotic pressure with glucose.

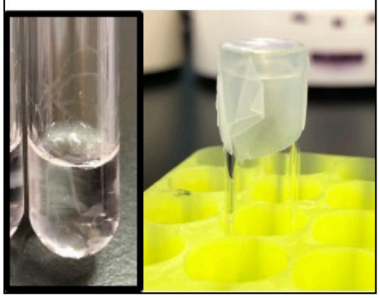
2. Lipid film preparation - POPC: POPG (70:30 mol%)

Put 70 µl of 100 mM POPC and 30 µl of 100 mM POPG into a glass test tube. Dry it up by a flow of N_2-gas with vortex.



3. Lipid-oil preparation - I

Immediately, add 500 µl of mineral oil to the tube, lid it by Parafilm, and then vortex 20 sec at maximum speed.



4. Lipid-oil preparation - II

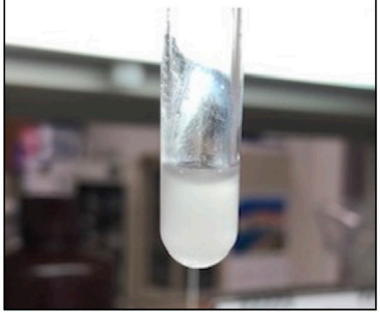
Heat the lipid-oil at 70 °C for 1min by heat-bath (fill each well with H_2O and preheat before use), then vortex 30s immediately.

5. Lipid-oil preparation - III

Heat the lipid-oil again at 70 °C for 1min, then vortex it until the sample reaches room temperature. (It takes 2-3 min)

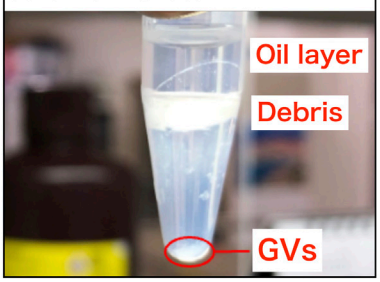
6. Droplet preparation

Immediately, add 20 µl of the inner solution to the tube, then vortex 30 sec.



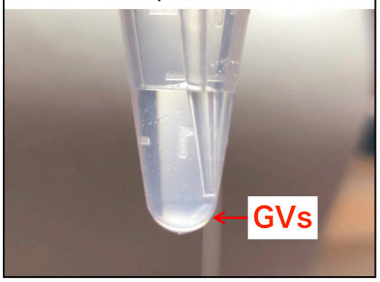
7. GVs formation

Put 300 µl of the outer solution into a microtube, then overlay the droplet solution on it. Centrifuge at 10,000 x G for 5 min, at 4 °C. A slight GV pellet appears at the bottom of the tube.



8. GVs collection

Remove the upper oil layer and debris at the middle layer using a micropipette, carefully and completely. Then, dip a fresh tip to the GVs pellet and slowly collect 20-40 µl of the solution.



9. GVs observation

Transfer the GVs solution into a fresh microtube. Check the quality and population of GVs by microscope.

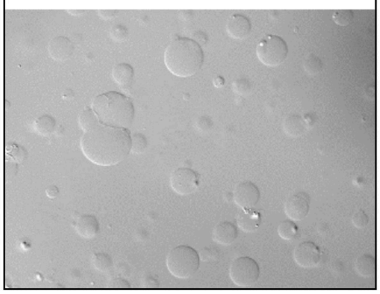


FIGURE 2 | Stepwise protocol for the preparation of artificial cells. The details of the inner and outer solutions are provided in **Tables 2**, **3**, respectively.

STEPWISE PROCEDURES

1. Inner solution preparation (5 min)		
1.1	Prepare 20 μL of the inner solution containing 200 mM sucrose (e.g., a reaction mixture of a cell-free system containing the genes of interest and sucrose) (**Figure 2.1**)	When using Ficoll PM70, mix the prepared inner solution with 12% (w/v) Ficoll PM70 at this step
1.2	Keep on ice until the lipid-oil is ready	Keep at room temperature (r.t.) if the sample should avoid low temperature
2. Outer solution preparation (2 min)		
2.1	Prepare 500 μL of the outer solution containing 200 mM glucose in a 1.5 ml microtube	The prepared outer solution can be stored in a –20°C freezer for at least 1 month
2.2	Keep on ice	Keep at r.t. if the internal solution is r.t
3. Lipid film preparation (5 min)		
3.1	Transfer 100 μL of 100 mM phospholipid solution into a glass tube which was washed with pure $CHCl_3$ in advance (e.g., for making POPC: POPG (70:30 mol%) membrane, mix 70 μL of 100 mM POPC and 30 μL of 100 mM POPG)	An inactive gas such as N_2 or argon should be filled in the bottle of lipid powder or solution before restoring in a freezer **CAUTION**: $CHCl_3$ is toxic for the respiratory tract and skin. Work within a hood wearing gloves and glasses
3.2	Dry up the solvent by gently flowing nitrogen gas from the top of the tube with vortex (**Figure 2.2**)	Dried lipids should be used immediately in the next step because it absorbs moisture easily **CAUTION**: The vapor of $CHCl_3$ is toxic for the respiratory tract and eyes. Work within a hood wearing gloves and glasses
3.3	(Option) Completely dry up the solvent and remove the moisture under low pressure in a desiccator for overnight (or more) while avoiding light	If a fluorescent-labeled lipid is used, the glass tubes of the lipid film should be covered with aluminum foil to avoid light while drying up
4. Lipid-oil preparation (8 min)		
4.1	Add 500 μL of mineral oil to the lipid film and vortex vigorously for 20 s (**Figure 2.3**)	
4.2	Heat at 70°C for 1 min, then vortex for 30 s immediately (**Figure 2.4**)	**CAUTION:** Beware of burns when using an incubator with a hot temperature
4.3	Heat at 70°C for 1 min again, then vortex immediately until the oil has cooled down to room temperature (**Figure 2.5**)	If the solution becomes turbid, repeat heating again
5. GV formation (8 min)		
5.1	Add 20 μL of the prepared inner solution to the lipid-oil	
5.2	Vortex for 30 s until W/O droplets are formed homogeneously (**Figure 2.6**)	The W/O droplets should be formed until they are fully emulsified
5.3	Transfer all W/O droplets onto the outer solution of step 2.1	
5.4	Centrifuge at 10,000 × *g* for 5 min at 4°C	When the sample need to avoid low temperature, centrifuge at a moderate temperature (15–25°C)
6. GV collection (2 min)		
6.1	After the centrifugation, remove the upper oil layer and the debris completely by pipetting (**Figure 2.7**)	Chang the pipette tips frequently while removing the oil and debris. Residual oil may break the formed GVs during the collection from the bottom of the tube
6.2	Dip a fresh tip down to the bottom of the tube and slowly collect 20–40 μL of the GVs pellet (**Figure 2.8**)	
6.3	Transfer into a new tube	
6.4	Observe the formation of GVs by microscopy (**Figure 2.9**)	1 or 2 μL of the sample is enough for the microscopy observation
7. Washing the GVs (8 min) (if necessary)		
7.1	Add 300–500 μL of the outer solution to the collected GVs and mix well	
7.2	Centrifuge at 10,000 × *g* for 5 min at 4°C	When the sample need to avoid low temperature, centrifuge at a moderate temperature (15–25°C)
7.3	Dip a fresh tip down to the bottom of the tube and slowly collect 20–40 μL of the solution	
7.4	Transfer into a new tube	

Reagent Preparation

- Phospholipids (e.g., POPC or POPG) are dissolved in $CHCl_3$ at 100 mM in a glass vial and briefly processed with bath sonication.
- Sucrose and glucose solutions are prepared at 2 M with MilliQ water.
- Sol. I (Buffer) of PURE*frex*2.0 is preheated at 37°C for 5 min before mixing and placed on ice until use, as per the manufacturer's guidelines.
- Sol. II (enzymes) and III (ribosome) of PURE*frex*2.0 are placed on ice until use, as per the manufacturer's guidelines.
- Genes of interest are used as plasmids or linear DNAs.

Equipment

- Aluminum block heating bath (e.g., TITEC Corp., cat. #DTU-2BN)
- Vortex (e.g., Scientific Industries, Inc., cat. #SI-T236)
- Nitrogen (N_2) gas
- Centrifuge (e.g., TOMY Digital Biology, cat. #MX-307)
- Ultrasonic bath (e.g., Elma Schmidbauer GmbH, cat #S15H)
- Inverted fluorescence microscope (e.g., Olympus Life Science IX73) or confocal microscopy system (e.g., Nikon AIR)

Anticipated Results

GVs encapsulating a cell-free system (PURE system) and DNA of *gfp* were prepared by following the protocol described herein (**Figure 2**), using the outer solution described in **Table 3**. The resulting GVs were incubated at 37°C for 1–3 h to perform protein synthesis inside. Before incubation, 20 ng/μL of RNaseA were added to the exterior of the GVs to eliminate possible protein synthesis outside the GVs. ATP was also added to the outside of the

Troubleshooting

Problem	Solution
(Step 1.1) A small amount of Ficoll PM70 cannot be measured accurately	Measure 240 mg of Ficoll PM70 powder and dissolve in 10 ml of Milli Q water, then aliquot 100 µL into PCR tubes. After lyophilization, store in a vacuum desiccator with low pressure. The resulting tubes containing 2.4 mg of Ficoll PM70 can be used with 20 µL of the inner solution
(Step 3.1) The lipids does not completely dissolve in $CHCl_3$ and looks turbid	Reduce the concentration of the lipid solution (e.g., 25 mM). Do not change the total amount of lipids used to prepare the lipid-oil solution
(Step 4.2 and 4.3) The lipid film does not dissolve well in oil	Raise the heating temperature or/and extend the heating time (e.g., 70–90°C for 1–5 min)
(Step 4.3) The lipid-oil solution became turbid after cooling down	An excessive cooling may result in insolubility of the lipids. Repeat step 4.3
(Step 6.2) Lipid debris appeared over the precipitated GVs layer	Remove the debris by gently pipetting or take only GVs avoiding the debris
(Step 6.4) (a) A large amount of lipid or oil debris appeared on the formed GV membrane	Moisture absorbed in the lipid-oil may reduce the quality of the GVs formed. The lipid films should be used immediately after their preparation (step 3.2), or should be stored in a vacuum desiccator until just before use (step 3.3). Additionally, mineral oil should be stored under low pressure at least overnight just before use (step 4.1)
(Step 6.4) (b) No GVs were observed or very few	Ensure that the osmotic pressure of the outer solution is equal to that of the inner solution. When an osmometer is not available, adjust the osmolarity of the outer solution equivalent to the inner one, by repeating the increase of glucose concentration by 100 mM
GVs are not stable and break within 1 h after the preparation	Adjust the osmotic pressure of the outer solution. See troubleshooting for step 6.4 (b)

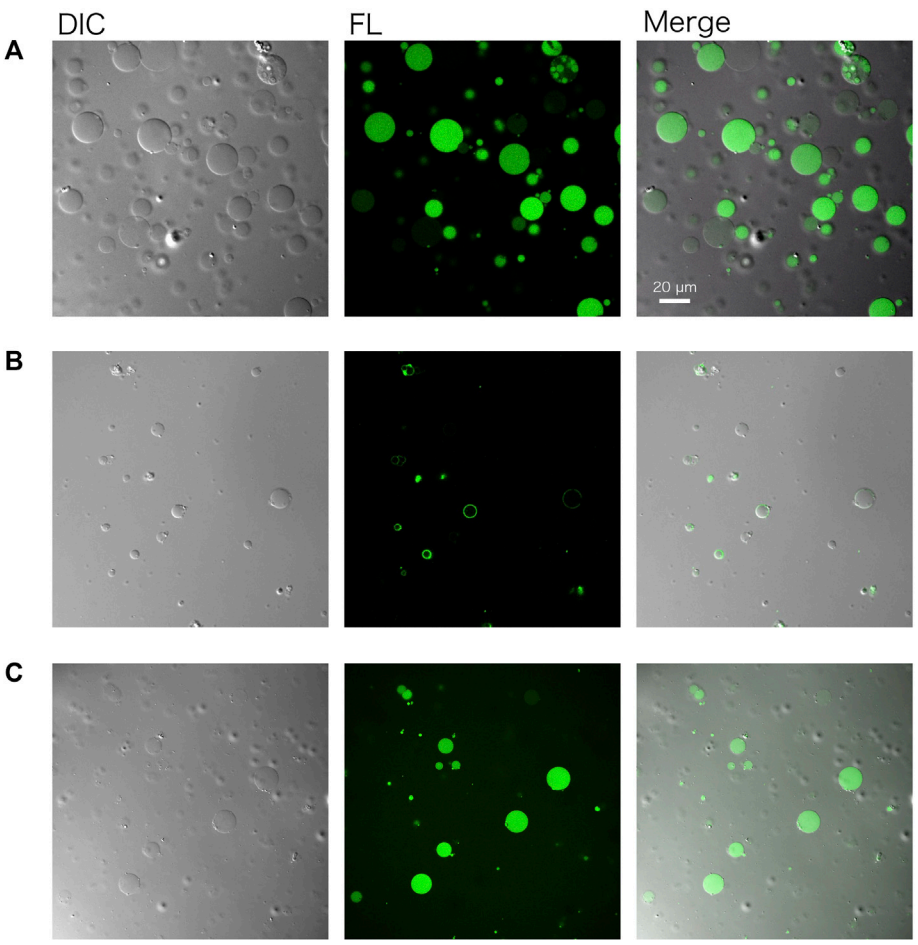


FIGURE 3 | Confocal microscopy images of artificial cells. **(A)** Green fluorescent protein (GFP) and **(B)** a fusion protein of PlsY-GFP were synthesized in GVs. **(C)** GFP was synthesized using the lyophilized and rehydrated PURE system inside GVs. In any case, the GV membrane was composed of POPC: POPG (50:50 mol%). Images were obtained using a Nikon confocal microscopy A1R system. DIC: differential interference contrast, FL: fluorescence image.

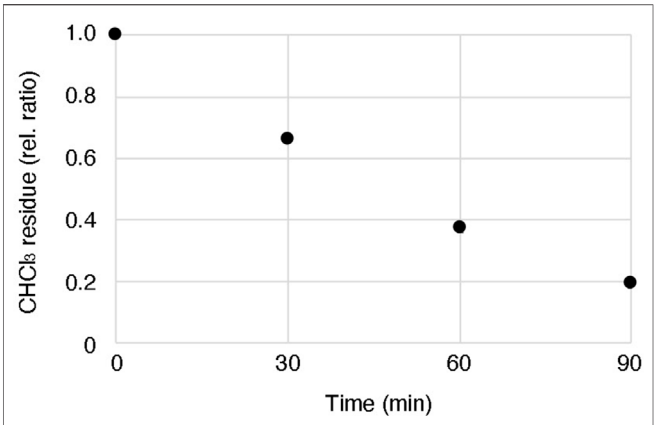


FIGURE 4 | Residual chloroform in oil after heating and stirring. 200 μL chloroform containing 100 mM lipid (POPC 50 mol%: POPG 50 mol%) was added into 1 ml mineral oil, and then the lipid-oil mixture was heated at 60°C with stirring. The weight of the mixture was measured every 30 min until 90 min. The data were based on three independent replicates of the experiments.

GVs to prevent the leakage of the encapsulated ATP. The reacted GVs were observed by a confocal microscope equipped with a differential interference contrast unit. Using a 488 nm laser, we observed the synthesized green fluorescent protein (GFP) within the GVs consist of 50 mol% POPC and 50 mol% POPG (**Figure 3A**).

A membrane protein was also synthesized using the same method by simply replacing the DNA. PlsY (glycerol-3-phosphate acyltransferase) is an integral membrane protein and was synthesized inside GVs. The gene encoding PlsY was fused to the GFP gene at the C terminus of the *plsY* DNA to visualize the product by microscopy. Note that modification of the N terminus of membrane proteins should be avoided because it may affect the co-translational spontaneous membrane localization of the proteins. **Figure 3B** shows successful protein synthesis in GVs and their localization onto the GV membrane.

Protein synthesis inside GVs was again performed using the lyophilized and rehydrated PURE system. All of the components of the reaction mixture including sucrose were mixed and lyophilized. After the rehydration of the PURE system, DNA encoding GFP was added and encapsulated inside GVs. The observed result was the same as that observed using the normal PURE system (**Figure 3C**).

DISCUSSION

Although the construction of artificial cells will provide a deepened understanding of the life system of cells, technical restrictions are preventing the development of this research field. Many protocols for GV formation have been published to date, but, in many cases, those are optimized within individual laboratories in detail and often contain unpublished tips and knowledge. This fact sometimes impedes the reproducibility of the results when other researchers work for GV formation by following the method reported in the article.

According to one of the standard methods, a mixture of lipid and oil is prepared by adding the lipids dissolved in chloroform directly to the mineral oil and evaporating the chloroform by heating and stirring (Fujii et al., 2014; Uyeda et al., 2021). However, in this method, we found that a certain amount of chloroform remains in the oil solution even after evaporating the solvent for 90 min at 60°C with stirring (**Figure 4**). Although the effect of the residual solvent on the cell-free system might be negligible, there is a possibility that the reproducibility of the obtained GV quality may not be stable depending on the degree of residual solvent. Our proposed method can circumvent this risk, as the solvent is evaporated before mixing with the oil. As the other possibility, moisture in the oil or oxidation of the lipids may reduce the quality of the formation of GVs, as suggested by Robinson's group (Moga et al., 2019) and Dekker's group (Van de Cauter et al., 2021). To avoid these problems, the dried lipid film should be stored in a desiccator with low pressure under a stable temperature.

The use of lyophilized cell-free mixture and dried lipid films not only enables further reduction of the preparation time but also has the potential to expand the versatility of artificial cell technology. This is extremely useful for working outside the laboratory and for shipping samples. Moreover, it may be applied as a ready-to-use biosensor kit for the detection of DNA sequences derived from specific viruses without the use of special equipment. For example, by combining with a signal amplification system (Sato et al., 2019), the presence of pathogenic viruses in an environment or biological samples can be detected outside the laboratory. We believe our method will be the basis for the development of artificial cell engineering.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

Experiments were designed by YS and YK. Data were generated by YS and analyzed by YS and YK. The manuscript was written by YK.

FUNDING

This work was supported by the Human Frontier Science Program (RPG0029/2020 to YK), JST PRESTO (JPMJPR18K5 to YK), JSPS KAKENHI (16H06156, 16KK0161, 16H00797, 21H05156 to YK).

ACKNOWLEDGMENTS

We thank Rumie Matsumura (JAMSTEC) and Yuuki Haruyama (Masason Foundation) for assisting the experiments, and Dr. Takashi Kanamori (GeneFrontier) for valuable discussion.

REFERENCES

Altamura, E., Albanese, P., Marotta, R., Milano, F., Fiore, M., Trotta, M., et al. (2021). Chromatophores Efficiently Promote Light-Driven ATP Synthesis and DNA Transcription inside Hybrid Multicompartment Artificial Cells. *Proc. Natl. Acad. Sci. USA* 118 (7), e2012170118. doi:10.1073/pnas.2012170118

Altamura, E., Milano, F., Tangorra, R. R., Trotta, M., Omar, O. H., Stano, P., et al. (2017). Highly Oriented Photosynthetic Reaction Centers Generate a Proton Gradient in Synthetic Protocells. *Proc. Natl. Acad. Sci. USA* 114(15), 3837–3842. doi:10.1073/pnas.1617593114

Angelova, M. I., and Dimitrov, D. S. (1986). Liposome Electroformation. *Faraday Discuss. Chem. Soc.* 81 (0), 303–311. doi:10.1039/DC9868100303

Bashirzadeh, Y., Wubshet, N., Litschel, T., Schwille, P., and Liu, A. P. (2021). Rapid Encapsulation of Reconstituted Cytoskeleton inside Giant Unilamellar Vesicles. *JoVE* 177. doi:10.3791/63332

Berhanu, S., Ueda, T., and Kuruma, Y. (2019). Artificial Photosynthetic Cell Producing Energy for Protein Synthesis. *Nat. Commun.* 10 (1), 1325. doi:10.1038/s41467-019-09147-4

Bhattacharya, A., Brea, R. J., Niederholtmeyer, H., and Devaraj, N. K. (2019). A Minimal Biochemical Route towards De Novo Formation of Synthetic Phospholipid Membranes. *Nat. Commun.* 10 (1), 300. doi:10.1038/s41467-018-08174-x

Blanken, D., Foschepoth, D., Serrão, A. C., and Danelon, C. (2020). Genetically Controlled Membrane Synthesis in Liposomes. *Nat. Commun.* 11 (1), 4317. doi:10.1038/s41467-020-17863-5

Chakraborty, S., Doktorova, M., Molugu, T. R., Heberle, F. A., Scott, H. L., Dzikovski, B., et al. (2020). How Cholesterol Stiffens Unsaturated Lipid Membranes. *Proc. Natl. Acad. Sci. USA* 117 (36), 21896–21905. doi:10.1073/pnas.2004807117

Fujii, S., Matsuura, T., Sunami, T., Nishikawa, T., Kazuta, Y., and Yomo, T. (2014). Liposome Display for *In Vitro* Selection and Evolution of Membrane Proteins. *Nat. Protoc.* 9 (7), 1578–1591. doi:10.1038/nprot.2014.107

Fujiwara, K., and Yanagisawa, M. (2014). Generation of Giant Unilamellar Liposomes Containing Biomacromolecules at Physiological Intracellular Concentrations Using Hypertonic Conditions. *ACS Synth. Biol.* 3 (12), 870–874. doi:10.1021/sb4001917

Furusato, T., Horie, F., Matsubayashi, H. T., Amikura, K., Kuruma, Y., and Ueda, T. (2018). De Novo Synthesis of Basal Bacterial Cell Division Proteins FtsZ, FtsA, and ZipA inside Giant Vesicles. *ACS Synth. Biol.* 7 (4), 953–961. doi:10.1021/acssynbio.7b00350

Gracià, R. S., Bezlyepkina, N., Knorr, R. L., Lipowsky, R., and Dimova, R. (2010). Effect of Cholesterol on the Rigidity of Saturated and Unsaturated Membranes: Fluctuation and Electrodeformation Analysis of Giant Vesicles. *Soft Matter* 6 (7), 1472–1482. doi:10.1039/B920629A

Hunt, J. P., Yang, S. O., Wilding, K. M., and Bundy, B. C. (2017). The Growing Impact of Lyophilized Cell-free Protein Expression Systems. *Bioengineered* 8 (4), 325–330. doi:10.1080/21655979.2016.1241925

Jimbo, T., Sakuma, Y., Urakami, N., Ziherl, P., and Imai, M. (2016). Role of Inverse-Cone-Shape Lipids in Temperature-Controlled Self-Reproduction of Binary Vesicles. *Biophysical J.* 110 (7), 1551–1562. doi:10.1016/j.bpj.2016.02.028

Kamiya, K. (2020). Development of Artificial Cell Models Using Microfluidic Technology and Synthetic Biology. *Micromachines* 11 (6), 559. doi:10.3390/mi11060559

Kurihara, K., Tamura, M., Shohda, K.-i., Toyota, T., Suzuki, K., and Sugawara, T. (2011). Self-reproduction of Supramolecular Giant Vesicles Combined with the Amplification of Encapsulated DNA. *Nat. Chem* 3 (10), 775–781. doi:10.1038/nchem.1127

Lee, K. Y., Park, S.-J., Lee, K. A., Kim, S.-H., Kim, H., Meroz, Y., et al. (2018). Photosynthetic Artificial Organelles Sustain and Control ATP-dependent Reactions in a Protocellular System. *Nat. Biotechnol.* 36 (6), 530–535. doi:10.1038/nbt.4140

Lee, M. S., Raig, R. M., Gupta, M. K., and Lux, M. W. (2020). Lyophilized Cell-free Systems Display Tolerance to Organic Solvent Exposure. *ACS Synth. Biol.* 9 (8), 1951–1957. doi:10.1021/acssynbio.0c00267

Litschel, T., Kelley, C. F., Holz, D., Adeli Koudehi, M., Vogel, S. K., Burbaum, L., et al. (2021). Reconstitution of Contractile Actomyosin Rings in Vesicles. *Nat. Commun.* 12 (1), 2254. doi:10.1038/s41467-021-22422-7

Litschel, T., Ramm, B., Maas, R., Heymann, M., and Schwille, P. (2018). Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations. *Angew. Chem. Int. Ed.* 57 (50), 16286–16290. doi:10.1002/anie.201808750

Lowe, L. A., Loo, D. W. K., and Wang, A. (2022). Methods for Forming Giant Unilamellar Fatty Acid Vesicles. *Methods Mol. Biol.* 2402, 1–12. doi:10.1007/978-1-0716-1843-1_1

Lu, Y., Allegri, G., and Huskens, J. (2021). Vesicle-based Artificial Cells: Materials, Construction Methods and Applications. *Mater. Horiz.* 9, 892–907. doi:10.1039/d1mh01431e

Méléard, P., Bagatolli, L. A., and Pott, T. (2009). Giant Unilamellar Vesicle Electroformation from Lipid Mixtures to Native Membranes under Physiological Conditions. *Methods Enzymol.* 465, 161–176. doi:10.1016/s0076-6879(09)65009-6

Moga, A., Yandrapalli, N., Dimova, R., and Robinson, T. (2019). Optimization of the Inverted Emulsion Method for High-Yield Production of Biomimetic Giant Unilamellar Vesicles. *Chembiochem* 20 (20), 2674–2682. doi:10.1002/cbic.201900529

Morita, M., Katoh, K., and Noda, N. (2018). Direct Observation of Bacterial Growth in Giant Unilamellar Vesicles: A Novel Tool for Bacterial Cultures. *ChemistryOpen* 7 (11), 845–849. doi:10.1002/open.201800126

Nakamura, S., Suzuki, S., Saito, H., and Nishiyama, K.-i. (2018). Cholesterol Blocks Spontaneous Insertion of Membrane Proteins into Liposomes of Phosphatidylcholine. *J. Biochem.* 163 (4), 313–319. doi:10.1093/jb/mvx083

Nishiyama, K.-i., Ikegami, A., Moser, M., Schiltz, E., Tokuda, H., and Müller, M. (2006). A Derivative of Lipid A Is Involved in Signal Recognition particle/SecYEG-dependent and -independent Membrane Integrations. *J. Biol. Chem.* 281 (47), 35667–35676. doi:10.1074/jbc.M608228200

Noireaux, V., and Libchaber, A. (2004). A Vesicle Bioreactor as a Step toward an Artificial Cell Assembly. *Proc. Natl. Acad. Sci. U.S.A.* 101 (51), 17669–17674. doi:10.1073/pnas.0408236101

Noireaux, V., and Liu, A. P. (2020). The New Age of Cell-free Biology. *Annu. Rev. Biomed. Eng.* 22, 51–77. doi:10.1146/annurev-bioeng-092019-111110

Oberholzer, T., Albrizio, M., and Luisi, P. L. (1995). Polymerase Chain Reaction in Liposomes. *Chem. Biol.* 2 (10), 677–682. doi:10.1016/1074-5521(95)90031-4

Oda, A., Watanabe, C., Aoki, N., and Yanagisawa, M. (2020). Liposomal Adhesion via Electrostatic Interactions and Osmotic Deflation Increase Membrane Tension and Lipid Diffusion Coefficient. *Soft Matter* 16 (18), 4549–4554. doi:10.1039/d0sm00416b

Pautot, S., Frisken, B. J., and Weitz, D. A. (2003). Engineering Asymmetric Vesicles. *Proc. Natl. Acad. Sci.* 100(19), 10718–10721. doi:10.1073/pnas.1931005100

Pöyry, S., and Vattulainen, I. (2016). Role of Charged Lipids in Membrane Structures - Insight Given by Simulations. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1858 (10), 2322–2333. doi:10.1016/j.bbamem.2016.03.016

Reeves, J. P., and Dowben, R. M. (1969). Formation and Properties of Thin-Walled Phospholipid Vesicles. *J. Cell. Physiol.* 73 (1), 49–60. doi:10.1002/jcp.1040730108

Robinson, T. (2019). Microfluidic Handling and Analysis of Giant Vesicles for Use as Artificial Cells: A Review. *Adv. Biosys.* 3 (6), 1800318. doi:10.1002/adbi.201800318

Sato, Y., Komiya, K., Kawamata, I., Murata, S., and Nomura, S.-i. M. (2019). Isothermal Amplification of Specific DNA Molecules inside Giant Unilamellar Vesicles. *Chem. Commun.* 55 (62), 9084–9087. doi:10.1039/C9CC03277K

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* 19 (8), 751–755. doi:10.1038/90802

Toparlak, Ö. D., Zasso, J., Bridi, S., Serra, M. D., Macchi, P., Conti, L., et al. (2020). Artificial Cells Drive Neural Differentiation. *Sci. Adv.* 6(38), eabb4920. doi:10.1126/sciadv.abb4920

Tsumoto, K., Matsuo, H., Tomita, M., and Yoshimura, T. (2009). Efficient Formation of Giant Liposomes through the Gentle Hydration of Phosphatidylcholine Films Doped with Sugar. *Colloids Surf. B: Biointerfaces* 68 (1), 98–105. doi:10.1016/j.colsurfb.2008.09.023

Uyeda, A., Reyes, S. G., Kanamori, T., and Matsuura, T. (2022). Identification of Conditions for Efficient Cell-Sized Liposome Preparation Using Commercially Available Reconstituted *In Vitro* Transcription-Translation System. *J. Biosci. Bioeng.* 133 (2), 181–186. doi:10.1016/j.jbiosc.2021.10.008

Van de Cauter, L., Fanalista, F., van Buren, L., De Franceschi, N., Godino, E., Bouw, S., et al. (2021). Optimized cDICE for Efficient Reconstitution of Biological Systems in Giant Unilamellar Vesicles. *ACS Synth. Biol.* 10 (7), 1690–1702. doi:10.1021/acssynbio.1c00068

Walde, P., Cosentino, K., Engel, H., and Stano, P. (2010). Giant Vesicles: Preparations and Applications. *Chem. Eur. J. Chem. Bio.* 11 (7), 848–865. doi:10.1002/cbic.201000010

Yanagisawa, M., Iwamoto, M., Kato, A., Yoshikawa, K., and Oiki, S. (2011). Oriented Reconstitution of a Membrane Protein in a Giant Unilamellar Vesicle: Experimental Verification with the Potassium Channel KcsA. *J. Am. Chem. Soc.* 133 (30), 11774–11779. doi:10.1021/ja2040859

Yang, J., Cui, Y., Cao, Z., Ma, S., and Lu, Y. (2021). Strategy Exploration for Developing Robust Lyophilized Cell-free Systems. *Biotechnol. Notes* 2, 44–50. doi:10.1016/j.biotno.2021.08.004

Yoshida, A., Kohyama, S., Fujiwara, K., Nishikawa, S., and Doi, N. (2019). Regulation of Spatiotemporal Patterning in Artificial Cells by a Defined Protein Expression System. *Chem. Sci.* 10 (48), 11064–11072. doi:10.1039/c9sc02441g

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 | Frontiers in Bioengineering and Biotechnology

TYPE General Commentary
PUBLISHED 14 October 2022
DOI 10.3389/fbioe.2022.1037809

OPEN ACCESS

EDITED BY
Simon J. Moore,
University of Kent, United Kingdom

REVIEWED BY
Aleksander Czogalla,
University of Wrocław, Poland

*CORRESPONDENCE
Pasquale Stano,
pasquale.stano@unisalento.it

SPECIALTY SECTION
This article was submitted to Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and Biotechnology

RECEIVED 06 September 2022
ACCEPTED 26 September 2022
PUBLISHED 14 October 2022

CITATION
Stano P (2022), Commentary: Rapid and facile preparation of giant vesicles by the droplet transfer method for artificial cell construction.
Front. Bioeng. Biotechnol. 10:1037809.
doi: 10.3389/fbioe.2022.1037809

Commentary: Rapid and facile preparation of giant vesicles by the droplet transfer method for artificial cell construction

Pasquale Stano*

Department of Biological and Environmental Sciences and Technologies (DiSTeBA), University of Salento, Lecce, Italy

KEYWORDS

synthetic cells, artificial cells, protocells, lipid vesicles, liposomes, liposome technology, droplet transfer method, inverted emulsion method

A Commentary on
Rapid and facile preparation of giant vesicles by the droplet transfer method for artificial cell construction

by Shimane, Y. and Kuruma Y. (2022). Front. Bioeng. Biotechnol. 10:873854. doi: 10.3389/fbioe.2022.873854

1 The "droplet transfer" (or "inverted emulsion" or "emulsion transfer" or "phase transfer") method

When considered within the well-established field of liposome technology, the droplet transfer method (DTm) might appear as a rather niche technique, but it has literally revolutionized experimental approaches to bottom-up Artificial Cells (ACs) (Stano, 2019). The DTm was firstly used by Vincent Noireaux and Albert Libchaber in their renowned 2004 "bioreactor" paper (Noireaux and Libchaber, 2004) and since then it has been widely applied to the construction of several types of ACs. This method is at the basis of several successful studies, which are not commented here for space limitations. A general discussion on the method can be found in Walde et al. (2010) and in Chapter 1 of "*The Giant Vesicle Book*" (Dimova et al., 2020).

The timely investigation presented in *Frontiers on Bioengineering and Biotechnology* by Shimane and Kuruma (2022), to which this commentary is dedicated, proves the importance and the utility of the DTm in order to prepare ACs based on giant liposomes (giant lipid vesicles, GVs). In particular, the Authors have illustrated in a quite detailed and practical way the typical operations required to encapsulate complex macromolecular mixtures–such as the PURE system (Shimizu et al., 2001; Shimizu et al., 2005)—inside GVs in physiological and bioactive conditions, and in high yields. The resulting ACs are then competent for *gene expression* (a key AC functionality) thanks to the presence of

about 100 different macromolecules and dozens of small MW compounds co-captured in the GV lumen. Proteins will be produced in cell-like fashion, i.e., from within.

This commentary is an opportunity to briefly add some considerations on the DTm. We will highlight its major historical developments, the pros and cons, and the still missing mechanistic details.

2 Major historical developments

Demetrios Papahadjopoulos (1934–1998), one of the fathers of liposome research, reported that attempts to prepare liposomes from water-in-oil emulsion droplets date back to the 1970s (Träuble and Grell, 1971; Szoka and Papahadjopoulos, 1980). The aim was the production of conventional small or large unilamellar vesicles (SUVs or LUVs) with an *asymmetric membrane*. This is in principle possible when the two lipid leaflets of a vesicle membrane are assembled sequentially. Membrane asymmetry was also the goal of more recent studies, due to the Weitz group (Pautot et al., 2003a; Pautot et al., 2003b) and to less cited, but prior, investigators (Zhang et al., 1997; Xiao et al., 1998a; Xiao et al., 1998b).

The shift of interest from LUVs to GVs, and from membrane asymmetry to high encapsulation yield coincided with the above-mentioned Noireaux and Libchaber (2004) paper. This double-leap actually made the difference, and brought the DTm to the attention of the AC community. Indeed the DTm, when the *isotonic density gradient* strategy is employed[1], allows the facile preparation of solute-filled GVs, solving "once for all" (and simultaneously) long standing issues such as efficient macromolecule (co)encapsulation, employment of physiological conditions (concentrated/salty buffers), reduction of the volume of precious solution to be entrapped, reduction of the preparation time, and avoidance of difficult to handle equipment. Moreover, in most cases the resulting GVs appear to be gracefully unilamellar (GUVs) (Chiba et al., 2014).

Subsequent investigations have refined, optimized (Fujii et al., 2014; Moga et al., 2019; Uyeda et al., 2022), scaled-up (Rampioni et al., 2018) and extended the DTm, including the structural analysis of the resulting GVs by flow cytometry (Nishimura et al., 2009), its implementation in microfluidic (Matosevic and Paegel, 2011) or in rotatory capillary devices (Abkarian et al., 2011), but essentially there have been no major noteworthy findings.

3 Pros and cons

The advantages of the DTm have been mentioned above (see also Figure 1A). A further merit consists in the possibility of encapsulating very large particles (e.g., nanoparticles, SUVs/LUVs, organellae, bacteria, etc.) inside GVs. This feature has lead to the easy preparation of nested multi-compartmentalized systems that mimic eukaryotic cells. A pregnant example refers to ACs endowed with natural or artificial energy-producing organellae in their lumen (Berhanu et al., 2019; Altamura et al., 2021). Another valuable feature concerns the reconstitution of membrane proteins from within, in order to obtain otherwise unfeasible physiological-like orientations (Yanagisawa et al., 2011; Altamura et al., 2017).

The most-often repeated criticism pertains to the possible residual presence of hydrocarbons in the GVs membrane. Researchers interested in studying phenomena strictly dependent on biophysical/biochemical properties of membranes are understandably worried about this issue (e.g., membrane permeability and mechanical properties, dynamics of proteins embedded into the membrane) (e.g., Campillo et al., 2013; Kamiya et al., 2016; Faizi et al., 2022). The "mineral oils" typically used in the DTm actually are cheap heterogeneous mixture of linear, branched, and possibly cyclic hydrocarbons, which can be adsorbed on or absorbed in the membrane (McIntosh et al., 1980; Richens et al., 2015). Pautot et al. (2003b) already recognized this issue, recalling that it could be alleviated by using specific hydrocarbons (e.g., squalene). Fortunately, there are types of investigations that can be safely carried out disregarding the potential presence of residual oil in the GV membrane (clearly, it depends on the aim of the experiment).

A less frequently mentioned aspect is that DTm is not the best choice in studies intended to explore the *spontaneous* formation of cell-like systems from lipids and solutes. Therefore, when ACs are intended as primitive cell models and the experimental focus is on spontaneous assembly mechanisms in an aqueous environment, other preparation methods should be considered.

4 Mechanistic details

The ever increasing number of reports that exploit the production of ACs based on the DTm contrasts with the sporadic study about its mechanistic details. Pautot *et al.* (2003ab) reported several insightful observations, but because

1 The isotonic density gradient was introduced, to the best of our knowledge, by Takagi and collaborators (Hamada et al., 2008; but see also Hadorn and Eggenberger Hotz, 2009), providing a key modification that has definitely led a reliable GV production method. It consists in the usage of isotonic sucrose and glucose solutions (e.g., 200 mM or higher), respectively, as inner and outer aqueous solutions, in order to facilitate the droplet transfer from the top oil phase to the bottom aqueous phase. High concentrations of sugars, however, can affect membrane properties (Andersen et al., 2011). Träuble and Grell (1971) employed instead CsCl gradients, but the resulting vesicles were osmotically sensitive.

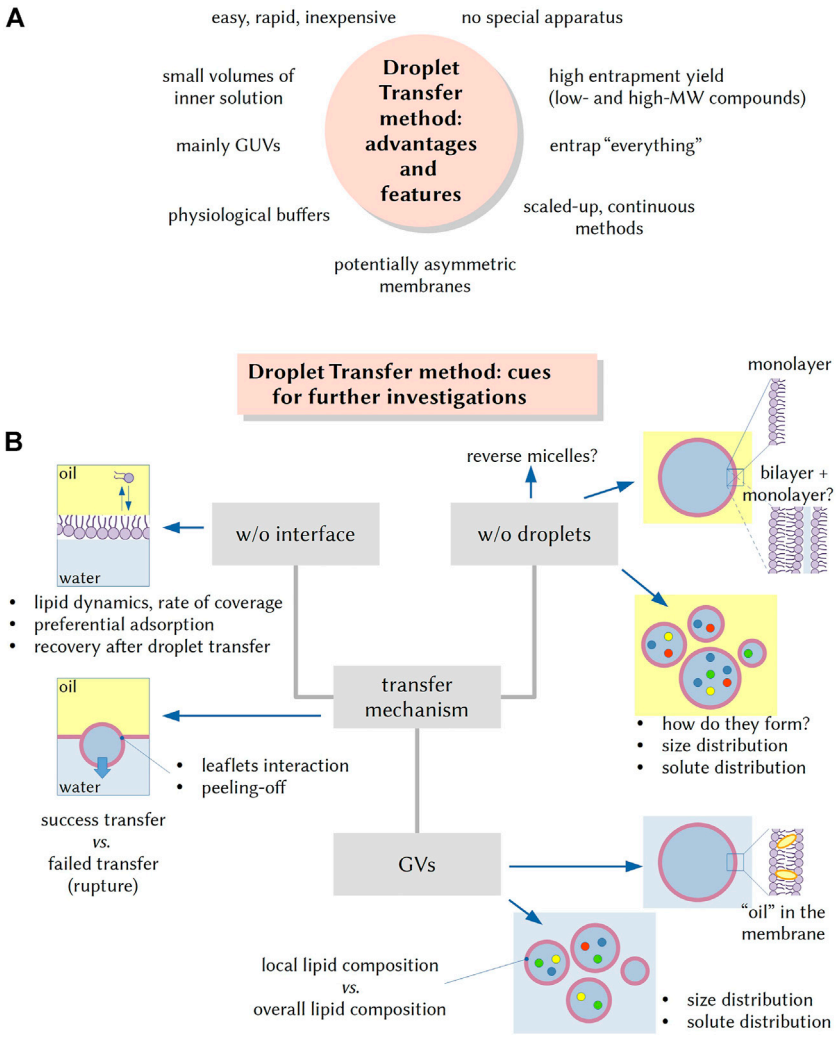


FIGURE 1
The droplet transfer method (DTm): advantageous features and cues for further investigations. **(A)** The advent of the DTm in AC research is quite attractive as it allows the formation of solute-filled GVs in a quite easy and inexpensive manner. Additional advantages are the rapidity (few minutes), no need of specific equipment, employment of small volumes (especially important when complex mixtures need to be encapsulated in the vesicle lumen), the possibility of using physiological buffers, and the possibility of entrapping very large particles. **(B)** Several experimental steps for carrying out the DTm have been recently investigated and optimized, but open questions remain, and they can be the focus of future investigations. For example, it would be important to know more about the lipid dynamics on the oil/water interface, the rate of interface coverage by lipids and the recovery of lipids in the "naked" area after a w/o droplet has been transfer. What factors affect the fraction of w/o droplets that become GVs? Similarly, the very mechanism of w/o droplet formation (*via* stochastic fragmentation and coagulation) is not well known, as well as the factors determining the droplet size and the solute distribution. Other questions can be: are droplets always covered by a lipid monolayer? What is the amount of residual "oil" in the membrane of resulting GVs? In the case of lipid mixtures, preferential localization of some lipid species on all interfaces (vs those remaining in bulk solution) can occur, and thus, do the obtained GVs have a lipid composition that mirrors the overall lipid composition? Is there any correlation between w/o droplets and GVs size and solute distributions?

their protocol is quite different from the ones commonly used today, more investigations are needed (see Figure 1B). An elegant study made with a 90° tilted microscope (in order to observe the dynamic of droplet transfer) came from the group of Yoshikawa (Ito et al., 2013).

For example, open questions concern the mechanism of formation and the structure of the w/o droplets (always surrounded by a monolayer?) and the distribution of solutes therein. The dynamics of lipids at the interface between the emulsion phase and the aqueous outer phase and the very mechanism of bilayer formation *via* adhesion of the two lipid leaflets, during the transfer, are key aspects too. Moreover, the case when lipid mixtures are used is not well studied (as well as the usage of non-phospholipid amphiphiles): does the composition of the resulting GV membranes mirrors the overall bulk composition of lipids used in the preparation?

(Some lipids could be preferentially included in the membrane).

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The Author is grateful to all students and collaborators that have contributed to the Author's research on this subject.

Conflict of interest

The author declares 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.

References

Abkarian, M., Loiseau, E., and Massiera, G. (2011). Continuous droplet interface crossing encapsulation (cDICE) for high throughput monodisperse vesicle design. *Soft Matter* 7, 4610–4614. doi:10.1039/C1SM05239J

Altamura, E., Albanese, P., Marotta, R., Milano, F., Fiore, M., Trotta, M., et al. (2021). Chromatophores efficiently promote light-driven ATP synthesis and DNA transcription inside hybrid multicompartment artificial cells. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2012170118. doi:10.1073/pnas.2012170118

Altamura, E., Milano, F., Tangorra, R. R., Trotta, M., Omar, O. H., Stano, P., et al. (2017). Highly oriented photosynthetic reaction centers generate a proton gradient in synthetic protocells. *Proc. Natl. Acad. Sci. U.S.A.* 114, 3837–3842. doi:10.1073/pnas.1617593114

Andersen, H. D., Wang, C., Arleth, L., Peters, G. H., and Westh, P. (2011). Reconciliation of opposing views on membrane-sugar interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1874–1878. doi:10.1073/pnas.1012516108

Berhanu, S., Ueda, T., and Kuruma, Y. (2019). Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun.* 10, 1325. doi:10.1038/s41467-019-09147-4

Campillo, C., Sens, P., Köster, D., Pontani, L.-L., Lévy, D., Bassereau, P., et al. (2013). Unexpected membrane dynamics unveiled by membrane nanotube extrusion. *Biophysical J.* 104, 1248–1256. doi:10.1016/j.bpj.2013.01.051

Chiba, M., Miyazaki, M., and Ishiwata, S. (2014). Quantitative analysis of the lamellarity of giant liposomes prepared by the inverted emulsion method. *Biophysical J.* 107, 346–354. doi:10.1016/j.bpj.2014.05.039

Dimova, R., Stano, P., Marques, C. M., and Walde, P. (2020). "Preparation methods for giant unilamellar vesicles," in *The giant vesicle Book*. Editors R. Dimova, and C. M. Marques (Boca Raton, FL: Taylor & Francis Group), 3–20.

Faizi, H. A., Tsui, A., Dimova, R., and Vlahovska, P. M. (2022). Bending rigidity, capacitance, and shear viscosity of giant vesicle membranes prepared by spontaneous swelling, electroformation, gel-assisted, and phase transfer methods: A comparative study. *Langmuir* 38, 10548–10557. doi:10.1021/acs.langmuir.2c01402

Fujii, S., Matsuura, T., Sunami, T., Nishikawa, T., Kazuta, Y., Yomo, T., et al. (2014). Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* 9, 1578–1591. doi:10.1038/nprot.2014.107

Hadorn, M., and Eggenberger Hotz, P. (2009). "Multivesicular assemblies as real-world testbeds for embryogenic evolutionary systems," in *Artificial life: Borrowing from Biology. Lecture notes in computer science*. Editors K. Korb, M. Randall, and T. Hendtlass (Berlin, Heidelberg: Springer), 169–178. doi:10.1007/978-3-642-10427-5_17

Hamada, T., Miura, Y., Komatsu, Y., Kishimoto, Y., Vestergaard, M., Takagi, M., et al. (2008). Construction of asymmetric cell-sized lipid vesicles from lipid-coated water-in-oil microdroplets. *J. Phys. Chem. B* 112, 14678–14681. doi:10.1021/jp807784j

Ito, H., Yamanaka, T., Kato, S., Hamada, T., Takagi, M., Ichikawa, M., et al. (2013). Dynamical formation of lipid bilayer vesicles from lipid-coated droplets across a planar monolayer at an oil/water interface. *Soft Matter* 9, 9539–9547. doi:10.1039/c3sm51766g

Kamiya, K., Kawano, R., Osaki, T., Akiyoshi, K., and Takeuchi, S. (2016). Cell-sized asymmetric lipid vesicles facilitate the investigation of asymmetric membranes. *Nat. Chem.* 8, 881–889. doi:10.1038/nchem.2537

Matosevic, S., and Paegel, B. M. (2011). Stepwise synthesis of giant unilamellar vesicles on a microfluidic assembly line. *J. Am. Chem. Soc.* 133, 2798–2800. doi:10.1021/ja109137s

McIntosh, T. J., Simon, S. A., and MacDonald, R. C. (1980). The organization of n-alkanes in lipid bilayers. *Biochimica Biophysica Acta (BBA) - Biomembr.* 597, 445–463. doi:10.1016/0005-2736(80)90219-9

Moga, A., Yandrapalli, N., Dimova, R., and Robinson, T. (2019). Optimization of the inverted emulsion method for high-yield production of biomimetic giant unilamellar vesicles. *ChemBioChem* 20, 2674–2682. doi:10.1002/cbic.201900529

Nishimura, K., Hosoi, T., Sunami, T., Toyota, T., Fujinami, M., Oguma, K., et al. (2009). Population analysis of structural properties of giant liposomes by flow cytometry. *Langmuir* 25, 10439–10443. doi:10.1021/la902237y

Noireaux, V., and Libchaber, A. (2004). A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17669–17674. doi:10.1073/pnas.0408236101

Pautot, S., Frisken, B. J., and Weitz, D. A. (2003a). Engineering asymmetric vesicles. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10718–10721. doi:10.1073/pnas.1931005100

Pautot, S., Frisken, B. J., and Weitz, D. A. (2003b). Production of unilamellar vesicles using an inverted emulsion. *Langmuir* 19, 2870–2879. doi:10.1021/la026100v

Rampioni, G., D'Angelo, F., Messina, M., Zennaro, A., Kuruma, Y., Tofani, D., et al. (2018). Synthetic cells produce a quorum sensing chemical signal perceived byPseudomonas aeruginosa. *Chem. Commun.* 54, 2090–2093. doi:10.1039/C7CC09678J

Richens, J. L., Lane, J. S., Mather, M. L., and O'Shea, P. (2015). The interactions of squalene, alkanes and other mineral oils with model membranes; effects on membrane heterogeneity and function. *J. Colloid Interface Sci.* 457, 225–231. doi:10.1016/j.jcis.2015.06.052

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802

Shimizu, Y., Kanamori, T., and Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods* 36, 299–304. doi:10.1016/j.ymeth.2005.04.006

Shimane, Y., and Kuruma, Y. (2022). Rapid and Facile Preparation of Giant Vesicles by the Droplet Transfer Method for Artificial Cell Construction. *Front. in Bioeng. Biotechnol.* 10, 873854

Stano, P. (2019). Gene expression inside liposomes: From early studies to current protocols. *Chem. Eur. J.* 25, 7798–7814. doi:10.1002/chem.201806445

Szoka, F., and Papahadjopoulos, D. (1980). Comparative properties and methods of preparation of lipid vesicles (liposomes). *Annu. Rev. Biophys. Bioeng.* 9, 467–508. doi:10.1146/annurev.bb.09.060180.002343

Träuble, H., and Grell, E. (1971). Carriers and specificity in membranes. IV. Model vesicles and membranes. The formation of asymmetrical spherical lecithin vesicles. *Neurosci. Res. Program Bull.* 9, 373–380.

Uyeda, A., Reyes, S. G., Kanamori, T., and Matsuura, T. (2022). Identification of conditions for efficient cell-sized liposome preparation using commercially available reconstituted *in vitro* transcription-translation system. *J. Biosci. Bioeng.* 133, 181–186. doi:10.1016/j.jbiosc.2021.10.008

Walde, P., Cosentino, K., Engel, H., and Stano, P. (2010). Giant vesicles: Preparations and applications. *Chem. Eur. J. Chem. Bio.* 11, 848–865. doi:10.1002/cbic.201000010

Xiao, Z., Huang, N., Xu, M., Lu, Z., and Wei, Y. (1998a). Novel preparation of asymmetric liposomes with inner and outer layer of different materials. *Chem. Lett.* 27, 225–226. doi:10.1246/cl.1998.225

Xiao, Z., Xu, M., Li, M., Lu, Z., and Wei, Y. (1998b). Preparation of asymmetric bilayer-vesicles with inner and outer monolayers composed of different amphiphilic molecules. *Supramol. Sci.* 5, 619–622. doi:10.1016/S0968-5677(98)00088-1

Yanagisawa, M., Iwamoto, M., Kato, A., Yoshikawa, K., and Oiki, S. (2011). Oriented reconstitution of a membrane protein in a giant unilamellar vesicle: Experimental verification with the potassium channel KcsA. *J. Am. Chem. Soc.* 133, 11774–11779. doi:10.1021/ja2040859

Zhang, L., Hu, J., and Lu, Z. (1997). Preparation of liposomes with a controlled assembly procedure. *J. Colloid Interface Sci.* 190, 76–80. doi:10.1006/jcis.1997.4820

frontiers | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fbioe.2022.896751

The Potential of Eukaryotic Cell-Free Systems as a Rapid Response to Novel Zoonotic Pathogens: Analysis of SARS-CoV-2 Viral Proteins

Franziska Ramm[1,2], *Srujan K. Dondapati*[1], *Hoai Anh Trinh*[1,3], *Dana Wenzel*[1], *Ruben M. Walter*[1,3], *Anne Zemella*[1] *and Stefan Kubick*[1,2,4]*

[1]Fraunhofer Institute for Cell Therapy and Immunology (IZI), Branch Bioanalytics and Bioprocesses (IZI-BB), Potsdam, Germany, [2]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany, [3]Department of Applied Biochemistry, Institute of Biotechnology, Technical University Berlin, Berlin, Germany, [4]Faculty of Health Sciences, Joint Faculty of the Brandenburg University of Technology Cottbus–Senftenberg, The Brandenburg Medical School Theodor Fontane, The University of Potsdam, Potsdam, Germany

The ongoing pandemic caused by the novel coronavirus (SARS-CoV-2) has led to more than 445 million infections and the underlying disease, COVID-19, resulted in more than 6 million deaths worldwide. The scientific world is already predicting future zoonotic diseases. Hence, rapid response systems are needed to tackle future epidemics and pandemics. Here, we present the use of eukaryotic cell-free systems for the rapid response to novel zoonotic diseases represented by SARS-CoV-2. Non-structural, structural and accessory proteins encoded by SARS-CoV-2 were synthesized by cell-free protein synthesis in a fast and efficient manner. The inhibitory effect of the non-structural protein 1 on protein synthesis could be shown *in vitro*. Structural proteins were quantitatively detected by commercial antibodies, therefore facilitating cell-free systems for the validation of available antibodies. The cytotoxic envelope protein was characterized in electrophysiological planar lipid bilayer measurements. Hence, our study demonstrates the potential of eukaryotic cell-free systems as a rapid response mechanism for the synthesis, functional characterization and antibody validation against a viral pathogen.

Keywords: eukaryotic cell-free systems, SARS-CoV-2, viral proteins, rapid response, protein analytics, planar lipid bilayer measurements

OPEN ACCESS

Edited by:
Simon J. Moore,
University of Kent, United Kingdom

Reviewed by:
Karen Marie Polizzi,
Imperial College London,
United Kingdom
Dong-Myung Kim,
Chungnam National University,
South Korea

***Correspondence:**
Stefan Kubick
stefan.kubick@izi-bb.fraunhofer.de

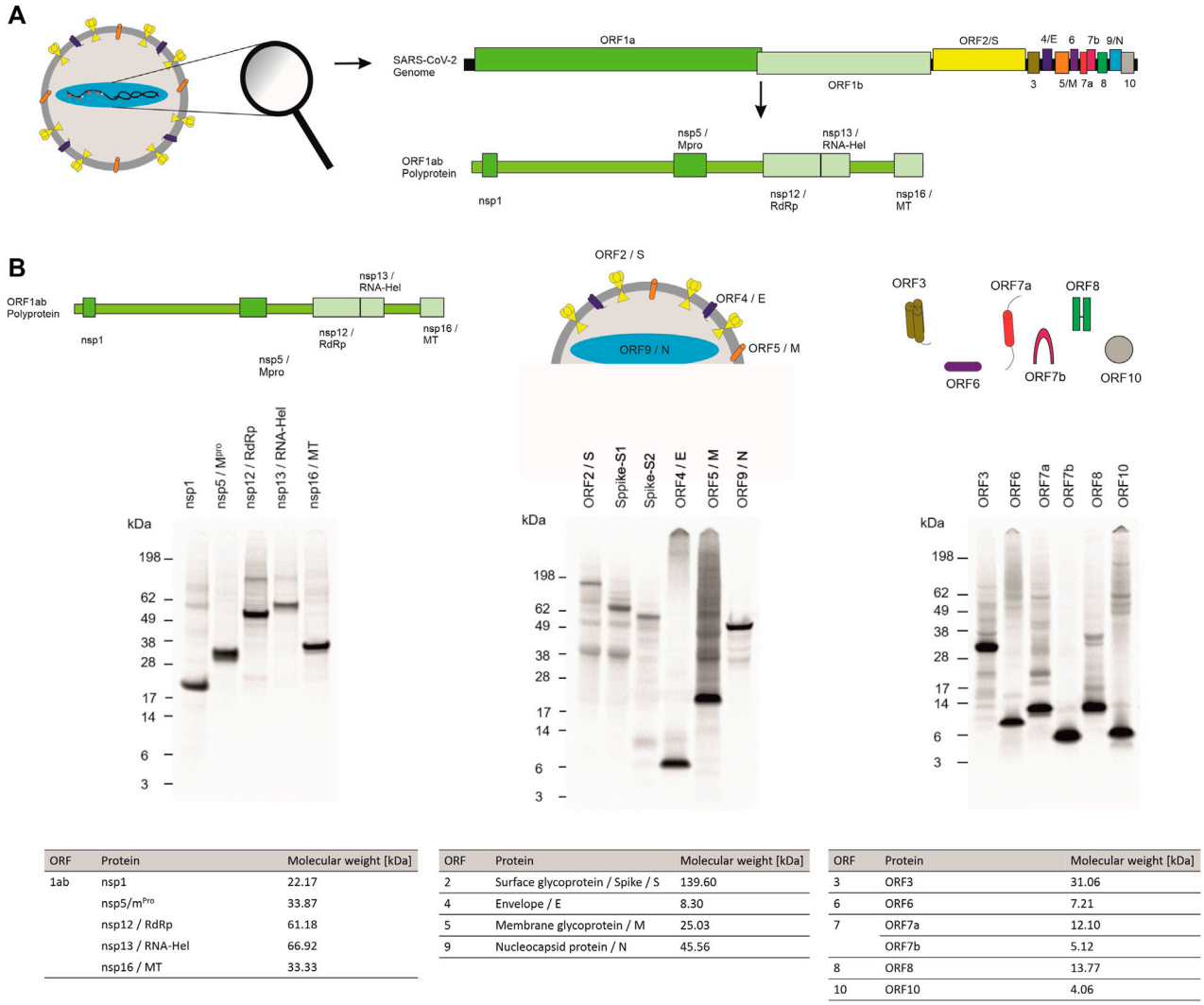
Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *15 March 2022*
Accepted: *01 April 2022*
Published: *19 April 2022*

Citation:
Ramm F, Dondapati SK, Trinh HA, Wenzel D, Walter RM, Zemella A and Kubick S (2022) The Potential of Eukaryotic Cell-Free Systems as a Rapid Response to Novel Zoonotic Pathogens: Analysis of SARS-CoV-2 Viral Proteins. Front. Bioeng. Biotechnol. 10:896751. doi: 10.3389/fbioe.2022.896751

INTRODUCTION

Infectious diseases that are transferred from an animal to a human being, so-called zoonoses, can lead to devastating health issues around the world as can be seen from the example of the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). The close interaction with animals, such as in agriculture and with domesticated animals (pets), the increasing consumption of different meats as well as the intrusion of humans into the natural habitat of animals, causes a high risk for the development of novel zoonoses that might lead to short lived disease outbreaks, epidemics or even pandemics. Standard techniques such as the detection of viral antigens by polymerase chain reactions (PCR) as well as rapid antigen tests accelerate prompt responses such as quarantines and shutting down social contacts. Unfortunately, some pathogens, such as airborne viruses, are persistent and have to be counteracted with vaccines and therapeutics. Consequently, the thorough characterization of the virus itself and its mode of action, including the viral assembly, cell attack, pathogenesis of the



ORF	Protein	Molecular weight [kDa]
1ab	nsp1	22.17
	nsp5/m^{Pro}	33.87
	nsp12 / RdRp	61.18
	nsp13 / RNA-Hel	66.92
	nsp16 / MT	33.33

ORF	Protein	Molecular weight [kDa]
2	Surface glycoprotein / Spike / S	139.60
4	Envelope / E	8.30
5	Membrane glycoprotein / M	25.03
9	Nucleocapsid protein / N	45.56

ORF	Protein	Molecular weight [kDa]
3	ORF3	31.06
6	ORF6	7.21
7	ORF7a	12.10
	ORF7b	5.12
8	ORF8	13.77
10	ORF10	4.06

FIGURE 1 | Cell-free synthesis of SARS-CoV-2 proteins in a CHO cell-free system. **(A)** Schematic figure showing the SARS-CoV-2 virus particle and its structure as well as the coding sequences of the genome. **(B)** Qualitative analysis of cell-free synthesized SARS-CoV-2 proteins by autoradiography. Autoradiograph showing ^{14}C-leucine labeled proteins synthesized in a CHO system from the translation mixture (TM).

underlying disease and characterization of the viral proteins are necessary to tackle these tasks and to identify efficient drugs. It is essential to understand the different modes of action of the individual proteins as each particular component plays a specific role in the viral assembly, host infection and immune invasion as well as viral replication. The novel coronavirus expresses 10 different open reading frames (ORFs) encoding the ORF1ab polyprotein, eight single ORFs corresponding to eight single proteins as well as ORF7 which can be further separated into ORF7a and b which encode two different proteins (**Figure 1A**) (Kim et al., 2020; Wang et al., 2020; Yoshimoto, 2020). The ORF1ab polyprotein encodes 16 non-structural proteins (nsp) including proteases (nsp3 and nsp5), a RNA-helicase (nsp13) and a RNA-dependent RNA-polymerase (RdRp, nsp12) that are mainly responsible for viral replication (Gao et al., 2020; Jang et al., 2020; Shin et al., 2020; Shu et al., 2020; Yoshimoto, 2020; Zhang et al., 2020). This polyprotein is encoded by the two ORFs ORF1a and ORF1b. A frameshift before the stop codon in ORF1a facilitates the translation to be continued to ORF1b and therefore resulting in the polyprotein ORF1ab (Kim et al., 2020). The other ORFs can be divided into two major classes, namely the structural proteins and the accessory proteins. Structural proteins are well known and well characterized as these proteins assemble to the viral capsid. The core structure of SARS-CoV-2 virus is maintained by the ORF2 surface glycoprotein (S) otherwise known as the Spike protein, the ORF4 envelope protein (E), the ORF5 membrane glycoprotein (M) and the ORF9 nucleocapsid protein (N) (**Figure 1A**). The accessory proteins of SARS-CoV-2 include membrane proteins like the putative ion channel encoded by ORF3 (Kern et al., 2021), the type I transmembrane protein coded by ORF7a (Rosenthal et al., 2020) as well as the integral membrane protein ORF7b

TABLE 1 | SARS-CoV-2 viral proteins and their characteristics.

ORF	Protein	Short Form	Function
1ab	Nsp1	—	Inhibition of protein translation
	Nsp5	m^{Pro}	Protease
	Nsp12	RdRp	RNA-dependent RNA polymerase
	Nsp13	RNA-Hel	RNA-helicase
	Nsp16	MT	Methyltransferase
2	Surface glycoprotein	Spike/S	Binding of host cells
3	ORF3	—	Ion channel
4	Envelope	E	Facilitates assembly
5	Membrane protein	M	Interacts with E, S and N to form stable assembly
6	ORF6	—	Immune invasion
7	ORF7a	—	Virus-host interaction, the type I transmembrane protein
	ORF7b	—	Virus-host interaction, integral membrane protein
8	ORF8	—	Immune invasion
9	Nucleocapsid protein	N	Genome packaging
10	ORF10	—	Unclear

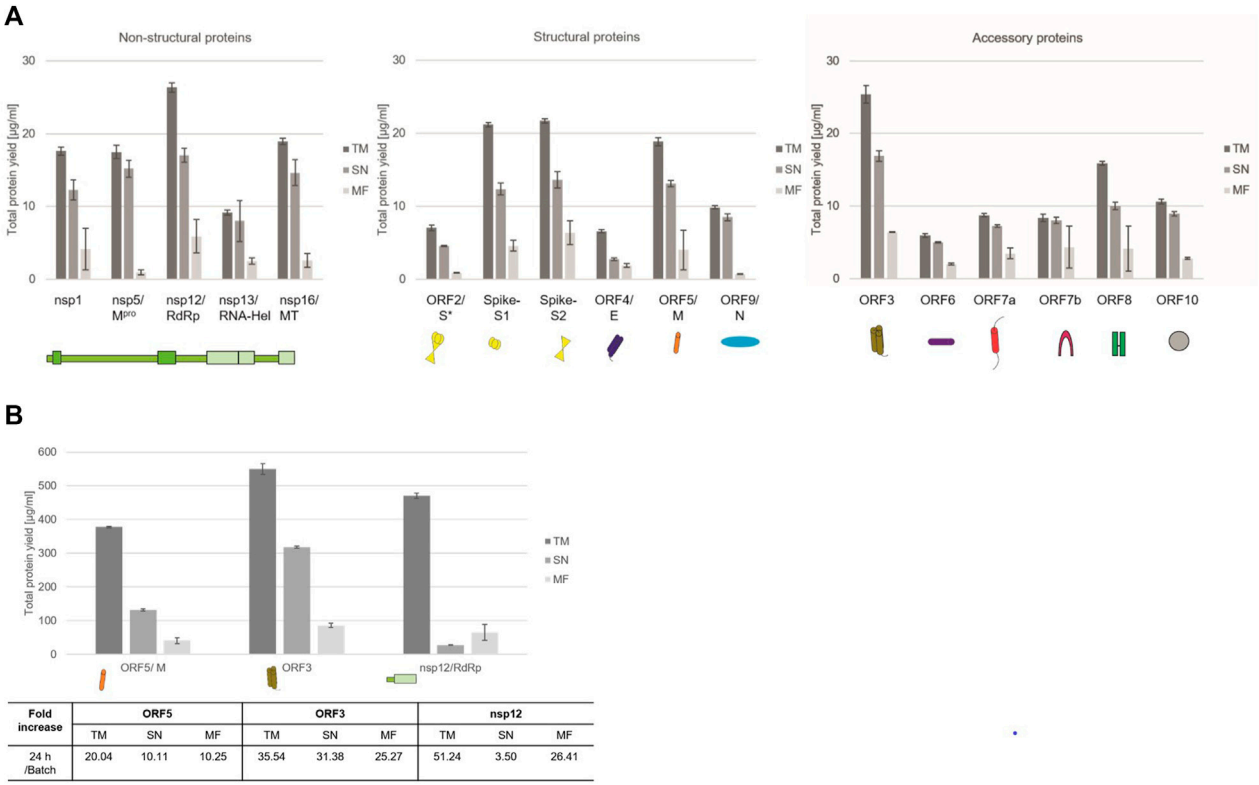
(Schaecher et al., 2008). Accessory proteins ORF6 and ORF8 are known to be involved in the host cell immune invasion and in interferon signaling (Miorin et al., 2020; Flower et al., 2021; Zhang et al., 2021). In some infections the protein encoded by ORF10 is present whilst in others this protein cannot be found which might be caused due to a read through of this ORF (Pancer et al., 2020; Hassan et al., 2021). The proteins encoded by SARS-CoV-2 and their characteristics are summarized in **Table 1**.

In order to characterize such a versatile set of proteins, a system that can produce and characterize all kinds of proteins is needed. Therefore, we present eukaryotic cell-free systems as a promising methodology for the fast and efficient synthesis and characterization of viral proteins based on the example of SARS-CoV-2 proteins. In cell-free protein synthesis a crude cell lysate rather than viable, intact cells are used, allowing the synthesis of "difficult-to-express proteins" such as membrane proteins or even cytotoxic proteins (Chalmeau et al., 2011; Orth et al., 2011; Henrich et al., 2015; Thoring et al., 2017). With the utilization of a cell lysate, the production of genetically modified organisms becomes obsolete. As a result no high laboratory safety standards are necessary for the cell-free production of toxic and viral proteins. In order to avoid cloning procedures for the generation of templates encoding viral proteins, PCR templates can be used for cell-free synthesis (Sawasaki et al., 2002). Thus, a fast screening of different mutants is possible. Another advantage of some eukaryotic cell-free systems are endogenous microsomal vesicles derived from the endoplasmic reticulum (ER) that are present in the lysate. These vesicles enable post-translational modifications (PTMs) and are a natural surrounding for membrane proteins (Brödel et al., 2014). Here, we demonstrate that the whole set of viral proteins derived from SARS-CoV-2 including membrane proteins, enzymes as well as modulatory proteins can be synthesized in a lysate based on Chinese hamster ovary (CHO) cells (Brödel et al., 2014; Thoring et al., 2017). The synthesis of functional protein was verified by cell-free synthesized nsp1 downregulating the *in vitro* synthesis of a model protein and the demonstration of cytotoxic events as well as single channel events in planar lipid bilayer measurements induced by the cell-free synthesized envelope protein. Further, the cell-free synthesized nucleocapsid protein was used to demonstrate the possibility to quantitatively evaluate the binding of commercially available antibodies. Taken together eukaryotic cell-free systems, including but not limited to the use of CHO lysate, can be applied to characterize viral proteins and might facilitate the screening of antibodies as well as pharmaceuticals and blockers against these viral proteins.

RESULTS

Viral pathogens such as SARS-CoV-2 induce cytotoxic effects often associated with severe damage to the host cell. This might be one of the major factors in the pathology and disease caused by viruses. A valid system to characterize novel viral pathogens should be able to synthesize and characterize structural as well as non-structural proteins. Therefore, we used a eukaryotic cell-free system to synthesize non-structural, structural and accessory proteins encoded by SARS-CoV-2 (**Figure 1A**). Qualitative analysis of viral proteins synthesized in a CHO cell-free system showed that all viral proteins tested, could be synthesized. Additionally, multimerization of proteins such as ORF3 and ORF7a as well as ORF8 and ORF10 was visualized and defined cleavage products as seen for Spike proteins and nucleocapsid protein were detected by autoradiography (**Figure 1B**).

Quantitative analysis by hot TCA precipitation and subsequent liquid scintillation verified the acquired data for the qualitative analysis. Total protein yields for non-structural proteins showed that these proteins were mainly present in a soluble form as higher protein yields were detected in the supernatant fraction (SN) compared to the microsomal fraction (MF). Cell-free protein synthesis of the full length Spike protein (ORF2/S) was conducted using a PCR template which resulted in a lower template concentration used. This reduced template concentration led to lower total protein yields. Nonetheless, the high molecular weight protein could be synthesized in an equal amount to the comparably small



Fold increase	ORF5			ORF3			nsp12		
	TM	SN	MF	TM	SN	MF	TM	SN	MF
24 h /Batch	20.04	10.11	10.25	35.54	31.38	25.27	51.24	3.50	26.41

FIGURE 2 | Quantitative analysis of cell-free synthesized SARS-CoV-2 proteins. Viral proteins were synthesized in CHO lysate in **(A)** batch-based reaction and **(B)** a CECF reaction. The fold increase of total protein yield from a batch reaction to a CECF reaction is shown. Quantitative analysis of ^{14}C-labeled cell-free synthesized proteins was performed by liquid scintillation counting. Standard deviations were calculated from triplicate analysis. The translation mixture (TM) was separated into the soluble proteins in the supernatant (SN) and the microsomal fraction (MF). The template for the full length Spike protein was based on a PCR-template (*).

envelope protein (ORF4/E). The initial data for the synthesis of the accessory proteins showed that all of these proteins could be synthesized as well. The immunomodulatory protein encoded by ORF6 showed the overall lowest protein yield of 6 µg/ml while the transmembrane protein encoded by ORF3 showed the highest protein yields with 25 µg/ml (**Figure 2A**).

In order to increase the protein yields in a continuous-exchange cell-free (CECF) system, one representative protein of the three protein groups was synthesized for 24 h. The nsp12 coding for an RNA-dependent RNA polymerase (RdRp), the ORF5 membrane glycoprotein and the channel-like ORF3 were chosen (**Figure 2B**). The protein yields from a batch-based synthesis could be increased by about 20, 35 and 50 fold for ORF5, ORF3 and nsp12, respectively, in a 24 h CECF reaction. Apparently, the nsp12 enzyme was not suitable for a CECF reaction as the soluble protein aggregated in the MF which suggests that a batch-based reaction was more suitable for this enzyme. These data indicate that CFPS offers a platform for the rapid synthesis and analysis of SARS-CoV-2 proteins. As each protein showed different requirements for the cell-free synthesis, the open cell-free system offers an easy way to adapt the synthesis conditions to the need of each individual protein.

To further show the applicability of CFPS as a rapid response system for viral pathogens, we analyzed the individual protein groups and analyzed the functionality of proteins of interest. In a first step, cell-free synthesized nsp1 protein was characterized. This protein is also known as the leader protein responsible for the inhibition of host protein translation (Banerjee et al., 2020; Thoms et al., 2020; Lapointe et al., 2021). It was further shown that nsp1 did not decrease the translation of viral mRNA (Banerjee et al., 2020), thus nsp1 was synthesized without any alterations in the cell-free synthesis scheme. The nsp1 protein was pre-synthesized in a cell-free manner and was added to the cell-free synthesis of the model protein enhanced yellow fluorescent protein (eYFP). Subsequently, the fluorescence intensity of eYFP was measured during a 3 h synthesis time (schematic representation in **Figure 3A**). The fluorescence signal of eYFP without the supplementation of nsp1 was set as a baseline value of 100%. All other data were normalized to this intensity. In an initial experiment pre-synthesized nsp1 from a CECF reaction, a volume equivalent NTC and ORF6 as a protein control were added to the eYFP synthesis. These data showed that the addition of nsp1 at concentrations of 1000 and 600 nM reduced the fluorescence intensity of eYFP in a concentration dependent manner. 1000 nM of nsp1 led to a complete inhibition of eYFP fluorescence. Unfortunately, the addition of the NTC and ORF6 at 600 nM showed interactions with the eYFP fluorescence. Nonetheless, 600 nM nsp1 induced the highest

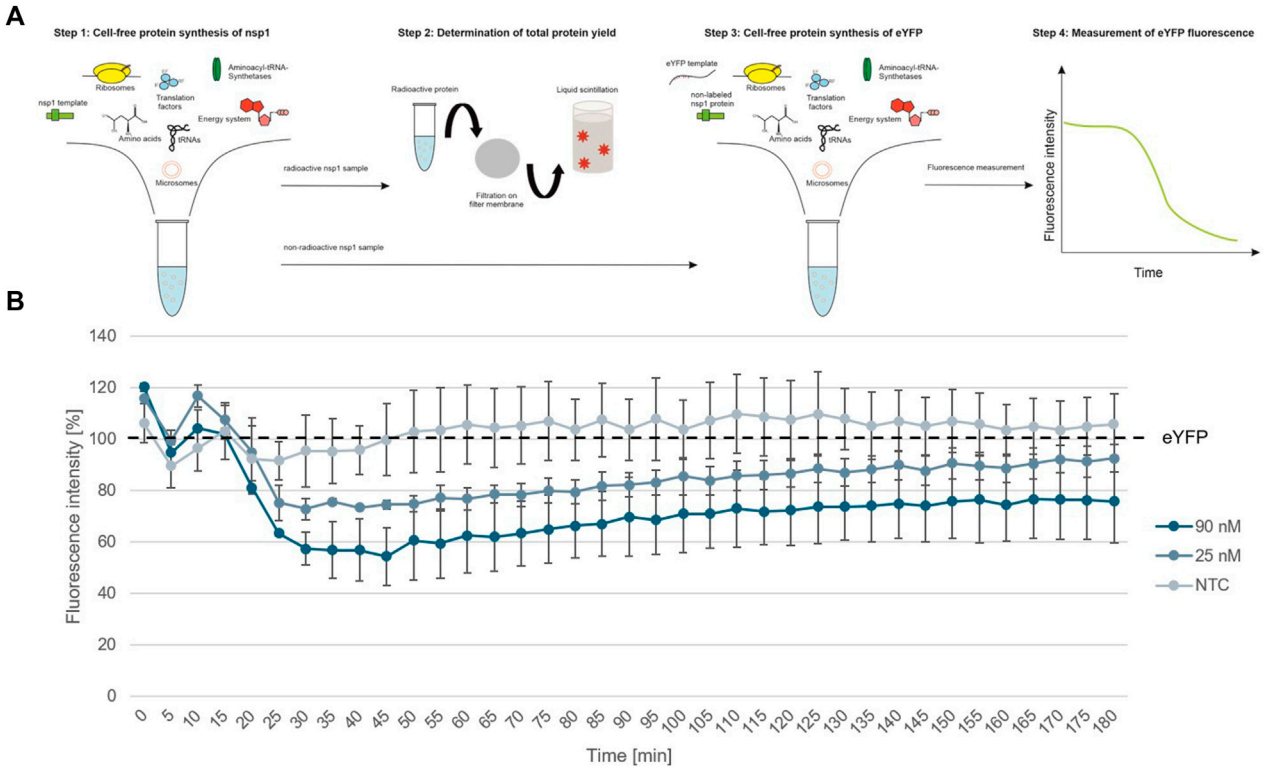


FIGURE 3 | Functional analysis of nsp1. **(A)** Schematic scheme of nsp1 functionality assay. **(B)** Application of cell-free synthesized nsp1 at 90 and 25 nM to the cell-free protein synthesis of the model protein eYFP. Data were normalized to eYFP fluorescence signal. Standard deviation was calculated from two experiments.

fluorescence decline (**Supplementary Figure S1**). Therefore, the reaction conditions were optimized and a batch-based reaction was used. Two concentrations (25 and 90 nM) of nsp1 protein were added to the eYFP synthesis. The NTC was administered in a volume equivalent to the highest nsp1 concentration. Starting after about 20 min reaction time, the fluorescence intensity of the eYFP slowly decreased when nsp1 was supplemented to the reaction but no specific effect could be seen for an NTC supplementation (**Figure 3B**). A concentration of 90 nM of nsp1 decreased the eYFP fluorescence to less than 60%, while 25 nM nsp1 decreased the eYFP fluorescence to about 75%. This suggested a concentration dependent effect. Strikingly, the inhibitory effect of nsp1 weakened over time so that the fluorescence intensity of eYFP increased again (**Figure 3B**). Nonetheless, the translation inhibition effect of nsp1 could be shown.

The characterization of non-structural viral proteins is essential for a rapid response to a virus. Nonetheless, the virus needs accessory proteins to stably infect the host. In SARS-CoV-2 these are known to trigger a variety of different interactions within the host such as the involvement in the interferon signaling pathway (Miorin et al., 2020; Yuen et al., 2020), causing an immune invasion (Flower et al., 2021) or acting as membranous channel-like proteins in order to disrupt the host`s cell homeostasis (Yoshimoto, 2020; Kern et al., 2021). As shown in **Figure 1**, all accessory proteins could be synthesized in a cell-free manner and showed that cell-free protein synthesis can be a tool to rapidly analyze the synthesis of such proteins, to identify optimal synthesis conditions as well as to assess their solubility. In the beginning of a pandemic, the role of such proteins is not yet fully known, therefore we tested whether a signal peptide interferes with a defined protein or might even inhibit the translation of such a protein. Accordingly, ORF3, ORF6, ORF7b, ORF8, and ORF10 were synthesized with and without a Melittin (Mel) signal peptide, which typically allows for the co-translational translocation. As some proteins might inhibit the protein translation machinery, the Mel signal peptide allows for the translocation of the protein into the vesicles present in the cell-free lysates, enabling protein translation. All constructs with and without a signal peptide were fractionated into the soluble proteins (SN) and the proteins in the microsomal fraction (MF). In general, our data indicated that all ORFs were more stably expressed in the presence of a signal peptide (**Figure 4A**). The NCM-ORF3 construct was based on a PCR template which might have resulted in a lower overall translation efficiency due to a lower template concentration. Strikingly, ORF10 protein translation could not be quantitatively detected without a signal peptide (**Figure 4A**, red box). In the presence of a Mel signal peptide, protein yields of 13.7 μg/ml were detected for ORF10 suggesting a better translation initiation. ORF7a is known to harbor a native signal peptide as it encodes a type I transmembrane protein (Rosenthal et al., 2020). Comparing the native signal peptide (Nat-SP) with the Mel signal peptide, the native signal peptide showed a higher total amount of protein

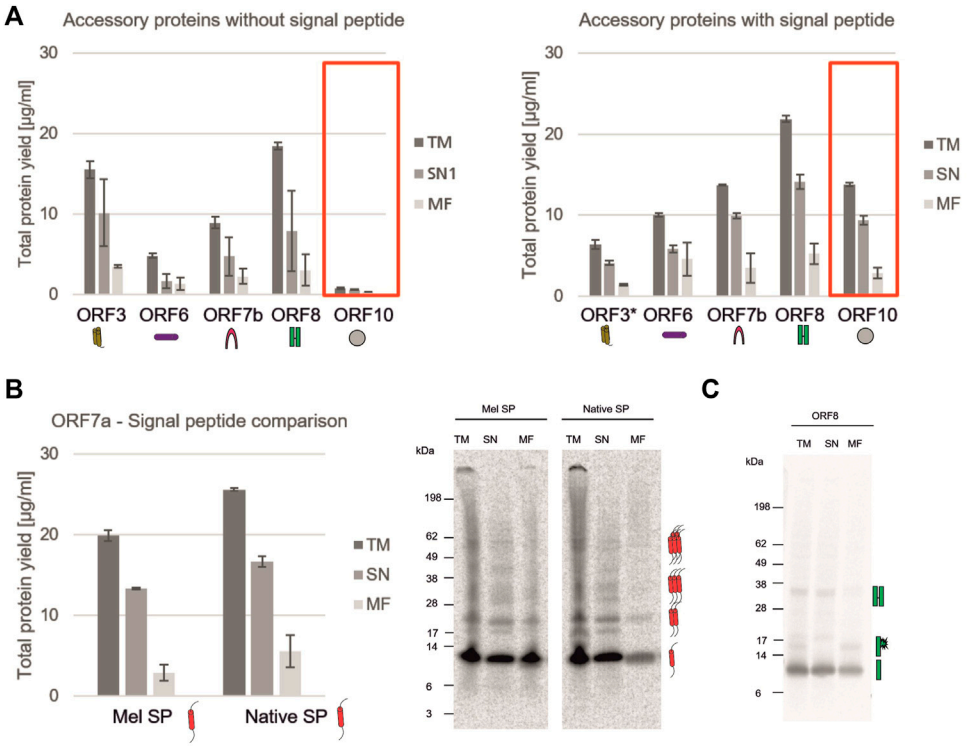


FIGURE 4 | Analysis of SARS-CoV-2 accessory proteins. **(A)** Accessory proteins ORF3, ORF6, ORF7b, ORF8 and ORF10 with and without a Mel signal peptide were synthesized in a CHO batch-based reaction. Quantitative analysis of cell-free synthesized proteins was performed by liquid scintillation counting. Standard deviations were calculated from triplicate analysis. The template for ORF3 with a Mel signal peptide was based on a PCR-template (*). **(B)** ORF7a was synthesized in a CHO batch-based cell-free system with Mel and a native signal peptide. Quantitative analysis of cell-free synthesized proteins as performed by liquid scintillation counting. Standard deviations were calculated from triplicate analysis. Qualitative analysis by autoradiography. **(C)** Autoradiograph showing ORF8 synthesized in a CHO cell-free system. In all experiments, the translation mixture (TM) was separated into the soluble proteins in the supernatant (SN) and the microsomal fraction (MF).

and a higher protein yield in the MF suggesting efficient targeted embedding into the microsomal membrane. Qualitative analysis showed potential multimerization with the Mel and the native signal peptide. Protein bands of the potential monomer (~12 kDa), dimer (~24 kDa), trimer (~36 kDa) and pentamer (~60 kDa) could be detected (**Figure 4B**). The ORF8 protein showed high protein yields with and without a Mel signal peptide. Recent studies have shown that ORF8 contains a sequence homology to the SARS CoV ORF8ab native signal peptide (Tan et al., 2020), indicating that in the Mel-ORF8 construct two signal peptides were present. Hence, later work only focused on the ORF8 template without the Mel signal peptide as it contained the putative native signal peptide. Our data depict that potential SDS-stable multimerization was possible even in the presence of DTT. The autoradiograph also depicts a protein band with a higher molecular weight (~15 kDa) in the TM and MF fraction but not in the SN fraction indicating a glycosylated ORF8 (**Figure 4C**). Taken together these data clearly qualify cell-free protein synthesis for the in depth analysis of the translation of viral proteins and the analysis of signal peptide use, protein translocation and translation initiation.

During a pandemic, structural proteins play a major role in the rapid response to the novel pathogen. Typically, they are used for the detection of the virus itself by PCR techniques as well as antigen detection. Hence, our study aimed to identify the possibility of cell-free synthesized viral antigens for the characterization and detection of structural proteins. As the small envelope protein is reported to form a homo-pentamer (Mandala et al., 2020), the multimerization of the envelope protein was investigated. The multimerization during a batch-based cell-free protein synthesis showed that a monomeric state was preferred but multimers could also be detected (**Figure 5A**). In comparison to the envelope protein, the surface glycoprotein Spike is a large protein. It is cleaved into the S1 and S2 domain which were individually analyzed. As seen in **Figure 1B**, the full-length protein as well as the cleavage products can be synthesized in a cell-free manner. On an exemplary basis, the S2 domain was investigated for its N-linked glycosylation. Hence, the S2 domain was synthesized, fractionated and the MF was digested by PNGaseF and EndoH. Autoradiography showed that the TM and MF depicted an additional protein band with a higher molecular weight in comparison to the SN. When the MF was digested with PNGaseF and EndoH, this additional protein band was not detectable anymore, indicating glycosylation in the cell-free system (**Figure 5B**).

The nucleocapsid protein (N) is a major contributor to the structure and humoral response of the virus (Smits et al., 2021) but is also a key structure for diagnostic purposes such as rapid

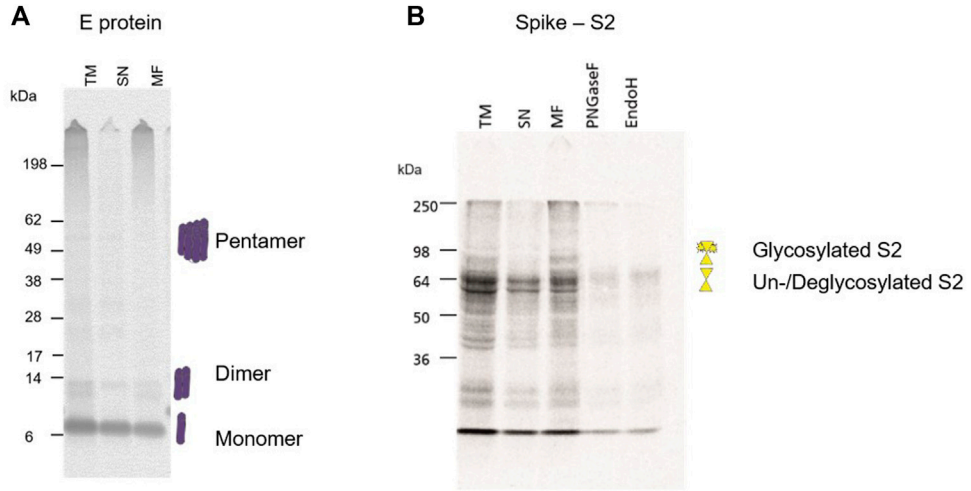


FIGURE 5 | Structural protein analysis of E and S2. **(A)** Autoradiograph showing the envelope protein synthesized in a CHO cell-free system showing multimerization. The translation mixture (TM) was separated into the soluble proteins in the supernatant (SN) and the microsomal fraction (MF). **(B)** Autoradiograph showing the S2 domain from the Spike protein synthesized in a CHO cell-free system. The translation mixture (TM) was separated into the soluble proteins in the supernatant (SN) and the microsomal fraction (MF). The MF was digested with PNGaseF and EndoH.

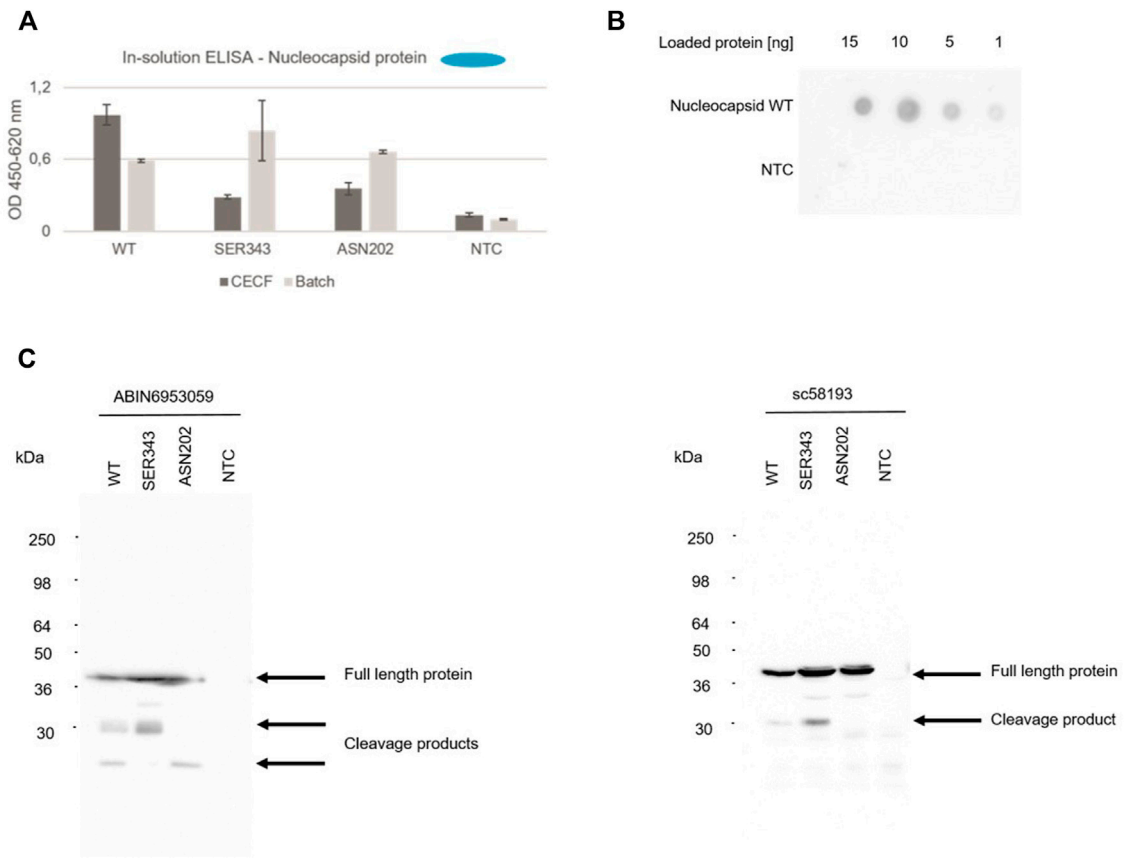


FIGURE 6 | Detection of N protein. The nucleocapsid WT protein as well as both mutants (SER343 and ASN202) were synthesized in a CHO cell-free system. Detection of cell-free synthesized nucleocapsid protein in an in-solution ELISA **(A)**, a Dot Blot **(B)** and Western Blots **(C)**.

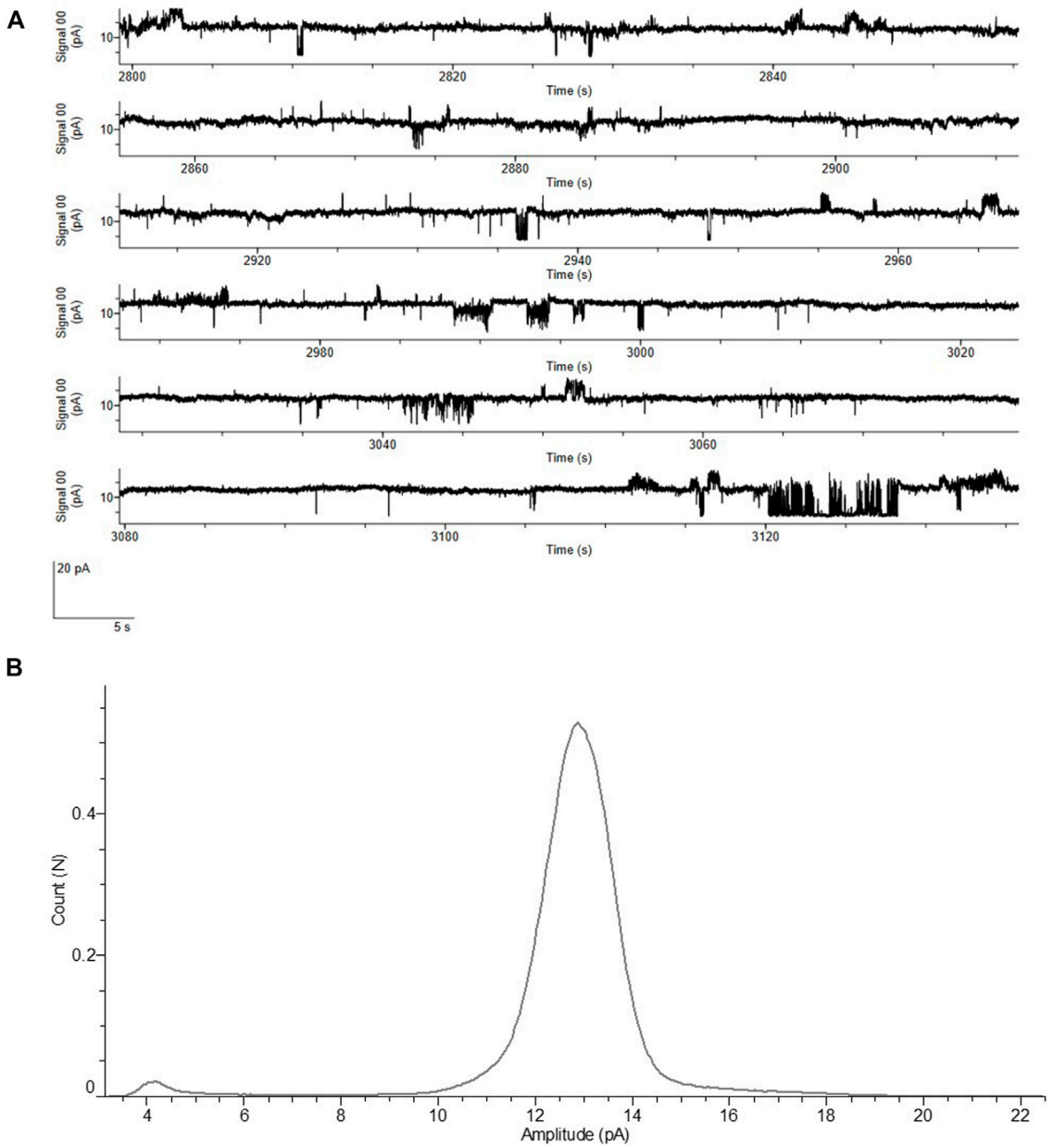


FIGURE 7 | Functional analysis of ORF4 envelope protein reconstituted into lipid bilayer. **(A)** Current recordings at a voltage clamp of +100 mV from the ORF4 reconstituted into DPhPC bilayer. **(B)** All point Histogram plotted from the current recordings showing the large peak at 13 pA indicating a stable pore. All measurements of ORF4 were done in the presence of 150 mM NaCl, 10 mM HEPES, pH 7.0 buffer (n = 5).

antigen tests (Grant et al., 2021). In this context, an in-solution ELISA was used to detect the cell-free synthesized nucleocapsid wild type (WT) protein as well as two early onset mutants (SER343, ASN202), which were detected in April 2020, in order to validate cell-free synthesis for testing antibodies as well as providing viral antigens for the diagnostic use. N protein, WT as well as mutants, synthesized in a batch and in a 24 h CECF reaction could be detected in an in-solution ELISA (**Figure 6A**). To further demonstrate the use of cell-free synthesized proteins for the validation of antibodies we performed a quantitative dot blot analysis. N WT protein was blotted onto a nitrocellulose membrane at four different concentrations (15, 10, 5, 1 ng). Our data showed that even concentrations of 1 ng could be detected (**Figure 6B**). In the early onset of a pandemic antibodies and antivirals against similar viral strains are tested for their efficacy. Here, the detection with an anti-SARS-CoV-2 N antibody (ABIN6953059) as used in the prior experiments was compared to an anti-SARS N antibody (sc58193). Corresponding to the prior data, a Western Blot analysis showed that the N protein was detected by both

antibodies. Further, cleavage products could be detected. Strikingly, a protein band at about 30 kDa in the WT and SER343 mutant but not in the ASN202 mutant was identified. The anti-SARS-CoV-2 N antibody showed less by products at an exposure time of 20 s (**Figure 6C**). These data highlight the efficiency of cell-free protein synthesis for the validation of antibodies detecting viral proteins.

Vaccine development is a key process once a novel pathogen presents itself, but the characterization of proteins harming the host as well as identifying potential blockers to the virus is of utmost importance. The ORF4 encoding the envelope protein might be a candidate for such blocking events as it is a viroporin (Xia et al., 2021). Therefore, we performed an in depth analysis of the envelope protein including the optimized cell-free protein synthesis (data not shown) as well as the characterization of the envelope protein as a cytotoxic channel like protein using planar lipid bilayer measurements. Different voltages were applied in order to measure the stable recordings from the protein reconstituted lipid bilayer. All the measurements were performed with the MF of the envelope protein. In order to avoid activity from the endogenous proteins in the microsomal vesicles, we used 150 mM NaCl as an electrolyte. Control measurements were performed with the same experimental conditions using the MF from the NTC without any expressed protein. Electrophysiological data of the envelope protein was compared to the controls. We noticed that there are small bilayer disruption currents leading to the lipid bilayer breaking in both cases, which we carefully ignored. There are two types of currents observed in the ORF4 reconstituted bilayers and completely absent in the controls. There is a single channel activity followed by pore formation in the ORF4 reconstituted lipid bilayer (**Figure 7A** and **Supplementary Figure S2**).

ORF4 reconstituted lipid bilayers showed different types of response ranging from cytotoxic events with large undefined current levels (>50 pA), pore forming stable currents of around 12–13 pA and single channel events with frequent flickering between open and closed states (5–7 pA). NTCs showed only few cytotoxic events and complete absence of pore and single channel events. The frequency of events is presented in **Supplementary Table S1**. Single channel events are further shown in **Supplementary Figure S3**. The histogram identified a large current peak at 13 pA suggesting a stable pore formation (**Figure 7B**). These data highlight the straightforward approach to analyze functionally active channel-like viral proteins derived from cell-free systems.

Overall, our data demonstrate that the CHO cell-free system can be applied to characterize the complete set of SARS-CoV-2 viral proteins and facilitates the screening of antibodies as well as pharmaceuticals and blockers against these viral proteins.

DISCUSSION

The rise of SARS-CoV-2 in late 2019 and the development to a global health issue in early 2020 continuing until today showed the dramatic results a viral outbreak could have. This has led to an increasing awareness worldwide to novel zoonotic diseases. The development of new detection techniques and a fast generation of adapted vaccines against virus variants are absolutely necessary as close interactions with wildlife and domesticated pets as well as climate change might pave the way for future pandemics (Scudellari, 2020; Beyer et al., 2021; Skegg et al., 2021; Thoradeniya and Jayasinghe, 2021). Until today some studies have evaluated the possibility of synthesizing viral proteins and virus-like proteins in cell-free systems. A virus like particle derived from the norovirus (Sheng et al., 2017) and Qβ phage (Smith et al., 2012) were synthesized in an *E. coli* based cell-free system. Using eukaryotic cell-free systems the capsid protein (Spearman and Ratner, 1996) and the envelope protein (Mikami et al., 2006) of the human immunodeficiency virus, the gp67 envelope protein (Merk et al., 2015), a virus like particle from the human papillomavirus (Wang et al., 2008) and the capsid protein of hepatitis E virus (Kubickova et al., 2021) have been synthesized and evaluated. These data show that cell-free protein synthesis can help to characterize single viral proteins and this knowledge can be used to further elucidate the complete virus. As cell-free protein synthesis uses a cell lysate rather than viable cells, no higher safety standards are needed for the synthesis of these viral proteins (Zemella et al., 2015). Our study aimed to demonstrate that cell-free protein synthesis can be applied to gain fast knowledge about individual novel viral proteins during a pandemic or epidemic, thus facilitating preparedness and responses.

Altincekic et al. (2021) demonstrated results on cell-free protein synthesis of ORF3, ORF6, ORF7a, ORF8, ORF9, the E and M protein in a wheat germ system and thus presented the first results on cell-free synthesis of SARS-CoV-2 proteins. These proteins were further purified and used for NMR spectroscopy. These data are extended by the findings acquired in this study as we showed that all different viral protein classes and groups could be synthesized the CHO lysate (**Figures 1**, **2**). When looking at the nsp1, also called the leader protein, the binding of the 40S ribosome within the entry channel of the mRNA was shown which results in the inactivation of protein translation within the host (Thoms et al., 2020). Further data suggested the disruption of tRNA recruitment to the 80S ribosome (Banerjee et al., 2020). Banerjee et al. further showed that viral mRNA translation is not inhibited. As the *in vitro* protein translation was possible in the CHO-based system, this could also indicate that viral translation is not inhibited by nsp1. The activity of the protein was tested *in vitro* in HeLa and rabbit reticulocyte cell lysates as well as *in vivo* in HEK293T cells showing a downregulation in host protein translation of different proteins (Banerjee et al., 2020; Thoms et al., 2020). Our results depicted an inhibition of protein translation in an *in vitro* system based on CHO lysate indicating that different eukaryotic systems can be applied for the analysis of such a protein. Nonetheless, the effect on the protein synthesis of the model protein eYFP decreased over time and the fluorescence intensity was regained (**Figure 3**). In our system, the internal ribosomal entry site (IRES) from the cricket paralysis virus (CrPV) was used to initiate the translation independently from cap structures. Strikingly, a previous study showed that nsp1 associated less efficiently with the ribosomes when the CrPV IRES

was used. The IRES from hepatitis C virus favored a binding of nsp1 (Lapointe et al., 2021). Thus, our data align with these prior findings by indicating that the effect of nsp1 was reduced by the CrPV IRES. Accordingly, future projects studying novel viral proteins should evaluate the protein synthesis and characterization in a cap-dependent and independent manner, as a protein translation could be detected in the CHO-based cell-free system.

After the successful synthesis and characterization of the nsp1 as a representative of the non-structural proteins, our study investigated the accessory proteins. Whilst the structure or function for some of the proteins is already known, other proteins yet have to be analyzed in detail. The ORF7b protein was already synthesized in a wheat germ cell-free system and multimerization could be shown (Fogeron et al., 2021). Here we demonstrate that ORF7a forms potential multimers up to a pentamer (**Figure 4B**). As this has not been shown before, detailed analysis of the multimeric state of the ORF7a has to be performed. ORF8 was shown to be involved in the immune invasion process (Flower et al., 2021) and structural analysis suggested a multimerization of the ORF8 via disulfide bridges (Flower et al., 2021). Additionally, potential glycosylation sites that stabilize the protein structure as was already seen in SARS ORF8ab were described (Mohammad et al., 2020). Accordingly, we analyzed ORF8 in our cell-free system in depth. Our data clearly depicted the formation of multimers even up to a tetramer of about 55 kDa suggesting stable disulfide bridges. In our CHO system a possible glycosylation of the monomer could be detected (**Figure 4C**). Future studies should evaluate whether the SARS-CoV-2 ORF8 is unstable without glycosylation as this was seen for SARS-CoV ORF8ab (Mohammad et al., 2020).

Until today, data on ORF10 are conflicting. Initial work on ORF10 identified epitopes for cytotoxic T lymphocytes and a possible alteration of SARS-CoV-2 pathogenicity by mutations in the ORF10 protein was described (Hassan et al., 2021). Other studies reflected that the protein can be terminated prematurely but also read through was emphasized (Pancer et al., 2020; Schuster, 2020). Within our work we could show the synthesis of ORF10 in the presence of a signal peptide (**Figure 4**). We also present oligomerization of ORF10 (**Figure 1B**) which correlates to previous work where it was hypothesized that ORF10 might oligomerize to form a pore for ion fluctuations co-localizing with other accessory proteins (Schuster, 2020). The data seen in this study clearly show that the translation initiation was improved in the presence of a signal peptide. This could indicate a better sequence context for protein translation. In prior studies, it was further suggested that ORF10 binds to an E3 ubiquitin ligase and the N-terminus is essential for that matter, but it was still not important for the infection by SARS-CoV-2 (Mena et al., 2021). Thus, our data depict a stable synthesis of ORF10 only in the presence of a N-terminal signal peptide supporting the importance of the N-terminus in ORF10 for structure and function.

Structural proteins of a viral pathogen are the first proteins to be analyzed in detail as these proteins are mainly used for primary response mechanisms. Therefore, it was mandatory to show that our cell-free system can be used to study and evaluate these proteins. In particular, we chose the nucleocapsid protein to validate our cell-free system for the detection of viral proteins in an in-solution ELISA, a Western Blot and a Dot Blot (**Figure 6**). Prior work has shown that Dot Blots are regularly used for the detection of SARS-CoV-2 antibodies in patient samples. In a Dot Blot, the complete nucleocapsid protein as well as nucleocapsid protein fragments were able to identify SARS-CoV-2 positive patient samples (Smits et al., 2021). In our study, we showed that cell-free synthesized proteins can be spotted and detected at different protein concentrations and therefore we can qualify cell-free protein synthesis for the validation of antibodies. We showed that the N protein as well as early onset mutations could be detected. The results were obtained using an anti-SARS-CoV-2 and anti-SARS antibody reflecting on the possibility of cell-free systems to assist in tackling future pandemics right from the beginning.

Finally, our study aimed to further characterize the envelope protein. As the envelope protein from the SARS outbreak in 2003 is a homo-pentameric cation channel (Wilson et al., 2004; Verdiá-Báguena et al., 2012), it was suggested that the SARS-CoV-2 envelope protein reflects a similar structure and mechanism of action. In our study, we could also show the pentameric state of the SARS-CoV-2 envelope protein after the cell-free synthesis (**Figure 5A**). A channel-like structure forming a pore of 2.1 Å has already been depicted (Mandala et al., 2020) and cytotoxic activity on various cell lines was shown (Xia et al., 2021) in initial studies. In a recent study the activity of the envelope protein as a cation channel potentially selective for potassium, sodium, calcium and magnesium was demonstrated in planar lipid bilayer recordings (Xia et al., 2021). The cation channel like behavior and pH sensitivity was further observed in *Xenopus* oocytes (Cabrera-Garcia et al., 2021). These data reflect upon the envelope protein as a drug target. In addition to the just mentioned studies, the envelope protein showed different types of electrophysiological activities in the presence of a lipid bilayer. Single channel activity typical for ion channels with transitions between open and closed states and with a stable baseline current could be detected. Sometimes these single channel events were followed by a pore formation with stable currents of around 13 pA at +100 mV without any flickering activity. This could be due to the complete pentameric assembly of the protein leading to an active open pore. Currents, which are large and abrupt and do not resemble the typical ion channel activity, were frequently observed. They are typically measured for cytotoxic proteins. These large cytotoxic currents (>50 pA) might be due to the increase in the surface density of the envelope protein assembly leading to membrane destabilization. Therefore, our data combine the knowledge gathered in previous studies. It was already shown that the envelope protein from other viral infections such as MERS and SARS demonstrated single-channel activity in lipid bilayers suggesting a pentameric ion channel (Surya et al., 2015). Whole cell recordings of SARS envelope protein expressed in HEK293 cells indentified a cation selective channel like behavior (Pervushin et al., 2009). Data on the SARS-CoV-2 envelope protein in combination with data on MERS and SARS envelope proteins reflect upon the fact that the ORF4 envelope protein from SARS-CoV-2 is a cation

selective ion channel that triggers cytotoxic events. Thus, blocking the envelope protein of SARS-CoV-2 could inhibit pathogenic events in the host. To our knowledge, this is the first study where a viroporin could directly be analyzed in planar lipid bilayer recordings when synthesized in a cell-free system without any additional purifications.

As discussed, novel pathogens could develop easily provoking challenges in modelling the emergence of novel pathogens. Moreover, the length of a pandemic cannot be predicted in general as diverse factors play a role in the progression of the infection (Scudellari, 2020). It is necessary to identify the most promising prevention methods, but therapeutic approaches are further needed (Meganck and Baric, 2021). The results obtained in this study qualify cell-free systems as a rapid response methodology for evidence-informed decision making in health policy and research. All viral proteins of SARS-CoV-2 could be synthesized, studied and proteins like the envelope protein could be functionally characterized. Antibody interaction with cell-free produced viral proteins was demonstrated and we have further shown that mutants of the nucleocapsid protein, can be assessed. The omicron variant of SARS-Cov-2 has shown that mutations can alter the course of the pandemic (Karim and Karim, 2021). Therefore, a rapid system for the characterization of viral variants is necessary. As presented in this study, 2-step PCR reactions were used to modify the templates for the cell-free reactions. Thereby demonstrating that a simple cloning-free method to efficiently generate templates suitable for coupled cell-free protein production was applied for viral genes. Based on previous studies and on the data presented here (Sawasaki et al., 2002; Bechlars et al., 2013), templates for viral mutants can be designed by fast and efficient 2-step PCR schemes thus allowing for the rapid and parallel high-throughput screening of the functional activity of mutants. This procedure will further allow the screening of antibodies and potential inhibitors against mutant viral proteins without using BSL-1 through 4 laboratories.

Taken together, we identified eukaryotic cell-free systems as a versatile technology to synthesize and characterize the viral pathogen SARS-CoV-2. The data presented here demonstrate that eukaryotic cell-free systems facilitate the rapid response to novel zoonotic diseases.

MATERIALS AND METHODS

Template Design

Plasmids encoding each ORF or single protein of SARS-CoV-2 genome (MN908947.3) were designed according to Brödel et al. (2013) containing the internal ribosomal entry site (IRES) of the cricket paralysis virus (CrPV) and the T7-promoter (Brödel et al., 2013). Nucleocapsid mutations were designed accordingly based on MT081067.1 (ASN202, exchange at amino acid 202 from serine to asparagine) and MT123290.1 (SER343, exchange at amino acid 343 from proline to serine) Genes were obtained by *de novo* gene synthesis (Biocat GmbH) in a pUC57-1.8K vector backbone. These plasmids were directly used for cell-free protein synthesis.

The full length Spike protein was also obtained by *de novo* gene synthesis (Biocat GmbH) in a pCI vector. This gene was amplified using an expression PCR (E-PCR) with the HiFidelity polymerase (Qiagen). Standard PCR protocol was applied using the forward N0 (5′ ATG ATA TCT CGA GCG GCC GCT AGC TAA TAC GAC TCA CTA TAG GGA GAC CAC AAC GGT TTC CCT CTA GAA ATA ATT TTG TTT AAC TTT AAG AAG GAG ATA AAC AAT G 3′) and reverse C0 (5′ ATG ATA TCA CCG GTG AAT TCG GAT CCA AAA AAC CCC TCA AGA CCC GTT TAG AGG CCC CAA GGG GTA CAG ATC TTG GTT AGT TAG TTA TTA 3′) primers.

As ORF3 is a putative channel-like protein the influence of the Mel signal peptide on the possible membrane embedding and multimerization was investigated. As only an NC construct was available for the ORF3, a 2-step EPCR was performed to fuse the Mel signal peptide to the ORF3 gene based on Bechlars et al. (2013). Shortly, in a first step the gene specific forward primer X-CoV-2-ORF3-oe-Mel-F (5′ TAC ATT TCT TAC ATC TAT GCG GAC GAT TTG TTT ATG AGA ATC TT 3′) and the reverse adapter primer C0 were used. During this step an overlap from the ORF3 DNA template to the Mel signal peptide was generated. In the second step, the adapter primers NCM-F (5′ ATG ATA TCT CGA GCG GCC GCT AGC TAA TAC GAC TCA CTA TAG GGA GAC CAC AAC GGT TTC CCT CTA GAA ATA ATT TTG TTT AAC TTT AAG AAG GAG ATA AAC AAA AGC AAA AAT GTG ATC TTG CTT GTA AAT ACA ATT TTG AGA GGT TAA TAA ATT ACA AGT AGT GCT ATT TTT GTA TTT AGG TTA GCT ATT TAG CTT TAC GTT CCA GGA TGC CTA GTG GCA GCC CCA CAA TAT CCA GGA AGC CCT CTC TGC GGT TTT TCA GAT TAG GTA GTC GAA AAA CCT AAG AAA TTT ACC TGC TAA ATT CTT AGT CAA CGT TGC CCT TGT TTT TAT GGT CGT ATA CAT TTC TTA CAT CTA TGC GGA C 3′) and the C0 reverse primer were used according to standard protocol. As a final construct, the NCM-ORF3-C0 was generated where a Mel signal peptide was fused to the ORF3 gene.

Cell-Free Protein Synthesis

Cell-free synthesis reactions using translationally active lysate derived from eukaryotic cells were performed as previously described by utilizing Chinese hamster ovary cells (CHO-K1) (Brödel et al., 2014; Thoring et al., 2016). Further, each template was applied individually and a no template control (NTC) consisting of a translation mixture without any DNA template was used as a background control.

Batch-Based Reactions

Protein synthesis was conducted in coupled transcription/translation reactions in a final volume of 25–80 µl. Cell-free synthesis reactions were composed of 40% (v/v) translationally active CHO lysate supplemented with HEPES-KOH (pH 7.6, f.c. 30 mM, Carl Roth GmbH), sodium acetate (f.c. 100 mM, Merck), $Mg(OAc)_2$ (f.c. 3.9 mM, Merck), KOAc (f.c. 150 mM, Merck), amino acids (complete 100 µM, Merck), spermidin (f.c. 0.25 mM; Roche), Dithiothreitol (DTT, 2.5 mM, Life technologies GmbH)

and energy regenerating components including creatine phosphokinase (f.c. 0.1 mg/ml, Roche), creatine phosphate (20 mM, Roche), ATP (1.75 mM, Roche) and GTP (0.3 mM, Roche). To allow for DNA transcription during cell-free protein synthesis, 1 U/µl T7 RNA polymerase, 0.3 mM of UTP (Roche) and CTP (Roche) and 0.1 mM of the cap analogue m7G(ppp)G (Prof. Edward Darzynkiewicz, Warsaw University, Poland) were added to the reaction. Additionally, PolyG primer (f.c. 12 µM, IBA) was supplemented. To monitor the protein quantity and quality, cell-free protein synthesis reactions were supplemented with radioactive ^{14}C-leucine (f.c. 50 µM, specific radioactivity 66.67 dpm/pmol, Perkin Elmer). Batch synthesis reactions were incubated at 30°C for 3 h at 500 rpm (Thermomixer comfort, Eppendorf).

Continuous-Exchange Cell-Free Reactions

CECF-reactions were performed as previously described (Thoring et al., 2017). Briefly, reactions took place in commercially available dialysis devices (Biotechrabbit GmbH) and incubated in a thermomixer (Eppendorf) for 24–48 h, at 30°C at 600 rpm. Two mixtures were individually prepared. Reaction mixtures were composed similar to batch-based reactions (see above). PolyG was added at a final concentration of 4.5 µM and Mg(OAc)2 was added at a final concentration of 18.5 mM. The feeding mixture was composed of HEPES-KOH (f.c. 30 mM, pH 7.6), Mg(OAc)2 (f.c. 3.9 mM), KOAc (f.c. 150 mM), amino acids (complete 100 µM f.c.), spermidine (f.c. 0.25 mM), energy regenerating components (f.c. 1.75 mM ATP, 0.3 mM GTP), CTP (f.c. 0.3 mM), UTP (f.c. 0.3 mM) and the cap analogue G(ppp)G (f.c. 0.33 mM). Further, the caspase inhibitor AC-DEVD-CMK (f.c. 30 µM; Promega) and sodium azide (f.c. 0.02%, Merck) were added to both reaction and feeding mixture. ^{14}C-leucine (f.c. 50 µM, specific radioactivity 10 dpm/pmol) for radio-labeling of *de novo* produced proteins was added to the reaction, when necessary.

Fractionation

After the incubation time the crude translation mixture (TM) was centrifuged (16,000xg, 10 min, 4°C) resulting in the supernatant (SN), containing the soluble subunits, and the pelleted microsomes containing putative membrane bound subunits. The pellet was resuspended in phosphate buffered saline (PBS) resulting in the microsomal fraction (MF).

Analysis of Radio-Labeled Proteins

Total protein yields of cell-free synthesized proteins were determined by incorporation of ^{14}C-leucine and subsequent precipitation by hot trichloro acetic acid (TCA, Carl Roth GmbH). Briefly, 3–5 µl aliquots of the fraction of interest were mixed with 3 ml of 10% TCA/2% casein hydrolysate (Carl Roth GmbH) solution and incubated at 80°C for 15 min. After a 30 min incubation on ice, ^{14}C-labeled proteins were transferred to membrane filters (VWR) using a vacuum filtration system (Hoefer). Filters were washed with 5% TCA to remove non-incorporated ^{14}C-leucine and dried with acetone. Subsequently, filters were placed in 3 ml scintillation cocktail (Quicksafe A, Zinsser Analytik), incubated for at least 1 h and measured by liquid scintillation counting using the Hidex 600 SL (Hidex). Total protein yields of *de novo* synthesized proteins were calculated based on the molecular weight and number of leucines of the respected protein.

For qualitative analysis of the proteins' homogeneity and molecular size, 3–5 µl aliquots were precipitated in cold acetone (Carl Roth GmbH) as described previously (Thoring et al., 2017). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using precast gels (NuPAGE, 10% Bis-Tris, Life technologies) and self-prepared 10% gels using SureCast resolving and stacking buffer (Thermo Fisher Scientific) was performed. The SeeBlue Pre-Stained marker (Thermo Fisher) was used as a standard for the molecular weight measurement. Precast gels were run at 185 V for 35 min while self-prepared gels were run at 150 V for 55 min. Gels were dried at 70°C (Unigeldryer 3545D, Uniequip, Planegg), placed on a phosphor screen (GE Healthcare) and radioactively labeled proteins were visualized using a Typhoon Trio + variable mode imager (GE Healthcare).

Protein Translation Inhibition Caused by nsp1

Nsp1 was pre-synthesized in a cell-free manner in a CHO lysate and subsequently added to the synthesis of a non-viral model protein, namely an enhanced yellow fluorescent protein (eYFP). Defined concentrations of the nsp1 were administered to the cell-free synthesis of eYFP. The NTC was administered as a volume equivalent to the highest nsp1 concentration. A 50 µl reaction of the eYFP was pipetted in a black 96-well plate and placed into a microplate reader (Mithras Tristar^2 LB 943 Berthold Technologies). The fluorescence intensity of the eYFP was measured every 10 min during the 3 h synthesis time.

Deglycosylation Assay

Protein N-glycosylation was investigated using PNGaseF (peptide N-glycosidase F, NEB) or EndoH (endoglycosidase H, NEB) according to the manufacturer's protocol. Therefore, proteins were synthesized in a cell-free manner in the presence of ^{14}C-leucine. Subsequently, 5 µl of the protein sample were used for the deglycosylation assay and precipitated in acetone for SDS-PAGE analysis as described above.

In-Solution ELISA

Cell-free synthesized nucleocapsid proteins were mixed with Ni-NTA Magnetic Agarose Beads (Qiagen) and handled according to the manufacturer's protocol. Shortly, samples were incubated with beads over night at 4°C on a rotator. All following steps were performed on ice. The next day, beads were washed three times and 200 µl of primary anti-nucleocapsid antibody (1:1000 in binding buffer, Antibodies online ABIN6953059) were added to the beads and incubated at 4°C on a rotator for 2 h. This washing step and antibody incubation was repeated for the secondary HRP-linked anti-rabbit antibody (1:3000 in binding buffer, Cell Signaling Technologies). After a final washing procedure, 100 µl TMB-solution was added to the beads and incubated on ice for 1 min. The TMB solution was separated from the beads and mixed with

100 µl H_2SO_4. The absorbance at 450 and 620 nm was measured using the Mithras Tristar^2 LB 943 (Berthold Technologies).

Western Blot

For qualitative binding analysis, the SN fractions of the nucleocapsid WT proteins and the two mutants were analyzed. Proteins were synthesized in a cell-free manner using the CHO eukaryotic lysate. Samples were precipitated with ice cold acetone and run on a self-cast SDS-PAGE as described above. The SeeBlue Plus 2 Pre-Stained marker (Thermo Fisher) was used as a weight measurement. Subsequently, Western blotting was performed using the iBlot Gel transfer device (Invitrogen) where proteins were blotted onto a PVDF membrane with 20 V for 10 min. Membranes were washed with Tris buffered saline with 0.1% Tween (TBS/T) for 5 min, repeated twice and incubated overnight at 4°C with blocking buffer (2% BSA in TBS/T). On the next day, the membranes were washed three times in TBS/T before incubating the membrane with the commercially available antibody against the SARS nucleocapsid protein sc58193 and SARS-Cov-2 N protein ABIN6953059 (antikoerper-online.de and Santa Cruz Biotechnology, respectively, 1:1,000, 1% BSA in TBS/T) for 3 h at room temperature on an orbital shaker at 60 rpm. The washing procedure in TBS/T was repeated again, followed by an incubation of 2 h with the secondary HRP-linked anti-mouse IgG antibody (1: 3000, 1% BSA in TBS/T, Cell Signaling Technologies). The binding was visualized with chemiluminescent WesternBright ECL HRP substrate (Advanstra) and detected with the Azure c600 Gel Imaging System (Azure Biosystems).

Dot Blot Analysis

The total protein yield of cell-free synthesized nucleocapsid protein was determined by hot TCA precipitation as described above. Defined yields of 15, 10, 5, 1 ng protein were precipitated in ice cold acetone (see above) and the dried pellet was solubilized in 4 µl LDS buffer and 1 µl protein binding buffer (G-Biosciences, VWR). Subsequently, the Enhancer Dot Blot System (G-Biosciences, VWR) was used to spot the protein samples onto a nitrocellulose membrane (GE Healthcare). After that the same washing, blocking and antibody incubation procedure as for the Western Blot analysis was performed. Primary antibody against the SARS-Cov-2 N protein ABIN6953059 (antikoerper-online.de) in a 1:1000 dilution. The HRP-linked anti-mouse secondary antibody was used in a 1:3000 dilution. Again, the binding was visualized with chemiluminescent WesternBright ECL HRP substrate (Advanstra) and detected with the Azure c600 Gel Imaging System (Azure Biosystems).

Electrophysiological Analysis of the ORF4 Envelope Protein

Planar bilayer experiments were performed as explained previously (Thoring et al., 2017). Lipid bilayers were formed from 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC, (Avanti Polar Lipids) which were dissolved in octane (Sigma Aldrich) at a concentration of 10 mg/ml. Concentrations of 150 mM sodium chloride (NaCl, Sigma Aldrich), 10 mM HEPES (Sigma Aldrich), buffered at pH 7.0 were used as an electrolyte. For current measurements, voltage ramp protocol, as well as individual voltage steps were applied, to analyze the functional properties. A single channel amplifier (EPC-10, HEKA Electronic Dr. Schulze GmbH) was connected to the multiplexer electronics port of the Orbit16 system (Nanion). Recordings were done at a sampling rate of 50 kHz with a 10 kHz Bessel. Data were analyzed with Clampfit (Molecular Devices). 5 µl of MF of the cell-free synthesized envelope protein were added into the chamber containing 180 µl of the buffer and mixed gently by pipetting for a better fusion of microsomes with the underlying lipid bilayer. After waiting for 20 min electrophysiological measurements were started.

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/nuccore/MN908947.3.

AUTHOR CONTRIBUTIONS

FR: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft, and writing—review and editing; HT: investigation, data curation, formal analysis, and witing—review and editing; RW: investigation, formal analysis, and writing—review and editing; DW: investigation; SD: investigation, data curation, formal analysis, writing—original draft, and writing—review and editing; AZ: methodology and writing—review and editing; SK: conceptualization, data curation, funding acquisition, methodology, project administration, resources, supervision, validation, visualization, and writing—review and editing.

FUNDING

This project is funded within the framework of the Fraunhofer-Gesellschaft's internal programs Anti-CoV and the German Ministry of Education and Research (BMBF No. L1FHG42421).

ACKNOWLEDGMENTS

For the CHO lysate preparation, the authors would like to thank DW, Dr. Ing. Lena Thoring and Dipl. Ing. (FH) Doreen A. Wüstenhagen (Fraunhofer IZI-BB, Potsdam-Golm, Germany).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.896751/full#supplementary-material

REFERENCES

Altincekic, N., Korn, S. M., Qureshi, N. S., Dujardin, M., Ninot-Pedrosa, M., Abele, R., et al. (2021). Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Front. Mol. Biosci.* 8, 653148. doi:10.3389/fmolb.2021.653148

Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., et al. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* 183 (5), 1325–1339. doi:10.1016/j.cell.2020.10.004

Bechlars, S., Wüstenhagen, D. A., Drägert, K., Dieckmann, R., Strauch, E., and Kubick, S. (2013). Cell-free Synthesis of Functional Thermostable Direct Hemolysins of Vibrio Parahaemolyticus. *Toxicon* 76, 132–142. doi:10.1016/j.toxicon.2013.09.012

Beyer, R. M., Manica, A., and Mora, C. (2021). Shifts in Global Bat Diversity Suggest a Possible Role of Climate Change in the Emergence of SARS-CoV-1 and SARS-CoV-2. *Sci. Total Environ.* 767, 145413. doi:10.1016/j.scitotenv.2021.145413

Brödel, A. K., Sonnabend, A., Roberts, L. O., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2013). IRES-mediated Translation of Membrane Proteins and Glycoproteins in Eukaryotic Cell-free Systems. *PLoS One* 8 (12), e82234. doi:10.1371/journal.pone.0082234

Brödel, A. K., Sonnabend, A., and Kubick, S. (2014). Cell-free Protein Expression Based on Extracts from CHO Cells. *Biotechnol. Bioeng.* 111 (1), 25–36. doi:10.1002/bit.25013

Cabrera-Garcia, D., Bekdash, R., Abbott, G. W., Yazawa, M., and Harrison, N. L. (2021). The Envelope Protein of SARS-CoV-2 Increases Intra-golgi pH and Forms a Cation Channel that Is Regulated by pH. *J. Physiol.* 599 (11), 2851–2868. doi:10.1113/JP281037

Chalmeau, J., Monina, N., Shin, J., Vieu, C., and Noireaux, V. (2011). α-Hemolysin Pore Formation into a Supported Phospholipid Bilayer Using Cell-free Expression. *Biochim. Biophys. Acta Biomembr.* 1808 (1), 271–278. doi:10.1016/j.bbamem.2010.07.027

Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., and Hurley, J. H. (2021). Structure of SARS-CoV-2 ORF8, a Rapidly Evolving Immune Evasion Protein. *Proc. Natl. Acad. Sci. U.S.A.* 118 (2), e2021785118. doi:10.1073/pnas.2021785118

Fogeron, M.-L., Montserret, R., Zehnder, J., Nguyen, M.-H., Dujardin, M., Brigandat, L., et al. (2021). SARS-CoV-2 ORF7b: Is a Bat Virus Protein Homologue a Major Cause of COVID-19 Symptoms? *bioRxiv*.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., et al. (2020). Structure of the RNA-dependent RNA Polymerase from COVID-19 Virus. *Science* 368 (6492), 779–782. doi:10.1126/science.abb7498

Grant, B. D., Anderson, C. E., Alonzo, L. F., Garing, S. H., Williford, J. R., Baughman, T. A., et al. (2021). A SARS-CoV-2 Coronavirus Nucleocapsid Protein Antigen-Detecting Lateral Flow Assay. *PLoS One* 16 (11), e0258819. doi:10.1371/journal.pone.0258819

Hassan, S. S., Attrish, D., Ghosh, S., Choudhury, P. P., Uversky, V. N., Aljabali, A. A. A., et al. (2021). Notable Sequence Homology of the ORF10 Protein Introspects the Architecture of SARS-CoV-2. *Int. J. Biol. Macromol.* 181, 801–809. doi:10.1016/j.ijbiomac.2021.03.199

Henrich, E., Hein, C., Dötsch, V., and Bernhard, F. (2015). Membrane Protein Production inEscherichia Colicell-free Lysates. *FEBS Lett.* 589 (15), 1713–1722. doi:10.1016/j.febslet.2015.04.045

Jang, K.-J., Jeong, S., Kang, D. Y., Sp, N., Yang, Y. M., and Kim, D.-E. (2020). A High ATP Concentration Enhances the Cooperative Translocation of the SARS Coronavirus Helicase nsP13 in the Unwinding of Duplex RNA. *Sci. Rep.* 10 (1), 4481. doi:10.1038/s41598-020-61432-1

Karim, S. S. A., and Karim, Q. A. (2021). Omicron SARS-CoV-2 Variant: a New Chapter in the COVID-19 Pandemic. *Lancet* 398 (10317), 2126–2128. doi:10.1016/s0140-6736(21)02758-6

Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., et al. (2021). Cryo-EM Structure of SARS-CoV-2 ORF3a in Lipid Nanodiscs. *Nat. Struct. Mol. Biol.* 28 (7), 573–582. doi:10.1038/s41594-021-00619-0

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181 (4), 914–921. doi:10.1016/j.cell.2020.04.011

Kubickova, B., Schenk, J. A., Ramm, F., Markuškienė, K., Reetz, J., Dremsek, P., et al. (2021). A Broadly Cross-Reactive Monoclonal Antibody against Hepatitis E Virus Capsid Antigen. *Appl. Microbiol. Biotechnol.* 105 (12), 4957–4973. doi:10.1007/s00253-021-11342-7

Lapointe, C. P., Grosely, R., Johnson, A. G., Wang, J., Fernández, I. S., and Puglisi, J. D. (2021). Dynamic Competition between SARS-CoV-2 NSP1 and mRNA on the Human Ribosome Inhibits Translation Initiation. *Proc. Natl. Acad. Sci. U.S.A.* 118 (6), e2017715118. doi:10.1073/pnas.2017715118

Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020). Structure and Drug Binding of the SARS-CoV-2 Envelope Protein Transmembrane Domain in Lipid Bilayers. *Nat. Struct. Mol. Biol.* 27 (12), 1202–1208. doi:10.1038/s41594-020-00536-8

Meganck, R. M., and Baric, R. S. (2021). Developing Therapeutic Approaches for Twenty-First-century Emerging Infectious Viral Diseases. *Nat. Med.* 27 (3), 401–410. doi:10.1038/s41591-021-01282-0

Mena, E. L., Donahue, C. J., Vaites, L. P., Li, J., Rona, G., O'Leary, C., et al. (2021). ORF10-Cullin-2-ZYG11B Complex Is Not Required for SARS-CoV-2 Infection. *Proc. Natl. Acad. Sci. U.S.A.* 118 (17), e2023157118. doi:10.1073/pnas.2023157118

Merk, H., Rues, R.-B., Gless, C., Beyer, K., Dong, F., Dötsch, V., et al. (2015). Biosynthesis of Membrane Dependent Proteins in Insect Cell Lysates: Identification of Limiting Parameters for Folding and Processing. *Biol. Chem.* 396 (9-10), 1097–1107. doi:10.1515/hsz-2015-0105

Mikami, S., Kobayashi, T., Yokoyama, S., and Imataka, H. (2006). A Hybridoma-Based *In Vitro* Translation System that Efficiently Synthesizes Glycoproteins. *J. Biotechnol.* 127 (1), 65–78. doi:10.1016/j.jbiotec.2006.06.018

Miorin, L., Kehrer, T., Sanchez-Aparicio, M. T., Zhang, K., Cohen, P., Patel, R. S., et al. (2020). SARS-CoV-2 Orf6 Hijacks Nup98 to Block STAT Nuclear Import and Antagonize Interferon Signaling. *Proc. Natl. Acad. Sci. U.S.A.* 117 (45), 28344–28354. doi:10.1073/pnas.2016650117

Mohammad, S., Bouchama, A., Mohammad Alharbi, B., Rashid, M., Saleem Khatlani, T., Gaber, N. S., et al. (2020). SARS-CoV-2 ORF8 and SARS-CoV ORF8ab: Genomic Divergence and Functional Convergence. *Pathogens* 9 (9), 677. doi:10.3390/pathogens9090677

Orth, J. H. C., Schorch, B., Boundy, S., Ffrench-Constant, R., Kubick, S., and Aktories, K. (2011). Cell-free Synthesis and Characterization of a Novel Cytotoxic Pierisin-like Protein from the Cabbage Butterfly Pieris Rapae. *Toxicon* 57 (2), 199–207. doi:10.1016/j.toxicon.2010.11.011

Pancer, K., Milewska, A., Owczarek, K., Dabrowska, A., Kowalski, M., Łabaj, P. P., et al. (2020). The SARS-CoV-2 ORF10 Is Not Essential *In Vitro* or *In Vivo* in Humans. *Plos Pathog.* 16 (12), e1008959. doi:10.1371/journal.ppat.1008959

Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F. L., Yu, D., et al. (2009). Structure and Inhibition of the SARS Coronavirus Envelope Protein Ion Channel. *Plos Pathog.* 5 (7), e1000511. doi:10.1371/journal.ppat.1000511

Rosenthal, S. H., Kagan, R. M., Gerasimova, A., Anderson, B., Grover, D., Hua, M., et al. (2020). Identification of Eight SARS-CoV-2 ORF7a Deletion Variants in 2,726 Clinical Specimens. *bioRxiv*.

Sawasaki, T., Ogasawara, T., Morishita, R., and Endo, Y. (2002). A Cell-free Protein Synthesis System for High-Throughput Proteomics. *Proc. Natl. Acad. Sci. U.S.A.* 99 (23), 14652–14657. doi:10.1073/pnas.232580399

Schaecher, S. R., Diamond, M. S., and Pekosz, A. (2008). The Transmembrane Domain of the Severe Acute Respiratory Syndrome Coronavirus ORF7b Protein Is Necessary and Sufficient for its Retention in the Golgi Complex. *J. Virol.* 82 (19), 9477–9491. doi:10.1128/JVI.00784-08

Schuster, N. A. (2020). Characterization and Structural Prediction of the Putative ORF10 Protein in SARS-CoV-2. *bioRxiv*.

Scudellari, M. (2020). How the Pandemic Might Play Out in 2021 and beyond. *Nature* 584 (7819), 22–25. doi:10.1038/d41586-020-02278-5

Sheng, J., Lei, S., Yuan, L., and Feng, X. (2017). Cell-free Protein Synthesis of Norovirus Virus-like Particles. *RSC Adv.* 7 (46), 28837–28840. doi:10.1039/C7RA03742B

Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like Protease Regulates SARS-CoV-2 Viral Spread and Innate Immunity. *Nature* 587 (7835), 657–662. doi:10.1038/s41586-020-2601-5

Shu, T., Huang, M., Wu, D., Ren, Y., Zhang, X., Han, Y., et al. (2020). SARS-Coronavirus-2 Nsp13 Possesses NTPase and RNA Helicase Activities that Can Be Inhibited by Bismuth Salts. *Virol. Sin.* 35 (3), 321–329. doi:10.1007/s12250-020-00242-1

Skegg, D., Gluckman, P., Boulton, G., Hackmann, H., Karim, S. S. A., Piot, P., et al. (2021). Future Scenarios for the COVID-19 Pandemic. *Lancet* 397 (10276), 777–778. doi:10.1016/S0140-6736(21)00424-4

Smith, M. T., Varner, C. T., Bush, D. B., and Bundy, B. C. (2012). The Incorporation of the A2 Protein to Produce Novel Qβ Virus-like Particles Using Cell-free Protein Synthesis. *Biotechnol. Prog.* 28 (2), 549–555. doi:10.1002/btpr.744

Smits, V. A. J., Hernández-Carralero, E., Paz-Cabrera, M. C., Cabrera, E., Hernández-Reyes, Y., Hernández-Fernaud, J. R., et al. (2021). The Nucleocapsid Protein Triggers the Main Humoral Immune Response in COVID-19 Patients. *Biochem. Biophys. Res. Commun.* 543, 45–49. doi:10.1016/j.bbrc.2021.01.073

Spearman, P., and Ratner, L. (1996). Human Immunodeficiency Virus Type 1 Capsid Formation in Reticulocyte Lysates. *J. Virol.* 70 (11), 8187–8194. doi:10.1128/JVI.70.11.8187-8194.1996

Surya, W., Li, Y., Verdià-Bàguena, C., Aguilella, V. M., and Torres, J. (2015). MERS Coronavirus Envelope Protein Has a Single Transmembrane Domain that Forms Pentameric Ion Channels. *Virus. Res.* 201, 61–66. doi:10.1016/j.virusres.2015.02.023

Tan, Y., Schneider, T., Leong, M., Aravind, L., and Zhang, D. (2020). Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related Viruses. *MBio* 11 (3). doi:10.1128/mBio.00760-20

Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., et al. (2020). Structural Basis for Translational Shutdown and Immune Evasion by the Nsp1 Protein of SARS-CoV-2. *Science* 369 (6508), 1249–1255. doi:10.1126/science.abc8665

Thoradeniya, T., and Jayasinghe, S. (2021). COVID-19 and Future Pandemics: a Global Systems Approach and Relevance to SDGs. *Glob. Health* 17 (1), 59. doi:10.1186/s12992-021-00711-6

Thoring, L., Wüstenhagen, D. A., Borowiak, M., Stech, M., Sonnabend, A., and Kubick, S. (2016). Cell-Free Systems Based on CHO Cell Lysates: Optimization Strategies, Synthesis of "Difficult-To-Express" Proteins and Future Perspectives. *PLoS One* 11 (9), e0163670. doi:10.1371/journal.pone.0163670

Thoring, L., Dondapati, S. K., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2017). High-yield Production of "Difficult-To-Express" Proteins in a Continuous Exchange Cell-free System Based on CHO Cell Lysates. *Sci. Rep.* 7 (1), 11710. doi:10.1038/s41598-017-12188-8

Verdiá-Báguena, C., Nieto-Torres, J. L., Alcaraz, A., DeDiego, M. L., Torres, J., Aguilella, V. M., et al. (2012). Coronavirus E Protein Forms Ion Channels with Functionally and Structurally-Involved Membrane Lipids. *Virology* 432 (2), 485–494. doi:10.1016/j.virol.2012.07.005

Wang, X., Liu, J., Zheng, Y., Li, J., Wang, H., Zhou, Y., et al. (2008). An Optimized Yeast Cell-free System: Sufficient for Translation of Human Papillomavirus 58 L1 mRNA and Assembly of Virus-like Particles. *J. Biosci. Bioeng.* 106 (1), 8–15. doi:10.1263/jbb.106.8

Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., et al. (2020). The Establishment of Reference Sequence for SARS-CoV-2 and Variation Analysis. *J. Med. Virol.* 92 (6), 667–674. doi:10.1002/jmv.25762

Wilson, L., McKinlay, C., Gage, P., and Ewart, G. (2004). SARS Coronavirus E Protein Forms Cation-Selective Ion Channels. *Virology* 330 (1), 322–331. doi:10.1016/j.virol.2004.09.033

Xia, B., Shen, X., He, Y., Pan, X., Liu, F.-L., Wang, Y., et al. (2021). SARS-CoV-2 Envelope Protein Causes Acute Respiratory Distress Syndrome (ARDS)-like Pathological Damages and Constitutes an Antiviral Target. *Cell Res* 31, 847–860. doi:10.1038/s41422-021-00519-4

Yoshimoto, F. K. (2020). The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or N-COV19), the Cause of COVID-19. *Protein J.* 39 (3), 198–216. doi:10.1007/s10930-020-09901-4

Yuen, C.-K., Lam, J.-Y., Wong, W.-M., Mak, L.-F., Wang, X., Chu, H., et al. (2020). SARS-CoV-2 Nsp13, Nsp14, Nsp15 and Orf6 Function as Potent Interferon Antagonists. *Emerging Microbes Infections* 9 (1), 1418–1428. doi:10.1080/22221751.2020.1780953

Zemella, A., Thoring, L., Hoffmeister, C., and Kubick, S. (2015). Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem* 16 (17), 2420–2431. doi:10.1002/cbic.201500340

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-ketoamide Inhibitors. *Science* 368 (6489), 409–412. doi:10.1126/science.abb3405

Zhang, Y., Chen, Y., Li, Y., Huang, F., Luo, B., Yuan, Y., et al. (2021). The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Down-Regulating MHC-I. *Proc. Natl. Acad. Sci. U.S.A.* 118 (23), e2024202118. doi:10.1073/pnas.2024202118

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 | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fbioe.2022.896763

A Cell-free Expression Pipeline for the Generation and Functional Characterization of Nanobodies

Lisa Haueis[1,2†], *Marlitt Stech*[1†] *and Stefan Kubick*[1,3,4*]

[1]Fraunhofer Institute for Cell Therapy and Immunology (IZI), Branch Bioanalytics and Bioprocesses (IZI-BB), Potsdam, Germany, [2]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany, [3]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany, [4]Faculty of Health Sciences, Joint Faculty of the Brandenburg University of Technology Cottbus—Senftenberg, The Brandenburg Medical School Theodor Fontane and the University of Potsdam, Potsdam, Germany

Cell-free systems are well-established platforms for the rapid synthesis, screening, engineering and modification of all kinds of recombinant proteins ranging from membrane proteins to soluble proteins, enzymes and even toxins. Also within the antibody field the cell-free technology has gained considerable attention with respect to the clinical research pipeline including antibody discovery and production. Besides the classical full-length monoclonal antibodies (mAbs), so-called "nanobodies" (Nbs) have come into focus. A Nb is the smallest naturally-derived functional antibody fragment known and represents the variable domain (V_HH, ~15 kDa) of a camelid heavy-chain-only antibody (HCAb). Based on their nanoscale and their special structure, Nbs display striking advantages concerning their production, but also their characteristics as binders, such as high stability, diversity, improved tissue penetration and reaching of cavity-like epitopes. The classical way to produce Nbs depends on the use of living cells as production host. Though cell-based production is well-established, it is still time-consuming, laborious and hardly amenable for high-throughput applications. Here, we present for the first time to our knowledge the synthesis of functional Nbs in a standardized mammalian cell-free system based on Chinese hamster ovary (CHO) cell lysates. Cell-free reactions were shown to be time-efficient and easy-to-handle allowing for the "on demand" synthesis of Nbs. Taken together, we complement available methods and demonstrate a promising new system for Nb selection and validation.

Keywords: cell-free protein synthesis, *In vitro* transcription/translation, nanobody, VHH, camelid, CHO cell lysate

OPEN ACCESS

Edited by:
Yong-Chan Kwon,
Louisiana State University,
United States

Reviewed by:
Javin P. Oza,
California Polytechnic State University,
United States
Joongoo Lee,
Pohang University of Science and Technology, South Korea

***Correspondence:**
Stefan Kubick
Stefan.Kubick@izi-bb.fraunhofer.de

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *15 March 2022*
Accepted: *11 April 2022*
Published: *28 April 2022*

Citation:
Haueis L, Stech M and Kubick S (2022) A Cell-free Expression Pipeline for the Generation and Functional Characterization of Nanobodies. Front. Bioeng. Biotechnol. 10:896763. doi: 10.3389/fbioe.2022.896763

INTRODUCTION

Nanobodies (Nbs) have a molecular mass of approximately 12–15 kDa and are the smallest natural antigen-binding entities known so far (Genst et al., 2006). They originate from unique functional heavy-chain-only antibodies (HCAbs) circulating in the blood of Camelidae. Unlike conventional antibodies which consist of two identical heavy and two identical light chains forming a heterotetramer, HCAbs are devoid of light chains and consist exclusively of two heavy chains which assemble to a homodimer (Hamers-Casterman et al., 1993) (**Figure 1**). The variable antigen binding domain of these HCAbs, the so-called V_HH domain or Nb, can be recombinantly expressed as separate entity, while retaining full antigen binding capacity (van der Linden et al., 1999). Whereas the term nanobody© is the registered trademark of Ablyx (Ghent, Belgium), V_HHs are also known as single-domain antibodies (sdAbs).

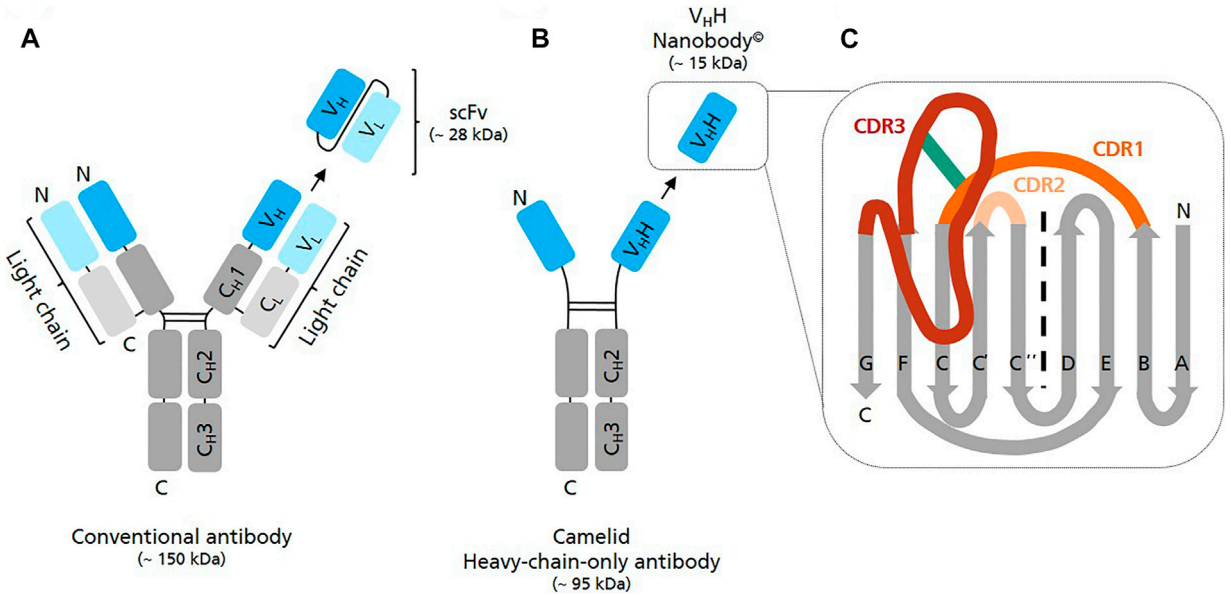


FIGURE 1 | Graphical representation of different antibody structures and recombinantly expressible formats. **(A)** A conventional full length IgG antibody consists of two identical light (variable—V_L and constant—C_L domain) and two identical heavy polypeptide chains (variable V_H and constant—$C_H1/2/3$ domains) which assemble into a heterotetramer. Single chain variable fragments (scFv) are recombinant binders consisting of the variable domain of antibody heavy chain (V_H) and antibody light chain (V_L) fused by a flexible peptide linker. **(B)** A camelid heavy-chain-only antibody (HCAb) consists of two identical heavy polypeptide chains (variable—V_HH and constant $C_H2/3$ domains) forming a homodimer. The antigen binding region of HCAbs is called V_HH domain or Nanobody. **(C)** Schematic view of the V_HH domain consisting of 2 ß-sheets, one with 4 and one with 5 ß-strands. Scaffold ß strands are depicted in grey (A, B, E, D, C″, C′, C, F, G), antigen-binding loops in dark orange (CDR1), light orange (CDR2) and in red (CDR3). The frequently observed interloop disulfide bond betwenn CDR1 and CDR3 is depicted in green.

In comparison to other monomeric binders such as affibodies, monobodies or anticalins Nbs have the advantage that they originate from naturally-occurring antibodies whose genes have evolved over millions of years (Flajnik et al., 2011). Accordingly, Nbs are extremely robust, highly soluble and strictly monomeric thereby representing well optimized high-affinity binders (Muyldermans 2013). In comparison to single-chain variable fragments (scFv), which are also relatively small binders (~25 kDa) but derived from conventional antibodies (~160 kDa), they are less prone to denaturation/aggregation and spontaneous dimerization (Muyldermans 2021a).

Nbs have found numerous applications as research tools, in diagnostics and also therapeutics (Muyldermans 2021b). For example, Nbs contributed to the elucidation of multiple G protein coupled receptor (GPCR) structures, as they can bind conformational and non-linear epitopes better than conventional antibodies (Rasmussen et al., 2011; Groof et al., 2019). Further, Nbs have been shown to be excellent tools for the generation of intrabodies targeting intracellular structures (reviewed in (Soetens et al., 2020; Wagner and Rothbauer 2020). Nb-based intrabodies have been used as imaging probes (Rothbauer et al., 2006; Traenkle and Rothbauer 2017), biosensors (Irannejad et al., 2013; Stoeber et al., 2018; Cao et al., 2019) and even as intracellular modulators of signaling pathways and cellular targets (van Impe et al., 2013; Gulati et al., 2018; Singh et al., 2018) (reviewed in (Wagner and Rothbauer 2020). Moreover Nbs have been shown to be well-suited for *in vivo* imaging approaches as they show superior properties with regard to tissue distribution, rapid tumor accumulation, tumor penetration and fast clearance from the blood circulation (Huang et al., 2008; Chanier and Chames 2019). One example is a ^{68}Ga-labeled HER2-targeting Nb which is in clinical trials for imaging of HER2 in breast cancer patients (Keyaerts et al., 2016). Caplacizumab is the first EMA and FDA-approved Nb specific for von Willebrand factor for the treatment of acquired thrombotic thrombocytopenic purpura (Callewaert et al., 2012; Duggan 2018; Scully et al., 2019).

Due to their small size and the lack of posttranslational modifications, Nbs are easy to tailor and can be cost-efficiently produced in microbial systems including *Escherichia coli* (*E. coli*) and yeast with high yields (reviewed in (Marco 2020)). In the last decade, cell-free translation systems have become popular as they allow for the efficient, versatile and high-throughput-compatible synthesis of a wide variety of target proteins *in vitro* and independently of living cells (Brookwell et al., 2021). In addition, cell-free systems have demonstrated great potential for the *in vitro* evolution of proteins and in particular the selection of antibodies based on *in vitro* display technologies such as mRNA (Roberts and Szostak 1997), cDNA (Yamaguchi et al., 2009) and ribosome display (Hanes and Plückthun 1997; He and Taussig 1997). *In vitro* display based on *E. coli* (Doshi et al., 2014; Chen et al., 2021), rabbit reticulocyte (Fenderico et al., 2019) and *Leishmania torentolae* (Bencurova et al., 2015) cell-free systems has already demonstrated its feasibility for Nb enrichment and identification. Using a eukaryotic cell-free system for *in vitro* translation and evolution of proteins is

tempting because these systems are known for a lower RNase activity and superior capacities to fold and functionally display eukaryotic proteins (Hanes et al., 1999). For these reasons we were highly encouraged to probe the capabilities of the previously published mammalian cell-free system based on translationally active Chinese hamster ovary (CHO) cells (Brödel et al., 2013; Brödel et al., 2014) for Nb synthesis *in vitro*. Cell-free synthesis of conventional antibodies and scFv fragments was already reported by using this system; however, functionally active binders were only obtained upon their signal-peptide induced translocation into the lumen of the CHO lysate-contained endogenous microsomes (Stech et al., 2017).

The aim of the present study was to investigate first, if functionally active Nbs could be synthesized within the CHO cell-free system at all and second, to analyze if the translocation step was required to achieve functionality. As model proteins, we chose two Nbs, one being specific for human epidermal growth factor receptor (EGFR) (Schmitz et al., 2013), while the other one was recognizing green fluorescent protein (GFP) (Kubala et al., 2010). In order to find the best conditions for cell-free Nb synthesis with regard to protein yield and functionality, we analyzed each Nb with and without melittin signal sequence and investigated the Nbs in the "cytosolic" fraction of the lysate, which comprised the non-translocated Nbs, and in the microsomal fraction, comprising co-translationally translocated Nbs. Importantly, our results show that the Nbs we analyzed did not require a translocation step to achieve binding activity. Thus, we present the CHO cell-free system as novel platform for the "on demand" synthesis and analysis of soluble and functionally active Nbs, thereby providing a time-saving and convenient alternative to classical cell-based production.

MATERIALS AND METHODS

Template Generation

In this study two different Nbs known from the literature were used as models to analyze the feasibility of the CHO cell-free system for Nb synthesis and functional analysis. First, a Nb specific for GFP was chosen as model. This Nb is known to bind specifically to GFP and YFP, but without affecting the fluorescence of these reporter proteins (Kubala et al., 2010). Secondly, an EGFR-specific Nb (7D12) was chosen according to Schmitz et al. (2013). V_HH domain 7D12 is reported to bind to EGFR domain III and to block ligand binding to EGFR similar to the monoclonal antibody cetuximab (Roovers et al., 2011; Schmitz et al., 2013). Coding sequences of the anti-GFP-Nb and the anti-EGFR-Nb were codon-optimized for *Cricetulus griseus* and *homo sapiens*, respectively, and equipped with the regulatory sequences necessary to enable *in vitro* transcription and translation according to Stech et al., 2017 (5′ untranslated region (UTR): T7 promotor sequence, multiple cloning site (MCS), internal ribosomal entry site (IRES) from the intergenic region (IGR) of the Cricket paralysis virus (CrPV), GCT as start codon; 3′ UTR: T7 terminator sequence, MCS) (Stech et al., 2017). Different templates were generated per Nb candidate: One with melittin signal sequence (MKFLVNVALVFMVVYISYIYAD (Tessier et al., 1991)), designated NCM-anti-GFP-Nb and NCM-anti-EGFR-Nb, and the other without melittin signal sequence, designated NC-anti-GFP-Nb and NC-anti-EGFR-Nb. The signal sequence of honey bee mellittin was chosen, as it has been shown before to allow for efficient translocation into microsomal vesicles in various cell-free systems (Brödel et al., 2013). Further, Nb sequences were C-terminally fused to a c-Myc and 6x-His-Tag sequence to enable detection by anti-c-Myc monoclonal antibodies and for purification purposes, respectively. DNA templates were synthesized *de novo* and cloned into appropriate vectors (pUC57-1.8k) by Biocat GmbH (Biocat GmbH, Heidelberg). Plasmid preparations for cell-free protein synthesis were prepared using the PureLink®HiPure Plasmid Midiprep Kit (Thermo Fisher Scientific, Waltham, United States) according to the manufacturer's instructions and subsequently control digested as well as sequenced to verify the correct DNA sequence (LGC Genomics GmbH, Berlin, Germany).

Cell-free Protein Synthesis

Coupled transcription-translation reactions were performed according to (Brödel et al., 2013; Thoring and Kubick 2018). In brief, translation reactions were composed of 40% (v/v) S7 nuclease-treated CHO lysate containing endogenous microsomal vesicles originating from the ER, HEPES-KOH (pH 7.6, f.c. 30 mM, BioMol GmbH, Hamburg, Germany), complete amino acids (f.c. 100 µM), $Mg(OAc)_2$ (f.c. 3.9 mM), KOAc (f.c. 135 mM, Merck, Darmstadt, Germany), spermidine (f.c. 0.25 mM, Sigma-Aldrich, St. Louis, United States), energy components (f.c. 1.75 mM ATP, f.c. 0.3 mM GTP, f.c. 0.3 mM CTP, f.c. 0.3 mM UTP), 0.1 mM (f.c.) m7G(ppp)G (Prof. Edward Darzynkiewicz, Warsaw University, Poland), creatine phospate (f.c. 20 mM), T7 polymerase (f.c. 1 U/µl) (Agilent Technologies, Santa Clara, United States) and ^{14}C-leucine (f.c. of 30 µM; specific radioactivity 46.15 dpm/pmol (PerkinElmer LAS (Germany) GmbH, Rodgau, Germany) in order to allow for the subsequent quantitative and qualitative analysis of cell-free synthesized proteins. Protein synthesis was initiated by addition of DNA template (f.c. 60 ng/µL). Reactions were incubated for 3 h at 30°C at 600 rpm in a standard thermomixer (Eppendorf Thermomixer Comfort). Background translational activity was monitored by performing a translation reaction without supplementation of plasmid, called "no template control" (NTC).

CHO lysates containing endogenous microsomal vesicles derived from the ER were prepared as described previously (Brödel et al., 2013; Thoring and Kubick 2018). In order to promote disulfide bond formation within the cell-free system Dithiothreitol (DTT) which is usually added at 4 mM f.c. was omitted from the lysate preparation procedure. In brief, CHO cells were grown exponentially in well-controlled fermenters at 37°C up to 18×10^6 cells/ml using chemically defined, serum-free media (PowerCHO™ 2 CD Medium, Lonza, Basel, Switzerland). After harvesting the cells by centrifugation at 200 × g for 15 min, pellets were washed twice and resuspended in buffer containing 30 mM HEPES-KOH (pH 7.5) and 100 mM NaOAc. Subsequently, the cell suspension was passed through a 20-

gauge needle using a syringe which resulted in the mechanical disruption of the cells. Nuclei and cell debris were removed by a centrifugation step at 6,500 × g for 10 min. The obtained supernatant was subjected to a gel filtration step using Sephadex G-25 column (GE Healthcare, Freiburg, Germany) equilibrated in a buffer containing 30 mM HEPES-KOH (pH 7.5) and 100 mM. Elution fractions (1 ml each) with an RNA content above an absorbance of 100 at 260 nm were pooled. In order to remove endogenous mRNA, cell lysates were treated with S7 micrococcal nuclease (Roche, Mannheim, Germany) (f.c. 10 U/ml) and $CaCl_2$ (f.c. 1 mM) and incubated for 20 min at room temperature (RT). Micrococcal S7 nuclease was inactivated by adding EGTA (f.c. 6.7 mM) and CHO lysates were further supplemented with creatine kinase (f.c. 100 µg/ml). Lysates were shock frozen in liquid nitrogen and subsequently stored at −80°C until further usage.

After completing the translation reaction, samples were centrifuged at 16,000 × g for 10 min at 4°C in order to separate the microsomes from the soluble fraction of the translation mixture. The resulting supernatant (designated as SN1) was transferred to a fresh reaction tube and stored on ice until further analysis, while the microsomal pellet was resuspended in 1x PBS containing 0.2% n-Dodecyl-β-D-Maltoside (DDM) in order to enable release of translocated and microsome-contained Nbs. Microsomes were resuspended manually by repeated up and down pipetting followed by vortexing and shaking on a vibrax for approximately 45 min. To separate the released proteins from microsomal membrane remnants, a second centrifugation step was performed. The resulting supernatant (designated as SN2) was transferred to a fresh reaction tube and stored on ice until further analysis. Reactions supplemented with ^{14}C-leucine were analyzed by SDS-PAGE followed by autoradiography and liquid scintillation counting, while non-radioactive samples were subjected to functional analysis by enzyme-linked immunosorbent assay (ELISA).

SDS-PAGE and Autoradiography

5 µL aliquots of SN1 or SN2 were mixed with 45 µL ddH_2O and 150 µL pre-cooled acetone. Samples were incubated for a minimum of 15 min on ice and subsequently centrifuged at 16,000 × g and 4°C for 10 min. The resulting supernatant was discarded and the protein pellet was dried for at least 30 min at 45°C and resuspended in 20 µL 1x LDS sample buffer (NuPAGE™ LDS sample buffer (f.c. 1x), Thermo Fisher Scientific, Waltham, United States) supplemented with 50 mM (f.c.) DTT (Roche). Samples were heated for 10 min at 70°C and analyzed by sodium dodecylsulfate (SDS) polyacrylamide gel electrophoresis (PAGE) using self-made 14% Tris-Glycin gels (SureCast Gel Handcast System, Thermo Fisher Scientific, Waltham, United States) and Tris-Glycin running buffer. The electrophoretic separation was conducted at 155 V for 65 min. SDS-PAGE gels were subsequently stained using Coomassie blue solution (SimplyBlue SafeStain, Thermo Fisher Scientific, Waltham, United States) and dried on Whatman paper for 70 min at 70 °C (Unigeldryer, 3545D, Uniequip Laborgerätebau-und Vertriebs GmbH, Planegg, Germany). ^{14}C-leucine-labeled protein bands were visualized using the Amersham RGB Imager (GE healthcare) after incubation of dried gels for at least three days on a phosphor screen (GE healthcare).

Quantitative Protein Analysis

Protein yields of *de novo* synthesized ^{14}C-leucine labeled Nbs were determined by liquid scintillation counting as described previously (Stech et al., 2012). In brief, 5 µL aliquots of SN1 or SN2 were precipitated in 3 ml of 10% (v/v) TCA—2% (v/v) casein hydrolysate (Carl Roth GmbH, Karlsruhe, Germany), boiled for 15 min at 80°C and subsequently cooled on ice for at least 30 min. Protein solutions were filtered using a vacuum filtration system (Hoefer, Holliston, United States) and concentrations of ^{14}C-leucine labeled proteins which were retained on the membrane filters (VWR, Darmstadt, Germany) were calculated based on liquid scintillation counting using the Hidex SL600 scintillation counter.

ELISA

Binding activities of cell-free synthesized Nbs in translation mixture fraction SN1 and SN2 were analyzed by ELISA. For the analysis of the anti-EGFR-Nb 96-well microtitre plates (Greiner Bio-one, Kremsmünster, Austria) were coated over night at 4°C with 50 µL/well recombinant EGFR extracellular domain (ECD) (f.c. 0.7 µg/ml in 1x PBS, pH 6). Recombinantly expressed EGFR ECD was purchased from Acro biosystems (Newark, United States) (EGFR, ECD L25-S645, His-Tag, 70.5 kDa, ABIN6253725). Accordingly, 96-well microtitre plates were coated with 50 µL/well of recombinant GFP (f.c. 1.3 µg/ml in sodium carbonate buffer, pH 9.6) for the analysis of the anti-GFP-Nb. Recombinant GFP-GST was kindly provided by the preclinics Gesellschaft für präklinische Forschung mbH (Potsdam, Germany). Subsequently, ELISA plates were washed 3x using 1x PBS-T containing 0.05% Tween-20 (200 µL/well). For the anti-EGFR-Nb all of the applied buffers were adjusted to pH 6, while for the anti-GFP-Nb the pH was adjusted to 7.4. Plates were blocked with 1xPBS containing 2% BSA (200 µL/well) for 2 h at RT and 250 rpm. After washing the plates 3x with washing buffer, wells were incubated for 2 h at RT and 250 rpm with translation mixture fractions SN1 and SN2 containing cell-free synthesized Nbs in serial dilutions (50 µL/well) using 1x PBS containing 1% BSA for SN1 and 1x PBS containing 1% BSA and 0.2% DDM for SN2 as dilution buffer. Plates were washed again 3x and incubated with mouse anti-c-Myc-tag antibody (Life technologies, Carlsbad, United States) diluted 1:1000 in PBS containing 1% BSA in (50 µL/well) for 2.5 h at RT and 250 rpm. Upon washing the plates for 3x, plates were incubated with anti-mouse-IgG, HRP-linked antibody (Cell signaling, Danvers, United States and advansta, San Jose, United States) diluted 1:3000 in PBS containing 1% BSA (50 µL/well) for 1 h at RT and 250 rpm. After washing 3x, bound antibodies were detected by adding TMB substrate solution (Life technologies, Carlsbad, United States) (50 µL/well). Color development was stopped by addition of 0.5 M H_2SO_4 (50 µL/well) after incubating for 10–15 min Absorbance was measured at 450 and 620 nm (as reference)

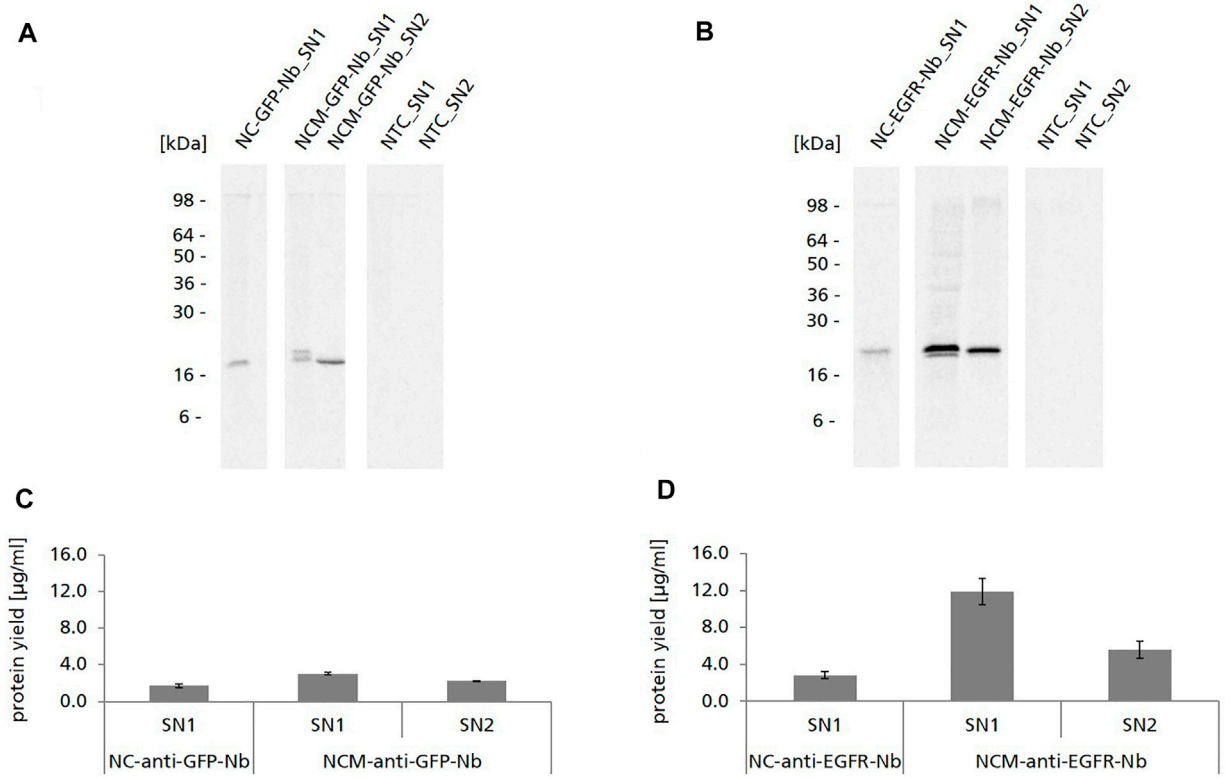


FIGURE 2 | Cell-free synthesis of anti-GFP-Nb and anti-EGFR-Nb based on CHO lysate. **(A,B)** Autoradiographs showing ^{14}C-leucine labeled Nanobodies in both soluble fractions of the translation mixture (SN1, SN2). **(C,D)**. Quantitative analysis of cell-free synthesized Nbs was performed by liquid scintillation counting. Standard deviations were calculated from quadruplicate analysis.

using the FLUOstar Omega (BMG Labtech, Ortenberg, Germany).

RESULTS

In order to analyze the influence of a signal peptide induced translocation of *de novo* synthesized Nbs to the lumen of the endogenous microsomes contained in the CHO lysate, coding sequences of anti-GFP-Nb and anti-EGFR-Nb with melittin signal sequence (NCM-anti-GFP-Nb/NCM-anti-EGFR-Nb) and without signal sequence (NC-anti-GFP-Nb/NC-anti-EGFR-Nb) were compared regarding Nb yield and functionality. Nb yield and binding activity were analyzed in two different preparations of the completed cell-free reaction, designated as SN1 and SN2. Nbs which have not been translocated to the microsomes, either because of the lack of a signal sequence (NC constructs) or because of insufficient translocation, were expected in translation mixture fraction SN1, while Nbs which have been translocated to the lumen of the microsomes were expected in fraction SN2 upon detergent-mediated microsome disintegration.

Nb constructs without signal sequence were detected as one distinct protein band in the autoradiograph showing the expected molecular mass (~15 kDa for NC-anti-GFP-Nb, ~18 kDa for NC-anti-EGFR-Nb) (**Figures 2A,B**). Nb constructs with signal sequence (NCM-anti-GFP-Nb/NCM-anti-EGFR-Nb) were detected as double band in SN1, while in SN2 only one band appeared which was on the level of the lower band of SN1. To exclude glycosylation as a reason for the observed double band, samples were digested with endoglycosidase H (Endo H) and Peptide N-glycosidase F (PNGase F) and analyzed by autoradiography (data not shown). As expected, the double band did not disappear by the digestion suggesting that it represented a mixture of Nb species with signal sequence and without, probably as a consequence of incomplete signal peptide cleavage within the cell-free system. The melittin signal peptide has a molecular mass of 2.5 kDa which corresponded to the band shift detected in the autoradiograph (**Figures 2A,B**).

For both Nbs we observed significantly higher protein yields when using the constructs with melittin signal sequence in comparison to the constructs without signal sequence (~3 µg/ml NCM-anti-GFP-Nb and ~12 µg/ml NCM-anti-EGFR-Nb *vs.* ~1.7 µg/ml NC-anti-GFP-Nb and ~2.9 µg/ml for NC-anti-EGFR-Nb) (**Figures 2C,D**). Strikingly, protein yields of the anti-GFP-Nb were considerably lower compared to the yields of the anti-EGFR-Nb. In general, protein yields of the anti-GFP-Nb were lower than expected for the CHO cell-free system which is reputed for its high productivity (Thoring et al., 2017). Interestingly, also different optimization strategies such as

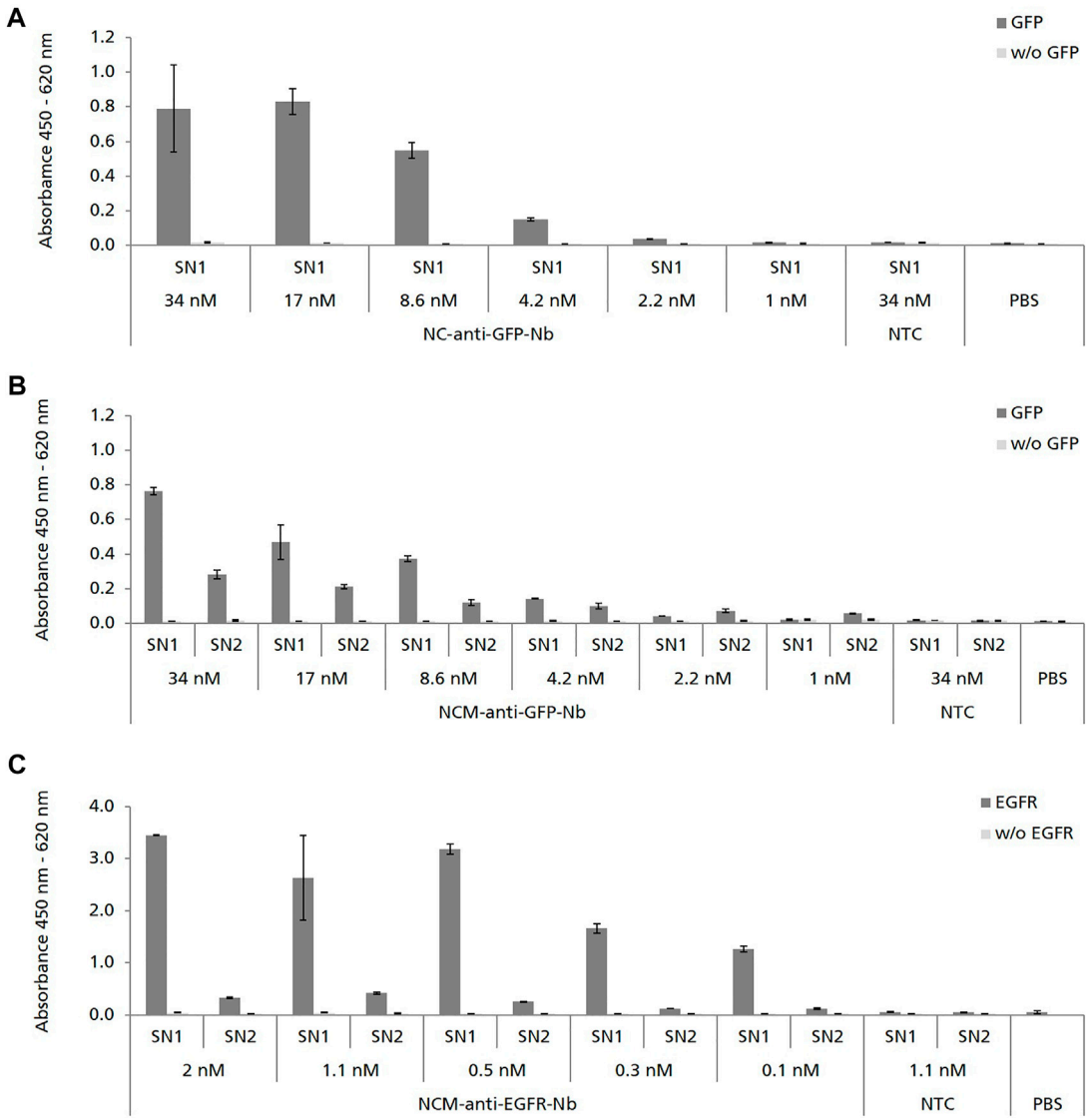


FIGURE 3 | ELISA analysis showing the specific binding of cell-free synthesized Nbs to their corresponding antigens. Nbs without signal sequence (NC) were analyzed in SN1 (comprising non-translocated Nbs). Nbs with signal sequence (NCM) were analyzed in SN1 and additionally in SN2 (comprising Nbs which were solubilized from the lumen of the endogenous CHO microsomes). No template control (NTC) and PBS were analyzed in parallel as negative controls. To monitor background binding Nb samples were also analyzed on wells coated without (w/o) antigen. **(A,B)** Binding of anti-GFP-Nb to its antigen green fluorescent protein (GFP) was analyzed in serial dilutions in the range of 34–1 nM. **(C)** Binding of anti-EGFR-Nb to epidermal growth factor receptor (EGFR) was analyzed in serial dilutions in the range of 2–0.1 nM. Standard deviations were calculated from quadruplicate analysis **(A,B)** and duplicate analysis **(C)**.

variation of the template concentration, incubation temperature and time as well as the addition of supplements (PolyG) did not result in considerably higher protein yields for this particular Nb (data not shown).

We next analyzed the binding capabilities of cell-free synthesized Nbs by using ELISA. NC-anti-GFP-Nb was analyzed in SN1, while NCM-anti-GFP-Nb was analyzed in both, SN1 and SN2 in serial dilutions starting at 34 nM cell-free synthesized Nb. ELISA analysis revealed specific binding of NC-anti-GFP-Nb and NCM-anti-GFP-Nb in SN1 to its antigen GFP (**Figures 3A,B**). The binding signal was shown to be dependent on the applied Nb concentration as the OD values decreased with increasing dilution of the Nb sample. Although applied in the same molar concentration NCM-anti-GFP-Nb in SN2 showed a lower binding signal compared to NCM-anti-GFP-Nb in SN1 (**Figure 3B**).

Due to the observation that higher protein yields were observed for the constructs with melittin signal sequence, NCM-anti-EGFR-Nb was preferred over NC-anti-EGFR-Nb (**Figure 2D**). Consistent with the ELISA data obtained for NCM-anti-GFP-Nb, melittin signal peptide fused anti-EGFR-Nb showed a concentration-dependent specific binding signal in SN1, while a lower signal was observed for SN2 (**Figure 3C**).

Taken together, our results demonstrate the specific binding of cell-free synthesized, melittin signal peptide fused anti-GFP-Nb and anti-EGFR-Nb to their corresponding antigens. Although the presence of the signal sequence was not mandatory for Nb functionality, it was shown to increase the synthesis efficiency.

DISCUSSION

In view of the therapeutic potential of Nbs and upcoming strategies to further functionalize Nbs eukaryotic selection and production systems are of highest relevance. Mammalian systems in particular are more demanding and expensive, and therefore their application for Nb expression must be justified. This is for example the case if the Nb is fused to the Fc part of IgG antibodies to restore ADCC (antibody dependent cell-mediated cytotoxicity) and CDC (complement-dependent cytotoxicity) effector functions *in vivo* requiring glycosylation of N297 in the C_H2 domain and to increase affinity by avidity (Bobkov et al., 2018; Godakova et al., 2019). In addition, the generation of Nbs fused to horse radish peroxidase (HRP) may justify the use of a mammalian expression system, since HRP is known to be difficult-to-express in bacteria (Sheng et al., 2019; Marco 2020). A further indication can be seen when Nbs are produced as fusions with proteinacous toxin moieties (Yu et al., 2017) which might demand for an eukaryotic and even mammalian expression system. Further, the development of intrabodies requires efficient expression of Nbs in mammalian systems as intrabodies must be able to maintain functionality even when they are located in the reducing environment of the cytoplasm (Wagner and Rothbauer 2020). In order to meet the demand for efficient mammalian Nb selection and production systems, we focused on the well-established CHO cell-free system (Brödel et al., 2013; Thoring et al., 2016). As previously reported, the CHO cell-free system allows for the folding and assembly of complex proteins (Thoring et al., 2017; Zemella et al., 2019) and protein multimers (Thoring et al., 2017; Ramm et al., 2020; Ramm et al., 2021) and enables several types of posttranslational modifications on target proteins, such as N-glycosylation (Thoring and Kubick 2018), lipidation (Fenz et al., 2014), phosphorylation (Thoring et al., 2016) and disulfide bridge formation (Martin et al., 2017; Stech et al., 2017). Though Nbs are not posttranslationally modified, except for their conserved intradomain disulfide bridge, the CHO cell-free system offers the advantages of cell-free systems in general: speed, open reaction design, high-throughput and on-demand synthesis (Thoring et al., 2016). At the same time, it provides a eukaryotic folding apparatus, which can be beneficial for diverse applications, such as the *in vitro* evolution of Nb fusion proteins and intrabodies as described above.

Here we show the proof-of-concept that functional Nbs can be synthesized within a 3 h incubation time by using the cell-free CHO system. As far as we know this is the first time that functional Nbs have been synthesized in a eukaryotic cell-free system apart from rabbit reticulocyte and *Leishmania* systems. The cell-free CHO system was capable of producing the selected model Nbs in yields up to ~12 μg/ml (**Figure 2**) which was sufficient for a straight-forward functional analysis of Nbs by ELISA, directly following cell-free synthesis and without requiring a purification step. In comparison, scFv and Fab fragments have been produced in yields of 50–200 μg/ml in a cell-free *E. coli* system (Wu et al., 2022). Unlike *E. coli*, the cell-free reticulocyte system is related to rather low protein yields (Carlson et al., 2012; Gregorio et al., 2019). However, reticulocyte lysate systems are based on an ethically questionable preparation strategy as the lysates are prepared from rabbits which have been made anemic and bled by heart puncture (Darbnbrough et al., 1973). In contrast, the mammalian cell-free system used here is prepared from the immortalized CHO-K1 cell line which can be cultivated in bioreactors using chemically defined and serum-free media, underlying a very strict process control (Thoring and Kubick 2018; Brödel et al., 2013). Based on the results we obtained by ELISA (**Figure 3**) the EC_{50} value of cell-free synthesized NCM-anti-EGFR-Nb analyzed in SN1 was estimated to be in the lower nanomolar range which corresponded to the data obtained for the cell-based produced Nb (Gainkam et al., 2008; Roovers et al., 2011). However, binding activities of cell-free versus cell-based produced Nbs were not compared side-by-side, which does not allow for a valid conclusion on that point.

The CHO cell-free system used in this study contained endogenous microsomal vesicles derived from the ER of the cells used for lysate preparation. By using an appropriate signal sequence target proteins can be co-translationally translocated to the lumen of these microsomes or in the case of membrane proteins embedded into the lipid bilayer (Fenz et al., 2014). In this study we found that the fusion of the melittin signal sequence to the Nb led to translocation to the lysate-contained microsomes. However, translocation was not mandatory for Nb functionality in contrast to scFv and IgG antibody formats synthesized in a *Spodoptera frugiperda* 21 (*Sf*21) and CHO cell-free system (Stech et al., 2017; Stech et al., 2014). In fact, translocated and microsome released Nbs were found to be less active compared to non-translocated Nbs (**Figure 3**). On the other hand the fusion to the signal sequence exhibited a positive effect on the protein yields of cell-free synthesized Nbs (1.8-fold for anti-GFP-Nb and 4-fold for anti-EGFR-Nb in comparison to the Nb without signal sequence) (**Figure 2**). It is known that signal sequences can stimulate protein synthesis in *E. coli* cell-free systems (Ahn et al., 2006) and the same observation has been made for a microsome-containing *Sf*21 cell-free system (Sonnabend et al., 2015). We provide evidence that this holds true also for the CHO cell-free system. It can be speculated that the presence of the melittin signal sequence at the 5′ end of the Nb gene stabilized translation initiation and translational processivity, which resulted in increased protein yields.

In previous reports, the CHO cell-free system has been shown to enable N-glycosylation of target proteins, which frequently resulted in the detection of glycosylated and non-glycosylated protein species (Brödel et al., 2014; Thoring et al., 2016). However, Nbs in general were not expected to be glycosylated and accordingly, the particular Nbs we analyzed did not contain a N-glycosylation consensus sequence (N-X-S/T) within their coding sequence. Referring to previous studies (Kubick et al., 2009) there is strong evidence that we observed signal peptide

cleavage within the CHO cell-free system as we detected a single lower band for the cell-free synthesized signal peptide fused Nbs in SN2, but a double band in SN1. Signal peptide cleavage seemed to be almost complete for the translocated Nbs located in SN2, while in SN1 a mixture of cleaved and non-cleaved Nbs was detected (**Figure 2**). The observation that Nbs with cleaved signal peptide were detected in SN1 may be explained by the presence of small-sized microsomal vesicles which were not depleted during the preparation of fraction SN1 by centrifugation as previously reported by Thoring et al. (Thoring et al., 2017). In accordance to previous studies we observed that the translation efficiency of the cell-free system exceeded its capacity for protein translocation to the microsomal vesicles as a considerable amount of melittin signal peptide fused Nbs was detected in lysate fraction SN1 (Stech et al., 2017). For the future, it seems evident that for cell-free Nb synthesis CHO lysates can be used which are depleted from their endogenous microsomes. Alternatively, the lysate preparation procedure could be modified in a way, to quantitatively deplete the microsomes from the lysate.

The disulfide bond within the V_HH framework region is evolutionally conserved and has been shown to be critical for the thermal stability of the V_HH. However, it is not strictly required for maintaining antigen binding ability (Rudikoff and Pumphrey 1986; Pleiner et al., 2015). Several reports were published demonstrating that V_HH retained antigen binding affinity even after removal of the canonical disulfide (Akazawa-Ogawa et al., 2016; Li et al., 2012; Liu et al., 2019). In addition, the production of functional, disulfide bond-independent Nbs (intrabodies) in the reducing environment of the *E. coli* cytoplasm for intrabody application has been demonstrated (Even-Desrumeaux et al., 2010; Olichon and Surrey 2007). The CDR3 of the anti-EGFR-Nb analyzed in this study is stabilized by van der Waals contacts and hydrogen bonds, but not by the interloop disulfide bond (**Figure 1**) which is frequently present in Nbs (Schmitz et al., 2013). The fact that the paratope of the anti-EGFR-Nb was found to be stable without the disulfide bond might be an explanation why this Nb was functional even without translocation to the microsomes. More difficult are conclusions on the meaning of disulfide bond formation within the anti-GFP-Nb. Kubala *et al.* revealed the structure of the GFP:GFP-Nb complex by X-ray crystallography, but due to the presence of the reducing agent DTT in the protein solution they found GFP-Nb with oxidized and reduced cysteines (Kubala et al., 2010). The CHO cell-free system used within this study was not specifically redox-optimized, but the reducing agent DTT was omitted from the coupled transcription-translation reaction. Clearly, a larger number of V_HH candidates needs to be analyzed for a definite conclusion whether the conditions in the applied CHO cell-free system are sufficient to support disulfide bond formation within V_HH in general. Balancing the redox potential in cell-free systems, e.g. by adding glutathione and/or disulfide bond isomerases and other chaperones to promote correct disulfide bond formation and isomerization, is common practice in eukaryotic (Bulleid et al., 1992; Kawasaki et al., 2003; Ezure et al., 2007; Stech et al., 2012; Martin et al., 2017) and prokaryotic (Ryabova et al., 1997; Bundy and Swartz 2011) cell-free systems. It is therefore very likely that these options will also be applicable to the CHO cell-free system used in this study. Based on the possibility to fine-tune the redox potential in open cell-free systems, it is conceivable that our system could be helpful to screen in particular for V_HH which are intrinsically stable high affinity binders even in the absence of a disulfide bond which can be in particular useful for the *in vitro* evolution of intrabodies.

To summarize, here we demonstrate the proof-of-concept that soluble and functionally active Nbs can be synthesized using a CHO cell-free system. Due to the tremendously shortened time for Nb synthesis (in the batch-based system only 3 h) and the option for reaction parallelization, information on the binding activity of a multitude of different Nbs can be obtained within one day of work by using the proposed system. Thus, we present the CHO cell-free system as timesaving alternative to classical cell-based approaches. For the future we hypothesize that the system can become a valuable new source for *in vitro* display technologies enabling the *in vitro* selection of novel Nb sequences against desired targets. In particular, we envision the mammalian CHO cell-free system as a promising technology platform when eukaryotic and even mammalian systems are required to set up an effective screening and production pipeline for difficult-to-express Nb fusion proteins. The fruitful combination of *in vitro* display-based antibody selection technologies on cell-free synthesized difficult-to-express targets has been recently demonstrated (Nakai et al., 2022) emphasizing its great potential not only for antibody selection but also for difficult-to-express targets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LH and MS contributed to conception and design of the study. LH planned and performed the experiments. MS wrote the first draft of the manuscript. LH wrote sections of the manuscript. SK provided funding. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the European Regional Development Fund (EFRE) (ILB, ProFIT Brandenburg, No. 80176625) and the German Ministry of Education and Research (BMBF, Nos. 031B0078A and 031B0831C).

ACKNOWLEDGMENTS

The authors would like to thank Doreen Wüstenhagen (Fraunhofer IZI-BB, Potsdam-Golm, Germany) regarding her input on the preparation of microsome containing eukaryotic lysates with respect to cell cultivation and lysate preparation.

Further, we thank Dana Wenzel (Fraunhofer IZI-BB, Potsdam-Golm, Germany) for her excellent support in preparing CHO lysates and Simon Krebs (Fraunhofer-IZI-BB, Potsdam-Golm, Germany) for the careful revision of this manuscript and support in Nb template design. In addition, we want to express our appreciation and thanks to our cooperation partner preclinics Gesellschaft für präklinische Forschung mbH (Potsdam, Germany) for their input regarding the selected model Nbs, fruitful discussions and for providing recombinantly expressed GFP-GST as antigen for the ELISA analysis.

REFERENCES

Ahn, J.-H., Hwang, M.-Y., Lee, K.-H., Choi, C.-Y., and Kim, D.-M. (2006). Use of Signal Sequences as an *In Situ* Removable Sequence Element to Stimulate Protein Synthesis in Cell-free Extracts. *Nucleic Acids Res.* 35 (4), e21. doi:10.1093/nar/gkl917

Akazawa-Ogawa, Y., Uegaki, K., and Hagihara, Y. (2016). The Role of Intra-domain Disulfide Bonds in Heat-Induced Irreversible Denaturation of Camelid Single Domain VHH Antibodies. *J. Biochem.* 159 (1), 111–121. doi:10.1093/jb/mvv082

Bencurova, E., Pulzova, L., Flachbartova, Z., and Bhide, M. (2015). A Rapid and Simple Pipeline for Synthesis of mRNA-Ribosome-VHH Complexes Used in Single-Domain Antibody Ribosome Display. *Mol. Biosyst.* 11 (6), 1515–1524. doi:10.1039/c5mb00026b

Bobkov, V., Zarca, A. M., van Hout, A., Arimont, M., Doijen, J., Bialkowska, M., et al. (2018). Nanobody-Fc Constructs Targeting Chemokine Receptor CXCR4 Potently Inhibit Signaling and CXCR4-Mediated HIV-Entry and Induce Antibody Effector Functions. *Biochem. Pharmacol.* 158, 413–424. doi:10.1016/j.bcp.2018.10.014

Brödel, A. K., Sonnabend, A., and Kubick, S. (2014). Cell-free Protein Expression Based on Extracts from CHO Cells. *Biotechnol. Bioeng.* 111 (1), 25–36. doi:10.1002/bit.25013

Brödel, A. K., Sonnabend, A., Roberts, L. O., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2013). IRES-mediated Translation of Membrane Proteins and Glycoproteins in Eukaryotic Cell-free Systems. *PloS one* 8 (12), e82234. doi:10.1371/journal.pone.0082234

Brookwell, A., Oza, J. P., and Caschera, F. (2021). Biotechnology Applications of Cell-free Expression Systems. *Life* 11 (12), 1367. doi:10.3390/life11121367

Bulleid, N. J., Bassel-Duby, R. S., Freedman, R. B., Sambrook, J. F., and Gething, M. J. (1992). Cell-free Synthesis of Enzymically Active Tissue-type Plasminogen Activator. Protein Folding Determines the Extent of N-Linked Glycosylation. *Biochem. J.* 286 (Pt 1), 275–280. doi:10.1042/bj2860275

Bundy, B. C., and Swartz, J. R. (2011). Efficient Disulfide Bond Formation in Virus-like Particles. *J. Biotechnol.* 154 (4), 230–239. doi:10.1016/j.jbiotec.2011.04.011

Callewaert, F., Roodt, J., Ulrichts, H., Stohr, T., van Rensburg, W. J., Lamprecht, S., et al. (2012). Evaluation of Efficacy and Safety of the Anti-VWF Nanobody ALX-0681 in a Preclinical Baboon Model of Acquired Thrombotic Thrombocytopenic Purpura. *Blood* 120 (17), 3603–3610. doi:10.1182/blood-2012-04-420943

Cao, J., Zhong, N., Wang, G., Wang, M., Zhang, B., Fu, B., et al. (2019). Nanobody-based sandwich Reporter System for Living Cell Sensing Influenza A Virus Infection. *Sci. Rep.* 9 (1), 15899. doi:10.1038/s41598-019-52258-7

Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. (2012). Cell-free Protein Synthesis: Applications Come of Age. *Biotechnol. Adv.* 30 (5), 1185–1194. doi:10.1016/j.biotechadv.2011.09.016

Chanier, T., and Chames, P. (2019). Nanobody Engineering: Toward Next Generation Immunotherapies and Immunoimaging of Cancer. *Antibodies* 8 (1), 13. doi:10.3390/antib8010013

Chen, X., Gentili, M., Hacohen, N., and Regev, A. (2021). A Cell-free Nanobody Engineering Platform Rapidly Generates SARS-CoV-2 Neutralizing Nanobodies. *Nat. Commun.* 12 (1), 5506. doi:10.1038/s41467-021-25777-z

Darbnbrough, C., Legon, S., Hunt, T., and Jackson, R. J. (1973). Initiation of Protein Synthesis: Evidence for Messenger RNA-independent Binding of Methionyl-Transfer RNA to the 40 S Ribosomal Subunit. *J. Mol. Biol.* 76 (3), 379–403. doi:10.1016/0022-2836(73)90511-1

De Genst, E., Saerens, D., Muyldermans, S., and Conrath, K. (2006). Antibody Repertoire Development in Camelids. *Developmental Comp. Immunol.* 30 (1-2), 187–198. doi:10.1016/j.dci.2005.06.010

De Groof, T. W. M., Bobkov, V., Heukers, R., and Smit, M. J. (2019). Nanobodies: New Avenues for Imaging, Stabilizing and Modulating GPCRs. *Mol. Cell. Endocrinol.* 484, 15–24. doi:10.1016/j.mce.2019.01.021

de Marco, A. (2020). Recombinant Expression of Nanobodies and Nanobody-Derived Immunoreagents. *Protein Expr. Purif.* 172, 105645. doi:10.1016/j.pep.2020.105645

Doshi, R., Chen, B. R., VibatCecile Rose, C. R. T., Huang, N., Lee, C.-W., and Chang, G. (2014). *In Vitro* nanobody Discovery for Integral Membrane Protein Targets. *Sci. Rep.* 4, 6760. doi:10.1038/srep06760

Duggan, S. (2018). Caplacizumab: First Global Approval. *Drugs* 78 (15), 1639–1642. doi:10.1007/s40265-018-0989-0

Even-Desrumeaux, K., Baty, D., and Chames, P. (2010). Strong and Oriented Immobilization of Single Domain Antibodies from Crude Bacterial Lysates for High-Throughput Compatible Cost-Effective Antibody Array Generation. *Mol. Biosyst.* 6 (11), 2241–2248. doi:10.1039/c005279e

Ezure, T., Suzuki, T., Shikata, M., Ito, M., Ando, E., Nishimura, O., et al. (2007). Expression of Proteins Containing Disulfide Bonds in an Insect Cell-free System and Confirmation of Their Arrangements by MALDI-TOF MS. *Proteomics* 7 (24), 4424–4434. doi:10.1002/pmic.200700774

Fenderico, N., van Scherpenzeel, R. C., Goldflam, M., Proverbio, D., Jordens, I., Kralj, T., et al. (2019). Anti-LRP5/6 VHHs Promote Differentiation of Wnt-Hypersensitive Intestinal Stem Cells. *Nat. Commun.* 10 (1), 365. doi:10.1038/s41467-018-08172-z

Fenz, S. F., Sachse, R., Schmidt, T., and Kubick, S. (2014). Cell-free Synthesis of Membrane Proteins: Tailored Cell Models Out of Microsomes. *Biochim. Biophys. Acta (Bba) - Biomembranes* 1838 (5), 1382–1388. doi:10.1016/j.bbamem.2013.12.009

Flajnik, M. F., Deschacht, N., and Muyldermans, S. (2011). A Case of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels? *Plos Biol.* 9 (8), e1001120. doi:10.1371/journal.pbio.1001120

Gainkam, L. O., Huang, L., Caveliers, V., Keyaerts, M., Hernot, S., Vaneycken, I., et al. (2008). Comparison of the Biodistribution and Tumor Targeting of Two 99mTc-Labeled Anti-EGFR Nanobodies in Mice, Using Pinhole SPECT/micro-CT. *J. Nucl. Med.* 49 (5), 788–795. doi:10.2967/jnumed.107.048538

Godakova, S. A., Noskov, A. N., Vinogradova, I. D., Ugriumova, G. A., Solovyev, A. I., and Esmagambetov, I. B. (2019). Camelid VHHs Fused to Human Fc Fragments Provide Long Term Protection against Botulinum Neurotoxin A in Mice. *Toxins* 11 (8), 464. doi:10.3390/toxins11080464

Gregorio, N. E., Levine, M. Z., and Oza, J. P. (2019). A User's Guide to Cell-free Protein Synthesis. *MPs* 2 (1), 24. doi:10.3390/mps2010024

Gulati, S., Jin, H., Masuho, I., Orban, T., Cai, Y., Pardon, E., et al. (2018). Targeting G Protein-Coupled Receptor Signaling at the G Protein Level with a Selective Nanobody Inhibitor. *Nat. Commun.* 9 (1), 1996. doi:10.1038/s41467-018-04432-0

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E. B., et al. (1993). Naturally Occurring Antibodies Devoid of Light Chains. *Nature* 363 (6428), 446–448. doi:10.1038/363446a0

Hanes, J., Jermutus, L., Schaffitzel, C., and Plückthun, A. (1999). Comparison of *Escherichia coli* and Rabbit Reticulocyte Ribosome Display Systems. *FEBS Lett.* 450 (1-2), 105–110. doi:10.1016/S0014-5793(99)00475-5

Hanes, J., and Plückthun, A. (1997). *In Vitro* selection and Evolution of Functional Proteins by Using Ribosome Display. *Proc. Natl. Acad. Sci. U.S.A.* 94 (10), 4937–4942. doi:10.1073/pnas.94.10.4937

He, M., and Taussig, M. J. (1997). Antibody-ribosome-mRNA (ARM) Complexes as Efficient Selection Particles for *In Vitro* Display and Evolution of Antibody Combining Sites. *Nucleic Acids Res.* 25 (24), 5132–5134. doi:10.1093/nar/25.24.5132

Huang, L., GainkamTchouate, L. O. T., Caveliers, V., Vanhove, C., Keyaerts, M., De Baetselier, P., et al. (2008). SPECT Imaging with 99mTc-Labeled EGFR-specific

Nanobody for *In Vivo* Monitoring of EGFR Expression. *Mol. Imaging Biol.* 10 (3), 167–175. doi:10.1007/s11307-008-0133-8

Irannejad, R., Tomshine, J. C., Tomshine, J. R., Chevalier, M., Mahoney, J. P., Steyaert, J., et al. (2013). Conformational Biosensors Reveal GPCR Signalling from Endosomes. *Nature* 495 (7442), 534–538. doi:10.1038/nature12000

Kawasaki, T., Gouda, M. D., Sawasaki, T., Takai, K., and Endo, Y. (2003). Efficient Synthesis of a Disulfide-Containing Protein through a Batch Cell-free System from Wheat Germ. *Eur. J. Biochem.* 270 (23), 4780–4786. doi:10.1046/j.1432-1033.2003.03880.x

Keyaerts, M., Xavier, C., Heemskerk, J., Devoogdt, N., Everaert, H., Ackaert, C., et al. (2016). Phase I Study of 68Ga-HER2-Nanobody for PET/CT Assessment of HER2 Expression in Breast Carcinoma. *J. Nucl. Med.* 57 (1), 27–33. doi:10.2967/jnumed.115.162024

Kubala, M. H., Kovtun, O., Alexandrov, K., and Collins, B. M. (2010). Structural and Thermodynamic Analysis of the GFP:GFP-nanobody Complex. *Protein Sci.* 19 (12), 2389–2401. doi:10.1002/pro.519

Kubick, S., Gerrits, M., Merk, H., Stiege, W., and Erdmann, V. A. (2009). Chapter 2 *In Vitro* Synthesis of Posttranslationally Modified Membrane Proteins. *Curr. Top. Membranes* 63, 25–49. doi:10.1016/s1063-5823(09)63002-7

Li, T., Bourgeois, J. P., Celli, S., Glacial, F., Le Sourd, A. M., Mecheri, S., et al. (2012). Cell-penetrating anti-GFAP VHH and Corresponding Fluorescent Fusion Protein VHH-GFP Spontaneously Cross the Blood-brain Barrier and Specifically Recognize Astrocytes: Application to Brain Imaging. *FASEB j.* 26 (10), 3969–3979. doi:10.1096/fj.11-201384

Liu, H., Schittny, V., and Nash, M. A. (2019). Removal of a Conserved Disulfide Bond Does Not Compromise Mechanical Stability of a VHH Antibody Complex. *Nano Lett.* 19 (8), 5524–5529. doi:10.1021/acs.nanolett.9b02062

Martin, R. W., Majewska, N. I., Chen, C. X., Albanetti, T. E., Jimenez, R. B. C., Schmelzer, A. E., et al. (2017). Development of a CHO-Based Cell-free Platform for Synthesis of Active Monoclonal Antibodies. *ACS Synth. Biol.* 6 (7), 1370–1379. doi:10.1021/acssynbio.7b00001

Muyldermans, S. (2021a). A Guide to: Generation and Design of Nanobodies. *Febs J.* 288 (7), 2084–2102. doi:10.1111/febs.15515

Muyldermans, S. (2021b). Applications of Nanobodies. *Annu. Rev. Anim. Biosci.* 9, 401–421. doi:10.1146/annurev-animal-021419-083831

Muyldermans, S. (2013). Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* 82, 775–797. doi:10.1146/annurev-biochem-063011-092449

Nakai, H., Isshiki, K., Hattori, M., Maehira, H., Yamaguchi, T., Masuda, K., et al. (2022). Cell-Free Synthesis of Human Endothelin Receptors and its Application to Ribosome Display. *Anal. Chem.* 94, 3831–3839. doi:10.1021/acs.analchem.1c04714

Olichon, A., and Surrey, T. (2007). Selection of Genetically Encoded Fluorescent Single Domain Antibodies Engineered for Efficient Expression in *Escherichia coli*. *J. Biol. Chem.* 282 (50), 36314–36320. doi:10.1074/jbc.M704908200

Pleiner, T., Bates, M., Trakhanov, S., Lee, C.-T., Schliep, J. E., Chug, H., et al. (2015). Nanobodies: Site-specific Labeling for Super-resolution Imaging, Rapid Epitope-Mapping and Native Protein Complex Isolation. *eLife* 4, e11349. doi:10.7554/eLife.11349

Ramm, F., Dondapati, S. K., Thoring, L., Zemella, A., Wüstenhagen, D. A., Frentzel, H., et al. (2020). Mammalian Cell-free Protein Expression Promotes the Functional Characterization of the Tripartite Non-hemolytic Enterotoxin from Bacillus Cereus. *Sci. Rep.* 10 (1), 2887. doi:10.1038/s41598-020-59634-8

Ramm, F., Stech, M., Zemella, A., Frentzel, H., and Kubick, S. (2021). The Pore-Forming Hemolysin BL Enterotoxin from Bacillus Cereus: Subunit Interactions in Cell-free Systems. *Toxins* 13 (11), 807. doi:10.3390/toxins13110807

Rasmussen, S. G. F., Choi, H.-J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., et al. (2011). Structure of a Nanobody-Stabilized Active State of the β(2) Adrenoceptor. *Nature* 469 (7329), 175–180. doi:10.1038/nature09648

Roberts, R. W., and Szostak, J. W. (1997). RNA-peptide Fusions for the *In Vitro* Selection of Peptides and Proteins. *Proc. Natl. Acad. Sci. U.S.A.* 94 (23), 12297–12302. doi:10.1073/pnas.94.23.12297

Roovers, R. C., Vosjan, M. J. W. D., Laeremans, T., el Khoulati, R., de Bruin, R. C. G., Ferguson, K. M., et al. (2011). A Biparatopic Anti-EGFR Nanobody Efficiently Inhibits Solid Tumour Growth. *Int. J. Cancer* 129 (8), 2013–2024. doi:10.1002/ijc.26145

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., et al. (2006). Targeting and Tracing Antigens in Live Cells with Fluorescent Nanobodies. *Nat. Methods* 3 (11), 887–889. doi:10.1038/nmeth953

Rudikoff, S., and Pumphrey, J. G. (1986). Functional Antibody Lacking a Variable-Region Disulfide Bridge. *Proc. Natl. Acad. Sci. U.S.A.* 83 (20), 7875–7878. doi:10.1073/pnas.83.20.7875

Ryabova, L. A., Desplancq, D., Spirin, A. S., and Plückthun, A. (1997). Functional Antibody Production Using Cell-free Translation: Effects of Protein Disulfide Isomerase and Chaperones. *Nat. Biotechnol.* 15 (1), 79–84. doi:10.1038/nbt0197-79

Schmitz, K. R., Bagchi, A., Roovers, R. C., van Bergen en Henegouwen, P. M. P., and Ferguson, K. M. (2013). Structural Evaluation of EGFR Inhibition Mechanisms for Nanobodies/VHH Domains. *Structure* 21 (7), 1214–1224. doi:10.1016/j.str.2013.05.008

Scully, M., Cataland, S. R., Peyvandi, F., Coppo, P., Knöbl, P., Kremer Hovinga, J. A., et al. (2019). Caplacizumab Treatment for Acquired Thrombotic Thrombocytopenic Purpura. *N. Engl. J. Med.* 380 (4), 335–346. doi:10.1056/NEJMoa1806311

Sheng, Y., Wang, K., Lu, Q., Ji, P., Liu, B., Zhu, J., et al. (2019). Nanobody-horseradish Peroxidase Fusion Protein as an Ultrasensitive Probe to Detect Antibodies against Newcastle Disease Virus in the Immunoassay. *J. Nanobiotechnol* 17 (1), 35. doi:10.1186/s12951-019-0468-0

Singh, S., Murillo, G., Chen, D., Parihar, A. S., and Mehta, R. G. (2018). Suppression of Breast Cancer Cell Proliferation by Selective Single-Domain Antibody for Intracellular STAT3. *Breast Cancer (Auckl)* 12, 117822341775085. doi:10.1177/1178223417750858

Soetens, E., Ballegeer, M., and Saelens, X. (2020). An inside Job: Applications of Intracellular Single Domain Antibodies. *Biomolecules* 10 (12), 1663. doi:10.3390/biom10121663

Sonnabend, A., Nikolaeva, O., and Kubick, S. (2015). Templatoptimierung in der zellfreien Proteinsynthese. *GIT Labor-Fachzeitschrift* 11 (11), 35–38.

Stech, M., Hust, M., Schulze, C., Dübel, S., and Kubick, S. (2014). Cell-free Eukaryotic Systems for the Production, Engineering, and Modification of scFv Antibody Fragments. *Eng. Life Sci.* 14 (4), 387–398. doi:10.1002/elsc.201400036

Stech, M., Merk, H., Schenk, J. A., Stöcklein, W. F. M., Wüstenhagen, D. A., Micheel, B., et al. (2013). Production of Functional Antibody Fragments in a Vesicle-Based Eukaryotic Cell-free Translation System. *J. Biotechnol.* 164 (2), 220–231. doi:10.1016/j.jbiotec.2012.08.020

Stech, M., Nikolaeva, O., Thoring, L., Stöcklein, W. F. M., Wüstenhagen, D. A., Hust, M., et al. (2017). Cell-free Synthesis of Functional Antibodies Using a Coupled *In Vitro* Transcription-Translation System Based on CHO Cell Lysates. *Sci. Rep.* 7 (1), 12030. doi:10.1038/s41598-017-12364-w

Stoeber, M., Jullié, D., Lobingier, B. T., Laeremans, T., Steyaert, J., Schiller, P. W., et al. (2018). A Genetically Encoded Biosensor Reveals Location Bias of Opioid Drug Action. *Neuron* 98 (5), 963–976. e5. doi:10.1016/j.neuron.2018.04.021

Tessier, D. C., Thomas, D. Y., Khouri, H. E., Laliberté, F., and Vernet, T. (1991). Enhanced Secretion from Insect Cells of a Foreign Protein Fused to the Honeybee Melittin Signal Peptide. *Gene* 98 (2), 177–183. doi:10.1016/0378-1119(91)90171-7

Thoring, L., Dondapati, S. K., Stech, M., Wüstenhagen, D. A., and Kubick, S. (2017). High-yield Production of "Difficult-To-Express" Proteins in a Continuous Exchange Cell-free System Based on CHO Cell Lysates. *Sci. Rep.* 7 (1), 11710. doi:10.1038/s41598-017-12188-8

Thoring, L., and Kubick, S. (2018). Versatile Cell-free Protein Synthesis Systems Based on Chinese Hamster Ovary Cells. *Methods Mol. Biol. (Clifton, N.J.)* 1850, 289–308. doi:10.1007/978-1-4939-8730-6_19

Thoring, L., Wüstenhagen, D. A., Borowiak, M., Stech, M., Sonnabend, A., and Kubick, S. (2016). Cell-Free Systems Based on CHO Cell Lysates: Optimization Strategies, Synthesis of "Difficult-To-Express" Proteins and Future Perspectives. *PloS one* 11 (9), e0163670. doi:10.1371/journal.pone.0163670

Traenkle, B., and Rothbauer, U. (2017). Under the Microscope: Single-Domain Antibodies for Live-Cell Imaging and Super-resolution Microscopy. *Front. Immunol.* 8, 1030. doi:10.3389/fimmu.2017.01030

van der Linden, R. H. J., Frenken, L. G. J., de Geus, B., Harmsen, M. M., Ruuls, R. C., Stok, W., et al. (1999). Comparison of Physical Chemical Properties of Llama VHH Antibody Fragments and Mouse Monoclonal Antibodies. *Biochim. Biophys. Acta (Bba) - Protein Struct. Mol. Enzymol.* 1431 (1), 37–46. doi:10.1016/s0167-4838(99)00030-8

van Impe, K., Bethuyne, J., Cool, S., Impens, F., Ruano-Gallego, D., De Wever, O., et al. (2013). A Nanobody Targeting the F-Actin Capping Protein CapG

Restrains Breast Cancer Metastasis. *Breast Cancer Res.* 15 (6), R116. doi:10.1186/bcr3585

Wagner, T. R., and Rothbauer, U. (2020). Nanobodies Right in the Middle: Intrabodies as Toolbox to Visualize and Modulate Antigens in the Living Cell. *Biomolecules* 10 (12), 1701. doi:10.3390/biom10121701

Wu, Y., Cui, Z., Huang, Y.-H., de Veer, S. J., Aralov, A. V., Guo, Z., et al. (2022). Towards a Generic Prototyping Approach for Therapeutically-Relevant Peptides and Proteins in a Cell-free Translation System. *Nat. Commun.* 13 (1), 260. doi:10.1038/s41467-021-27854-9

Yamaguchi, J., Naimuddin, M., Biyani, M., Sasaki, T., Machida, M., Kubo, T., et al. (2009). cDNA Display: a Novel Screening Method for Functional Disulfide-Rich Peptides by Solid-phase Synthesis and Stabilization of mRNA-Protein Fusions. *Nucleic Acids Res.* 37 (16), e108. doi:10.1093/nar/gkp514

Yu, Y., Li, J., Zhu, X., Tang, X., Bao, Y., Sun, X., et al. (2017). Humanized CD7 Nanobody-Based Immunotoxins Exhibit Promising Anti-T-cell Acute Lymphoblastic Leukemia Potential. *Ijn* 12, 1969–1983. doi:10.2147/IJN.S127575

Zemella, A., Richter, T., Thoring, L., and Kubick, S. (2019). A Combined Cell-free Protein Synthesis and Fluorescence-Based Approach to Investigate GPCR Binding Properties. *Methods Mol. Biol. (Clifton, N.J.)* 1947, 5757–7777. doi:10.1007/978-1-4939-9121-1_4

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 | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fbioe.2022.900162

Translational Detection of Indole by Complementary Cell-free Protein Synthesis Assay

You Jin Lee[1], *Soojin Lee*[2] *and Dong-Myung Kim*[1]*

[1]Department of Chemical Engineering and Applied Chemistry, Daejeon, Korea, [2]Department of Microbiology and Molecular Biology, Chungnam National University, Daejeon, Korea

The information encoded in a single copy of DNA is processed into a plethora of protein molecules via the cascade of transcription and translation. Thus, the molecular process of gene expression can be considered an efficient biological amplifier from the viewpoint of synthetic biology. Cell-free protein synthesis (CFPS) enables the implementation of this amplification module for *in vitro* analysis of important biomolecules and avoids many of the problems associated with whole cell-based approaches. Here, we developed a method to analyze indole by using a combination of enzymatic conversion of indole and amino acid-dependent CFPS. In this method, indole molecules in the assay sample are used to generate tryptophan, which is incorporated into signal-generating proteins in the subsequent cell-free synthesis reaction. The activity of cell-free synthesized proteins was successfully used to estimate the indole concentration in the assay sample. In principle, the developed method could be extended to analyses of other important bioactive compounds.

Keywords: on-site analysis, indole, metabolites, cell-free protein synthesis, personal glucose meter

OPEN ACCESS

Edited by:
Jian Li,
ShanghaiTech University, China

Reviewed by:
Hideo Nakano,
Nagoya University, Japan
Fu-Xing Niu,
Guangxi University of Science and Technology, China

***Correspondence:**
Dong-Myung Kim
dmkim@cnu.ac.kr

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *20 March 2022*
Accepted: *28 April 2022*
Published: *13 May 2022*

Citation:
Lee YJ, Lee S and Kim D-M (2022) Translational Detection of Indole by Complementary Cell-free Protein Synthesis Assay. Front. Bioeng. Biotechnol. 10:900162. doi: 10.3389/fbioe.2022.900162

INTRODUCTION

Most presently used biosensors rely on target-specific binding of purified biomolecules and commonly require complicated fabrication steps to integrate the sensing surface with separate transducers and amplifiers in order to generate readable outputs (Cho et al., 2021; Lim et al., 2021). On the other hand, microbes can sense chemical compounds and generate amplified signals using the protein synthesis machinery and regulatory components. Various microbes can be genetically engineered into stand-alone microbial biosensors that generate biological signals in response to target analytes. However, the widespread application of microbial sensors has been restricted by the intrinsic limitations of using live cells, including the requirement for time-consuming cell culture and conditioning steps. The slow response due to decelerated membrane diffusion of analytes is also a major drawback of microbial sensors (Su et al., 2011; Chen et al., 2020). These limitations of whole cell-based biosensors can be addressed by employing cell-free protein synthesis (CFPS) as an *in vitro* module to express reporter proteins (Lee and Kim, 2019; Voyvodic and Bonnet, 2020; Zhang et al., 2020). Unlike cell-based gene expression methods, CFPS can be directly programmed with the reporter genes without complicated cloning procedures. A CFPS-based biosensor can also be readily modularized to generate reporter proteins in response to target molecules and interfaced with a wide array of analytical devices (Gräwe et al., 2019; Lopreside et al., 2019; Thavarajah et al., 2020; Nguyen et al., 2021). Due to these advantages, the implementation of CFPS as a signal generation module to measure important metabolic compounds has been explored (Liu et al., 2020; McNerney et al., 2020;

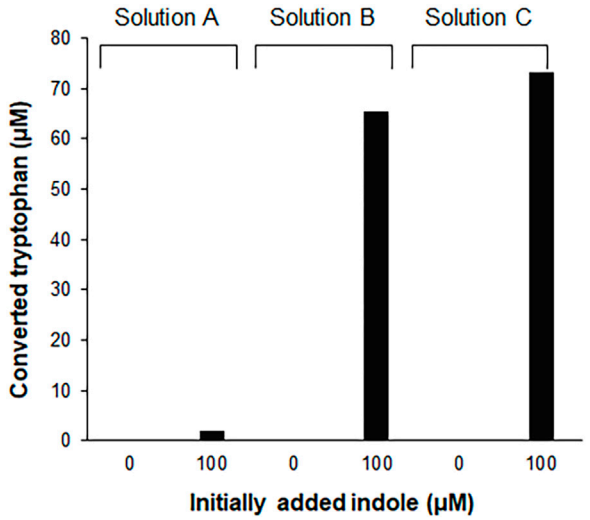


FIGURE 1 | Conversion of indole to tryptophan in the S12 extract solutions. 100 μM indole was incubated for 2 h at 30°C, in the S12 extract diluted in different solutions [27% (v/v)]. Solution A, S12 extract diluted in PBS; solution B, S12 extract diluted in the reaction mixture containing CFPS components prepared in the absence of tryptophan and template DNA; solution C, solution B without amino acids. Concentrations of tryptophan converted from indole were measured using a Hitachi L-8900 amino acid analyzer (Hitachi High-Technologies, Tokyo, Japan) following the manufacturer's protocols.

Silverman et al., 2020). For example, the complementary cell-free protein synthesis (CCFPS) assay was recently developed to detect amino acids and related metabolites (Jang et al., 2017; Jang et al., 2019). In this assay, a reaction mixture for CFPS is prepared in the absence of the amino acid to be analyzed. The lack of an amino acid prevents the reaction mixture from producing full-length proteins, while addition of an assay sample containing the missing amino acid completes the set of 20 amino acids and thereby allows the generation of the reporter protein encoded by the template DNA. The yield of the reporter protein linearly correlates with the titer of amino acids; therefore, the CCFPS assay enables precise measurement of amino acids without complicated chromatographic separation procedures. In addition to being the building blocks of proteins, the synthesis and degradation pathways of amino acids are closely interlinked with the metabolism of diverse compounds. Therefore, these compounds can be enzymatically converted into proteinogenic amino acids for their measurement by the CCFPS assay (Lim et al., 2020).

Indole is an important heterocyclic compound that works as a nucleus for the synthesis of many key compounds in the pharmaceutical (Kaushik et al., 2013; Sravanthi and Manju, 2016; Kumari and Singh, 2019), agricultural (Ye et al., 2021), plastics (Wang et al., 2018; Arza and Zhang, 2019), and perfumery industries (Rodrigues et al., 2021). In addition, indole functions as a major intercellular signaling molecule within the gut microbial ecosystem, and its biosynthesis is an important phenotypic characteristic that can be used to differentiate, identify, and diagnose enteric bacterial infections (Jaglin et al., 2018). Moreover, plasma levels of indole and its derivatives are associated with anxiety-like behaviors in rats, and inflammatory bowel disease and neurological disorders in human (Roager and Licht, 2018). Assays of indole presently involve the performance of gas chromatography or high-performance liquid chromatography (HPLC) in tandem with mass spectrometry, which requires expensive equipment with a large footprint, complicated operational procedures, and trained operators (Dehnhard et al., 1993). Indole is a metabolic precursor of tryptophan; therefore, we envisioned that it can be measured using a CCFPS assay after being converted to tryptophan. To this end, we incorporated an enzymatic reaction that converts indole to tryptophan into the CCFPS assay of tryptophan (Trp-CCFPS assay). The reaction mixture for the Trp-CCFPS assay could produce the reporter protein in response to indole, especially when the cell extract was supplemented with the recombinant β subunit of *Pyrococcus furiosus* tryptophan synthase (*Pf*TrpB) (Buller et al., 2015). When the CCFPS assay was programmed with different reporter genes, the activity of the cell-free synthesized proteins exhibited a linear correlation with the indole concentration. In particular, the CCFPS reaction could be programmed with the gene encoding invertase to generate glucose as the signaling molecule in the CCFPS assay. This allowed the indole titer to be read with a personal glucose meter (PGM), which markedly improved the convenience and accessibility of the assay. Taken together, using indole as a model compound, our results demonstrate that the principles of cell-free synthetic biology can be harnessed to build a customized molecular converter that transforms target analytes into readily measurable reporter proteins. The important advantage of this approach is that it can be readily configured to measure the target analytes using various devices available in laboratory settings.

MATERIALS AND METHODS

Materials

Oligonucleotides were synthesized by Macrogen (Seoul, Korea). The gene encoding *Pf*TrpB (Buller et al., 2015) was synthesized by Integrated DNA Technologies (Coralville, IA, United States). DNA polymerase and other reagents for PCR were purchased from Bioline (London, United Kingdom). Restriction enzymes and T4 ligase were obtained from Enzynomics (Daejeon, Korea). Luria-Bertani (LB) medium and ampicillin were purchased from Duchefa Biochemie (Haarlem, Netherlands). Creatine phosphate and creatine kinase were purchased from Roche Diagnostics (Mannheim, Germany). All other chemical reagents and human serum were obtained from Sigma-Aldrich (St. Louis, MO, United States). The S12 extract for CFPS was prepared from the *Escherichia coli* (*E. coli*) strain BL21star (DE3) (Invitrogen, Carlsbad, CA, United States) and diafiltered to remove residual amino acids. To remove residual amino acids, the S12 extract was centrifuged in a Vivaspin centrifugal concentrator (Sartorius Stedim Biotech GmbH, Göttingen, Germany) installed with a 50 kDa molecular weight cutoff membrane. After diluting 2 ml of the extract with 13 ml of the

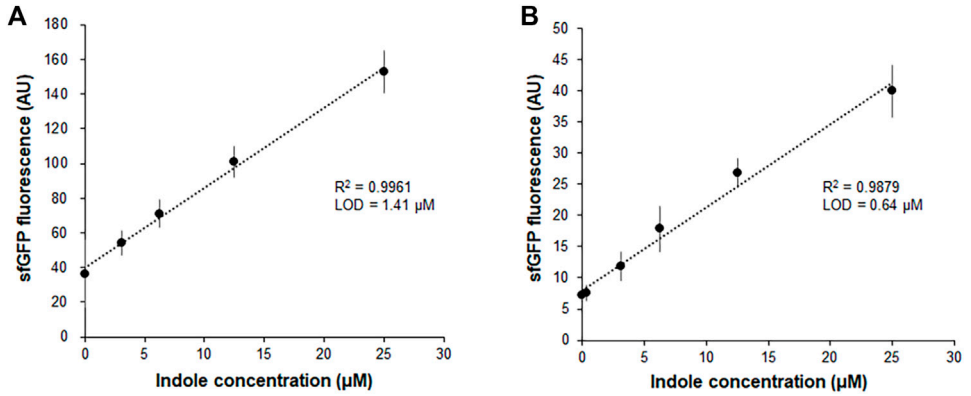


FIGURE 2 | Measurement of indole by the fluorescence of sfGFP synthesized from Indole-CCFPS assay. Indole-CCPFS assay was performed with varying concentrations of indole using standard S12 extract **(A)** or diafiltered S12 extract **(B)**. The fluorescence of the synthesized sfGFP was measured after 2 h incubation at 30°C. Measurements were performed in triplicate, and the error bars represent the standard deviations of three independent experiments.

S12 buffer (20 mM Tris–acetate, 28 mM magnesium acetate, and 120 mM potassium acetate, pH 8.2), the diluted S12 extract was concentrated back to its original volume by centrifuging the concentrator at 2,000 × *g*. This process was repeated five times, and the washed S12 extract was stored in a deep freezer in aliquots (Kim et al., 2006; Jang et al., 2019).

Preparation of Recombinant *Pyrococcus furiosus* Tryptophan Synthase

The chemically synthesized gene encoding *Pf*TrpB was cloned into the pET21a vector to create the plasmid pET21a TrpB. The *E. coli* strain BL21 (DE3) transformed with pET21a TrpB was grown at 37°C in 2.5 L baffled flasks containing 500 ml LB medium. Expression of the recombinant enzyme was induced by adding 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) when the OD600 of the culture broth reached 0.6. After overnight culture at 30°C following IPTG induction, the cells were harvested by centrifugation (4,500 × *g*, 20 min) and washed thrice with deionized water. After the final wash, cell pellets were resuspended in 20 ml equilibrium buffer (50 mM phosphate buffer, pH 8.0, and 300 mM NaCl) and disrupted by a single passage through a French press (Thermo Fisher Scientific, Waltham, MA, United States) at 12,000 psi. After centrifugation at 12,000 × g for 20 min, recombinant *Pf*TrpB in the supernatant was purified using a Ni-NTA agarose column.

Complementary Cell-Free Protein Synthesis Assay to Measure Indole

The reaction mixture for the indole assay consisted of the following components in a final volume of 30 μL: 57 mM HEPES-KOH (pH 8.2), 1.2 mM ATP, 0.85 mM each of CTP, GTP, and UTP, 2 mM DTT, 0.17 mg/mL *E. coli* total tRNA mixture (from the strain MRE600), 0.64 mM cAMP, 90 mM potassium glutamate, 80 mM ammonium acetate, 12 mM magnesium acetate, 34 μg/mL L-5-formyl-5,6,7,8-tetrahydrofolic acid, 2 mM each of 19 amino acids excluding tryptophan, 2% polyethylene glycol 8000, 67 mM creatine phosphate, 3.2 μg/ml creatine kinase, 27% (v/v) diafiltered S12 extract, and 6.7 μg/ml template DNA encoding *E. coli* invertase between the T7 promoter and T7 terminator (Dehnhard et al., 1993; Jang et al., 2017). In experiments using purified *Pf*TrpB, the assay mixture was supplemented with 100 mM Tris-HCl (pH 8.0), 180 mM NaCl, and 10 μM pyridoxal phosphate (PLP) (Cellini et al., 2020). After addition of indole at various concentrations, the assay mixture was incubated at 30°C for 90 min. To analyze superfolder green fluorescent protein (sfGFP) produced during the CCFPS assay, 15 μL of the completed assay reaction was diluted with 200 μL phosphate-

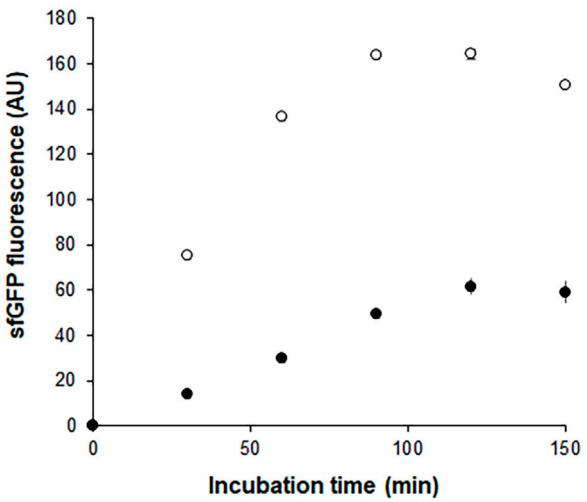


FIGURE 3 | Time courses of sfGFP synthesis during Indole-CCFPS and Trp-CCFPS assays. The CCFPS assay mixtures programmed with the template DNA encoding sfGFP were incubated at 30°C. 15 μL assay mixture was withdrawn at the indicated time points and measured for sfGFP fluorescence after being diluted in 200 μL PBS. Measurements were performed in triplicate, and the error bars represent the standard deviations of three independent experiments.

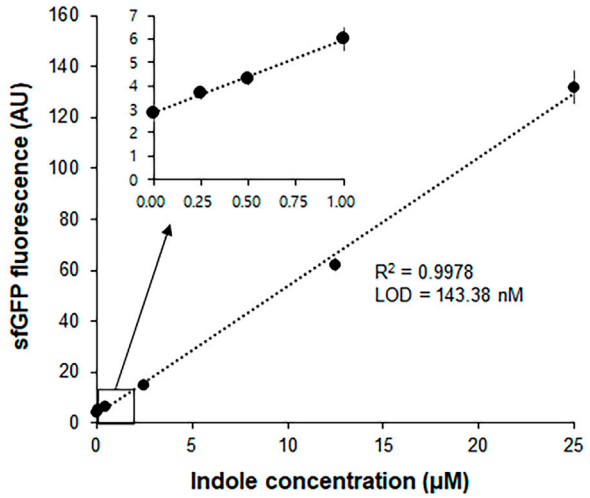


FIGURE 4 | Measurement of indole by Indole-CCPFS assay employing *Pf*TrpB-enriched S12 extract. The S12 extract was prepared after overexpression of the *Pf*TrpB enzyme during the cultivation of *E. coli* cells. The resulting *Pf*TrpB-enriched S12 extract was diafiltered to remove residual amino acids prior to its use for Indole-CCFPS assay programmed with the template DNA encoding sfGFP. Indole-CCFPS assay was performed with varying concentrations of indole, and measured for the fluorescence of synthesized sfGFP. Inset represents the enlarged graph of the boxed region. Measurements were performed in triplicate, and the error bars represent the standard deviations of three independent experiments.

buffered saline (PBS) and fluorescence was measured with a Victor X2 microplate reader (PerkinElmer, Waltham, MA, United States). In the CCFPS assay programmed with the invertase gene, 15 μL of the completed assay mixture was transferred to an Eppendorf tube containing 15 μL of 200 mM sucrose prepared in PBS and further incubated for 10 min at 37°C. After heat-inactivation of invertase, the glucose titer was measured in 5 μL of the supernatant from the briefly centrifuged reaction mixture using a PGM (Accu-Check Inform II, Roche Diagnostics). Limit of detection (LOD) for each experiment was calculated based on the standard deviation of the response (Sy) of the curve and the slope of the calibration curve (S) at levels approximating the LOD according to the formula: LOD = 3.3(Sy/S).

RESULTS AND DISCUSSION

Conversion of Indole to Tryptophan in *E. coli* Extract

Tryptophan contains an indole ring attached to the alanyl side chain. As shown in **Supplementary Figure S1**, biosynthesis of tryptophan from chorismate involves five enzymes. Among these enzymes, tryptophan synthase, a tetrameric enzyme consisting of αββα subunits, catalyzes the final two steps of tryptophan biosynthesis. The *α* subunit (TrpA) catalyzes the cleavage of indole-3-glycerol phosphate to form indole and glyceraldehyde-3-phosphate, while the *β* subunit is responsible for the condensation of indole with serine to produce tryptophan. This is an endogenous pathway in *E. coli*; therefore, we tested if the S12 extract used for the CCFPS assay can convert indole to tryptophan. As expected, HPLC analysis revealed that incubation of indole in the S12 solution led to accumulation of tryptophan (**Figure 1**). However, the conversion yield of indole to tryptophan varied dramatically depending on the solution in which the S12 extract was diluted. For example, the conversion yield in the S12 extract diluted in PBS (solution A in **Figure 1**) did not exceed 5% after incubation for 2 h. By contrast, the same amount of S12 extract converted more than 60% of indole to tryptophan when it was diluted in reaction mixture containing CFPS components prepared in the absence of tryptophan and template DNA (solution B in **Figure 1**). Tryptophan is a condensation product of indole and serine; therefore, we first reasoned that the lower conversion yield of indole by the PBS-diluted S12 extract was due to the lack of serine, which was abundantly supplied in the CFPS mixture. However, contrary to this expectation, removal of amino acids from solution B did not lower the conversion yield of indole (solution C in **Figure 1**). This indicates that there was a substantial amount of residual serine in the S12 extract and that conversion of indole by the S12 extract was affected by a different component of solution B. Although it was not further investigated which component(s) in solution B affected the conversion of indole to tryptophan, this result indicates that the assay mixture of CFPS provides an environment that substantially enhances the conversion of indole to tryptophan.

Measurement of the Indole Concentration by the Trp-Complementary Cell-Free Protein Synthesis Assay

Our previous report demonstrated that the CCFPS assay is a powerful alternative to existing chromatographic methods for amino acid analysis (Jang et al., 2017; Jang et al., 2019). For instance, without any chemical derivatization and chromatographic separation steps, the Trp-CCFPS assay could detect micromolar concentrations of tryptophan within 90 min (**Supplementary Figure S2**). HPLC analysis of the aforementioned experiments confirmed that indole was converted to tryptophan by the S12 extract under the conditions used for the CCFPS assay; therefore, we attempted to link the extract-catalyzed conversion of indole with the Trp-CCFPS assay (Indole-CCFPS assay). The initially designed assay mixture for the Indole-CCFPS assay consisted of the same components as were used for the Trp-CCFPS assay. When the Indole-CCFPS mixture programmed with DNA encoding sfGFP was supplied with 100 μM indole, incubation of the assay mixture led to generation of sfGFP fluorescence, as expected (**Supplementary Figure S3A**). However, the fluorescence intensity was already high without addition of indole. In principle, signal generation during a CCFPS assay should be dependent on exogenous addition of the amino acid missing from the assay solution. However, in practice, the S12 extract contains substantial amounts of amino acids from the cytoplasm of *E. coli*.

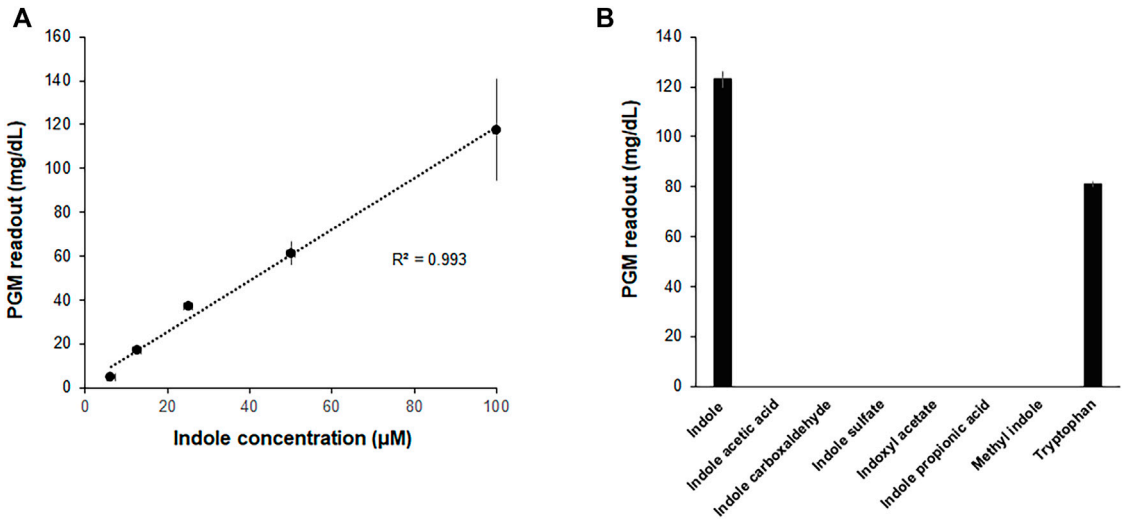


FIGURE 5 | Measurement of indole by a PGM. The assay mixture for Indole-CCFPS was programmed with the template DNA encoding bacterial invertase. 15 μL of the assay mixture was transferred to an Eppendorf tube containing 15 μL of 200 mM sucrose prepared in PBS and further incubated for 10 min at 37°C. After heat-inactivation of invertase, the glucose titer was measured in 5 μL of the supernatant from the briefly centrifuged reaction mixture using a PGM (Accu-Check Inform II, Roche Diagnostics). **(A)** The PGM readouts were measured for varying concentrations of indole. **(B)** The assay mixture for Indole-CCFPS was supplied with the same concentrations (100 μM) of indole, and measured for the glucose concentration using a PGM. Measurements were performed in triplicate, and the error bars represent the standard deviations of three independent experiments.

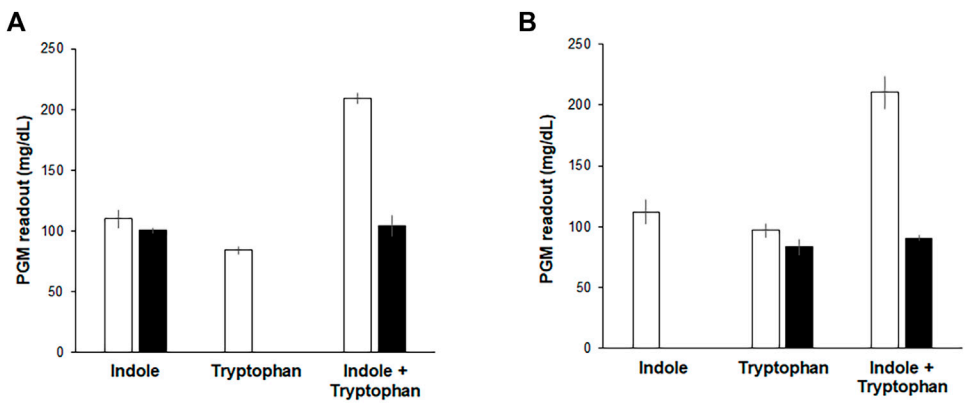


FIGURE 6 | Discriminative measurement of indole and tryptophan. Indole, tryptophan, and their mixture were analyzed by Indole-CCFPS assay using a PGM as a measuring device. **(A)** for the selective detection of tryptophan, the mixed solution of tryptophan and indole was treated with LAAO (filled bars), as described in Materials and Methods. Blank bars represent the results of Indole-CCFPS assays without the LAAO treatment. **(B)** for the selective detection of indole, the assay mixture was supplemented with AOA (filled bars), as described in Materials and Methods. Blank bars represent the results of Indole-CCFPS assays in the absence of AOA. Measurements were performed in triplicate, and the error bars represent the standard deviations of three independent experiments.

During the Indole-CCFPS assay, residual tryptophan would cause background synthesis of proteins. The high level of background protein synthesis might affect the sensitivity of the CCFPS assay; therefore, the S12 extract was diafiltered to remove residual tryptophan prior to its use in this assay. Diafiltration of the cell extract markedly lowered the background synthesis of sfGFP during the Indole-CCFPS assay (**Supplementary Figure S3B**). Although fluorescence observed upon addition of indole was also lowered, due to the markedly reduced background, the signal-to-noise ratio upon addition of 100 μM indole was improved from 4 to 7 (**Supplementary Figure S3B**). When the Indole-CCFPS assay was performed using the diafiltered S12 extract, the intensity of sfGFP fluorescence exhibited a strong correlation with the indole concentration. Compared with the same assay performed with the standard S12 extract, the use of diafiltered S12 extract substantially lowered the limit of detection for indole (**Figures 2A,B**).

In a separate experiment, we compared the time-courses of sfGFP generation between the Trp-CCFPS and Indole-CCFPS

assays. When the same concentrations (100 μM) of tryptophan and indole were tested, both the increase rate and maximum value of sfGFP fluorescence were markedly lower during the Indole-CCFPS assay than during the Trp-CCFPS assay (**Figure 3**). We reasoned that the rate of indole conversion limited the supply of tryptophan during the Indole-CCFPS assay, resulting in sluggish synthesis of sfGFP. Based on this assumption, to further increase the speed and sensitivity of indole detection, we attempted to accelerate the conversion of indole using recombinant tryptophan synthase. Only the activity of the β subunit, not the entire αββα complex of tryptophan synthase, is required for coupling of indole and serine; therefore, the assay mixture for the Indole-CCFPS assay was supplemented with *Pf*TrpB, an engineered β subunit of tryptophan synthase of *Pyrococcus furiosus* (Buller et al., 2015). While the activity of natural TrpB decreases when it is separated from the native complex, this variant enzyme carries activating mutations that restore the activity of stand-alone TrpB. As expected, the fluorescence signal from the Indole-CCFPS assay increased in proportion to the amount of supplemented *Pf*TrpB and peaked when 1.0 mg/ml of the recombinant enzyme was used (**Supplementary Figure S4**). Use of *Pf*TrpB also markedly reduced the amount of time required to perform the Indole-CCFPS assay because sfGFP fluorescence plateaued within 90 min. Separate preparation of purified *Pf*TrpB could be avoided by preparing the S12 extract after overexpressing the enzyme during cultivation of transformed *E. coli* cells. The *Pf*TrpB-enriched extract resulted in a similar rate and yield of indole conversion as the standard S12 extract supplemented with purified enzymes and thus was used in subsequent experiments (**Supplementary Figures S4 and S5**). With these modifications, the sensitivity of the Indole-CCFPS assay was further improved to measure nanomolar concentrations of indole (**Figure 4**).

Measurement of Indole by a Personal Glucose Meter

In an attempt to further improve the convenience and portability of the indole assay, the CCFPS assay was interfaced with glucose measurement by a PGM, the most widely distributed analytical device. To this end, the template DNA for the Indole-CCFPS assay was switched from pK7sfGFP to pK7-inv, which encodes *E. coli* invertase. In addition, 200 mM sucrose was included in the assay mixture to generate glucose in the presence of indole. When the glucose titer in the assay mixture was measured after incubation for 90 min, the readout of the PGM showed almost a linear correlation with the concentration of indole (**Figure 5A**), indicating that a PGM can be used as a portable device to measure indole concentrations. The Indole-CCFPS assay could also specifically discriminate indole from its derivatives. When seven compounds structurally related to indole were tested in parallel, none of the analogs could generate signals detectable with the PGM in the CCFPS assay (**Figure 5B**).

Discriminative Measurement of Tryptophan and Indole in a Mixed Solution

Although the above results indicate that the Indole-CCFPS assay can conveniently detect indole with high specificity and sensitivity, it cannot discriminate indole in a mixture with tryptophan, which may be required when analyzing indole in biological samples. This problem was addressed by pre-treating assay samples with L-amino acid oxidase (LAAO). LAAO oxidizes tryptophan to indole-3 pyruvate and thus prevents the incorporation of tryptophan in assay samples into the reporter protein during the Indole-CCFPS assay (**Supplementary Figure S6**). Incubation with LAAO from *Crotalus adamanteus* almost completely repressed incorporation of tryptophan into sfGFP, but did not affect signal generation by indole (**Figure 6A**). On the other hand, tryptophan can also be differentiated from indole by repressing the conversion of indole to tryptophan. Aminooxyacetate (AOA) is a potent inhibitor of PLP-dependent enzymes. TrpB is a PLP-dependent enzyme (Cellini et al., 2020); therefore, we expected that conversion of indole to tryptophan would be repressed by including AOA in the CCFPS assay mixture (**Supplementary Figure S6**). Indeed, when tryptophan and/or indole were added to the assay mixture containing AOA, only tryptophan could be used to synthesize sfGFP (**Figure 6B**). Taken together, these results demonstrate that tryptophan and indole can be discriminately measured by dividing the mixed solution into two and treating one sample with LAAO and the other with AOA.

CONCLUSION

By harnessing the translational machinery of cells as an *in vitro* signal generation module, we developed a method that enables rapid detection of nanomolar concentrations of indole. A few colorimetric methods have been developed to detect indole without using sophisticated instruments. For example, Kovác's assay uses para-dimethylaminobenzaldehyde, which reacts with indole to generate a red product (Turner, 1961). Darkho et al. also reported a colorimetric method that uses hydroxylamine and can discriminate indole from structurally related compounds (Darkoh et al., 2015). However, these methods can only detect millimolar concentrations and have limited selectivity for indole. In particular, the sensitivity of indole assay needs to be improved to differentiate subtle changes caused by the infection of pathogenic bacteria (Chappell et al., 2016). The amplification nature of protein synthesis employed in the presented method enables the generation of biologically converted and amplified signals, thereby markedly improving the sensitivity of detection. Furthermore, the modular nature of this approach provides great flexibility in terms of how the signal outputs generated from indole are read. For example, by programming the assay with a glucose-generating enzyme, the titer of indole can be measured with a PGM. PGMs are the most widely distributed analytical device around the world and are relatively cheap and easy to use. A method that uses a PGM to analyze analytes of interest will greatly enhance simplicity

and convenience, especially in low resource settings with limited laboratory services or where it is hard to access relevant facilities.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

D-MK conceived this project, and planned the experiments. D-MK and SL supervised the work. YL designed and performed the experiments. D-MK, SL, and YL analyzed the data, prepared the figures, and wrote the manuscript.

FUNDINGS

This work was supported by the National Research Foundation (NRF) (2020R1A2C2013114 and 2020R1A5A8017671).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.900162/full#supplementary-material

REFERENCES

Arza, C. R., and Zhang, B. (2019). Synthesis, Thermal Properties, and Rheological Characteristics of Indole-Based Aromatic Polyesters. *ACS Omega* 4, 15012–15021. doi:10.1021/acsomega.9b01802

Buller, A. R., Brinkmann-Chen, S., Romney, D. K., Herger, M., Murciano-Calles, J., and Arnold, F. H. (2015). Directed Evolution of the Tryptophan Synthase β-subunit for Stand-Alone Function Recapitulates Allosteric Activation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14599–14604. doi:10.1073/pnas.1516401112

Cellini, B., Zelante, T., Dindo, M., Bellet, M. M., Renga, G., Romani, L., et al. (2020). Pyridoxal 5′-Phosphate-Dependent Enzymes at the Crossroads of Host-Microbe Tryptophan Metabolism. *Ijms* 21, 5823. doi:10.3390/ijms21165823

Chappell, C. L., Darkoh, C., Shimmin, L., Farhana, N., Kim, D.-K., Okhuysen, P. C., et al. (2016). Fecal Indole as a Biomarker of Susceptibility to Cryptosporidium Infection. *Infect. Immun.* 84, 2299–2306. doi:10.1128/IAI.00336-16

Chen, H., Simoska, O., Lim, K., Grattieri, M., Yuan, M., Dong, F., et al. (2020). Fundamentals, Applications, and Future Directions of Bioelectrocatalysis. *Chem. Rev.* 120, 12903–12993. doi:10.1021/acs.chemrev.0c00472

Cho, S. W., Ko, H. J., and Park, T. H. (2021). Identification of a Lung Cancer Biomarker Using a Cancer Cell Line and Screening of Olfactory Receptors for Biomarker Detection. *Biotechnol. Bioproc E* 26, 55–62. doi:10.1007/s12257-020-0132-4

Darkoh, C., Chappell, C., Gonzales, C., and Okhuysen, P. (2015). A Rapid and Specific Method for the Detection of Indole in Complex Biological Samples. *Appl. Environ. Microbiol.* 81, 8093–8097. doi:10.1128/AEM.02787-15

Dehnhard, M., Claus, R., Hillenbrand, M., and Herzog, A. (1993). High-performance Liquid Chromatographic Method for the Determination of 3-methylindole (Skatole) and Indole in Adipose Tissue of Pigs. *J. of Chromatogr. B Biomed. Sci. and Appl.* 616, 205–209. doi:10.1016/0378-4347(93)80387-J

Gräwe, A., Dreyer, A., Vornholt, T., Barteczko, U., Buchholz, L., Drews, G., et al. (2019). A Paper-Based, Cell-free Biosensor System for the Detection of Heavy Metals and Date Rape Drugs. *PLoS One* 14, e0210940. doi:10.1371/journal.pone.0210940

Jaglin, M., Rhimi, M., Philippe, C., Pons, N., Bruneau, A., Goustard, B., et al. (2018). Indole, a Signaling Molecule Produced by the Gut Microbiota, Negatively Impacts Emotional Behaviors in Rats. *Front. Neurosci.* 12, 216. doi:10.3389/fnins.2018.00216

Jang, Y.-J., Lee, K.-H., Yoo, T. H., and Kim, D.-M. (2017). Complementary Cell-free Translational Assay for Quantification of Amino Acids. *Anal. Chem.* 89, 9638–9642. doi:10.1021/acs.analchem.7b01956

Jang, Y.-J., Lee, K.-H., Yoo, T. H., and Kim, D.-M. (2019). Interfacing a Personal Glucose Meter with Cell-free Protein Synthesis for Rapid Analysis of Amino Acids. *Anal. Chem.* 91, 2531–2535. doi:10.1021/acs.analchem.8b05526

Kaushik, N., Kaushik, N., Attri, P., Kumar, N., Kim, C., Verma, A., et al. (2013). Biomedical Importance of Indoles. *Molecules* 18, 6620–6662. doi:10.3390/molecules18066620

Kim, T.-W., Keum, J.-W., Oh, I.-S., Choi, C.-Y., Park, C.-G., and Kim, D.-M. (2006). Simple Procedures for the Construction of a Robust and Cost-Effective Cell-free Protein Synthesis System. *J. of Biotechnol.* 126, 554–561. doi:10.1016/j.jbiotec.2006.05.014

Kumari, A., and Singh, R. K. (2019). Medicinal Chemistry of Indole Derivatives: Current to Future Therapeutic Prospectives. *Bioorg. Chem.* 89, 103021. doi:10.1016/j.bioorg.2019.103021

Lee, K.-H., and Kim, D.-M. (2019). *In Vitro* use of Cellular Synthetic Machinery for Biosensing Applications. *Front. Pharmacol.* 10, 1166. doi:10.3389/fphar.2019.01166

Lim, H. J., Jang, Y. J., Lee, K.-H., and Kim, D.-M. (2020). Translational Detection of Nonproteinogenic Amino Acids Using an Engineered Complementary Cell-free Protein Synthesis Assay. *Anal. Chem.* 92, 11505–11510. doi:10.1021/acs.analchem.0c01978

Lim, S. H., Ryu, Y. C., and Hwang, B. H. (2021). Aptamer-immobilized Gold Nanoparticles Enable Facile and On-Site Detection of Staphylococcus aureus. *Biotechnol. Bioproc E* 26, 107–113. doi:10.1007/s12257-020-0161-z

Liu, X., Silverman, A. D., Alam, K. K., Iverson, E., Lucks, J. B., Jewett, M. C., et al. (2020). Design of a Transcriptional Biosensor for the Portable, On-Demand Detection of Cyanuric Acid. *ACS Synth. Biol.* 9, 84–94. doi:10.1021/acssynbio.9b00348

Lopreside, A., Wan, X., Michelini, E., Roda, A., and Wang, B. (2019). Comprehensive Profiling of Diverse Genetic Reporters with Application to Whole-Cell and Cell-free Biosensors. *Anal. Chem.* 91, 15284–15292. doi:10.1021/acs.analchem.9b04444

McNerney, M. P., Piorino, F., Michel, C. L., and Styczynski, M. P. (2020). Active Analyte Import Improves the Dynamic Range and Sensitivity of a Vitamin B12 Biosensor. *ACS Synth. Biol.* 9, 402–411. doi:10.1021/acssynbio.9b00429

Nguyen, P. Q., Soenksen, L. R., Donghia, N. M., Angenent-Mari, N. M., de Puig, H., Huang, A., et al. (2021). Wearable Materials with Embedded Synthetic Biology Sensors for Biomolecule Detection. *Nat. Biotechnol.* 39, 1366–1374. doi:10.1038/s41587-021-00950-3

Roager, H. M., and Licht, T. R. (2018). Microbial Tryptophan Catabolites in Health and Disease. *Nat. Commun.* 9 (1), 3294. doi:10.1038/s41467-018-05470-4

Rodrigues, A. E., Nogueira, I., and Faria, R. P. V. (2021). Perfume and Flavor Engineering: a Chemical Engineering Perspective. *Molecules* 26, 3095. doi:10.3390/molecules26113095

Silverman, A. D., Akova, U., Alam, K. K., Jewett, M. C., and Lucks, J. B. (2020). Design and Optimization of a Cell-free Atrazine Biosensor. *ACS Synth. Biol.* 9, 671–677. doi:10.1021/acssynbio.9b00388

Sravanthi, T. V., and Manju, S. L. (2016). Indoles - A Promising Scaffold for Drug Development. *Eur. J. of Pharm. Sci.* 91, 1–10. doi:10.1016/j.ejps.2016.05.025

Su, L., Jia, W., Hou, C., and Lei, Y. (2011). Microbial Biosensors: a Review. *Biosens. and Bioelectron.* 26, 1788–1799. doi:10.1016/j.bios.2010.09.005

Thavarajah, W., Silverman, A. D., Verosloff, M. S., Kelley-Loughnane, N., Jewett, M. C., and Lucks, J. B. (2020). Point-of-use Detection of Environmental

Fluoride via a Cell-free Riboswitch-Based Biosensor. *ACS Synth. Biol.* 9, 10–18. doi:10.1021/acssynbio.9b00347

Turner, J. M. (1961). A New Reagent for the Assay of Indole in the Tryptophanase Reaction. *Biochem. J.* 78, 790–792. doi:10.1042/bj0780790

Voyvodic, P. L., and Bonnet, J. (2020). Cell-free Biosensors for Biomedical Applications. *Curr. Opin. Biomed. Eng.* 13, 9–15. doi:10.1016/j.cobme.2019.08.005

Wang, P., Arza, C. R., and Zhang, B. (2018). Indole as a New Sustainable Aromatic Unit for High Quality Biopolyesters. *Polym. Chem.* 9, 4706–4710. doi:10.1039/C8PY00962G

Ye, M., Liu, M., Erb, M., Glauser, G., Zhang, J., Li, X., et al. (2021). Indole Primes Defence Signalling and Increases Herbivore Resistance in Tea Plants. *Plant Cell. Environ.* 44, 1165–1177. doi:10.1111/pce.13897

Zhang, L., Guo, W., and Lu, Y. (2020). Advances in Cell-Free Biosensors: Principle, Mechanism, and Applications. *Biotechnol. J.* 15, 2000187. doi:10.1002/biot.202000187

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 | Frontiers in Bioengineering and Biotechnology

REVIEW
published: 13 May 2022
doi: 10.3389/fbioe.2022.891808

Codon-Reduced Protein Synthesis With Manipulating tRNA Components in Cell-Free System

Jiaojiao Li[1,2], *Mengtong Tang*[1,2] *and Hao Qi*[1,2]*

[1]School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [2]Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin University, Tianjin, China

Manipulating transfer RNAs (tRNAs) for emancipating sense codons to simplify genetic codons in a cell-free protein synthesis (CFPS) system can offer more flexibility and controllability. Here, we provide an overview of the tRNA complement protein synthesis system construction in the tRNA-depleted Protein synthesis Using purified Recombinant Elements (PURE) system or S30 extract. These designed polypeptide coding sequences reduce the genetic codon and contain only a single tRNA corresponding to a single amino acid in this presented system. Strategies for removing tRNAs from cell lysates and synthesizing tRNAs *in vivo*/*vitro* are summarized and discussed in detail. Furthermore, we point out the trend toward a minimized genetic codon for reducing codon redundancy by manipulating tRNAs in the different proteins. It is hoped that the tRNA complement protein synthesis system can facilitate the construction of minimal cells and expand the biomedical application scope of synthetic biology.

Keywords: cell-free system, transfer RNA, codon-reduced, *in vitro*-transcription, protein synthesis

OPEN ACCESS

Edited by:
Jian Li,
ShanghaiTech University, China

Reviewed by:
Kensaku Sakamoto,
RIKEN, Japan
Yixin Huo,
Beijing Institute of Technology, China

***Correspondence:**
Hao Qi
haoq@tju.edu.cn

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *08 March 2022*
Accepted: *25 April 2022*
Published: *13 May 2022*

Citation:
Li J, Tang M and Qi H (2022) Codon-Reduced Protein Synthesis With Manipulating tRNA Components in Cell-Free System. Front. Bioeng. Biotechnol. 10:891808. doi: 10.3389/fbioe.2022.891808

1 INTRODUCTION

A codon constitutes three consecutive nucleotide bases that are recognized by a specific tRNA to integrate a single amino acid into a growing peptide chain inside the ribosome, or a stop signal that terminates protein synthesis. In general, there are 64 different combinations of the four nucleotides that encode a pool of 20 amino acids and translation stop signals in the genetic codon. Therefore, there is redundancy in the genetic codon, so that some amino acids are encoded by several so-called synonymous codons. This degeneracy of the genetic codon provides favorable conditions for us to rearrange the orthogonal sense codons (Li, 2021). As carriers of amino acids, tRNAs play a vital role in converting the genetic information from messenger RNA (mRNA) into the polypeptide chain (proteins). During protein synthesis, the aminoacylated tRNAs can bring the specific amino acid to a cognate codon of the mRNA inside the ribosome, which then elongates the polypeptide (Su et al., 2020). It is well known that different codons with their cognate tRNAs and aminoacyl-tRNA synthetases (aaRS) act orthogonally during protein translation. To reduce the redundancy of the genetic codon, it is necessary to expand the coding scope by engineering these orthogonal triplets inside a cell or *in vitro*. To establish a simplified codon set, two conditions are required in the protein synthesis reaction system: 1) the expression template is simplified to only one codon corresponding to one tRNA; 2) the orthogonality of the tRNA/aaRS/AA system needs to be maintained.

Simplified codon protein synthesis offers several profound advantages for the bioengineering and study of protein synthesis (**Figure 1**). First of all, the accuracy of decoding mRNA can be precisely regulated by manipulating the tRNA population added to the CFPS. This enables plasticity and

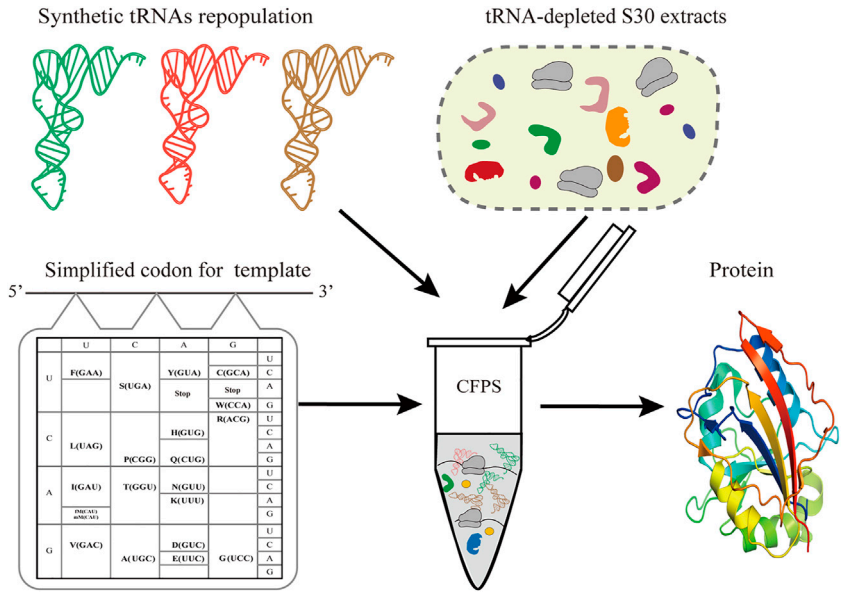


FIGURE 1 | Schematic of simplified codon protein synthesis in cell-free systems. In the sfGFP template, each amino acid is encoded by one codon and simultaneously decoded by one tRNA. In a tRNA-free CFPS system, the protein synthesis of sfGFP is performed by controlling the types and concentration of synthetic tRNAs according to the designed template.

flexibility of protein translation beyond codon limits or interspecies differences. Undoubtedly, this can free up more sense codons for more unnatural amino acids at multiple sites, avoiding the competition for intracellular isoacceptor tRNAs. Ideally, we only need to prepare one mRNA template to generate a variety of protein products by fine-tuning the types of added tRNAs in the CFPS.

However, reprogramming and simplifying the genetic codon is a challenging process that requires the creation of new connections between codons and amino acids. For example, strain C321. ΔA, which has a deletion of the release factor 1 (RF1) gene and is genomically recoded at all UAG stop codons to RF1-independent UAA stop codons, can be used to reassign the blank UAG codon for the incorporation of non-canonical amino acids (ncAAs), which is favorable for industrial protein production (Lajoie et al., 2013). To reduce competition effects in living cells when expanding the genetic codon, it is necessary to use the multiplex genome editing approach, which is time-consuming and labor-intensive, while only one codon is released. Moreover, due to the existence of the wobble base pair, effective decoding by native tRNAs is limited. In one study, Phe and naphthylalanine were respectively assigned at the UUC and UUU codons in a Phe-auxotrophic *Escherichia coli* (*E. coli*) strain (Link and Tirrell, 2005). When using the heterologous tRNA/aaRS/ncAAs system, there is a 20% false incorporation rate in decoding that results from some wobble effects that do not follow the Watson-Crick base pair rules (Kwon et al., 2003). The feasibility of the strategy above depends on the deletion of competing host tRNA(s). Some practical approaches arose when considering strategies to increase the reassignment efficiency for removing the native tRNAs in *E. coli*. It has been proved that deletion of a competing endogenous Arg-tRNA by genetic manipulation and complementation in *E. coli* could successfully reassign the rare AGG (Arg) codon to different ncAAs (Lee et al., 2015; Mukai et al., 2015). However, the entire genome of the cell needs to be re-synthesized, including the knock-out of tRNA genes with sense codons, while the removal of redundant tRNAs may reduce the viability of the cells, leading to a decreased protein production capacity. In addition, it is unwise to engineer the edition and activation sites of the two aaRS, since this requires large mutant libraries and tedious screening. Due to the precise regulation of the native translation system and the high cost of genome reassignment, it is difficult to simplify the codon table and reassign blank codons to new amino acids used in protein synthesis with the simplest codon form.

Compared to the bottleneck problems of *in vivo* systems, an alternative approach is cell-free protein synthesis (CFPS), wherein the components related to transcription and translation can be handled easily according to the experimental requirements. In the last 2 decades, the CFPS system has been developed into a promising technology and established a new field of protein synthesis beyond cells (Hodgman and Jewett, 2012; Lu, 2017). Moreover, it is an effective approach for the extensive reassignment of sense codons *in vitro*. Recently, cell-free gene expression has shown advantages in the production of membrane proteins (Schoborg et al., 2018), therapeutic proteins (Lu et al., 2014; Min et al., 2016), natural products (Li et al., 2018; Moore, 2019), and other value-added chemicals (Yang et al., 2021), and many emerging applications (Liu et al., 2019). Compared with the intact cell system, CFPS generally produces higher

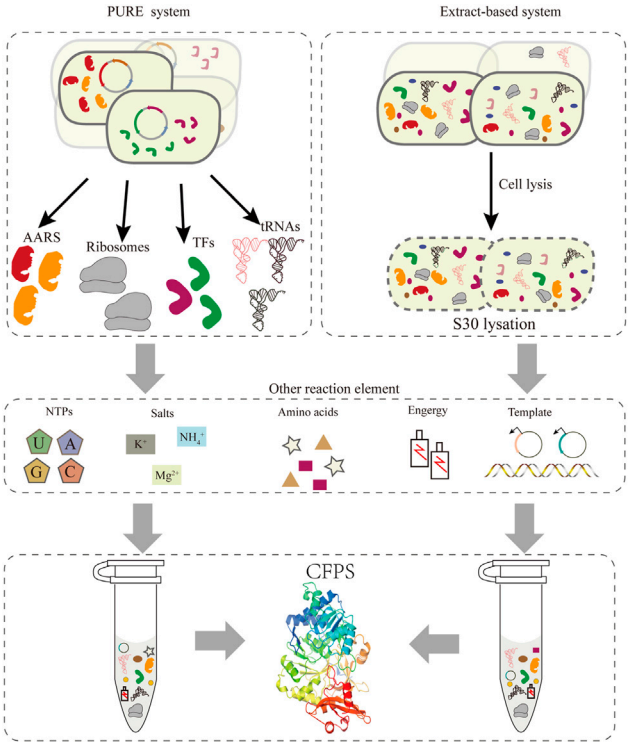


FIGURE 2 | Two strategies exist for implementing CFPS: the PURE system and the extract-based system. In the PURE system (left), each essential factor associated with protein translation is purified separately from overexpressing cells, including aaRS, ribosomes, translation factors, tRNAs, etc. In addition, the corresponding template and reaction buffer components are added to construct a functional translation system. The extract-based system (right) is simple and rapid, since it relies on preparing the native cellular extract by lysing cells.

yields and breaks the limitation of the cell by providing more energy for protein synthesis. At present, cell-free systems can be divided into two major classes (**Figure 2**). One relies on protein synthesis using purified recombinant elements, called the PURE system, which can be reconstructed easily by purifying all translation factors *via* histidine (His)-tags (Shimizu et al., 2001). This system allows the control of the concentrations of all the translation elements, offering greater flexibility for protein synthesis (Shimizu et al., 2005). However, its high cost of purification and concentration-tuning work compared with the extraction system impeded its broader application. Two approaches to improve the scalability of the PURE system were one-pot nickel nitrilotriacetic acid (Ni-NTA) purifications, which use a single batch culture at low cost (Lavickova and Maerkl, 2019), and PURE 3.0 consisting of three high-copy expression plasmids, which enables the bulk purification of necessary translation factors (Shepherd et al., 2017). Extract-based systems, which provide the whole translation machinery in a cell lysate, are another class of CFPS. Extract-based systems are simple and rapid, while also containing factors that contribute to correct folding of functional proteins, such as glycosyltransferases (Jaroentomeechai et al., 2018; Kightlinger et al., 2018). Taken together, CFPS as an open platform in which almost any molecule in the reaction system can be controlled subtly for different experimental purposes. A promising application of this approach is editing of the genetic codon and reassigning sense codons for new information. Exerting better control over the cell-free system entails extract processing and the selective depletion of components of the translation machinery. However, it is necessary to include effective measures to delete the native total tRNA or individual specific tRNAs in this cell-free system to preventing decoding errors.

Here, we provide an overview of the emerging methods for removing the redundancy of the genetic codon to synthesize functional proteins based on a reduced set of tRNAs in the cell-free system. Firstly, we review the published tRNA-depleted S30 extracts including the depletion of total tRNA and specific tRNAs. Additionally, we summarize emerging methods for the purification of specific individual tRNAs *in vivo* and *in vitro*. Finally, we focus on the different reduced codons of different protein syntheses in the PURE system without total tRNA or the tRNA-depleted S30 extract system. More importantly, we point out current trends in the development of minimal codon protein synthesis systems for synthetic biology, biotechnology, and other potential fields.

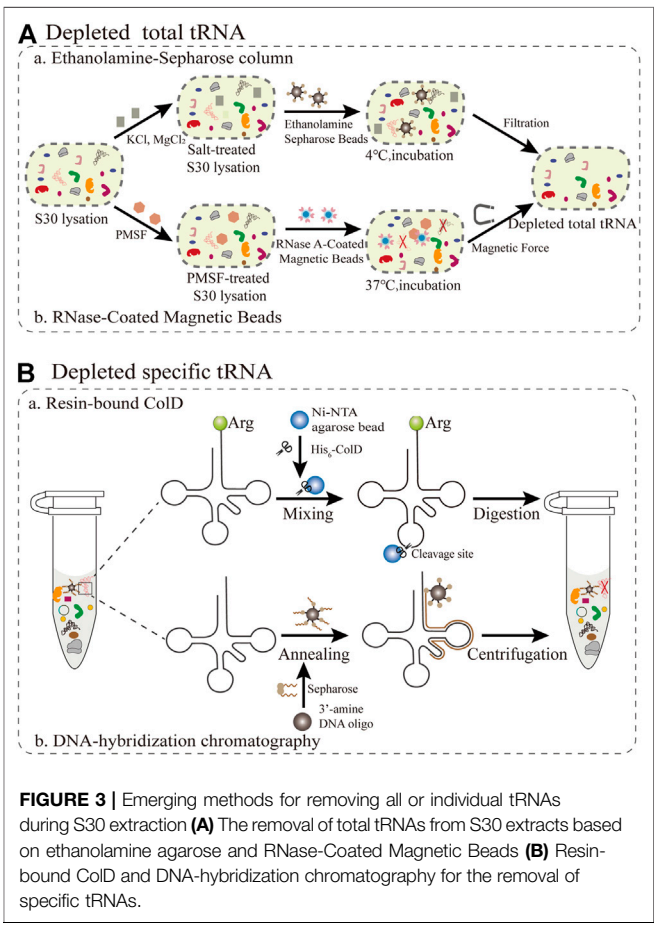


FIGURE 3 | Emerging methods for removing all or individual tRNAs during S30 extraction **(A)** The removal of total tRNAs from S30 extracts based on ethanolamine agarose and RNase-Coated Magnetic Beads **(B)** Resin-bound ColD and DNA-hybridization chromatography for the removal of specific tRNAs.

2 TRNA-DEPLETED S30 EXTRACTS

As adapter molecules, tRNAs play a key role in converting the genetic codon information to the corresponding amino acid sequence in the growing polypeptide chain. Since the molecular weights of tRNAs and small molecules are relatively close, direct dialysis is not suitable for the removal of tRNAs. The main obstacle is the competition between synthetic and endogenous tRNAs, resulting in the production of truncated or inactive proteins. Theoretically, this competition can be averted as long as native tRNAs are eliminated from the cell-free system. Therefore, the target protein can be synthesized intrinsically in a one-to-one relationship among tRNAs, amino acids, and codons. Because all components of the PURE system, such as total tRNA and translation cofactors, are controllable and commercially available, it is simple and convenient to inject single tRNAs (Tuckey et al., 2014). In this section, we mainly summarize and focus on the current methods for removing total tRNAs and specific tRNAs in S30 extraction (**Figure 3**).

2.1 Depletion of Total tRNA

2.1.1 Ethanolamine-Sepharose Column

The first reported method for the removal of total tRNA from S30 extracts is based on ethanolamine agarose (**Figure 3A**). It was discovered by accident that 90% of native tRNA in rabbit reticulocyte lysates could be separated by covalent interactions using the chemical groups of ethanolamine anchored to the resin. (Jackson et al., 2001). For this method, a column of epoxy-activated Sepharose® 6B is more sensitive than ω-aminododecyl-Sepharose. This lysate system depends on supplementation with tRNA and is only applicable for the synthesis of smaller and medium-size proteins, but not for large proteins. The ethanolamine-Sepharose column procedure can expand opportunities for manipulating the tRNA population for reorganizing the genetic codon. However, this method is not universally applicable for all extracts. The most critical step is to optimize the concentration of different salts in the reaction, which influence the interaction between the tRNAs and the ethanolamine-Sepharose resin. Furthermore, there are still several challenges redesigning sense codon for ncAAs owing to the residual total tRNAs. Subsequently, the buffer used to equilibrate the resin was substituted with pure water, resulting in the elimination of about 95% of the total endogenous tRNAs in *E. coli* extracts (Ahn et al., 2006). Although this treatment process is simple and the removal efficiency is improved, a small amount of tRNA is still present, which increases the risk of residual tRNA coupling with amino acids. To alleviate the negative effects, reaction conditions including salts, ionic strength, temperature and retention time could be optimized to increase the interaction between tRNA and ethanolamine agarose so as to facilitate the construction of a tRNA complement protein synthesis system.

2.1.2 RNase-Coated Magnetic Beads

The reassignment for sense codons requires the complete removal of native tRNAs from the cell extract. A promising emerging approach is based on RNase-coated magnetic beads and a phenylmethylsulfonyl fluoride (PMSF)-treated cell extract, which results in near-complete tRNA depletion (Salehi et al., 2017). For this approach (**Figure 3A**), ribonuclease A (RNase A) attached to superparamagnetic beads was more convenient than an immobilized RNase A Resin in controlling the RNase degradation ability and subsequently removing the RNase from the extract (Kunda et al., 2000). In addition, this method can make full use of the protective effect of nucleoproteins, so that it not only degrades tRNA but also ensures the activity of rRNA (Suhasini and Sirdeshmukh, 2006). PMSF treatment has a two-sided effect, since it inhibits proteases fivefold less than using RNase inhibitor, while also preventing RNaseA from leaching into the cell extract. Notably, the rate of tRNA removal was directly measured using quantitative real-time PCR (qPCR) with tRNA-specific primers (Kralik and Ricchi, 2017). The average removal ratio for all assessed native tRNAs was 99.3% following RNase A treatment for 60 min. However, there are still some problems that need to be addressed. The cell extract treated with RNase A-beads was able to produce a designer peptide containing 40 Val residues with the addition of a synthetic tRNA, but it remains unclear if proteins with larger molecular weights can be synthesized. The removal of tRNA by the magnetic bead method is only applicable for the cell extract of *E. coli*, and must be further optimized for other bacteria or eukaryotes. In general, this method is more practical than most current methods for tRNA removal. The direct quantification by PCR is a valuable tool to assess cell extracts for high-fidelity codon reassignment. Looking forward, this robust platform has the potential to be applied in emerging synthetic biology and biomedical engineering applications.

2.2 Depletion of Specific tRNAs

2.2.1 Resin-Bound Colicin D

For specific tRNA removal, the primary issue is codon choice, while excluding Trp and Met with only one encoding codon. According to previous studies, the tRNase colicin D (ColD) can specifically recognize and degrade four different $tRNA^{Arg}$ species of *E. coli*, including the $tRNA_{ICG}$, $tRNA_{CCG}$, $tRNA_{UCU}$, and $tRNA_{CCU}$ (Tomita et al., 2000). This method (**Figure 3B**) takes full advantage of this property to inactivate all the $tRNA^{Arg}$ from the *E. coli* cell extract (S12) (Kim et al., 2006), resulting in a inability to incorporate Arg in this system. It was further demonstrated that ColD-treated lysates did not affect the ability of protein synthesis, under the condition of supplementing $tRNA^{Arg}$. More importantly, this creates favorable conditions for rearrangements of the remaining Arg codons to expand the genetic codon (Lee et al., 2016). There are two reasons for choosing AGG as the sole codon encoding Arg. Firstly, this is a rare codon, while $tRNA^{UCU}$ and $tRNA^{CCU}$ are relatively resistant to ColD digestion, resulting in a residual amount remaining in the treated S12 lysates. For specific sense codon rearrangement, it is also important to take into account the flexible effect of wobble base pairs in the *in vivo* decoding process (Das and Duncan Lyngdoh, 2014). Although this method is simple and feasible in reconstructing translation systems, it can only be used to rearrange Arg codons and not for other types of tRNAs. Therefore, there are more other tRNases to be explored for use in similar approaches, such as the $tRNA^{Lys}$

anticodon nuclease PrrC (Blanga-Kanfi, 2006). Overall, this method expands further implementations for building a platform for a small number of sense codon rearrangements, and it creatively broadens new horizons and ideas for reconstructing a more general platform.

2.2.2 DNA-Hybridization Chromatography

Emerging attempts have been made to free up more sense codons for the incorporation of multiple ncAAs in the genetic codon simultaneously. Alexandrov et al. demonstrated that isoacceptor tRNAs can be removed through DNA-hybridization chromatography of the standard *E. coli* S30 lysate (Cui et al., 2017). For this approach (**Figure 3B**), DNA oligos with 3′-amine modifications were designed complementary to the sequence spanning the D-arm down to the anticodon loop of the targeted native isoacceptor tRNA ($tRNA^{Arg}_{UCU}$), and were immobilized on an NHS-activated matrix. Then, the treated DNA oligos are mixed with the extract to carry out the hybridization reaction for the chromatographic depletion. The authors introduced a novel criterion for tRNA removal efficiency as follows:

$$\textbf{tRNA removal efficiency} = 1 - \left(\frac{\textbf{RFU}\left(\textbf{Depleted tRNA}\right)}{\textbf{RFU}\left(\textbf{Depleted tRNA} + \textbf{t7 tRNA}\right)}\right)$$

According to this formula, there are many factors contributing to the more precisely quantified tRNA removal. First of all, the feasibility of this method was demonstrated since the protein expression level could be restored with the addition of the targeted synthetic tRNA in the tRNA-depleted lysate. The depletion efficiency is also related to the number of corresponding decoded codons contained in the template of protein synthesis. For example, using super-folder green fluorescent protein (sfGFP) as the template, the deletion rates of one and six AGG codons were nearly 100 and 90%, respectively. Therefore, the depletion efficiency is determined depending on codon composition in different templates, which further illustrates the role of a specific tRNA and the ability of protein synthesis. Early studies identified a biased distribution of tRNA abundance for bacteria growing at different rates (Dong et al., 1996). This could directly affect the quality of the prepared cell lysate. To ensure basic levels of protein synthesis, several new technical means can be used to accurately measure the amount of specific tRNAs to determine the optimized time for collection of bacteria, such as the modification-induced misincorporation tRNA sequencing (mim-tRNAseq) method (Behrens et al., 2021).

Subsequently, several implemented effective measures were taken to abrogate the targeted tRNA activity and liberate the corresponding sense codons (Cui et al., 2018). This strategy offers several improvements over the above-mentioned methods. Firstly, the antisense oligonucleotides are not affected by the high-level complex molecular structures in tRNAs that reform the L-shaped conformation. Secondly, 2′O-Me modifications of the antisense oligonucleotide perform better than other types in the entropic stabilization to the hybrid base pairs (Yildirim et al., 2014). It should be noted that methylated oligonucleotides display slow dissociation kinetics from the target tRNA, resulting in better RNA strand displacement. This is particularly important for the designed sequence targeting a specific tRNA in the *E. coli* S30 lysate. The hybridization ability shows inconsistencies in the sequence of the same tRNAs between species (*E. coli* and *Leishmania tarentolae*), even though they are located in the anticodon or variable loop region in the same cell. A few heterologous factors need to be further explored to improve the general applicability of the method, such as the input ratio of tRNA and oligonucleotides, incubation time, and temperature, which requires further work in the future. In addition, the two-step protein synthesis reaction system was divided into the treated extract and the total RNA. The described standard *L. tarentolae* extract (LTE) (Mureev et al., 2009) was generated by removing the total RNA by the ethanolamine-sepharose column method. The total RNA was isolated from the lysates. Then, individual isoacceptor tRNAs were selectively sequestered from the purified total RNA by the treatment with DNA oligos with 2′O-Me modifications. Finally, the above two parts are added to the whole reaction system to minimize the background level. Although a somewhat cumbersome, this method is worth learning and considering to reduce the background activity. Treatment with the RNase-related compound angiogenin (ANG) (Su et al., 2019), which can cleave tRNA anticodons to inactivate specific codons, may be a feasible strategy for freeing up sense codons. More effective strategies for the tRNA denaturation approach are expected to be developed in the near future.

3 SPECIFIC PURIFICATION OF INDIVIDUAL TRNAS

To construct a different reduced-codon set to support protein translation, the source of tRNA preparation and tuning the composition of the tRNA pool is critical. Recent studies have demonstrated various methods for the preparation of tRNAs for *in vitro* translation systems. Nevertheless, current methods for the isolation of specific tRNAs from host cells involve complicated procedures. In the following section, we discuss the growing research on the preparation of individual tRNAs *in vivo* and *in vitro*, with an emphasis on three methods *in vivo* (**Table 1**).

3.1 *In vivo*

The posttranscriptional modifications of tRNAs play vital roles in the accurate decoding of mRNAs inside the ribosome (Agris et al., 2017). In particular, the tRNA anticodon domain (positions 34 and 37) contains various specific modifications that facilitate the recognition of cognate and wobble codons, which directly influence the translational fidelity (Pereira et al., 2018). Although intracellular purification remains a challenge, there are several novel advances in the purification of specific tRNAs produced *in vivo* (**Figure 4**).

TABLE 1 | Advantages and disadvantages of different tRNA production methods.

Approach	Method	Advantages	Disadvantages	References
Purification *in vivo*	Hydrophobic tagging	Modified nucleotides, low cost, simple operation	Multiple complicated procedures, cannot distinguish isoacceptor tRNAs	Kothe et al. (2006)
	DNA probe-elution	Cost-effective, modified nucleotides, simple operation	Low yield, low purity, low binding efficiency	Kaneko et al. (2003)
	DNA probe-digestion	High purity, high specificity, well-established method	Extensive downstream purification, using toxic reagents	Nilsen (2013)
In vitro biosynthesis	Direct	Easy purification, fast, the mutagenesis of the tRNA is easy	Addition of nucleotides at the 3′end, the first base must be a guanine, without modified nucleotides	Li et al. (1999)
	Hammerhead	Fully active in aminoacylation; Not limited to the first base	No 5′phosphorylation, without modified nucleotides, addition of nucleotides at the 3′ end	Fechter et al. (1998)
	RNase P	Not limited to the first base; no addition of nucleotides at the 3′ end	Without modified nucleotides	Hibi et al. (2020)
Chemical synthesis	Solid-phase chemical synthesis	Modifications possible, easy purification, fast, no sequence-specific optimization	Expensive equipment required, length limited, limited availability of labeled or modified phosphoramidites	Ogilvie et al. (1988)

3.1.1 Hydrophobic Tagging Method

In protein translation systems, the requirement for purifying individual tRNAs *in vivo* has become increasingly prominent. A prominent approach relies on the introduction of a hydrophobic tag for the specifically charged aminoacyl-tRNAs (aa-tRNA) followed by hydrophobic interaction chromatography (**Figure 4A**) (Kothe et al., 2006). The used hydrophobic tag was 9-fluorenylmethyl-succinimidyl carbonate (FmocOSu), which can react with the free amino group of a selected aa-tRNA. Therefore, the hydrophobicity of this complex is significantly increased compared with other tRNAs, which in turn increases the retention time in the chromatographic column, and finally results in its isolation from total tRNA. Although the used raw material is available at a low cost, and the operation is relatively simple, the entire purification process requires multiple steps, including aminoacylation, modification, purification, and deacylation. In addition, this hydrophobic tagging method has broader applicability for diverse tRNAs than previous approaches (Cayama et al., 2000), because it solely depends on the specific aa-tRNA synthetases and the specific amino acids. The aaRS can only aminoacylate tRNA with one specific amino acid, but it cannot recognize and distinguish the isoacceptor tRNAs. According to this principle, the purified species are a complex of tRNAs encoding the same amino acid. This method is therefore not suitable for the fine-tuning of transfer RNA (tRNAs) in the cell-free lysate. Recent research found that a two-dimensional liquid chromatography (2D-LC) integrating a weak anion-exchange method could be used to isolate $tRNA^{Val}_{UAC}$ and $tRNA^{Leu}_{CAG}$ (Cao et al., 2020). These purified individual tRNAs might be applied to the minimal codon protein synthesis system. In terms of the cost and feasibility of purification, these methods will open new avenues in the process of simplified chromatographic purification.

3.1.2 DNA Probe-Elution Method

In an early study on probe-elution, biotinylated DNA oligonucleotides were immobilized onto streptavidin agarose beads to isolate individual tRNAs from the *Leishmania tarentolae* (Kaneko et al., 2003). Recently, Söll et al. incubated 5′biotinylated oligo-beads with yeast tRNAs in the equilibrated buffer, and this complex was washed several times to remove the nonspecifically bound or loosely bound molecules. The pure $tRNA_m^{Glu}$ was eluted by a low-salt buffer at high temperature (Rinehart et al., 2005). Although the principle and operation of this procedure are relatively simple, verification of the tRNAs from a cell by northern blot may be necessary (**Figure 4B**). However, the current purified volume and purity are relatively low for the distribution and regulation of different species of tRNA for *in vitro* protein synthesis, which hinders its application and development. The following different factors are worth exploring to produce high-quality tRNA. First of all, the individual tRNAs could be over-expressed by constructing recombinant plasmids with a strong constitutive *E. coli* promoter (Cervettini et al., 2020). The main purpose of this process is to enrich enough tRNA for initial extraction and purification, greatly reducing the proportion of non-target tRNAs. In addition, the combination of a chromatographic spin column and streptavidin-coated agarose beads may have more obvious advantages in improving the binding and purification efficiency for bulk purification. The continuous circulation process in the Chaplet Column Chromatography (CCC) method circumvents these drawbacks to a certain extent (Suzuki and Suzuki, 2007). Similarly, the choice of synthetic 3′-biotinylated DNA probes and chromatographic columns can be changed according to the amount of the *Bos taurus* mitochondrial tRNA mixture, exhibiting great flexibility and scalability.

3.1.3 DNA Probe-Digestion Method

With easy availability, biotinylated DNA oligonucleotide probes complementary to the targeted tRNA were successfully used to extract the tRNA-derived stress-induced RNAs (tiRNAs) fraction by the gel purification process (Nilsen, 2013). The hybridized tiRNA-oligo complex was pulled down using streptavidin agarose beads (**Figure 4C**). With the addition of the Trizol reagent, the targeted tRNA fraction can be freed *via* digestion with DNase I (Akiyama et al., 2020). While initially met with some

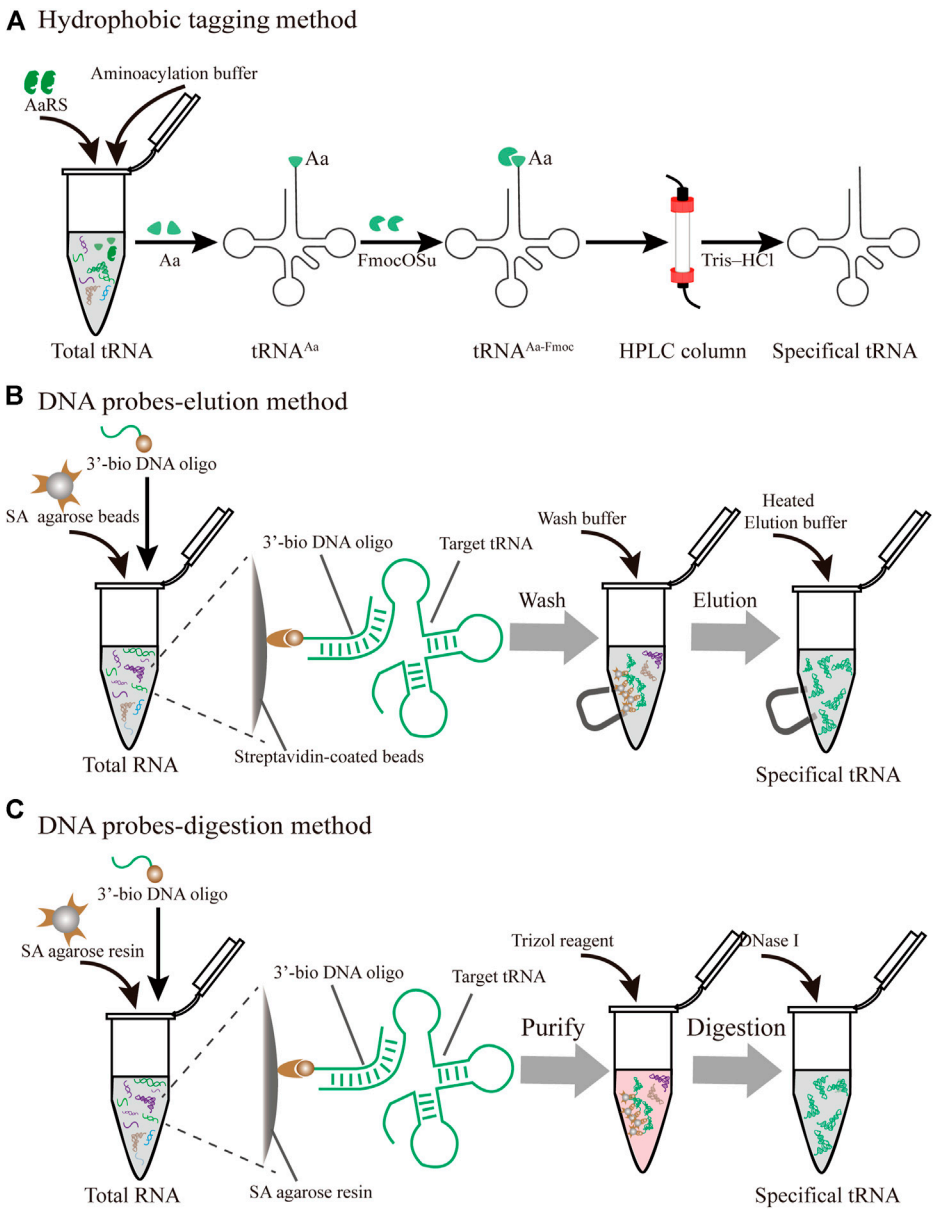


FIGURE 4 | Different strategies for the purification of specific individual tRNAs produced *in vivo* **(A)** Hydrophobic tagging method. The hydrophobic tag (FmocOSu) can react with the free amino group on the aminoacylated tRNA, which can be separated from other tRNAs due to its high molecular weight, then purified and de-acylated **(B)** DNA probe-elution method. Biotinylated DNA oligonucleotides are immobilized on streptavidin sepharose beads, and individual tRNAs are isolated from total RNA using a magnet **(C)** DNA probe-digestion method. The hybridized tRNAs-oligo complex was purified from the tRNAs-oligo-resin complex using Trizol. Individual tRNAs can be released using DNase I

skepticism, this approach was developed into feasible methods for the isolation of endogenous mature tRNAs, such as $tRNA^{Ala}_{AGC}$ and $tRNA^{Gly}_{GCC}$. Since tRNAs have a complex molecular structure and are heavily modified, this approach conquered the current limitation of tRNA purification from host cells and opened up a new Frontier in tRNA biology. Nevertheless, several disadvantages of this approach should be addressed. To obtain high-purity target tRNA, toxic Trizol was used repeatedly throughout the purification process, which should be circumvented by seeking alternative reagents. The designed sequences of DNA probes need to be optimized due to the high sequence similarity between the selected tRNA and its isoacceptor tRNAs. Finally, there is a lack of scientific evaluation criteria for the purity and activity of purified tRNA. Emerging technologies, including high-throughput tRNA sequencing (Shigematsu et al., 2017; Erber et al., 2020) and immuno-northern blotting (Mishima et al., 2015; Mishima and Abe, 2019), may be applied to push forward tRNA purification technology.

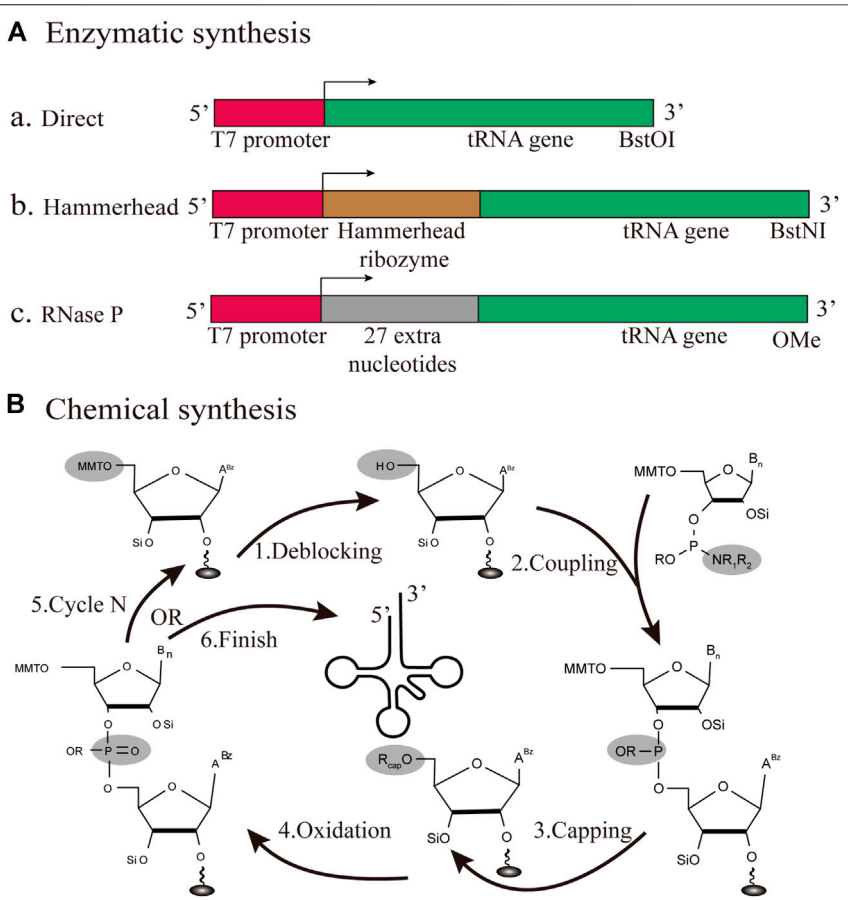


FIGURE 5 | Different strategies for the production of individual tRNAs *in vitro* **(A)** Enzymatic synthesis. Three different patterns for tRNA transcription are based on the T7RNP transcription process, including the direct pattern, hammerhead pattern and RNase P pattern **(B)** Chemical synthesis. The production of tRNA is based on solid-phase chemical synthesis, with each cycle encompassing de-blocking, coupling, capping, and oxidation.

3.2 *In vitro*

In this section, we will provide a summary of *in vitro* tRNA production methods, mainly including enzymatic and chemical synthesis (**Figure 5**). This part aims to discuss several crucial points of the tRNA preparation literature, including the latest advances.

3.2.4 Enzymatic Synthesis

Among bacteriophages T3, T7, and SP6, T7 RNA polymerase (T7RNP) offers the largest yield of RNA transcription. Since the length of tRNA is between 75 and 95 nt, the T7RNP transcriptional synthesis of small RNA has been established in initial reports (Milligan and Uhlenbeck, 1989). Originally, there were two major flaws in the T7 transcription process. One defect is that its transcription efficiency mainly depends on the specific recognition of its cognate promoter sequence, incorporating an unfavorable sequence at the 5′-end. On the other hand, the runoff of T7 transcription results in a mixture of products with 3′-end heterogeneity, which usually results in multiple consecutive A nucleotides at the end of the product (Helm et al., 1999). Three different transcription patterns for tRNA transcription will also be discussed based on their applications in the field of transcriptional product heterogeneity (**Figure 5A**). In a direct pattern, the transcription template starts with the T7 promoter sequence, followed by the encoded tRNA gene sequence, which is directly transcribed (Li et al., 1999). This approach works for most mature tRNAs within cells, but it is limited to tRNA sequences that do not begin with a G, such as $tRNA^{Asn}$ and $tRNA^{Pro}$. The ribozyme pattern is a fusion between the T7 promoter sequence and the tRNA gene, encoding a ribozyme called hammerhead, which has a cis-acting, self-cleavage ability (Huang et al., 2019). This design intends to circumvent the above-mentioned deficits and release a tRNA transcript with the desired 5′- sequence (Korencic et al., 2002). The main advantage of this approach is that pure transcripts can be obtained directly in large quantities for all tRNAs. Although the 5′-OH tRNA transcript can be aminoacylated in this reaction, they proved to be active. Additionally, RNase P can catalyze tRNA maturation with the generation of tRNAs with homogeneous 3′and 5′ends (Fukunaga et al., 2006). RNase P has a catalytic RNA subunit, which can cleave homogeneous 3′-OH ends in tRNAs of interest (Gossringer et al., 2012). The 2′-methoxy modification of the second nucleotide 5′end at the primer prevents additional nucleotide amplification of the 3′-terminal transcripts.

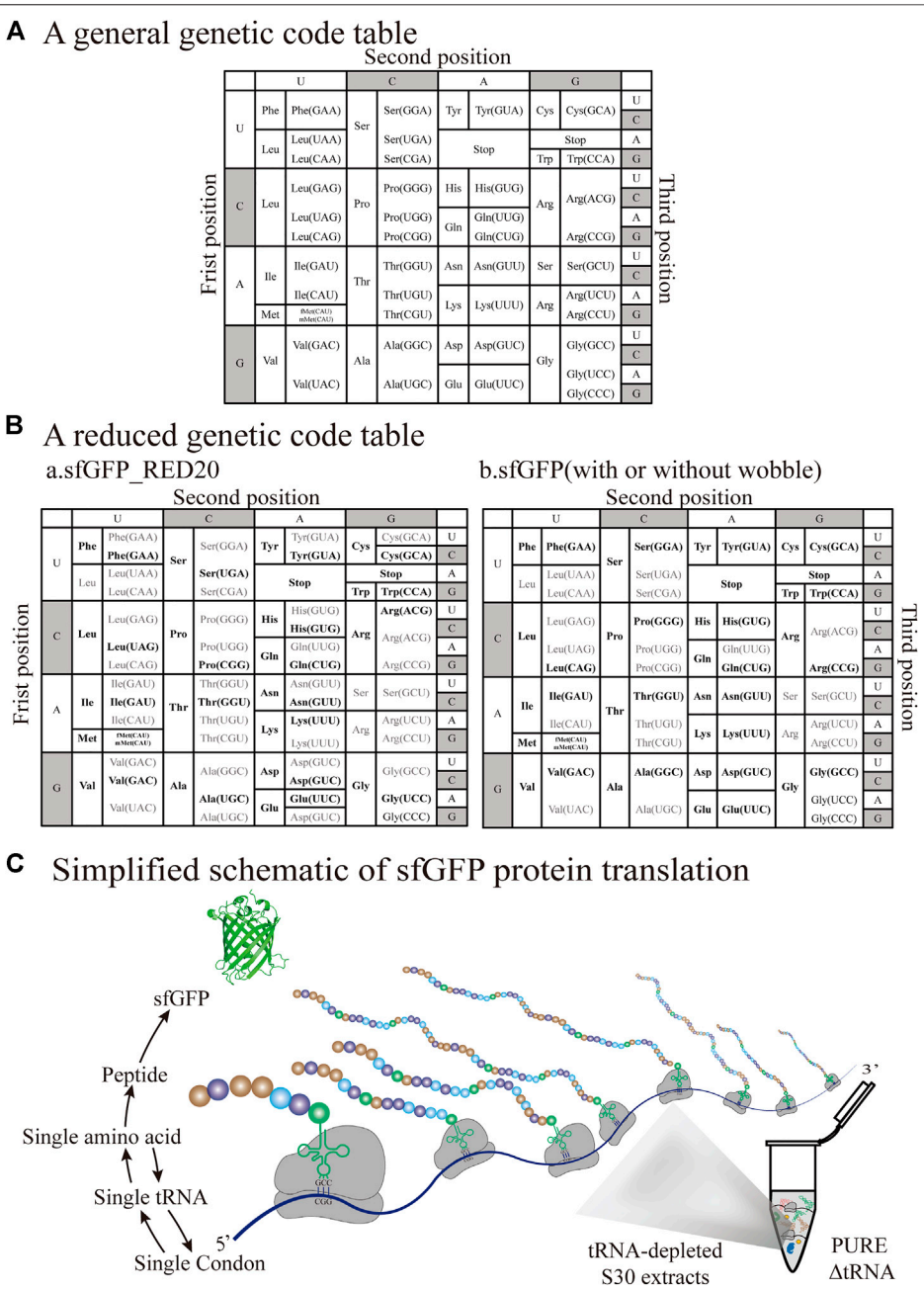


FIGURE 6 | Selection of different tRNA sets for the decoding of sfGFP templates with reduced-codons **(A)** The native genetic code table of *E. coli* **(B)** Two reduced genetic code tables for the production of different versions of sfGFP (sfGFP_RED20 and sfGFP; with or without wobble, respectively, Unused codons are represented in gray, while frequently used codons are represented in black **(C)** Simplified schematic of sfGFP protein translation based on the tRNA-depleted S30 extract and PURE ΔtRNA system. This sfGFP mRNA clearly shows the one-to-one correspondence in the codon-tRNA-amino acid relationship.

Therefore, the latest research fully integrated the above two characteristics, to achieve the production of transcripts of high quality (Hibi et al., 2020). Using this approach, 21 different tRNAs from *E. coli* were analyzed by urea-PAGE, and the homogeneity of the major synthesis was successfully verified. What's more, the purified iVTtRNA can be sufficiently aminoacylated by the corresponding aaRS. This breakthrough demonstrates that despite lacking any modifications *in vivo*, tRNA transcribed using RNase P was capable of protein synthesis comparable to the native tRNA mixture. The *in vitro* transcribed transfer RNA (iVTtRNA) lacks the modifications at position 37, which possibly allow undesirable wobbling at codon first positions and frameshifts. Indeed, generating the complex base modifications in the anticodon loop regions of iVTtRNA will be an important perspective in decoding accuracy required for the appropriate modification enzymes like the TsaD and TrmD

TABLE 2 | A reduced set of tRNA used for protein synthesis based on the cell free system.

Amino Acids	Codons	tRNAs	Product	tRNA source	Reaction system	References
11	11	11	RGS peptide	*in vitro* transcribed	tRNA-depleted S30 extracts	Cui et al. (2015)
20	45	41	sfGFP_T1			
20	30	26	sfGFP_T2			
20	44	39	sfGFP_T3			
23	23	32	P9-Naa peptide	*in vitro* transcribed	PURE ΔtRNA	Iwane et al. (2016)
2	2	2	(Val)40	*in vitro* transcribed	tRNA-depleted S30 extracts	Salehi et al. (2017)
20	20	21	sfGFP-RED20	chemically synthesized	PURE ΔtRNA	Calles et al. (2019)
20	20	21	sfGFP	*in vitro* transcribed	PURE ΔtRNA	Hibi et al. (2020)

modification enzymes. The introduction of modified nucleotides into the anticodon loop of iVTtRNAs shows a significant increase in the aminoacylation, the protein yield, and the fidelity. Overall, this is a fast, inexpensive and efficient method for the production of specific tRNAs of good purity, providing favorable conditions for *in vitro* protein synthesis based on the iVTtRNA system.

3.2.5 Chemical Synthesis

In early studies, solid-phase chemical synthesis was used for the production of 77 nt $tRNA^{Met}$ (Ogilvie et al., 1988). The whole production cycle involves de-blocking, coupling, capping, and oxidation (**Figure 5B**). According to the sequence of chemical synthesis from the 3′- to the 5′-terminus, the synthetic target tRNA is finally obtained through reversed-phase high-performance liquid chromatography (HPLC) purification (Baronti et al., 2018). From the cost perspective, the upper limit of the synthesizable length is approximately 80 nt. In this approach, the synthesis mainly depends on the used phosphoramidite monomers and the spatial structure of target sequences. Accordingly, synthetic production is a challenge for all tRNAs, which are generally structurally complex molecules. However, it is still feasible and effective to adopt chemical methods for the synthesis of tRNAs with specifically modified bases (Vare et al., 2017). Chemical synthesis has been used for decades in the production of small RNAs, and it still plays a major role. For future studies, there is an urgent need to develop a lower-cost, high-efficiency protocol, focusing on the manipulation of tRNAs.

4 SIMPLIFIED CODON PROTEIN SYNTHESIS WITH TRNA COMPLEMENT

The general codon table is a genetic coding rule shared by almost all organisms (**Figure 6A**). Since the 20 common proteinogenic amino acids are decoded by 61 codons (triplets), there is a great redundancy in the standard code. To achieve protein synthesis with reduced codons, two major factors need to be considered in a cell-free system. The first factor is to exclude the influence of endogenous tRNAs in the extract, so it is necessary to establish a cell-free system with completely removed native tRNAs. The second factor is the synthesis and purification method of added tRNA for the manipulation, so it is necessary to seek an alternative way for the synthesis of tRNA. Generally, these two conditions have been discussed above. Numerous attempts to produce different proteins by reducing codon redundancy have been published. Here we present an overview of the designed pattern of different protein synthesis approaches based on the tRNA-depleted S30 extract and PURE ΔtRNA system (**Table 2**). Although the complete removal of endogenous tRNAs is currently still a challenge, some of them have important reference value and milestone significance in the application of simplified and minimal codon synthesis.

4.1 tRNA-Depleted S30 Extracts

This novel semisynthetic tRNA complement system was primarily used for reassigning sense codons in protein synthesis based on the depleted endogenous tRNAs of BL21 (DE3)GOLD S30 cell extract (Cui et al., 2015). The *in vitro* peptide expression assay takes advantage of synthesizing a short peptide reporter named the RGS-peptide (peptide RGSIDTWV) (Stein and Alexandrov, 2014). An advantage of this method is that it not only reduces the pressure of the amount of codons used in the expression template, but also efficiently evaluates the efficiency of iVTtRNA decoding. This approach provides a simple and efficient alternative to the preparation of purified tRNA for verification of aminoacylation activity, without the complex process of radiolabeling employed previously (Schwartz and Pan, 2017). Importantly, it was found that 17 amino acids can be efficiently decoded by the corresponding iVTtRNAs, with the exception of Glu, Asn, and Ile. Therefore, a tRNA mixture containing 20 natural amino acids was constructed by combining *in vivo* purification with an *in vitro* transcription strategy. What's more, sfGFP templates containing changed sets of codons (Pedelacq et al., 2006) were used to confirm that the tRNA mixture retains the decoding ability of the synthesized full-length protein, as expected. It is worth noting that different versions of the sfGFP gene have been simplified to free up more sense codons. At present, the least tested is the use of 26 different tRNAs and 30 amino acids with the template sfGFP_T2 (**Figure 6B**). Although the complete removal of native tRNAs remains challenging, this approach has broken the existing bottleneck for the redundancy of the general genetic codon, significantly broadening the artificially designed platform for protein synthesis using the smallest number of codons. However, recent work has taken advantage of the real-time PCR method for the accurate quantitative analysis of the remaining amount of tRNA in the treated cell extract, rather than relying solely on protein synthesis ratios (Salehi et al., 2017). A synthetic $tRNA^{fMet}$ and $tRNA^{Val}$ were successfully used to

produce a polypeptide named $(Val)_{40}$, achieving artificially the one-to-one correspondence between the codon-tRNA-amino acid relationship. Although there are only two amino acids in this polypeptide, the feasibility of this method has been preliminarily verified, which provides the conditions for the synthesis of macromolecular proteins. It is clear that residual tRNAs substantially decrease the codon reassignment efficacy of macromolecular protein synthesis. As mentioned above, advances in this direction have been made in the evaluation and optimization of the complete deletion of the native tRNA mixture to simplify the genetic codon.

4.2 PURE ΔtRNA

The reconstituted PURE system (PURE ΔtRNA), which lacks all native tRNAs, is an appropriate system for the manipulation of different tRNA sets for protein expression. The introduction of a new concept of artificial codon-box division is based on blanking a sense codon, which is then reassigned to a non-natural amino acid (Iwane et al., 2016). In this methodology, the flexible *in vitro* translation (FIT) system (Hipolito and Suga, 2012), which contains only 32 *in vitro* tRNA transcripts, was referred to as the FIT-32t system. Using this FIT-32t system, the authors synthesized the P9-Naa peptide, which contained the 20 natural amino acids and three non-natural amino acids (^{Ac}K, ^{Iodo}F, and Cit), which was demonstrated to be the major product by MALDI-TOF-MS. Although the relative yield of this P9-Naa peptide was approximately 15% of that obtained using the native tRNA mixture, it demonstrated the feasibility of reducing the redundancy of the genetic codons. In addition, the established RED20 code system is defined that the map containing the 20 sense codons corresponding to 20 amino acids with the addition of the three stop codons (Calles et al., 2019). In this sfGFP-RED20 template system, there are significant differences in the total protein-expression capacity between the 21 synthetic tRNAs and the full set of natural tRNAs. This difference may be caused by many different factors, such as a lack of modification on the synthetic tRNAs, which affects their ability to decode the mRNA, as well as the frequency of codon usage or the promiscuous decoding of null codons, among others. This reduced set of tRNAs encoding RED20 will play a foundational role in reassigning sense codons and simplifying the general codon box. Consistent with this approach, the developed 21 iVTtRNAs without nucleotide modifications were able to restore the sfGFP and dihydrofolic acid reductase (DHFR) production capacity based on the PURE system (Hibi et al., 2020). In consideration of the effect of wobble base pairings, an sfGFP template with or without wobble bases was designed to demonstrate the fidelity of protein translation using this system (**Figure 6B**). A single aminoacyl-tRNA can decode multiple codons since "wobble" base pairs in natural translation system. Therefore, the genetic codon box cannot be simply divided out of the blank codons for the rearrangement of novel amino acids. Due to the wobble base pair, the designed any set of 20 tRNA species can cover more than 20 codons, leaving significantly less than 41 codons to novel amino acids. An ideal approach would be to artificially assign each codon to a specific tRNA and a specific amino acid to achieve a one-to-one correspondence (**Figure 6C**). This feature is not applicable in natural organisms due to biological and artificial isolation phenomena. This artificial design pattern provides a powerful platform for the exploration of biosafety issues (Schmidt, 2019). However, a new direction of these redesigned codon systems should be explored in the future. Firstly, it is worthwhile to construct a universal database that optimizes the best tRNA set for different functional proteins to achieve the highest possible yield and efficiency. Furthermore, ensuring the fidelity of the translated opened reading frame is necessarily important in the protein translation process. It is well known that particular codons and sequence contexts are "shifty" and frequently cause frameshifts. This follows that a particular set of 20 codons would better maintain the reading frame than other possible sets. The iVTtRNA transcription and tRNA modification enzymes need to be coupled to maintain the fidelity of decoding. Finally, advances have been made in the incorporation of multiple non-natural amino acids and the construction of a minimal cell system.

5 CONCLUSIONS AND PROSPECTS

Simplified codon protein synthesis with tRNA complement has unique advantages, which were summarized and discussed in this review. To obtain the tRNA-depleted S30 extract, several approaches have been developed to remove the native tRNA set, including the ethanolamine-sepharose column and RNase-coated magnetic bead methods. In addition, further methods for the depletion of specific tRNAs from S30 lysates were also generalized, including the enzymatic depletion method and DNA-hybridization chromatography. The advantages and disadvantages of these methods have been summarized and discussed in detail above. At the same time, the preparation of tRNA is still a very important factor, and this paper also comprehensively summarizes *in vivo* and *in vitro* approaches for targeted tRNA synthesis. The strategies of hydrophobic labeling, DNA probe-elution, and DNA probe-digestion were mainly summarized in the section on *in vivo* purification. These approaches can be used to purify a specific tRNA and have modification advantages compared with *in vitro* purification. What's more, compared with *in vivo* purification and chemical synthesis, the T7RNP-based method for enzymatic tRNA synthesis has many advantages, such as simple operation, low cost, and high yield. Overall, the use of a combination of *in vivo* purification and T7RNP-based enzymatic purification may be the most efficient method for tRNA purification applications. Finally, the key issues and challenges in the tRNA-depleted S30 extract and PURE ΔtRNA systems for different versions of the existing artificially designed genetic codon tables were discussed individually. By emphasizing these crucial elements for simplified codon protein synthesis in a cell-free protein system, this review is intended to provide current progress and explore new challenges in the field of *in vitro* synthetic biology.

Looking forward, this emphasized tRNA complement with protein synthesis system has the potential to be applied in various emerging fields, such as the construction of minimal cells (Noireaux et al., 2011; Stano and Luisi, 2013; Yue et al., 2019), the employment of microfluidic devices (Damiati et al., 2018; Weiss et al., 2018), the incorporation of multiple different ncAAs (Zheng et al., 2018; Chemla et al., 2019; Italia et al., 2019), the direct preparation of bio-conjugates (Bundy et al., 2008; Zimmerman et al., 2014) and other novel biotechnology applications. Aiming at controllably reconstituting a minimum set of compounds, the simplified genetic codon with the CFPS system will play a significant role in the bottom-up assembly of minimal cells. In particular, several efforts have been devoted to making breakthroughs in redesigning genomes by freeing up more sense codons. For instance, the genomic reassignment of the 57-codon *E. coli* MDS42 (*rE.coli-57*) was a major attempt at constructing a fully recoded organism (Ostrov et al., 2016). In this *rE.coli-57*, a total of seven selected codons were replaced with their synonymous codons on a genome-wide scale to expand the genetic codon. However, only four orthogonal nnAAs with novel chemical properties could be incorporated into the targeted protein. By contrast, the Syn61 with a 61-codon genome has a complete genome-wide deletion of the three target codons (amber-TAG, S-TCG, and S-TCA), with their defined synonymous substitution (Fredens et al., 2019). It was demonstrated that the reassignment of vacant TCG codons provides the possibility of further codon expansion. Despite the high synthetic cost and cumbersome construction process, this genomic engineering approach could help release sense codons for rearrangement to a certain extent. In future work, comprehensive research coupling the genomically recoded organisms and the CFPS system might be an alternative approach for enabling the recoding of much more sense codons. What's more, several emerging approaches such as flexizyme aminoacylation (Katoh et al., 2018; Lee et al., 2019) and the evolved orthogonal aaRS/tRNA pairs (Bryson et al., 2017; Willis and Chin, 2018) could be applied to the minimal codon system for the synthesis of non-canonical biopolymers. Taken together, construction of cell-free protein synthesis with manipulating tRNA, which reduces the use of codons and tRNA species, might provide favorable conditions for simplified cellular mimics or self-replicating systems (Van Nies et al., 2018). We firmly believe that further systematic exploration of the minimal codon protein synthesis system will advance the development of synthetic cells and broaden the field of synthetic biology.

AUTHOR CONTRIBUTIONS

JL collected the relevant subject literature, designed the frameworks and figures, and drafted the original manuscript. HQ conceived the presented idea and supervised the writing of the article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Foundation of China (Grant No. 21778039) and the National Key R&D Program of China (Grant No. 2019YFA0904103).

REFERENCES

Agris, P. F., Narendran, A., Sarachan, K., Väre, V. Y. P., and Eruysal, E. (2017). The Importance of Being Modified: The Role of RNA Modifications in Translational Fidelity. *Enzymes* 41, 1–50. doi:10.1016/bs.enz.2017.03.005

Ahn, J.-H., Hwang, M.-Y., Oh, I.-S., Park, K.-M., Hahn, G.-H., Choi, C.-Y., et al. (2006). Preparation Method forEscherichia Coli S30 Extracts Completely Dependent upon tRNA Addition to Catalyze Cell-free Protein Synthesis. *Biotechnol. Bioprocess Eng.* 11, 420–424. doi:10.1007/BF02932309

Akiyama, Y., Kharel, P., Abe, T., Anderson, P., and Ivanov, P. (2020). Isolation and Initial Structure-Functional Characterization of Endogenous tRNA-Derived Stress-Induced RNAs. *RNA Biol.* 17, 1116–1124. doi:10.1080/15476286.2020.1732702

Baronti, L., Karlsson, H., Marušič, M., and Petzold, K. (2018). A Guide to Large-Scale RNA Sample Preparation. *Anal. Bioanal. Chem.* 410, 3239–3252. doi:10.1007/s00216-018-0943-8

Behrens, A., Rodschinka, G., and Nedialkova, D. D. (2021). High-resolution Quantitative Profiling of tRNA Abundance and Modification Status in Eukaryotes by mim-tRNAseq. *Mol. Cell* 81, 1802–1815. doi:10.1016/j.molcel.2021.01.028

Blanga-Kanfi, S. (2006). PrrC-anticodon Nuclease: Functional Organization of a Prototypical Bacterial Restriction RNase. *Nucleic Acids Res.* 34, 3209–3219. doi:10.1093/nar/gkl415

Bryson, D. I., Fan, C., Guo, L.-T., Miller, C., Söll, D., and Liu, D. R. (2017). Continuous Directed Evolution of Aminoacyl-tRNA Synthetases. *Nat. Chem. Biol.* 13, 1253–1260. doi:10.1038/nchembio.2474

Bundy, B. C., Franciszkowicz, M. J., and Swartz, J. R. (2008). Escherichia Coli-Based Cell-free Synthesis of Virus-like Particles. *Biotechnol. Bioeng.* 100, 28–37. doi:10.1002/bit.21716

Calles, J., Justice, I., Brinkley, D., Garcia, A., and Endy, D. (2019). Fail-safe Genetic Codes Designed to Intrinsically Contain Engineered Organisms. *Nucleic Acids Res.* 47, 10439–10451. doi:10.1093/nar/gkz745

Cao, K.-Y., Pan, Y., Yan, T.-M., and Jiang, Z.-H. (2020). Purification, Characterization and Cytotoxic Activities of Individual tRNAs from *Escherichia coli*. *Int. J. Biol. Macromol.* 142, 355–365. doi:10.1016/j.ijbiomac.2019.09.106

Cayama, E., Yepez, A., Rotondo, F., Bandeira, E., Ferreras, A. C., and Triana-Alonso, F. J. (2000). New Chromatographic and Biochemical Strategies for Quick Preparative Isolation of tRNA. *Nucleic Acids Res.* 28, 64e–64. doi:10.1093/nar/28.12.e64

Cervettini, D., Tang, S., Fried, S. D., Willis, J. C. W., Funke, L. F. H., Colwell, L. J., et al. (2020). Rapid Discovery and Evolution of Orthogonal Aminoacyl-tRNA Synthetase-tRNA Pairs. *Nat. Biotechnol.* 38, 989–999. doi:10.1038/s41587-020-0479-2

Chemla, Y., Ozer, E., Shaferman, M., Zaad, B., Dandela, R., and Alfonta, L. (2019). Simplified Methodology for a Modular and Genetically Expanded Protein Synthesis in Cell-free Systems. *Synthetic Syst. Biotechnol.* 4, 189–196. doi:10.1016/j.synbio.2019.10.002

Cui, Z., Stein, V., Tnimov, Z., Mureev, S., and Alexandrov, K. (2015). Semisynthetic tRNA Complement Mediates *In Vitro* Protein Synthesis. *J. Am. Chem. Soc.* 137, 4404–4413. doi:10.1021/ja5131963

Cui, Z., Mureev, S., Polinkovsky, M. E., Tnimov, Z., Guo, Z., Durek, T., et al. (2017). Combining Sense and Nonsense Codon Reassignment for Site-Selective Protein Modification with Unnatural Amino Acids. *ACS Synth. Biol.* 6, 535–544. doi:10.1021/acssynbio.6b00245

Cui, Z., Wu, Y., Mureev, S., and Alexandrov, K. (2018). Oligonucleotide-mediated tRNA Sequestration Enables One-Pot Sense Codon Reassignment *In Vitro*. *Nucleic Acids Res.* 46, 6387–6400. doi:10.1093/nar/gky365

Damiati, S., Kompella, U., Damiati, S., and Kodzius, R. (2018). Microfluidic Devices for Drug Delivery Systems and Drug Screening. *Genes* 9, 103. doi:10.3390/genes9020103

Das, G., and Duncan Lyngdoh, R. H. (2014). Configuration of Wobble Base Pairs Having Pyrimidines as Anticodon Wobble Bases: Significance for Codon Degeneracy. *J. Biomol. Struct. Dyn.* 32, 1500–1520. doi:10.1080/07391102.2013.824822

Dong, H., Nilsson, L., and Kurland, C. G. (1996). Co-variation of tRNA Abundance and Codon Usage inEscherichia Coliat Different Growth Rates. *J. Mol. Biol.* 260, 649–663. doi:10.1006/jmbi.1996.0428

Erber, L., Hoffmann, A., Fallmann, J., Betat, H., Stadler, P. F., and Mörl, M. (2020). LOTTE-seq (Long Hairpin Oligonucleotide Based tRNA High-Throughput Sequencing): Specific Selection of tRNAs with 3'-CCA End for High-Throughput Sequencing. *RNA Biol.* 17, 23–32. doi:10.1080/15476286.2019.1664250

Fechter, P., Rudinger, J., Giege, R., and Theobald-Dietrich, A. (1998). Ribozyme Processed tRNA Transcripts With Unfriendly Internal Promoter for T7 RNA Polymerase: Production and Activity. *FEBS Lett.* 436, 99–103. doi:10.1016/s0014-5793(98)01096-5

Fredens, J., Wang, K., De La Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total Synthesis of *Escherichia coli* with a Recoded Genome. *Nature* 569, 514–518. doi:10.1038/s41586-019-1192-5

Fukunaga, J.-i., Gouda, M., Umeda, K., Ohno, S., Yokogawa, T., and Nishikawa, K. (2006). Use of RNase P for Efficient Preparation of Yeast tRNATyr Transcript and its Mutants. *J. Biochem.* 139, 123–127. doi:10.1093/jb/mvj005

Gossringer, M., Helmecke, D., and Hartmann, R. K. (2012). Characterization of RNase P RNA Activity. *Methods Mol. Biol.* 848, 61–72. doi:10.1007/978-1-61779-545-9_5

Helm, M., Brulé, H., Giegé, R., and Florentz, C. (1999). More Mistakes by T7 RNA Polymerase at the 5′ Ends of In Vitro-transcribed RNAs. *RNA* 5, 618–621. doi:10.1017/s1355838299982328

Hibi, K., Amikura, K., Sugiura, N., Masuda, K., Ohno, S., Yokogawa, T., et al. (2020). Reconstituted Cell-free Protein Synthesis Using *In Vitro* Transcribed tRNAs. *Commun. Biol.* 3, 350. doi:10.1038/s42003-020-1074-2

Hipolito, C. J., and Suga, H. (2012). Ribosomal Production and *In Vitro* Selection of Natural Product-like Peptidomimetics: the FIT and RaPID Systems. *Curr. Opin. Chem. Biol.* 16, 196–203. doi:10.1016/j.cbpa.2012.02.014

Hodgman, C. E., and Jewett, M. C. (2012). Cell-free Synthetic Biology: Thinking outside the Cell. *Metab. Eng.* 14, 261–269. doi:10.1016/j.ymben.2011.09.002

Huang, X., Zhao, Y., Pu, Q., Liu, G., Peng, Y., Wang, F., et al. (2019). Intracellular Selection of Trans-cleaving Hammerhead Ribozymes. *Nucleic Acids Res.* 47, 2514–2522. doi:10.1093/nar/gkz018

Italia, J. S., Addy, P. S., Erickson, S. B., Peeler, J. C., Weerapana, E., and Chatterjee, A. (2019). Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* 141, 6204–6212. doi:10.1021/jacs.8b12954

Iwane, Y., Hitomi, A., Murakami, H., Katoh, T., Goto, Y., and Suga, H. (2016). Expanding the Amino Acid Repertoire of Ribosomal Polypeptide Synthesis via the Artificial Division of Codon Boxes. *Nat. Chem.* 8, 317–325. doi:10.1038/nchem.2446

Jackson, R. J., Napthine, S., and Brierley, I. (2001). Development of a tRNA-dependent *In Vitro* Translation System. *RNA* 7, 765–773. doi:10.1017/s1355838201002539

Jaroentomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., et al. (2018). Single-pot Glycoprotein Biosynthesis Using a Cell-free Transcription-Translation System Enriched with Glycosylation Machinery. *Nat. Commun.* 9, 2686. doi:10.1038/s41467-018-05110-x

Kaneko, T., Suzuki, T., Kapushoc, S. T., Rubio, M. A., Ghazvini, J., Watanabe, K., et al. (2003). Wobble Modification Differences and Subcellular Localization of tRNAs in Leishmania Tarentolae: Implication for tRNA Sorting Mechanism. *EMBO J.* 22, 657–667. doi:10.1093/emboj/cdg066

Katoh, T., Iwane, Y., and Suga, H. (2018). tRNA Engineering for Manipulating Genetic Code. *RNA Biol.* 15, 453–460. doi:10.1080/15476286.2017.1343227

Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., Delisa, M. P., Mrksich, M., et al. (2018). Design of Glycosylation Sites by Rapid Synthesis and Analysis of Glycosyltransferases. *Nat. Chem. Biol.* 14, 627–635. doi:10.1038/s41589-018-0051-2

Kim, T.-W., Keum, J.-W., Oh, I.-S., Choi, C.-Y., Park, C.-G., and Kim, D.-M. (2006). Simple Procedures for the Construction of a Robust and Cost-Effective Cell-free Protein Synthesis System. *J. Biotechnol.* 126, 554–561. doi:10.1016/j.jbiotec.2006.05.014

Korencic, D., Soll, D., and Ambrogelly, A. (2002). A One-step Method for *In Vitro* Production of tRNA Transcripts. *Nucleic Acids Res.* 30, 105. doi:10.1093/nar/gnf104

Kothe, U., Paleskava, A., Konevega, A. L., and Rodnina, M. V. (2006). Single-step Purification of Specific tRNAs by Hydrophobic Tagging. *Anal. Biochem.* 356, 148–150. doi:10.1016/j.ab.2006.04.038

Kralik, P., and Ricchi, M. (2017). A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Front. Microbiol.* 8, 108. doi:10.3389/fmicb.2017.00108

Kunda, T., Takai, K., Yokoyama, S., and Takaku, H. (2000). An Easy Cell-free Protein Synthesis System Dependent on the Addition of Crude *Escherichia coli* tRNA. *J. Biochem.* 127, 37–41. doi:10.1093/oxfordjournals.jbchem.a022581

Kwon, I., Kirshenbaum, K., and Tirrell, D. A. (2003). Breaking the Degeneracy of the Genetic Code. *J. Am. Chem. Soc.* 125, 7512–7513. doi:10.1021/ja0350076

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically Recoded Organisms Expand Biological Functions. *Science* 342, 357–360. doi:10.1126/science.1241459

Lavickova, B., and Maerkl, S. J. (2019). A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-free System. *ACS Synth. Biol.* 8, 455–462. doi:10.1021/acssynbio.8b00427

Lee, B. S., Shin, S., Jeon, J. Y., Jang, K.-S., Lee, B. Y., Choi, S., et al. (2015). Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS Chem. Biol.* 10, 1648–1653. doi:10.1021/acschembio.5b00230

Lee, K. B., Hou, C. Y., Kim, C.-E., Kim, D.-M., Suga, H., and Kang, T. J. (2016). Genetic Code Expansion by Degeneracy Reprogramming of Arginyl Codons. *ChemBioChem* 17, 1198–1201. doi:10.1002/cbic.201600111

Lee, J., Schwieter, K. E., Watkins, A. M., Kim, D. S., Yu, H., Schwarz, K. J., et al. (2019). Expanding the Limits of the Second Genetic Code with Ribozymes. *Nat. Commun.* 10, 5097. doi:10.1038/s41467-019-12916-w

Li, Y., Chen, J., Wang, E., and Wang, Y. (1999). T7 RNA Polymerase Transcription of Escherichia Coli Isoacceptors tRNA (Leu). *Sci. China Ser. C. Life Sci.* 42, 185–190. doi:10.1007/BF02880055

Li, J., Zhang, L., and Liu, W. (2018). Cell-free Synthetic Biology for *In Vitro* Biosynthesis of Pharmaceutical Natural Products. *Synthetic Syst. Biotechnol.* 3, 83–89. doi:10.1016/j.synbio.2018.02.002

Li, D. J. (2021). Formation of the Codon Degeneracy during Interdependent Development between Metabolism and Replication. *Genes* 12, 2023–2038. doi:10.3390/genes12122023

Link, A. J., and Tirrell, D. A. (2005). Reassignment of Sense Codons *In Vivo*. *Methods* 36, 291–298. doi:10.1016/j.ymeth.2005.04.005

Liu, W.-Q., Zhang, L., Chen, M., and Li, J. (2019). Cell-free Protein Synthesis: Recent Advances in Bacterial Extract Sources and Expanded Applications. *Biochem. Eng. J.* 141, 182–189. doi:10.1016/j.bej.2018.10.023

Lu, Y., Welsh, J. P., and Swartz, J. R. (2014). Production and Stabilization of the Trimeric Influenza Hemagglutinin Stem Domain for Potentially Broadly Protective Influenza Vaccines. *Proc. Natl. Acad. Sci. U.S.A.* 111, 125–130. doi:10.1073/pnas.1308701110

Lu, Y. (2017). Cell-free Synthetic Biology: Engineering in an Open World. *Synthetic Syst. Biotechnol.* 2, 23–27. doi:10.1016/j.synbio.2017.02.003

Milligan, J. F., and Uhlenbeck, O. C. (1989). Synthesis of Small RNAs Using T7 RNA Polymerase. *Methods Enzym.* 180, 51. doi:10.1016/0076-6879(89)80091-6

Min, S. E., Lee, K.-H., Park, S.-W., Yoo, T. H., Oh, C. H., Park, J.-H., et al. (2016). Cell-free Production and Streamlined Assay of Cytosol-Penetrating Antibodies. *Biotechnol. Bioeng.* 113, 2107–2112. doi:10.1002/bit.25985

Mishima, E., and Abe, T. (2019). Immuno-Northern Blotting: Detection of Modified RNA Using Gel Separation and Antibodies to Modified Nucleosides. *Methods Mol. Biol.* 1870, 179–187. doi:10.1007/978-1-4939-8808-2_13

Mishima, E., Jinno, D., Akiyama, Y., Itoh, K., Nankumo, S., Shima, H., et al. (2015). Immuno-Northern Blotting: Detection of RNA Modifications by Using

Antibodies against Modified Nucleosides. *PLoS One* 10, e0143756. doi:10.1371/journal.pone.0143756

Moore, S. J. (2019). Enzyme Alchemy: Cell-free Synthetic Biochemistry for Natural Products. *Emerg. Top. Life Sci.* 3, 529–535. doi:10.1042/ETLS20190083

Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., et al. (2015). Reassignment of a Rare Sense Codon to a Non-canonical Amino Acid inEscherichia Coli. *Nucleic Acids Res.* 43, 8111–8122. doi:10.1093/nar/gkv787

Mureev, S., Kovtun, O., Nguyen, U. T. T., and Alexandrov, K. (2009). Species-independent Translational Leaders Facilitate Cell-free Expression. *Nat. Biotechnol.* 27, 747–752. doi:10.1038/nbt.1556

Nilsen, T. W. (2013). Gel Purification of RNA. *Cold Spring Harb. Protoc.* 2, 180–3. doi:10.1101/pdb.prot072942

Noireaux, V., Maeda, Y. T., and Libchaber, A. (2011). Development of an Artificial Cell, from Self-Organization to Computation and Self-Reproduction. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3473–3480. doi:10.1073/pnas.1017075108

Ogilvie, K. K., Usman, N., Nicoghosian, K., and Cedergren, R. J. (1988). Total Chemical Synthesis of a 77-Nucleotide-Long RNA Sequence Having Methionine-Acceptance Activity. *Proc. Natl. Acad. Sci. U.S.A.* 85, 5764–5768. doi:10.1073/pnas.85.16.5764

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., et al. (2016). Design, Synthesis, and Testing toward a 57-codon Genome. *Science* 353, 819–822. doi:10.1126/science.aaf3639

Pedelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* 24, 79–88. doi:10.1038/nbt1172

Pereira, M., Francisco, S., Varanda, A., Santos, M., Santos, M., and Soares, A. (2018). Impact of tRNA Modifications and tRNA-Modifying Enzymes on Proteostasis and Human Disease. *Int. J. Mol. Sci.* 19, 3738. doi:10.3390/ijms19123738

Rinehart, J., Krett, B., Rubio, M. A. T., Alfonzo, J. D., and Söll, D. (2005). *Saccharomyces cerevisiae* Imports the Cytosolic Pathway for Gln-tRNA Synthesis into the Mitochondrion. *Genes Dev.* 19, 583–592. doi:10.1101/gad.1269305

Salehi, A. S. M., Smith, M. T., Schinn, S. M., Hunt, J. M., Muhlestein, C., Diray-Arce, J., et al. (2017). Efficient tRNA Degradation and Quantification in *Escherichia coli* Cell Extract Using RNase-coated Magnetic Beads: A Key Step toward Codon Emancipation. *Biotechnol. Prog.* 33, 1401–1407. doi:10.1002/btpr.2511

Schmidt, M. (2019). A Metric Space for Semantic Containment: Towards the Implementation of Genetic Firewalls. *Biosystems* 185, 104015. doi:10.1016/j.biosystems.2019.104015

Schoborg, J. A., Hershewe, J. M., Stark, J. C., Kightlinger, W., Kath, J. E., Jaroentomeechai, T., et al. (2018). A Cell-free Platform for Rapid Synthesis and Testing of Active Oligosaccharyltransferases. *Biotechnol. Bioeng.* 115, 739–750. doi:10.1002/bit.26502

Schwartz, M. H., and Pan, T. (2017). Determining the Fidelity of tRNA Aminoacylation via Microarrays. *Methods* 113, 27–33. doi:10.1016/j.ymeth.2016.09.004

Shepherd, T. R., Du, L., Liljeruhm, J., SamudyataWang, J., Wang, J., Sjödin, M. O. D., et al. (2017). De Novo design and Synthesis of a 30-cistron Translation-Factor Module. *Nucleic Acids Res.* 45, 10895–10905. doi:10.1093/nar/gkx753

Shigematsu, M., Honda, S., Loher, P., Telonis, A. G., Rigoutsos, I., and Kirino, Y. (2017). YAMAT-seq: An Efficient Method for High-Throughput Sequencing of Mature Transfer RNAs. *Nucleic Acids Res.* 45, gkx005. doi:10.1093/nar/gkx005

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802

Shimizu, Y., Kanamori, T., and Ueda, T. (2005). Protein Synthesis by Pure Translation Systems. *Methods* 36, 299–304. doi:10.1016/j.ymeth.2005.04.006

Stano, P., and Luisi, P. L. (2013). Semi-synthetic Minimal Cells: Origin and Recent Developments. *Curr. Opin. Biotechnol.* 24, 633–638. doi:10.1016/j.copbio.2013.01.002

Stein, V., and Alexandrov, K. (2014). Protease-based Synthetic Sensing and Signal Amplification. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15934–15939. doi:10.1073/pnas.1405220111

Su, Z., Kuscu, C., Malik, A., Shibata, E., and Dutta, A. (2019). Angiogenin Generates Specific Stress-Induced tRNA Halves and is Not Involved in tRF-3-Mediated Gene Silencing. *J. Biol. Chem.* 294, 16930–16941. doi:10.1074/jbc.RA119.009272

Su, Z., Wilson, B., Kumar, P., and Dutta, A. (2020). Noncanonical Roles of tRNAs: tRNA Fragments and beyond. *Annu. Rev. Genet.* 54, 47–69. doi:10.1146/annurev-genet-022620-101840

Suhasini, A. N., and Sirdeshmukh, R. (2006). Transfer RNA Cleavages by Onconase Reveal Unusual Cleavage Sites. *J. Biol. Chem.* 281, 12201–12209. doi:10.1074/jbc.M504488200

Suzuki, T., and Suzuki, T. (2007). Chaplet Column Chromatography: Isolation of a Large Set of Individual RNAs in a Single Step. *Methods Enzymol.* 425, 231–239. doi:10.1016/s0076-6879(07)25010-4

Tomita, K., Ogawa, T., Uozumi, T., Watanabe, K., and Masaki, H. (2000). A Cytotoxic Ribonuclease Which Specifically Cleaves Four Isoaccepting Arginine tRNAs at Their Anticodon Loops. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8278–8283. doi:10.1073/pnas.140213797

Tuckey, C., Asahara, H., Zhou, Y., and Chong, S. (2014). Protein Synthesis Using a Reconstituted Cell-Free System. *Curr. Protoc. Mol. Biol.* 108, 16 31 1–22. doi:10.1002/0471142727.mb1631s108

Van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencía, M., and Danelon, C. (2018). Self-replication of DNA by its Encoded Proteins in Liposome-Based Synthetic Cells. *Nat. Commun.* 9, 1583. doi:10.1038/s41467-018-03926-1

Vare, V., Eruysal, E., Narendran, A., Sarachan, K., and Agris, P. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. *Biomolecules* 7, 29. doi:10.3390/biom7010029

Weiss, M., Frohnmayer, J. P., Benk, L. T., Haller, B., Janiesch, J.-W., Heitkamp, T., et al. (2018). Sequential Bottom-Up Assembly of Mechanically Stabilized Synthetic Cells by Microfluidics. *Nat. Mater* 17, 89–96. doi:10.1038/nmat5005

Willis, J. C. W., and Chin, J. W. (2018). Mutually Orthogonal Pyrrolysyl-tRNA Synthetase/tRNA Pairs. *Nat. Chem.* 10, 831–837. doi:10.1038/s41557-018-0052-5

Yang, C., Liu, Y., Liu, W.-Q., Wu, C., and Li, J. (2021). Designing Modular Cell-free Systems for Tunable Biotransformation of L-Phenylalanine to Aromatic Compounds. *Front. Bioeng. Biotechnol.* 9, 730663. doi:10.3389/fbioe.2021.730663

Yildirim, I., Kierzek, E., Kierzek, R., and Schatz, G. C. (2014). Interplay of LNA and 2′-O-Methyl RNA in the Structure and Thermodynamics of RNA Hybrid Systems: A Molecular Dynamics Study Using the Revised AMBER Force Field and Comparison with Experimental Results. *J. Phys. Chem. B* 118, 14177–14187. doi:10.1021/jp506703g

Yue, K., Zhu, Y., and Kai, L. (2019). Cell-Free Protein Synthesis: Chassis toward the Minimal Cell. *Cells* 8, 315. doi:10.3390/cells8040315

Zheng, Y., Gilgenast, M. J., Hauc, S., and Chatterjee, A. (2018). Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis. *ACS Chem. Biol.* 13, 1137–1141. doi:10.1021/acschembio.8b00021

Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., et al. (2014). Production of Site-specific Antibody-Drug Conjugates Using Optimized Non-natural Amino Acids in a Cell-free Expression System. *Bioconjugate Chem.* 25, 351–361. doi:10.1021/bc400490z

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 | Frontiers in Bioengineering and Biotechnology

BRIEF RESEARCH REPORT
published: 23 June 2022
doi: 10.3389/fbioe.2022.895069

Characterizing and Improving pET Vectors for Cell-free Expression

Kara Jew[1], Philip E. J. Smith[2], Byungcheol So[2], Jillian Kasman[2], Javin P. Oza[2] and Michael W. Black[1]**

[1]Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA, United States, [2]Chemistry & Biochemistry Department, California Polytechnic State University, San Luis Obispo, CA, United States

Cell-free protein synthesis (CFPS) is an *in vitro* process that enables diverse applications in research, biomanufacturing, point-of-care diagnostics, therapeutics, and education using minimal laboratory equipment and reagents. One of the major limitations of CFPS implementation is its sensitivity to plasmid type. Specifically, plasmid templates based on commonly used vector backbones such as the pET series of bacterial expression vectors result in the inferior production of proteins. To overcome this limitation, we have evaluated the effect of expression cassette elements present in the pET30 vector on protein production across three different CFPS systems: NEBExpress, PURExpress, and CFAI-based *E. coli* extracts. Through the systematic elimination of genetic elements within the pET30 vector, we have identified elements that are responsible for the poor performance of pET30 vectors in the various CFPS systems. As a result, we demonstrate that through the removal of the *lac* operator (*lacO)* and N-terminal tags included in the vector backbone sequence, a pET vector can support high titers of protein expression when using extract-based CFPS systems. This work provides two key advances for the research community: 1) identification of vector sequence elements that affect robust production of proteins; 2) evaluation of expression across three unique CFPS systems including CFAI extracts, NEBexpress, and PURExpress. We anticipate that this work will improve access to CFPS by enabling researchers to choose the correct expression backbone within the context of their preferred expression system.

Keywords: cell-free, protein synthesis, pET30, template, *in vitro*, translation

OPEN ACCESS

Edited by:
Jian Li,
ShanghaiTech University, China

Reviewed by:
Irene Castano,
Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), Mexico
Wan-Qiu Liu,
ShanghaiTech University, China

****Correspondence:***
Javin P. Oza
joza@calpoly.edu
Michael W. Black
mblack@calpoly.edu

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *12 March 2022*
Accepted: *26 May 2022*
Published: *23 June 2022*

Citation:
Jew K, Smith PEJ, So B, Kasman J, Oza JP and Black MW (2022) Characterizing and Improving pET Vectors for Cell-free Expression. Front. Bioeng. Biotechnol. 10:895069. doi: 10.3389/fbioe.2022.895069

INTRODUCTION

Cell-free protein synthesis (CFPS) provides an on-demand protein expression platform that is compatible with circular plasmids as well as linear DNA and RNA templates (Jewett and Swartz, 2004; Gregorio et al., 2019; Asahara et al., 2021; McSweeney and Styczynski, 2021; Batista et al., 2022). The use of CFPS bypasses the need to maintain living cells, therefore, all cellular energy and machinery can be directed toward protein synthesis. The open nature of the cell-free platform allows users greater control of the reaction conditions than *in vivo* expression platforms. CFPS also enables the expression of cytotoxic and complex proteins that may otherwise be difficult to express in living cells (Jewett and Swartz, 2004; Pardee et al., 2016; Dopp et al., 2019; Garenne et al., 2019; Jin et al., 2019; Kay and Jewett, 2020). Recently, improvements in the upstream and downstream processing of cell lysates from the widely adopted *E. coli* platform have led to more consistent results and an increased shelf life of the reaction mixtures (Smith et al., 2014; Kwon and Jewett, 2015; Cole et al., 2020; Gregorio et al.,

2020; Levine et al., 2020). Due to these benefits, CFPS systems are enabling a variety of academic research efforts, biotechnology innovations, and large scale biomanufacturing (Pardee et al., 2016; Huang et al., 2018; Khambhati et al., 2019; Kightlinger et al., 2019; Choi et al., 2020; Liu et al., 2020; Silverman et al., 2020; Williams et al., 2020; Burrington et al., 2021a; Brookwell et al., 2021; Si et al., 2021).

Barriers to access have reduced significantly as CFPS systems have become commercially available in the form of kits derived from lysates of a variety of chassis organisms, as well as reconstituted systems (Shimizu et al., 2005). For this study, the New England Biolab's NEBExpress and NEB PURExpress kits were used alongside our in-house *E. coli* lysate-based CFAI system to assess the effects that distinct vector elements have on protein synthesis (Levine et al., 2020; Smith et al., 2021; Mullin et al., 2022). Both the NEBExpress and CFAI systems utilize crude *E. coli* extracts. In contrast, the NEB PURExpress system is reconstituted with purified components of the *E. coli* translation machinery. The purified systems are an important part of the CFPS biotechnology portfolio since they provide protein expression conditions in which protease and nuclease activity is minimized (Shimizu et al., 2001) to preserve nucleic acid templates and protein products.

The CFPS community has systematically reduced many of the bottlenecks that limited the broad utility of CFPS ushering the biotechnology's renaissance over the last 20 years. However, compatibility of DNA templates in the CFPS system continues to remain a limit for the robust production of target proteins (Romantseva and Strychalski, 2020). In the *E. coli*–based CFPS platform, commonly used pET series expression vectors have been observed to result in significantly lower protein titers and yields when compared to the alternate vectors such as pJL1 (Zhang et al., 2018; Colant et al., 2021). The pJL1 vector (Addgene #69496 and #102634) is derived from the pY71 vector, which was a simplified version of the pK7 plasmid. This plasmid lineage has been successfully utilized for CFPS and has set the benchmarks for CFPS applications for over a decade (Swartz et al., 2004; Bundy and Swartz, 2010). The importance of expression vectors has also been demonstrated in Streptomyces-based cell-free systems (Xu et al., 2022). *In vivo* studies in *E. coli* have identified features within the pET series of expression vectors that hinder protein expression yields (Shilling et al., 2020). The *in vivo* study determined that an incomplete T7 promoter found in pET28a decreased sfGFP production. This truncated T7 promoter was also identified in 88 of the 103 pET expression vectors. Such efforts are needed for *in vitro* expression given the precedence for variation in sequence elements found in expression vectors being consequential for expression yields. We first established that the vector used in this study, pET30, contains the complete T7 promoter. The goal of this study was to examine additional features of the pET expression vector series that may have an impact on protein yields in CFPS systems. We assessed the effects of the pET30 *lacO* and the N-terminal tags (6x poly-histidine tag and S tag) on sfGFP expression. Four versions of the pET30 vector were constructed with and without *lacO* and N-terminal His-tag. The expression of sfGFP across three CFPS expression systems was then determined through fluorescence evaluate the impact of these sequence elements.

MATERIALS AND METHODS

Strains and Growth Conditions

E. coli strains BL21(DE3) and MC1061 were used in this study. Cultures were aerobically grown at 37°C in Luria Bertaini (LB) broth or plates. Kanamycin (30 μg/ml) was added to the media for cultures containing pET30-derived vectors and pJL1-sfGFP. The BL21(DE3) strain was used to prepare CFAI-based CFPS extracts as previously described (Levine et al., 2020; Smith et al., 2021; Mullin et al., 2022). The CFAI media auto-induces T7 RNAP expression during cell growth, and cells are harvested at high ODs. The MC1061 strain was used as the host for cloning variations of the pET30 expression plasmids. All transformations were performed via electroporation with 40 μl of electrocompetent cells and approximately 30 ng of DNA using the BTX Electro Cell Manipulator 600 (Harvard Apparatus Inc.; 2.45 kV, 129 Ω). Immediately after electroporation, cells were incubated with 500 μl SOC recovery medium for 1 h at 37°C, plated on LB-kanamycin plates, and incubated at 37°C for 18–24 h.

Molecular Techniques

The polymerase chain reactions (PCR) were performed in 20 μl volumes with Phusion Flash High-Fidelity PCR 2X Master Mix (Thermo Scientific, Rockford, IL, United States) containing 0.2 ng of template DNA and a final primer concentration of 0.1 μM. The vector and inserts used to construct the pET30 variations were amplified with forward and reverse primers noted in **Table 1**. The thermocycling parameters included a 1-min denaturation at 98°C followed by 30 cycles of 10 s at 98°C, 30 s at various annealing temperatures, and 15 s per kb of expected product at 72°C. The reaction ended with a final 5-min extension step at 72°C and hold at 4°C.

Gibson assembly was performed using 17 fmoles each of the amplified insert and vector fragments in a 6 μl reaction containing Taq ligase (4 U/μl), T5 exonuclease (0.02 U/μl), and Phusion DNA polymerase (0.025 U/μl) purchased from New England Biolabs (Ipswich, MA, United States). Each reaction was incubated 15 min at 50°C in 1X Gibson buffer (125 mM Tris-HCl pH 7.5, 6.25% PEG-8000 (w/v), 12.5 mM $MgCl_2$, 12.5 mM DTT, 2.5 mM dNTPs, and 1.25 mM NAD).

Cell-free protein synthesis reactions using in-house CFAI-based cell extracts were performed as described (Levine et al., 2020). Reactions using the PURExpress® and NEBExpress® kits (New England Biotech, Ipswich, MA, United States) were performed according to manufacturer's instructions. All CFPS reactions were run in triplicate using sfGFP as the reporter protein.

Quantification of Reporter Protein sfGFP

Fluorescence intensities of sfGFP from each CFPS reaction were measured in triplicate. Each measurement consisted of a solution of 48 μl of 0.05 M HEPES at pH 7.2 and 2 μl of the sfGFP CFPS reaction solution in a black 96 well plate. Each 50 μl solution's

TABLE 1 | Primers used to construct pET30 expression vectors. Primers for amplification of insert and vector backbones used in Gibson assembly to construct pET30-T7-sfGFP, pET30-lacO-sfGFP, pET30-His-sfGFP, and pET30-lacO-His-sfGFP.

Primer sequences		T_m	Annealing
pET30-T7-sfGFP			
Insert: T7-Pro-Gib-F	ccgcgaaattaatacgactcactatagg	59°C	63°C
Insert: T7-Term-Gib-R	ctttcagcaaaaaacccctcaag	56°C	
Vector: T7-Term-Gib-F	cttgaggggtttttgctgaaag	56°C	63°C
Vector: T7-Pro-Gib-R	cctatagtgagtcgtattaatttcgcgg	59°C	
pET30-T7-lacO-sfGFP			
Insert: RBS-sfGFP-F	ctttaagaaggagatatacatatgagcaaaggtgaagaactg	62°C	55°C
Insert: T7-Term-Gib-R	ctttcagcaaaaaacccctcaag	56°C	
Vector: T7-Term-Gib-F	cttgaggggtttttgctgaaag	56°C	63°C
Vector: pET-RBS-long-R	catatgtatatctccttcttaaagttaaacaaaattatttctagagg	58°C	
pET30-T7-His/S-sfGFP			
Insert: pET-RBS-F	gtttaactttaagaaggagatatacatatg	52°C	58°C
Insert: T7-Term-Gib-R	ctttcagcaaaaaacccctcaag	56°C	
Vector: T7-Term-Gib-F	cttgaggggtttttgctgaaag	56°C	58°C
Vector: pET-RBS-long-R	catatgtatatctccttcttaaagttaaacaaaattatttctagagg	58°C	
pET30-T7-lacO-His/S-sfGFP			
Insert: N-tag-sfGFP-F	cacatggacagcccagatctcatgagcaaaggtgaagaactg	69°C	63°C
Insert: T7-Term-Gib-R	ctttcagcaaaaaacccctcaag	56°C	
Vector: T7-Term-Gib-F	cttgaggggtttttgctgaaag	56°C	61°C
Vector: pET-No-Cut-DIC-R	agatctgggctgtccatgtg	58°C	

fluorescence was then measured at an excitation wavelength of 485 nm and an emission wavelength of 528 nm. The quantity of sfGFP was then calculated using a previously developed standard curve (Levine et al., 2019).

RESULTS

To systematically determine the effect of each constituent of the pET vector on CFPS, four vectors were constructed: pET30-T7-lacO-His/S-sfGFP (Addgene #180754), pET30-T7-lacO-sfGFP (Addgene #180756), pET30-T7-His/S-sfGFP (Addgene #180755), and pET30-T7-sfGFP (Addgene #180757). Graphic representations of the expression cassettes in each of the four plasmids are shown in **Figure 1**. As seen in **Figure 1B**, the gene encoding *sfGFP* is located downstream from both *lacO* and the N-terminal tags, and transcription is controlled by the T7 promoter. The effect on cell-free expression was measured as a function of removing the pET30 encoded N-terminal His & S tags (pET30-lacO-sfGFP), the *lacO* (pET30-His/S-sfGFP), or both N-terminal tags and the *lacO* sequences (pET30-T7-sfGFP). For this assessment, sfGFP expression from these plasmids were compared to the preferred pJL1-sfGFP reporter plasmid, which is also driven by the T7 promoter but does not include *lacO* or N-terminal His & S tags in the expression cassette. An alignment of the most relevant sequences that differ between the plasmids is provided in **Supplementary Figure S1**.

In the CFAI-based CFPS system, removing *lacO* (pET30-T7-His/S-sfGFP), the N-terminal tags (pET30-T7-lacO-sfGFP), or both *lacO* and the N-terminal tags (pET30-T7-sfGFP) resulted in increased sfGFP expression (**Figure 2A**). The removal of *lacO* alone (pET30-T7-His/S-sfGFP) had a greater effect on improving sfGFP expression compared to the removal of the N-terminal tags alone (pET30-T7-lacO-sfGFP). For the NEBExpress system, there was an increase in sfGFP expression when la*cO* (pET30-T7-His/S-sfGFP), the N-terminal tags (pET30-T7-lacO-sfGFP), or the combination of both *lacO* and the N-terminal tags (pET30-T7-sfGFP) were removed from the pET30 vector (**Figure 2B**). Similar to the CFAI-based CFPS system, the removal of *lacO* alone (pET30-T7-His/S-sfGFP) had a greater impact on improving sfGFP expression compared to the removal of N-terminal tags alone (pET30-T7-lacO-sfGFP).

As observed with extract-based CFPS systems, sfGFP expression could also be improved for the NEB PURExpress system upon removing elements upstream of the reporter gene. The PURExpress system was less sensitive to the *lacO* element, lacking the repressor is a likely advantage of the purified system. sfGFP expression increased notably upon removal of the N-terminal tags (pET30-T7-lacO-sfGFP) and when both *lacO* and the N-terminal tags were removed (pET30-T7-sfGFP) (**Figure 2C**).

Overall, removing both *lacO* and the N-terminal tags enhanced the expression of sfGFP; however, the individual effects of *lacO* and the N-terminal tags differed between the three CFPS systems. In general, the removal of *lacO* alone had a more substantial impact on sfGFP fluorescence and expression in the extract-based CFAI and NEB Express systems whereas, in the NEB PURExpress CFPS system, the removal of the N-terminal tags had the greatest effect. Consistent with prior observations, fluorescence data also revealed much higher maximum yields of sfGFP through the CFAI-based CFPS system (>1,000 μg/ml) than in the NEB Express (~800 μg/ml) and NEB PURExpress (~200 μg/ml) CFPS systems (Burrington et al., 2021b).

DISCUSSION

The choice of expression vectors plays a critical role in CFPS as vectors may contain elements that negatively impact protein

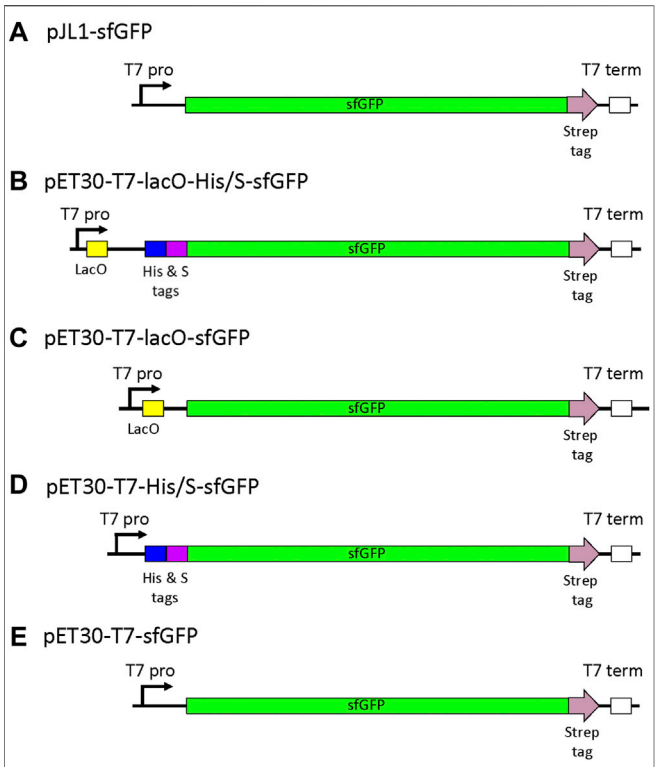


FIGURE 1 | Schematic of the expression vectors studied. **(A)** pJL1-sfGFP represents the reference vector, **(B)** the pET30-T7-lacO-His/S-sfGFP cassette contains the T7 promoter (T7 pro) followed by *lacO* and a N-terminal purification tags (His and S tags) included in the pET30 vector backbone. Modifications of the pET30 expression cassette included the deletion of **(C)** the N-terminal His/S-tags, **(D)** *lacO,* and **(E)** both *lacO* and the N-terminal tags to assess the effects on sfGFP expression. Similar to pET30-T7-sfGFP, the pJL1-T7-sfGFP cassette includes the same T7 promoter but not the *lacO* nor the N-terminal His-tag. All templates used in this study were in their circular plasmid forms. While the backbone sequences are not displayed here, pET30a includes the *lacI* gene that is not present in the pJL1 backbone.

expression. When implementing pET vectors in CFPS, the removal of *lacO* and the N-terminal tags resulted in increased sfGFP production across all three expression systems tested. The removal of *lacO* appears to play a more significant role in sfGFP expression in the lysate-based CFAI and the NEB Express CFPS systems. This is likely due to the presence of lactose operon repressor present in the *E. coli* cell extracts and absent in the reconstituted PURExpress system. As the pET30 plasmids include a *lacI* expression cassette, residual *E. coli* RNA polymerase in the extracts may also lead to expression of supplemental LacI repressor directly from the vector. When *lacO* was present in the pET30 expression vector (pET30-T7-lacO-sfGFP and pET30-T7-lacO-His/S-sfGFP), there was a larger decrease in sfGFP yield in the CFAI system (**Figure 2A**) compared to the decrease in the NEB Express CFPS system (**Figure 2B**). These differences may be due to the nuanced methods in extract preparation between the two systems. According to NEB, the NEBExpress CFPS system is supplied with an optimized quantity of wild-type T7 RNA polymerase, which may produce mRNA transcripts more efficiently in the presence of the *lac* repressor. The NEBExpress system is generated from an *E. coli* strain which has a single copy of the lacI gene, which presumably expresses low levels of the Lac repressor protein. Optimized quantities of DNA template are therefore expected to exceed the number of Lac repressor molecules. Increasing DNA template quantities may be one possible solution to the Lac-based repression. The benefits observed from the removal of the *lacO* sequence may also be due to additional, more nuanced factors such as mRNA structure and stability, and possible impacts on translation initiation. While these and other factors are worth further study, the results of this work provide practical and actionable set of insights for researchers to either sub-clone a gene of interest into a CFPS-compatible vector, or selectively eliminate regions that may adversely impact CFPS expression from their preferred vector.

Purification is a typical goal of recombinant protein expression, so purification tags are often integrated into frequently used expression vectors. The presence of N-terminal His and S tags suppressed protein expression in all

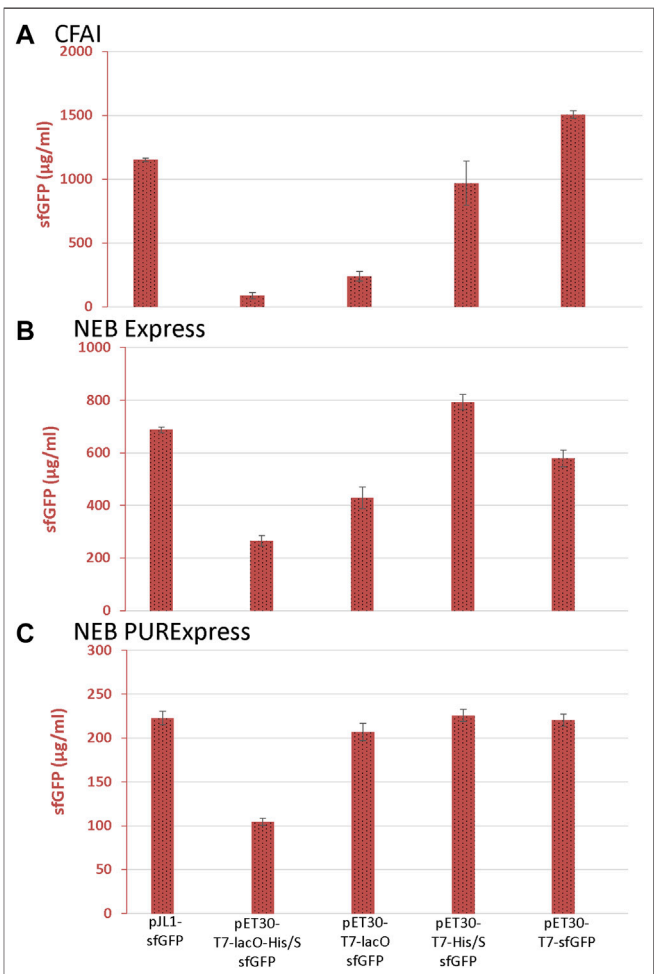


FIGURE 2 | sfGFP expression from the CFAI-based, NEB Express, and NEB PURExpress CFPS systems. The sfGFP yields from **(A)** CFAI-based, **(B)** NEBExpress, and **(C)** NEB PURExpress CFPS systems determined by fluorescence. Data are presented as mean ± s.d. (n = 3).

three CFPS systems to varying extents. Investigating the interplay between N-terminal tags and ribosome binding site (RBS) sequences may also be warranted (Salis et al., 2009; Zhang et al., 2021). To evaluate the importance of the RBS, we utilized the extremely useful tool De Novo DNA (www.denovodna.com) (Espah Borujeni et al., 2014; Espah Borujeni and Salis, 2016; Espah Borujeni et al., 2017). Notably, the RBS calculations do not correlate with the sfGFP expressions observed in our vectors, highlighting the importance of other molecular mechanisms must be at play for optimal expression in CFPS. Another consideration is that the presence of the His-tag may deplete the pool of L-histidine in the CFPS reactions, which could be further studied by either supplementing L-histidine to the reaction, or evaluating additional constructs in which the His-tag is moved to the C-terminus rather than removed. Given the importance of purification tags and the need to utilize them at the N-terminus, it will be important to further examine the effects of additional affinity tags. By investigating such effects, there may be an ideal purification tag that can be used to provide optimal protein expression in CFPS systems within the user's expression vector of choice. When possible, a C-terminal tag may be preferred, but due to the context dependencies of biomolecular systems, these data provide evidence that users should evaluate a variety of construct designs that vary in the type and location of purification tags to achieve optimal protein expression. This work demonstrates that vector elements have substantial effects on CFPS yields. Furthermore, the effects of a given element are dependent on the context of the CFPS system in which the vector will be utilized. Our results nullify the hypothesis that pET vectors result in inferior protein expression due to their significantly larger size and provides further support for the role of specific elements that interfere with expression. Based on our findings, we are optimistic that researchers utilizing CFPS for protein expression will achieve improved yields by pairing an optimized vector with their preferred expression system.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

KJ, MB, and JO wrote the manuscript. KJ, BS, and PS performed the experiments. KJ and JK performed statistical analysis. KJ and MB generated the figures. MB and JO conceived the project and led the effort. All authors helped revise the manuscript and agreed to the accuracy of the work reported.

ACKNOWLEDGMENTS

The authors acknowledge funding support from the Bill and Linda Frost Fund for supporting student research and publication costs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.895069/full#supplementary-material

REFERENCES

Asahara, H., Magnelli, P., Shi, X., Tuckey, C., Zhou, Y., and Samuelson, J. C. (2021). "Guidelines for Nucleic Acid Template Design for Optimal Cell-free Protein Synthesis Using an Escherichia coli Reconstituted System or a Lysate-Based System," in Methods In Enzymology *Recombinant Protein Expression: Prokaryotic Hosts and Cell-free Systems*. Editors Z. Kelman and W. B. O'Dell (Academic Press), 351–369. doi:10.1016/bs.mie.2021.07.005

Batista, A. C., Levrier, A., Soudier, P., Voyvodic, P. L., Achmedov, T., Reif-Trauttmansdorff, T., et al. (2022). Differentially Optimized Cell-free Buffer Enables Robust Expression from Unprotected Linear DNA in Exonuclease-Deficient Extracts. *ACS Synth. Biol.* 11, 732–746. doi:10.1021/acssynbio.1c00448

Brookwell, A., Oza, J. P., and Caschera, F. (2021). Biotechnology Applications of Cell-free Expression Systems. *Life* 11, 1367. doi:10.3390/life11121367

Bundy, B. C., and Swartz, J. R. (2010). Site-Specific Incorporation of P-Propargyloxyphenylalanine in a Cell-free Environment for Direct Protein–Protein Click Conjugation. *Bioconjugate Chem.* 21, 255–263. doi:10.1021/bc9002844

Burrington, L. R., Baryal, E., Hui, K., Lambert, E., Harding, S. T., and Oza, J. P. (2021a). The Fold-Illuminator: A Low-Cost, Portable, and Disposable Incubator-Illuminator Device. *Synthetic Syst. Biotechnol.* 6, 95–101. doi:10.1016/j.synbio.2021.04.003

Burrington, L. R., Watts, K. R., and Oza, J. P. (2021b). Characterizing and Improving Reaction Times for E. Coli-Based Cell-free Protein Synthesis. *ACS Synth. Biol.* 10, 1821–1829. doi:10.1021/acssynbio.1c00195

Choi, S. Y., Rhie, M. N., Kim, H. T., Joo, J. C., Cho, I. J., Son, J., et al. (2020). Metabolic Engineering for the Synthesis of Polyesters: A 100-year Journey from Polyhydroxyalkanoates to Non-natural Microbial Polyesters. *Metab. Eng.* 58, 47–81. doi:10.1016/j.ymben.2019.05.009

Colant, N., Melinek, B., Teneb, J., Goldrick, S., Rosenberg, W., Frank, S., et al. (2021). A Rational Approach to Improving Titer in Escherichia coli -based Cell-free Protein Synthesis Reactions. *Biotechnol. Prog.* 37, e3062. doi:10.1002/btpr.3062

Cole, S. D., Miklos, A. E., Chiao, A. C., Sun, Z. Z., and Lux, M. W. (2020). Methodologies for Preparation of Prokaryotic Extracts for Cell-free Expression Systems. *Synthetic Syst. Biotechnol.* 5, 252–267. doi:10.1016/j.synbio.2020.07.006

Dopp, B. J. L., Tamiev, D. D., and Reuel, N. F. (2019). Cell-free Supplement Mixtures: Elucidating the History and Biochemical Utility of Additives Used to Support *In Vitro* Protein Synthesis in E. coli Extract. *Biotechnol. Adv.* 37, 246–258. doi:10.1016/j.biotechadv.2018.12.006

Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014). Translation Rate Is Controlled by Coupled Trade-Offs between Site Accessibility, Selective RNA Unfolding and Sliding at Upstream Standby Sites. *Nucleic Acids Res.* 42, 2646–2659. doi:10.1093/nar/gkt1139

Espah Borujeni, A., and Salis, H. M. (2016). Translation Initiation Is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* 138, 7016–7023. doi:10.1021/jacs.6b01453

Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N., and Salis, H. M. (2017). Precise Quantification of Translation Inhibition by mRNA Structures that Overlap with the Ribosomal Footprint in N-Terminal Coding Sequences. *Nucleic Acids Res.* 45, 5437–5448. doi:10.1093/nar/gkx061

Garenne, D., Beisel, C. L., and Noireaux, V. (2019). Characterization of the All-E. coli Transcription-translation System myTXTL by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 33, 1036–1048. doi:10.1002/rcm.8438

Gregorio, N. E., Kao, W. Y., Williams, L. C., Hight, C. M., Patel, P., Watts, K. R., et al. (2020). Unlocking Applications of Cell-free Biotechnology through Enhanced Shelf Life and Productivity of E. coli Extracts. *ACS Synth. Biol.* 9, 766–778. doi:10.1021/acssynbio.9b00433

Gregorio, N. E., Levine, M. Z., and Oza, J. P. (2019). A User's Guide to Cell-free Protein Synthesis. *MPs* 2, 24. doi:10.3390/mps2010024

Huang, A., Nguyen, P. Q., Stark, J. C., Takahashi, M. K., Donghia, N., Ferrante, T., et al. (2018). BioBits Explorer: A Modular Synthetic Biology Education Kit. *Sci. Adv.* 4. doi:10.1126/sciadv.aat5105

Jewett, M. C., and Swartz, J. R. (2004). Mimicking theEscherichia Coli Cytoplasmic Environment Activates Long-Lived and Efficient Cell-free Protein Synthesis. *Biotechnol. Bioeng.* 86, 19–26. doi:10.1002/bit.20026

Jin, X., Kightlinger, W., and Hong, S. H. (2019). Optimizing Cell-free Protein Synthesis for Increased Yield and Activity of Colicins. *MPs* 2, 28. doi:10.3390/mps2020028

Kay, J. E., and Jewett, M. C. (2020). A Cell-free System for Production of 2,3-butanediol Is Robust to Growth-Toxic Compounds. *Metab. Eng. Commun.* 10, e00114. doi:10.1016/j.mec.2019.e00114

Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., and Singh, V. (2019). Exploring the Potential of Cell-free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* 7, 248. Available at: https://www.frontiersin.org/article/10.3389/fbioe.2019.00248 (Accessed March 4, 2022). doi:10.3389/fbioe.2019.00248

Kightlinger, W., Duncker, K. E., Ramesh, A., Thames, A. H., Natarajan, A., Stark, J. C., et al. (2019). A Cell-free Biosynthesis Platform for Modular Construction of Protein Glycosylation Pathways. *Nat. Commun.* 10, 5404. doi:10.1038/s41467-019-12024-9

Kwon, Y.-C., and Jewett, M. C. (2015). High-throughput Preparation Methods of Crude Extract for Robust Cell-free Protein Synthesis. *Sci. Rep.* 5, 8663. doi:10.1038/srep08663

Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R., and Oza, J. P. (2019). *Escherichia Coli*-Based Cell-Free Protein Synthesis: Protocols for a Robust, Flexible, and Accessible Platform Technology. *J. Vis. Exp.* (144), e58882. doi:10.3791/58882

Levine, M. Z., So, B., Mullin, A. C., Fanter, R., Dillard, K., Watts, K. R., et al. (2020). Activation of Energy Metabolism through Growth Media Reformulation Enables a 24-Hour Workflow for Cell-free Expression. *ACS Synth. Biol.* 9, 2765–2774. doi:10.1021/acssynbio.0c00283

Liu, R., Zhang, Y., Zhai, G., Fu, S., Xia, Y., Hu, B., et al. (2020). A Cell-Free Platform Based on Nisin Biosynthesis for Discovering Novel Lanthipeptides and Guiding Their Overproduction *In Vivo*. *Adv. Sci.* 7, 2001616. doi:10.1002/advs.202001616

McSweeney, M. A., and Styczynski, M. P. (2021). Effective Use of Linear DNA in Cell-free Expression Systems. *Front. Bioeng. Biotechnol.* 9, 662. doi:10.3389/fbioe.2021.715328

Mullin, A. C., Slouka, T., and Oza, J. P. (2022). "Simple Extract Preparation Methods for E. Coli-Based Cell-free Expression," in Cell-Free Gene Expression: Methods and Protocols *Methods in Molecular Biology*. Editors A. S. Karim and M. C. Jewett (New York, NY: Springer US), 51–64. doi:10.1007/978-1-0716-1998-8_2

Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., et al. (2016). Portable, On-Demand Biomolecular Manufacturing. *Cell* 167, 248–259. e12. doi:10.1016/j.cell.2016.09.013

Romantseva, E., and Strychalski, E. A. (2020). *CELL-FREE (Comparable Engineered Living Lysates for Research Education and Entrepreneurship) Workshop Report*. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.SP.1500-13

Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression. *Nat. Biotechnol.* 27, 946–950. doi:10.1038/nbt.1568

Shilling, P. J., Mirzadeh, K., Cumming, A. J., Widesheim, M., Köck, Z., and Daley, D. O. (2020). Improved Designs for pET Expression Plasmids Increase Protein Production Yield in Escherichia coli. *Commun. Biol.* 3, 1–8. doi:10.1038/s42003-020-0939-8

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* 19, 751–755. doi:10.1038/90802

Shimizu, Y., Kanamori, T., and Ueda, T. (2005). Protein Synthesis by Pure Translation Systems. *Methods* 36, 299–304. doi:10.1016/j.ymeth.2005.04.006

Si, Y., Kretsch, A. M., Daigh, L. M., Burk, M. J., and Mitchell, D. A. (2021). Cell-Free Biosynthesis to Evaluate Lasso Peptide Formation and Enzyme-Substrate Tolerance. *J. Am. Chem. Soc.* 143, 5917–5927. doi:10.1021/jacs.1c01452

Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020). Cell-free Gene Expression: an Expanded Repertoire of Applications. *Nat. Rev. Genet.* 21, 151–170. doi:10.1038/s41576-019-0186-3

Smith, M. T., Berkheimer, S. D., Werner, C. J., and Bundy, B. C. (2014). Lyophilized Escherichia Coli-Based Cell-free Systems for Robust, High-Density, Long-Term Storage. *BioTechniques* 56, 186–193. doi:10.2144/000114158

Smith, P. E. J., Slouka, T., Dabbas, M., and Oza, J. P. (2021). From Cells to Cell-Free Protein Synthesis Within 24 Hours Using Cell-Free Autoinduction Workflow. *J. Vis. Exp.* (173), e62866. doi:10.3791/62866

Swartz, J. R., Jewett, M. C., and Woodrow, K. A. (2004). "Cell-Free Protein Synthesis with Prokaryotic Combined Transcription-Translation," in Recombinant Gene Expression: Reviews and Protocols *Methods in Molecular Biology*. Editors P. Balbás and A. Lorence (Totowa, NJ: Humana Press), 169–182. doi:10.1385/1-59259-774-2:169

Williams, L. C., Gregorio, N. E., So, B., Kao, W. Y., Kiste, A. L., Patel, P. A., et al. (2020). The Genetic Code Kit: An Open-Source Cell-free Platform for Biochemical and Biotechnology Education. *Front. Bioeng. Biotechnol.* 8, 941. doi:10.3389/fbioe.2020.00941

Xu, H., Yang, C., Tian, X., Chen, Y., Liu, W.-Q., and Li, J. (2022). Regulatory Part Engineering for High-Yield Protein Synthesis in an All-Streptomyces-Based Cell-free Expression System. *ACS Synth. Biol.* 11, 570–578. doi:10.1021/acssynbio.1c00587

Zhang, L., Lin, X., Wang, T., Guo, W., and Lu, Y. (2021). Development and Comparison of Cell-free Protein Synthesis Systems Derived from Typical Bacterial Chassis. *Bioresour. Bioprocess.* 8, 58. doi:10.1186/s40643-021-00413-2

Zhang, Y., Huang, Q., Deng, Z., Xu, Y., and Liu, T. (2018). Enhancing the Efficiency of Cell-free Protein Synthesis System by Systematic Titration of Transcription and Translation Components. *Biochem. Eng. J.* 138, 47–53. doi:10.1016/j.bej.2018.07.001

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 | Frontiers in Bioengineering and Biotechnology

REVIEW
published: 01 July 2022
doi: 10.3389/fbioe.2022.918659

Biochemistry of Aminoacyl tRNA Synthetase and tRNAs and Their Engineering for Cell-Free and Synthetic Cell Applications

*Ragunathan Bava Ganesh and Sebastian J. Maerkl**

School of Engineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Cell-free biology is increasingly utilized for engineering biological systems, incorporating novel functionality, and circumventing many of the complications associated with cells. The central dogma describes the information flow in biology consisting of transcription and translation steps to decode genetic information. Aminoacyl tRNA synthetases (AARSs) and tRNAs are key components involved in translation and thus protein synthesis. This review provides information on AARSs and tRNA biochemistry, their role in the translation process, summarizes progress in cell-free engineering of tRNAs and AARSs, and discusses prospects and challenges lying ahead in cell-free engineering.

Keywords: AARS, tRNA, genetic code, synthetic biology, cell-free systems, molecular engineering

OPEN ACCESS

Edited by:
Simon J Moore,
University of Kent, United Kingdom

Reviewed by:
Yoav S. Arava,
Technion Israel Institute of Technology, Israel
Mireille Moutiez,
CEA Saclay, France

****Correspondence:***
Sebastian J. Maerkl
sebastian.maerkl@epfl.ch

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *12 April 2022*
Accepted: *18 May 2022*
Published: *01 July 2022*

Citation:
Ganesh RB and Maerkl SJ (2022) Biochemistry of Aminoacyl tRNA Synthetase and tRNAs and Their Engineering for Cell-Free and Synthetic Cell Applications. Front. Bioeng. Biotechnol. 10:918659. doi: 10.3389/fbioe.2022.918659

1 CELL-FREE SYNTHETIC BIOLOGY

Cell-free reactions were first used in 1897 when Buchner showed that yeast extract can be used for the fermentation process (Buchner, 1897). In the last century when the scientific community was exploring the molecular biology of the cell and identifying the key components involved in the central dogma of life, cell-free studies were used as a tool to understand their fundamental biochemical functions and served as an additional tool to validate or support their hypothesis. A classic example is the requirement of template RNA for amino acid incorporation during protein synthesis which was proved *in vitro* by Nirenberg in 1961 using a cell-free system extracted from *E. coli* (Nirenberg and Matthaei, 1961). Cell-free biology can be defined as, "the reproduction, study, and exploitation of complex biological processes without intact cells" (Swartz, 2006). Research that was not possible with intact cells or constrained by the limitations or complexities of the cell, was made possible by cell-free systems. For example, lysate-based systems have been successfully used for the study and implementation of synthetic gene regulatory networks (Swank, Laohakunakorn, and Maerkl, 2019) and forward engineering of genetic oscillators (Niederholtmeyer et al., 2015). Cell-free systems also provided a faster and more convenient way to synthesize proteins using linear rather than circular DNA templates. Cell-free systems opened up the possibility to work with a range of organisms from conventional model organisms such as *E. coli* and yeast to more un-conventional systems such as *Bacillus megaterium* (Moore et al., 2018), *Clostridium autoethanogenum* (Krüger et al., 2020), and eukaryote-derived systems such as rabbit reticulocyte lysate (Gagoski et al., 2016).

Initially, cell-free systems were prepared using cell lysates, where cells were lysed, chromosomal DNA and cell membrane debris were removed, and the rest of the cellular contents were used for studies. Lysate-based systems suffered from batch-to-batch variation, hampering the ability to obtain consistent results (Hunter et al., 2018; Dopp, Jo, and Reuel, 2019). These systems also often contained inhibitory factors, nucleases, and proteases which lowered protein yield. Moreover, and in the

context of molecular engineering of fundamental importance, lysates are complex and their composition is unknown.

More recently, a recombinant system called "protein synthesis using recombinant elements" (PURE) has been generated where all individual components required for transcription, translation, and energy regeneration are expressed, purified, and then reconstituted to create a cell-free system. PURE consists of 36 reconstituted proteins capable of cell-free transcription and translation. Proteins were purified using His-tag–based affinity chromatography (Shimizu et al., 2001). A recombinant-based cell-free system was made possible as a result of an improved understanding of cellular biochemistry and the molecular machinery involved in transcription and translation. Various studies on the PURE system were performed to make PURE preparation easier and to decrease its cost (Shepherd et al., 2017; Villarreal et al., 2018; Lavickova and Maerkl, 2019). Productivity and functionality was increased by adjusting the various PURE components (Li et al., 2014) and by supplementing additional factors to the system (Maddalena et al., 2016; Li et al., 2017b). With the advent of cell-free systems and the PURE system becoming more accessible and affordable, it is beginning to be used in various applications. The PURE system has been explored as a platform for producing therapeutic proteins (Cai et al., 2015; Dondapati et al., 2020) and for molecular diagnostics (Pardee et al., 2016). The modular nature of PURE has made this system also an appealing starting point for bottom-up synthetic cell approaches (Niederholtmeyer, Stepanova, and Maerkl, 2013; Lavickova, Laohakunakorn, and Maerkl, 2020). We have recently written a comprehensive review covering various aspects and applications of cell-free synthetic biology (Laohakunakorn et al., 2020), and the focus of this review lies on aminoacyl tRNA synthetase and tRNA biochemistry and engineering in the context of cell-free systems.

2 THE GENETIC CODE

2.1 Discovery of the Genetic Code

Genetic information in biology is decoded through transcription and translation steps to result in RNA and proteins, respectively. Scientific activities in the last century helped us reach our current understanding of the steps involved in processing genetic material. In the following, a brief history of the discovery of key steps and components involved in the central dogma is discussed. In the early 20th century, the scientific community believed that proteins should be the genetic material as they are structurally diverse being made from 20 different building blocks. However, nucleic acids have only 4 bases as their building blocks and were thus thought less likely to be the carrier of genetic information.

The nucleic acids (DNA and RNA) were first discovered in 1869 by Friedrich Miescher and were termed "nuclein" since they were found in the nucleus (Miescher, 1869). Many years after his discovery, Levene in 1919 identified the components of nucleic acid and the sugar group of nucleotides (Levene, 1919). Experiments by Avery and his colleagues in 1944 provided definitive proof that DNA is the genetic information with its transforming ability in bacteria (Avery, Macleod, and McCarty, 1944). On the other hand, proteins were conclusively proved not to be the genetic material by Hersley and Chase in 1952 (Hershey and Chase, 1952), and the work by Rosaling Franklin, Maurice Wilkins, James Watson, and Francis Crick leading to the 1953 paper describing the three-dimensional structure of the genetic material DNA (Watson and Crick, 1953), and its double-helical structure with base-pairing according to Chargaff's rules (Chargaff, 1950).

The involvement of messenger RNA (mRNA) in the central dogma was floated since 1947 but the experimental discovery came in the year 1961 by Francois Gros and Francois Jacob (Jacob and Monod, 1961). For transfer RNA (tRNA), Crick hypothesized the existence of "adaptor" molecules, which are unstable and help carry the amino acids to the ribosomes in the cytoplasm for protein synthesis (Crick, 1955). tRNA was discovered by Paul Zamecnik in 1958 as a soluble RNA intermediate in protein synthesis and was the first non-coding RNA to be discovered (Hoagland et al., 1958). The ribosomal complex consists of ribosomal RNA (rRNA) and proteins, which is the site of protein synthesis located in the cytoplasm. It was discovered by George Palade in 1955 as small cytoplasmic bodies (Palade, 1955). Aminoacyl tRNA synthetases (AARSs) were first identified as activating enzymes in 1958, responsible for activating amino acids in the presence of ATP. Only after undergoing this activation step were amino acids able to participate in protein synthesis (Hoagland, Keller, and Zamecnik, 1956).

The genetic code linkage between nucleic acids and proteins was discovered by Marshall Nirenberg in 1961 (Nirenberg and Matthaei, 1961). Nirenberg's work showed that codon triplets of RNA gave rise to amino acid sequences during protein synthesis. This work laid the foundation for establishing the codon–amino acid relationship in protein synthesis, which is referred to as the second genetic code.

2.2 Current Understanding of the Genetic Code

Many components are required to implement the genetic code and key components are DNA, DNA polymerase, RNA polymerase, mRNA, tRNA, ribosomal complex, amino acids, and AARS. Genetic code implementation is a result of specific interactions between these components. There is a nucleotide-based world (DNA and RNA) and an amino acid–based world (proteins). DNA replication duplicates and maintains the genetic code. The transcription step results in single-stranded RNA molecules using DNA as its template and the translation step uses RNA as its template to synthesize protein peptides and completes the decoding of genetic information. The translation process bridges the nucleotide world and the amino acid world. Specifically, AARS and tRNA connect the two worlds. Now we know that AARSs are the enzymes responsible for charging cognate amino acid onto its cognate tRNA. The fidelity of the translation process is hugely dependent on the specificity of the AARS enzymes. These enzymes have a direct influence on the protein synthesis process, segregating proteogenic amino acids

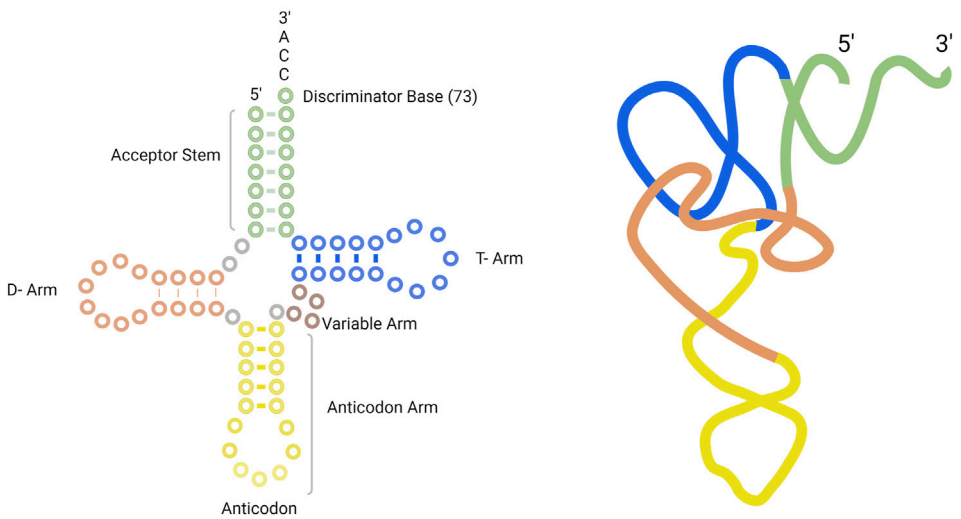


FIGURE 1 | Structure of tRNA presented in the 2-D cloverleaf model (left) and in the 3-D "L" shaped model (right). The tRNA consists of acceptor stem (green), T-arm (blue), D-arm (orange), anticodon arm (yellow), and variable arm (purple).

from non-proteogenic amino acids. The next section contains details on tRNAs and AARSs focusing on synthesis, biochemistry, function, and editing activity.

3 BIOCHEMISTRY OF TRNA AND AMINOACYL TRNA SYNTHETASES

This section focuses on providing basic information on the structure, biochemistry, mode of action, classification, and the role of tRNA and AARS in the protein synthesis process. We are exclusively discussing *E. coli* AARSs and tRNAs unless otherwise indicated.

3.1 tRNA

The tRNA molecule is a single-stranded, non-coding RNA molecule. The general structure of tRNA in 2-D and in 3-D is provided in **Figure 1**. The tRNA structure consists of the following: the anticodon arm, D-arm, T-arm, acceptor stem, and the variable arm. Bases in the anticodon arm of tRNA molecules read the genetic information in mRNA codons and contain the corresponding amino acid present in the 3′-CCA sequence of the acceptor stem. tRNAs from the same species can exhibit a difference in sequence length, size of the variable arm, and length of acceptor stem. For example, $tRNA^{Sel}$, the tRNA for selenocysteine amino acid in *E. coli*, is 95 nucleotides in length. On average, tRNA sequence length ranges from 75 to 90 nucleotides (Shepherd and Ibba, 2015).

Genes coding for tRNA are mostly arranged in groups in bacterial chromosomes with multiple copies present. Transcription is performed by RNA polymerase and results in tRNA precursor transcript having additional nucleotides on both 5′ and 3′ ends. Each tRNA transcript undergoes a maturation process where nucleotides are removed, specific nucleotide modifications occur, and structural integrity is gained, resulting in the cloverleaf shape. After maturation, a tRNA is available for amino acid charging at the 3′end by an AARS (Shepherd and Ibba, 2015).

The number of tRNAs present in an organism is dependent on codon usage. From a theoretical point of view, there are 64 (4^3) different codon sequences available. However, only 61 different tRNAs are used, each corresponding to a particular codon, and the remaining three codons (UAA, UAG, and UGA) are called stop codons and do not have a corresponding tRNA. These 61 tRNAs are shared amongst 20 amino acids. The number of tRNA acceptors for each amino acid is not the same and varies across amino acids. For example, there exists only one tRNA for the amino acid methionine with codon AUG. On the other hand, multiple tRNAs can carry the same amino acid at their 3' end and such tRNA groups are referred to as isoacceptors. For example, there are six isoacceptor tRNAs for the amino acid lysine with codons UUA, UUG, CUA, CUU, CUG, and CUC.

3.2 Aminoacyl tRNA Synthetases

AARSs are a family of enzymes responsible for adding an amino acid onto its cognate tRNA molecule. They are the enzyme implementing the genetic code. There are 21 AARS enzymes present, one for each amino acid except lysine, which has two AARSs. In addition to these 21 AARSs for proteogenic amino acids, there are AARSs for non-proteogenic amino acids such as pyrrolysyl tRNA synthetase and phosphoseryl tRNA synthetase. These additional AARSs are found in archaea and bacteria. Each tRNA has a particular AARS for its activation. In general, tRNA

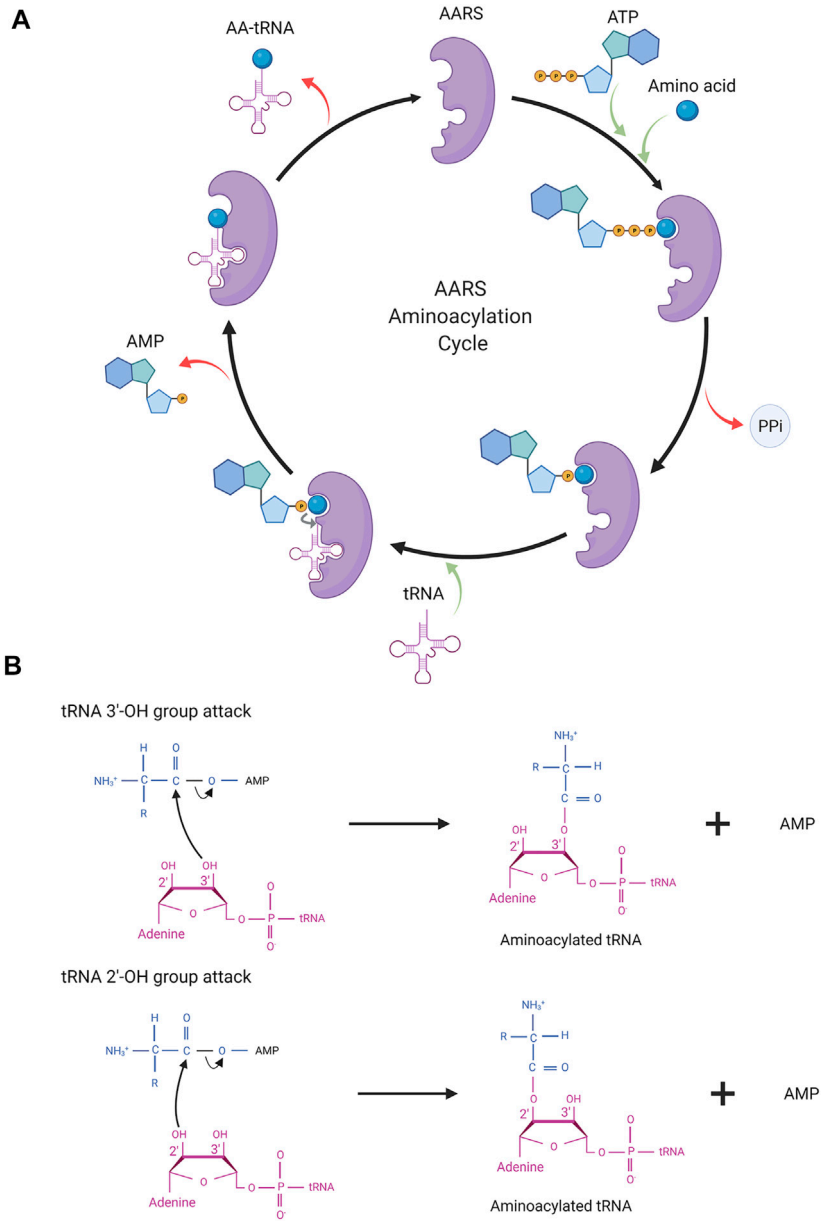


FIGURE 2 | (A) AARS aminoacylation process. In the first step, AARS binds to the ATP and amino acid to form an aminoacyl intermediate. In the last step, the amino acid is transferred onto tRNA resulting in activated tRNA being ready for the translation process. AARS becomes free for the next cycle of aminoacylation. **(B)** Depicting the molecular structure of terminal adenosine of tRNA and the attack of the hydroxyl group (2′/3′) in AA-AMP intermediates.

charging by AARSs with amino acid takes place in two steps (**Figure 2**).

In the first step, amino acid activation takes place. ATP and amino acid bind to the AARS enzyme triggering a nucleophilic attack of the amino acid carboxyl oxygen to the α-phosphate group of ATP. This results in amino acid adenylate intermediate (AA–AMP) and release of pyrophosphate (PPi). In the second step, one of the hydroxyl groups of adenosine (3′-OH/2′-OH) in tRNA attacks the carboxyl carbon of AA–AMP intermediate resulting in the transfer of amino acid to tRNA. The amino acids and tRNAs are linked by an ester bond. This step results in tRNA-AA, and AMP, which are released from the catalytic site of the enzyme and the AARS enzyme is free for the next cycle. In general, for the amino acid activation step, tRNA is not required but some AARSs such as GlnRS, GluRS, ArgRS, and class I LysRS require tRNA as a prerequisite for amino acid activation. Activated tRNA-AA binds with the elongation factor, EF-TU, and when reaching the ribosome participates in translation.

3.2.1 Aminoacyl tRNA Synthetases Classes

The 23 AARS enzymes are classified into two classes depending on the structure of the active site. In class I AARSs, the active site

contains the Rossmann fold with five parallel β sheets connected by α helices. The Rossmann fold contains highly conserved motifs: HIGH and KMSKS, and both motifs are connected by sequence stretches called CP1. Class II AARS active sites have several parallel β strands flanked by α helices. Class I enzymes are either monomeric or dimeric and class II enzymes are dimeric or tetrameric. There exist many differences between class I and class II AARSs. Amino acid charging takes place at the 3′-OH group of tRNA in class I AARSs and PheRS, and 2′-OH group in class II AARSs. ATP binding during amino acid activation differs in both classes. ATP binds in an extended conformation in class I AARSs and in a kink conformation in class II AARSs. AARSs differ in the way they bind tRNA with class I AARSs binding the minor groove of tRNA and class II AARSs binding the major groove. In terms of reaction catalysis, the rate-limiting steps between both classes differ as well. In class I AARSs, the release of activated tRNA (tRNA-AA) is the rate-limiting step whereas in class II AARSs it is the amino acid activation step. Each AARS class is further categorized into subclasses depending on the type of amino acid charged by the enzyme. Across both class I and class II, subclass A recognizes aliphatic and thiolated amino acids, subclass B recognizes charged polar amino acids, and subclass C recognizes aromatic amino acids (Rubio Gomez and Ibba, 2020).

3.2.2 Substrate Recognition by Aminoacyl tRNA Synthetases

Amino acids are much smaller in size than tRNAs, and so fewer chemical moieties are available for AARSs to distinguish cognate from non-cognate amino acids. Generally, amino acids are recognized based on their size, functional group, and ability to bind with metal ions present in the active site of the enzyme. In the enzyme PheRS, a conserved alanine residue helps discriminate phenylalanine over tyrosine (Reynolds et al., 2010). Glycine, the smallest amino acid, is recognized by the high negative charge in the binding pocket of its AARS. Five different conserved negative charges are used to identify glycine. Two threonine residues in the GlyRS binding pocket help to prevent other amino acids from activating. Crystallographic studies have shown that zinc metal ions in the active site of the enzyme help distinguish cognate from non-cognate amino acids (Valencia-Sánchez et al., 2016). In the case of ThrRS (Sankaranarayanan et al., 2000) and CysRS (Zhang et al., 2003), the cognate amino acid is selected by its ability to interact with zinc ions present in the active site whereas non-cognate amino acids fail to do so.

Each AARS has a specific binding pocket for tRNA. Selecting the cognate tRNA is crucial for ensuring translation fidelity. Initial binding of tRNA to AARS is fast and non-specific and governed by electrostatic interaction. Upon initial binding, specific interactions between tRNA and AARS ensure recognition of the correct tRNA. Specific interactions are formed more slowly, accompanied by conformational changes in the active site of AARS. Specific interactions are mediated by identity elements such as modified nucleotides, conserved residues, base stacking, and different tRNA arm lengths. The identity elements include determinants and anti-determinants. Determinants favor binding of cognate tRNAs with AARSs while anti-determinants disfavor binding of non-cognate tRNAs. The most commonly used identity elements are the anticodon bases 34, 35, and 36 in the anticodon arm and the discriminatory base 73 in the acceptor stem of the tRNA. In the case of AlaRS, tRNA is recognized exclusively based on the presence of G3-U70 base pair (McClain and Foss, 1988a). When non-cognate tRNAs were engineered *in vitro* containing the aforementioned base pair, AlaRS recognized those tRNAs and charged them with alanine (Hou and Schimmel, 1988; McClain and Foss, 1988a). For SerRS, the length of the variable arm is more crucial for its discrimination than sequence (Park and Schimmel, 1988; Asahara et al., 1993). The complete list of identity elements and their location in the tRNA for each AARS from *E. coli* is provided in **Table 1**. All AARSs are classified into three groups based on the location of the tRNA identity elements. AARSs in group 1 have identity elements located in all regions of the tRNA, namely, the acceptor stem, anticodon arm, other domains (T-arm, D-arm, and the variable arm). AARSs in group 2 have identity elements located only in the acceptor stem and anticodon arm. AARSs in group 3 have identity elements located in the acceptor stem and other domains of the tRNA but not in the anticodon arm. The list of AARSs in each group and the location of their identity elements on tRNA are provided in **Figure 3**.

Additional discrimination comes from the kinetic aspect of binding to discriminate cognate tRNAs. Aminoacylation with cognate tRNAs is more influenced by K_{cat} values than K_M values (Ebel et al., 1973). Evolutionary conservation of the identity elements in tRNAs suggests their importance, even though these elements do not directly contribute to protein synthesis.

3.2.3 Proofreading Mechanism by Aminoacyl tRNA Synthetases

Pauling in 1958, theoretically predicted that amino acid misincorporation during translation should be about 1 in 200 (Pauling, 1958). However, *in vivo* experiments showed that this error rate is about 1 in 3,000 (Loftfield, and Vanderjagt, 1972). Aminoacylation by AARS has an error rate of about 1 in 10,000. This led to the suggestion of some editing mechanism being in place to account for these observations. The low error rate for AARSs is due to better recognition of cognate substrates and a proofreading/editing mechanism. This section briefly describes the editing/proofreading mechanism used by AARSs to ensure faithful aminoacylation.

Fersht in 1977 proposed a “double sieve model” to explain the low error rate and presence of separate catalytic active and editing sites. According to this model, the active site of the enzyme acts as a first “coarse” sieve to filter out amino acids that are larger than the cognate amino acid. Amino acids which are similar to or smaller than cognate still will become activated in the active site. The second “fine” sieve is the editing site capable of hydrolysis, which has a pocket size smaller than the cognate amino acid. The editing site serves to de-acylate any of the mischarged smaller amino acids which passed through the first sieve (Fersht, 1977). This way, de-acylation of cognate amino acid is prevented as it cannot enter the editing site. Evidence for the presence of a separate editing site is seen in 10 AARS from both classes I and II.

TABLE 1 | List of tRNA identity elements and their location on tRNA for aminoacylation by AARS from *E. coli*. Identity elements for fMet are provided in italics and bold.

AARS	Identity element location				References
	Acceptor stem	Anticodon arm		Other domains (d-arm/ T-arm/variable arm)	
		Anticodon	Other location		
Alanine	A73, G2:C71, G3:U70, G4:C69			G20	Hou and Schimmel, 1988; McClain and Foss, 1988a; Francklyn et al., 1992
Arginine	A/G73	C35, U/G36		A20	McClain and Foss, 1988b; Schulman and Pelka, 1989; McClain et al., 1990; Tamura et al., 1992
Asparagine	G73	G34, U35, U36			Shimizu et al., 1992; Li et al., 1993
Aspartic acid	G73, G2:C71	G34, U35, C36,	C38	G10	Hasegawa et al., 1989; Nameki et al., 1992
Cysteine	U73, G2:C71, C3:G70	G34, C35, A36		G15: G48, A13:A22	Pallanck et al., 1992; Shimizu et al., 1992; Hou et al., 1993; McClain, 1993; Komatsoulis and Abelson, 1993; Hamann and Hou, 1997
Glutamine	G73, U1:A72, G2:C71, G3:C70	C/T34, U35, G36	A37, U38	G10	Rogers and Söll, 1988; Jahn et al., 1991; Hayase et al., 1992; Ibba et al., 1996
Glutamic acid	G1:C72, U2:A71	U34, U35,	A37	U11:A24, U13:G22-A46, 47	Normanly et al., 1990; Sylvers et al., 1993; Gregory and Dahlberg, 1995; Sekine et al., 1996
Glycine	U73, G1:C72, C2:G71, G3:C70	C35, C36			Shimizu et al., 1992; Francklyn et al., 1992; McClain et al., 1991
Histidine	C73, G1	Anticodon			Shimizu et al., 1992; Francklyn et al., 1992; Himeno et al., 1989; Francklyn and Schimmel, 1990; Yan and Francklyn, 1994; Yan et al., 1996
Isoleucine	A73, C4:G 69	G34, A35, U36	A37, A38	U12:A23, C29:G41	Pallanck and Schulman, 1991; Muramatsu et al., 1988; Nureki et al., 1993; Nureki et al., 1994
Leucine	A73			U8:A14	Normanly et al., 1992; Asahara et al., 1993
Lysine	A73	U34, U35, U36			Normanly et al., 1990; McClain et al., 1990; Tamura et al., 1992
Methionine ***fmet***	A73, U4:A69, A5:U68 ***G2:C71, C3:G70***	C34, A35, U36	***C32, U33, A37***		Uemura et al., 1982; Schulman and Pelka, 1988; Meinnel et al., 1993; Lee et al., 1992
Phenylalanine	A73	G34, A35, A36	G27:C43, G28:C42	U20, G44, U45, U59, U60	Pallanck and Schulman, 1991; McClain and Foss, 1988c; Peterson et al., 1993; Peterson and Uhlenbeck, 1992
Proline	A73, G72	G35, G36		G15:C48	Shimizu et al., 1992; McClain et al., 1994; Liu et al., 1995
Serine	G73, C72, G2:C71, A3:U70, C11:G24, G/A4: T/C69			C11:G24 (variable arm)	Normanly et al., 1992; Himeno et al., 1990; Rogers and Söll, 1988; Normanly et al., 1986; Sampson and Saks, 1993; Asahara et al., 1994; Saks and Sampson, 1996
Threonine	G1:C72, C2:G71	G34, G35, U36			Schulman and Pelka, 1990; Hasegawa et al., 1992
Tryptophan	G73, A1:U72, G2:C71, G3:C70	C34, C35, A36			Himeno et al., 1991; Pak et al., 1992; Rogers et al., 1992
Tyrosine	A73	U35			Hou and Schimmel, 1989; Bedouelle, 1990; Himeno et al., 1990; Sherman et al., 1992
Valine	A73, G3:C70, U4:A69	A35, C36			Chu and Horowitz, 1991; Tamura et al., 1991; Pallanck and Schulman, 1991

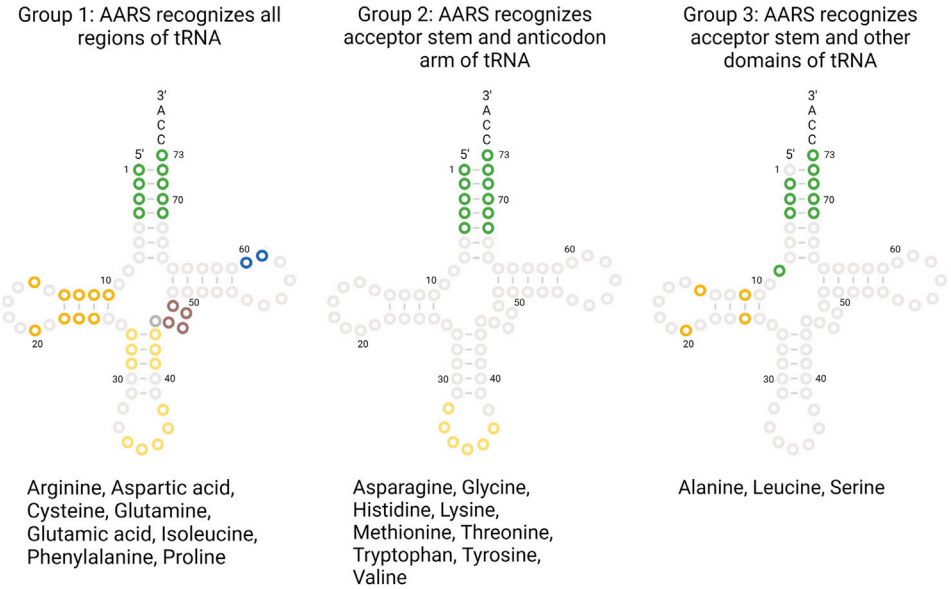


FIGURE 3 | AARS identity elements on tRNA. AARSs are grouped based on the location of the identity elements present on tRNA regions. The list of AARS under each group and the tRNA bases utilized for recognition are highlighted.

Editing activity can be divided into pre-transfer editing and post-transfer editing. In pre-transfer editing, editing occurs before the amino acid is transferred to tRNA, and in post-transfer editing, editing occurs after the amino acid is transferred to tRNA. Most AARSs use one of these editing mechanisms, but some AARS such as LeuRS and ValRS use both mechanisms.

Pre-transfer editing occurs after the formation of amino acid adenylate (AA–AMP) but before transfer to tRNA. Pre-transfer editing is seen in both AARS classes. Pre-transfer editing can occur in two methods. In the first method, AA–AMP is released by the enzyme to the cytosol and the phosphoester bond is spontaneously hydrolyzed. In the second method, enzymatic hydrolysis of AA–AMP occurs either in the active site or in a separate editing site. For example, thiolated non-proteogenic amino acids such as homocysteine and ornithine are cleared by pre-transfer editing in the active site of the enzyme by MetRS and LysRS.

Post-transfer editing occurs after the transfer of amino acid to tRNA and occurs in a separate editing site. This editing involves cleaving the ester bond between amino acid and tRNA. In general, once tRNA-AA is formed, amino acid triggers a conformational change in the 3′ end of tRNA and results in tRNA translocation. Translocation results in an amino acid being in the editing site where it is hydrolyzed. For class II AARS, mischarged tRNA is rapidly released and in those cases, enzymes are capable of recapturing these mischarged tRNA for editing. In the case of PheRS, PheRS competes for Tyr-tRNA^{Phe} with EF-TU to recapture and edit the tRNA.

The preference for the editing mechanism used is dependent on the rate of amino adenylate hydrolysis and transfer to tRNA. In case of a faster transfer rate to tRNA as in ValRS, post-transfer is preferred. For IleRS, both reaction rates are fairly equal and hence both editing mechanisms are used.

There also exists a separate group of editing proteins that act independently of AARSs called trans-editing factors. These enzymes provide additional quality control in the editing process. The role of the trans-editing factors is to clear mischarged tRNAs before they reach the ribosome. D-aminoacyl-tRNA deacetylases are another class of trans-editing factors targeting in particular tRNAs charged with D-amino acids. The presence of such multiple editing mechanisms signifies the importance of the aminoacylation process and fidelity of the translation process. A schematic representation of various editing mechanisms for aminoacylation is provided in **Figure 4**.

3.2.4 Aminoacylation Kinetics

The accuracy of protein synthesis relies on an AARS's ability to recognize cognate amino acids and tRNAs. Kinetic analysis is useful to develop the mechanism of action in the steps involved in aminoacylation. This section provides information about different kinetic approaches used to study tRNA aminoacylation.

The parameters widely used to describe the kinetics of AARSs are substrate affinity K_M, enzyme turnover K_{cat}, maximum velocity V_{max}, and enzyme specificity K_{cat}/K_M. K_M refers to affinity of the enzyme to the substrate. K_{cat} is the catalytic constant for substrate to product conversion. K_{cat}/K_M is the specificity constant or catalytic efficiency of the enzyme. Steady-state kinetics is useful for initial characterization of the enzyme and to measure kinetic parameters. Since steady-state kinetics are used generally, parameters obtained can be compared across systems. For example, enzyme specificity for cognate and non-cognate amino acids can be measured and compared across

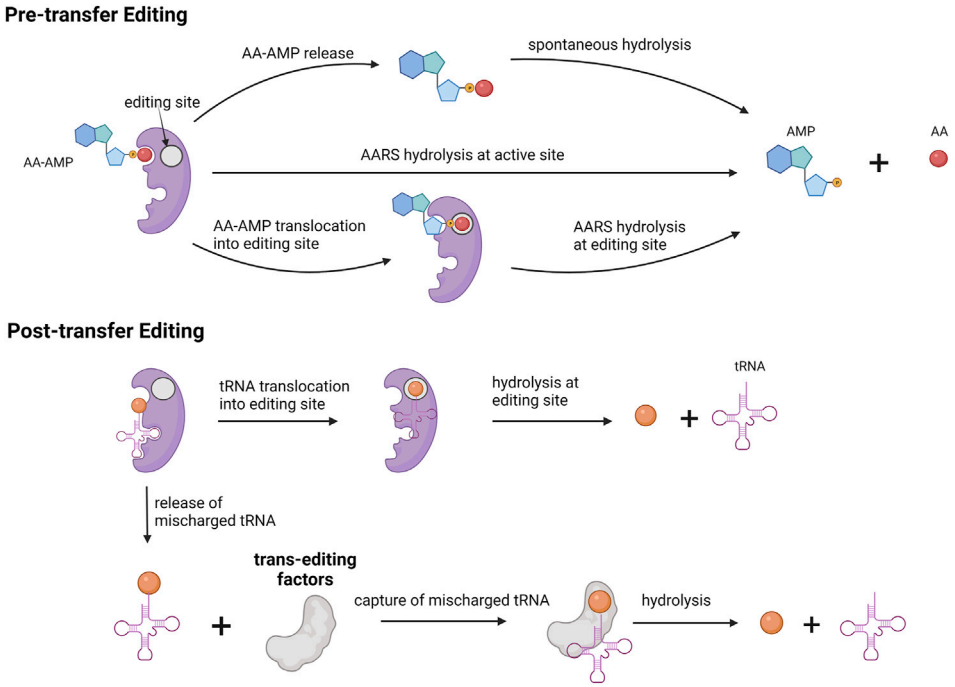


FIGURE 4 | Editing mechanism for aminoacylation process. Pre-transfer editing occurs before amino acid gets charged onto tRNA. Post-transfer editing occurs after amino acid is charged onto tRNA. Trans editing factors are independent editing proteins involved in hydrolysis of mischarged tRNA.

TABLE 2 | K_M and K_{cat} values of amino acids for AARS from *E. coli* unless mentioned otherwise. Unit for K_{cat} is s^{-1} unless mentioned otherwise.

AARS	Amino acid		References
	K_M (μm)	K_{cat} (s^{-1})	
AlaRS	240 ± 50	33 ± 7	Hill and Schimmel, 1989
ArgRS	12	2.2	Lin et al., 1988; Airas, 2006
AsnRS	32	1.6	Madern et al., 1992
AspRS	60	18	Martin et al., 1997
CysRS	0.4 ± 0.1	680 ± 60 (nmol min^{-1} mg $protein^{-1}$)	Komatsoulis and Abelson, 1993
GlnRS	0.114 ± 0.012	157 ± 7 min^{-1}	Liu et al., 1998
GluRS	5	5.5 ± 1.0	Lapointe and Söll, 1972; Campanacci et al., 2004
GlyRS	0.03	0.31 ± 0.02 (*Homo sapiens*)	Ostrem and Berg, 1974; Cader et al., 2007
HisRS	1.4 ± 0.6	2.6 ± 0.4	Augustine and Francklyn, 1997
IleRS	2.1 ± 0.2	3.1 ± 0.2	Xu et al., 1994
LeuRS	15	3	Chen et al., 2000
LysRS	230 ± 20	0.34 ± 0.009	Wang et al., 2006
MetRS	1.2 ± 0.2	3.2 ± 0.2	Ghosh et al., 1991
PheRS	1.8 ± 0.2	65 ± 3 min^{-1}	Moor et al., 2011
ProRS	250 ± 35	70 ± 25	Beuning and Musier-Forsyth, 2001
SerRS	0.56 ± 0.15	2.6 ± 0.4	Borel et al., 1994
ThrRS	12	0.3	Hirsh, 1968
TrpRS	0.53 ± 0.08	1.34 ± 0.26	Chan and Koeppe, 1994
TyrRS	3.3 ± 0.8	0.74 ± 0.06	Hamano-Takaku et al., 2000
ValRS	4.3	13.9	Tardif and Horowitz, 2004

AARSs. Steady-state measurements are usually performed with substrate concentrations much higher than enzyme since the assay follows product formation. Minimal material requirements and fast readout make the steady-state approach suitable for initial characterization. The drawback of steady-state kinetics is that elementary reactions cannot be characterized. To determine the rate of the aminoacylation process, ATP pyrophosphate exchange assays and aminoacylation assays are performed under steady-state kinetics.

The ATP–PPi exchange assay is based on the amino acid activation step. ATP and amino acids form aminoacyl adenylate intermediate (AA–AMP) with the release of pyrophosphate

TABLE 3 | Half-life values of tRNA-AA from *E. coli*. Values obtained based on ester bond hydrolysis under neutral or alkaline pH in a high ionic medium at 37°C (Hentzen, Mandel, and Garel, 1972).

tRNA-AA	$t_{1/2}$ (min)
Ala	6
Arg	12
Asn	11
Asp	11
Cys	16
Gln	9
Glu	9
Gly	8
His	16
Ile	65
Leu	7
Lys	14
Met	12
Phe	16
Pro	2
Ser	17
Thr	38
Trp	-
Tyr	15
Val	60

(PPi). In one approach, [^{32}P]-PPi is used for the reaction. The radioactive group reacts with AA–AMP resulting in [^{32}P]-ATP. This assay measures the exchange of [^{32}P]-PPi into ATP to provide rate of the activation step. In another approach, radioactive [^{32}P]-ATP is used for amino acid activation and the rate of ATP consumption is measured using activated charcoal or thin layer chromatography (TLC) plates.

Aminoacylation assays are dependent on the second step of amino acid transfer to tRNA. Amino acids radiolabeled with [^{3}H] or [^{14}C] are used to measure the rate of product AA-tRNA^{AA} formation over time. One drawback of using radiolabeled amino acids is that attaining saturating conditions is difficult. This can be challenging while determining the K_M value of tRNA for AARS. The aforementioned limitation can be overcome by using a radiolabeled [^{32}P] group at the 3′ end of tRNA and unlabeled amino acids. In such a way, saturating amino acid concentration can be used in the assay. **Table 2** contains the K_M and K_{cat} of amino acids for all AARSs from *E. coli* unless mentioned otherwise.

Pre-steady state kinetics is used to study elementary reaction steps. The pre-steady state kinetic approach is used to study fast reactions, in the order of a few milliseconds, present at an early stage of the interaction. This approach is best for understanding the mechanistic action of interaction. Parameters like individual rate constants of the reactants can be determined using pre-steady state kinetics. For example, substrate-binding order in the active site, formation, and consumption of intermediates can be studied by pre-steady state kinetics. Rapid chemical quench and stopped-flow fluorimetry are generally used to study AARSs. Rapid kinetic approaches were used to mechanistically distinguish the two classes of AARSs. As mentioned earlier, in class I AARSs, product release of AA-tRNA^{aa} is the rate-limiting step, and in class II AARSs, amino acid activation is the rate-limiting step.

Rapid chemical quench is a discontinuous assay providing a direct readout of the rate of the radiolabeled product formed. Stopped-flow fluorimetry is a continuous assay and provides an indirect readout of reaction progress. Progress is dependent on changes in intrinsic tryptophan fluorescence correlated to reaction chemistry.

Table 3 contains the half-life of activated tRNA-AA measured in *E. coli* (Hentzen, Mandel, and Garel, 1972). The value represents the spontaneous hydrolysis rate of tRNA-AA at neutral or alkaline pH in a high ionic condition at 37°C. Under the same conditions, the stability of the ester bond depends purely on the amino acid attached to tRNA. Half-life for all amino acids but tryptophan is presented and ranges from 2 to 65 min.

4 APPLICATIONS OF tRNA AND AMINOACYL tRNA SYNTHETASES IN CELL-FREE SYSTEMS

4.1 tRNA and Aminoacyl tRNA Synthetases *In Vitro* Synthesis

4.1.1 tRNA Synthesis

The ability to synthesize tRNA *in vitro* was demonstrated in 1973 when T4 DNA was transcribed to yield T4 tRNAs when incubated with *E. coli* extract obtained after infecting with T4 bacteriophage (Nierlich et al., 1973). Currently, a more sophisticated one-pot method for *in vitro* tRNA synthesis has been developed (Korencić, Söll, and Ambrogelly, 2002). In this method, T7 RNAP is used for tRNA transcription from a ds/ssDNA hybrid template. Transcribed tRNAs were shown to be produced in full length and aminoacylated by AARSs. This method did not have as high a yield as the control method using plasmid DNA as a template, but is a useful approach to produce functional tRNAs *in vitro* (Korencić, Söll, and Ambrogelly, 2002). *In vitro* synthesis of functional tRNA has made it easier to study interactions of tRNA with other components such as mRNA (Fauzi, Jack, and Hines, 2005) and AARSs (Wang et al., 2015), to investigate tRNA stability *in vitro* (Serebrov et al., 1998), and to determine mutational effects on tRNA function (Tamura et al., 1992).

The ability of *in vitro* transcribed tRNA to decode codons on mRNA during translation was studied as well. In one study, native tRNAs were depleted from the cell lysate and replaced with an *in vitro* transcribed tRNA subset. The cell lysate containing replenished tRNA was found to produce proteins and was able to decode all 61 codons with 48 transcribed tRNAs (Cui et al., 2015). This way, the tRNA pool directing the translation process can in principle be fully customized. In a similar study, efforts were taken to identify the minimal number of tRNAs required to support protein translation. The minimal set of tRNAs required for translation was explored by generating an entire set of 21 tRNAs *in vitro*, and they were shown to be able to synthesize proteins. A reduced number of tRNA was possible by using a single codon for each amino acid. The tRNAs were chosen such that they did not require modifications after synthesis. Furthermore, *in vitro* transcribed tRNAs provide flexibility in

redesigning the genetic code and facilitate site-specific incorporation of amino acids. Protein yield obtained using this reduced set of tRNA was shown to be up to 40% compared to the native system (Hibi et al., 2020).

4.1.2 Aminoacyl tRNA Synthetases Synthesis

From a synthetic biology perspective, the ability to express and sustain proteins is crucial for building a self-replicating cell. To this end, AARS proteins were expressed individually in the PURE system, and it was demonstrated that all AARSs except PheRS were expressed as soluble proteins (Awai, Ichihashi, and Yomo, 2015). These expressed proteins were functional with activity on par with their native counterparts purified from *E. coli*. The reason for the inactivity of PheRS was associated with insufficient formation of active dimers since dimer formation requires longer incubation at 4°C and low salt concentrations *in vitro* (Awai, Ichihashi, and Yomo, 2015). In a similar attempt, all 20 AARSs were expressed using a polycistronic plasmid encoding 32 proteins in total using the PURE system. The PURE system was able to synthesize all 32 proteins, including all 20 AARSs as confirmed by mass spectrometric analysis but functionality was not assessed (Doerr et al., 2021). Recently, in the bid to construct a self-replicating synthetic cell, a modified PURE system was used to demonstrate self-regeneration of up to 7 AARSs in a microfluidic reactor for more than 24 h (Lavickova, Laohakunakorn, and Maerkl, 2020).

4.2 tRNA and Aminoacyl tRNA Synthetase Engineering

4.2.1 tRNA Engineering

Major work on tRNA engineering was achieved *in vivo* in regard to genetic code expansion (GCE). GCE involves the increase in the genetic code alphabet by introducing new base nucleotides (Hoshika et al., 2019), unnatural base pairs (UBPs) (Mukba et al., 2020), creating new codons, and increasing the codon size to 4 (quadruplet codon) (Hohsaka et al., 2001). The aforementioned approaches were used for incorporating modified or non-canonical amino acids (NC-AAs). Another approach is to reassign existing codons for NC-AA incorporation. Incorporating a NC-AA requires an orthogonal translation system (OTS) with a tRNA/AARS pair that does not cross-react with the endogenous tRNA/AARS present in the system. There should not be any interference of the endogenous AARS with the exogenous AARS in its ability to recognize the exogenous tRNAs and NC-AAs, and vise-versa. Orthogonality can be achieved by utilizing tRNA/AARS pairs from phylogenetically distinct species, and this approach takes advantage of differences in codon usage. Another approach takes advantage of the difference in tRNA recognition by AARSs (Doctor and Mudd, 1963). Orthogonal pairs TyrRS/tRNA-Tyr obtained from archaea *Methanococcus jannaschii* and pyrrolysyl-(Pyl)RS/tRNA-Pyl from *Methanosarcina barkeri* are most commonly used for incorporating NC-AAs. The open nature of cell-free protein synthesis (CFPS) provides a higher degree of freedom for structurally and functionally diverse NC-AAs to be incorporated. Parameters like cellular toxicity, viability, and cross membrane transport are not constraints for CFPS allowing incorporation of many different NC-AAs.

Depending on the application, either site-specific or residue-specific incorporation is used. In site-specific incorporation, a NC-AA is incorporated into a specific location whereas in residue-specific incorporation a NC-AA is incorporated into all sites encoded by a specific codon. In site-specific incorporation, both sense codons and nonsense codons were utilized for NC-AA incorporation. Modifications are introduced in the identity elements present in the tRNA, mostly in the anti-codon arm and in the acceptor stem, to favor NC-AA incorporation. While using a nonsense codon, a stop codon (UAG, UGA, and UAA) is reassigned to incorporate NC-AA instead. The amber codon (UAG) is usually used since this codon is least used as a termination signal in *E. coli*. In the input DNA sequence, all instances of the amber codon are changed to either one of the other two stop codons which frees the amber codon for reassignment to a NC-AA. The stop codon UAG is then reassigned to the tRNA containing NC-AAs and the tRNA anticodon arm is modified to CUA. With this approach, NC-AAs can be site-specifically incorporated at multiple places (Hong et al., 2014; Martin et al., 2018). Similarly, tRNA suppressor targeted toward opal (UGA) and ochre (UAA) codons were developed and utilized for *in vitro* NC-AA incorporation (Gubbens et al., 2010). As for sense codons, amino acids such as leucine and arginine have up to 6 codons, therefore; 6 tRNAs code for the same amino acid. One of the least commonly used isoacceptor tRNAs is generally used for codon reassignment. A report from 2016 demonstrated site-specific incorporation using sense codon reassignment using *in vitro* transcribed tRNAs charged with NC-AA in the PURE system. One of the codons for valine (GUG), arginine (CGC), and glycine (GGC) was reassigned to a different NC-AA (Iwane et al., 2016). This approach suffered from low efficiency in NC-AA incorporation due to wobble decoding, the ability of the tRNA to recognize more than one codon. The commercially available FluoTech system has fluorescently modified lysine instead of lysine. The codon AAA for lysine is reassigned to BODIPY lysine. Here, tRNA pre-charged with BODIPY lysine is added to the *in vitro* protein expression system, incorporating fluorescent lysine into proteins. It should be noted that the native $tRNA_{AAA\text{-}lys}$ is still present in the system, leading to partial residue-specific incorporation of the fluorescent lysine. This system is used for easy detection of *in vitro* protein expression. For residue-specific incorporation, all of the sense codons for a particular amino acid are reassigned to the NC-AA. The native amino acid is replaced by a NC-AA in the amino acid pool. All tRNAs now carry the NC-AAs. Using an *E. coli* cell-free system canavanine amino acid, a toxic analog of arginine, was incorporated into the model protein at all locations of arginine (Worst et al., 2015).

As an alternative approach for NC-AA incorporation, quadruplet codons were used for tRNAs carrying the modified amino acids. The number of anticodons present in the anticodon arm is increased to 4 (UCCA) to match the corresponding 4-letter codon (AGGU). The modified amino acid was shown to be incorporated at multiple instances and up to two distinct

NC-AA were incorporated with this approach. A modified nitrophenylalanine-tRNA was used to decode the 4-letter codons. Such a modified tRNA allowed incorporating NC-AAs with better efficiency in an *E. coli in vitro* translation system (Hohsaka et al., 2001). So far, tRNAs were engineered to incorporate NC-AAs instead of a stop codon or amino acid (*via* sense codon). In an attempt to increase the amino acid diversity to more than 20, the codon table was split for amino acids arginine, glycine, and valine and, the free codons were assigned to three distinct NC-AAs, thereby increasing the total number of amino acids to 23, in addition to the 20 native amino acids (Iwane et al., 2016).

Charging of orthogonal tRNA with NC-AAs is a key step in the translation process and it is usually mediated by AARS enzymes. In addition to AARS enzymes, there are other methods such as the chemoenzymatic method, chemical method, and ribozyme-based approaches for acylation to charge tRNA with NC-AAs. These synthetic methods are utilized *in vitro* to generate pre-charged tRNAs with NC-AAs and can be directly supplemented into the cell-free system. Such pre-charged tRNAs are useful to learn more about single turnover translation. The key advantage of chemical acylation methods is that structurally and chemically diverse groups can be added onto tRNAs without the need to re-engineer AARS.

4.2.2 Aminoacyl tRNA Synthetases Engineering

The central theme for AARS engineering is in genetic code expansion and incorporation of unnatural or NC-AAs during peptide synthesis. AARSs can be engineered by modifying the amino acid binding pocket, tRNA binding pocket, and editing domain for acylating NC-AAs onto tRNAs. In a relatively simple approach, native AARSs are used to incorporate unnatural amino acids. Here, the lack of specificity in substrate recognition by native AARSs was exploited to charge tRNAs with unnatural amino acids to participate in protein synthesis. For example, research in non-ribosomal peptides utilized this approach to demonstrate simultaneous incorporation of 10 different amino acid analogs using the PURE-based recombinant system. The amino acid analogs used were substrates for 12 native AARSs from *E. coli* (Josephson, Hartman, and Szostak, 2005). Even though it is simple, this approach does suffer from low efficiency and therefore yield. Moreover, if cognate amino acids are present, they will compete with the same codon and lower the incorporation of NC-AAs. In such cases, the relative ratio of NC-AA to cognate amino acid should be tightly controlled to favor NC-AA incorporation.

The second approach uses engineered AARSs with altered specificity for amino acids or tRNAs. In the report mentioned earlier, AARS substrate diversity was further expanded by a mutation in the binding pocket domain and editing domain to accept previously unaccepted amino acids. For example, in PheRS, a specific mutation at binding pocket domain Ala294Gly can accept the phenylalanine analog p-iodo-Phe. Similarly, inactivating the editing domain of LeuRS by Asp345Ala accepted allylglycine (Josephson, Hartman, and Szostak, 2005). Furthermore, studies have systematically explored the diversity of NC-AAs as substrates for incorporation in native and engineered AARSs. It affirms the polyspecificity of AARS in substrate selection and expands substrate diversity for engineering proteins with novel functional groups (Hartman, Josephson, and Szostak, 2006; Fan et al., 2014).

Cell-free systems opened up the use of insoluble, non-canonical AARS for NC-AA incorporation. Pyrrolysyl synthetase (PylRS) is one such AARS which had limited use due to difficulties in purification. PylRS and its mutants are utilized in more than 100 unnatural amino acid incorporations (Wan, Tharp, and Liu, 2014) and are highly useful due to their minimal cross-reactivity with the *E. coli* system. The expression of PylRS/tRNA pair from *Methanosarcina mazei* in *E. coli* helped obtain a cell-free extract containing the insoluble PylRS. This cell-free extract was used to demonstrate the site-specific incorporation of two analogs of pyrrolysine into a reporter gene (Chemla et al., 2015). However site-specific incorporation requires a faster and more accurate aminoacylation reaction between NC-AA and tRNA, there is a need for a better strategy to identify mutants with better incorporation efficiency.

Identification of correct mutants of orthogonal AARS for NC-AA incorporation is largely benefited from methods such as multiplex automated genome engineering (MAGE) and phage-assisted continuous evolution (PACE). MAGE is an *in vivo* continuous platform for large-scale introduction of allelic replacement in chromosomal DNA achieved by inactivation of the mismatch repair system by the *mutS* gene and λ-Red recombinase systems containing exo, beta, and gam proteins. ssDNA or oligo nucleotides (90 bases) are delivered to cells using electroporation and are introduced at the lagging strand of the replication fork at target locations. Each MAGE cycle requires about 2–2.5 h and higher diversity is generated at multiple locations by increasing the number of MAGE cycles as required. The process facilitates accumulation of a large number of mutations and increases the possibility of producing a beneficial protein phenotype (Wang et al., 2009). PACE is an autonomous directed evolution approach, where a modified M13 bacteriophage contains the gene of interest and is subjected to continuous cycles of mutagenesis and selection. A large number of mutations are continuously generated and selected (Esvelt, Carlson, and Liu, 2011). These methods helped generate and screen a large library of AARS mutants in a short period, thereby saving time and effort. Despite the aforementioned methods being developed *in vivo*, orthogonal AARSs can also be used in *in vitro* systems for high efficiency NC-AA incorporation (Amiram et al., 2015; Bryson et al., 2017). Similarly, computational techniques have been used to generate and screen AARS mutant libraries specifically targeted for ortho-nitrobenzyl tyrosine, an analog of tyrosine. The advent of such methods will pave the way for better and faster engineering of novel AARSs (Baumann et al., 2019).

Furthermore, cell-free systems are advantageous to study non-canonical functions of AARS in addition to tRNA charging. For instance, it is well established that AARSs are also involved in mRNA binding and regulation in yeast (Levi and Arava, 2019), and AARSs can autoregulate their own expression by binding to specific DNA in bacteria (Putney and Schimmel, 1981). Cell-free

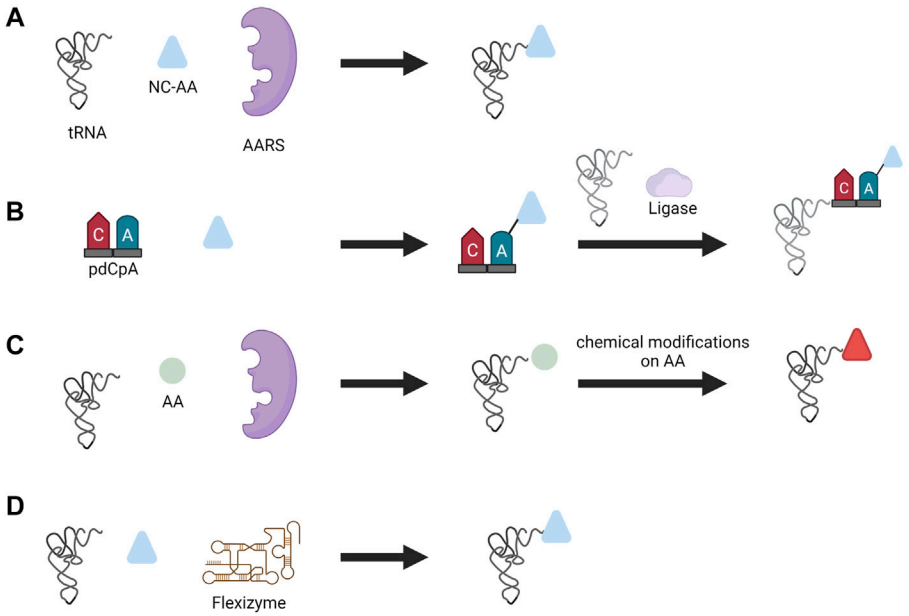


FIGURE 5 | Different aminoacylation methods used for NC-AA incorporation. **(A)** Enzymatic aminoacylation by AARS, **(B)** chemoenzymatic method, **(C)** chemical method, and **(D)** flexizyme method.

systems can be used to distinguish between canonical and non-canonical functions of AARSs. Readers interested in non-canonical functions of AARSs can refer to other comprehensive reviews (Ivanov, Moor and Lavrik, 2000; Guo and Schimmel, 2013).

4.2.2.1 Aminoacylation Methods

AARS enzymatic reactions are not the only way to charge tRNAs with a NC-AAs. Other methods to charge tRNAs include the chemoenzymatic method, chemical method, and ribozyme method (**Figure 5**). In the chemoenzymatic method, the first step is the chemical acylation of hybrid dinucleotide 5′-phospho-2′-deoxyribocytidylylriboadenosine (pdCpA) using an activated amino acid donor with an N-protected group. In the following step, the acylated dinucleotide is enzymatically ligated via T4 ligase to truncated tRNA lacking 3′-CA dinucleotide (Hecht et al., 1978). Many functional groups can be ligated to tRNA with this approach, but it is cumbersome due to the laborious chemical process involved in the preparation of acylated nucleotides.

In the chemical method, the side chains of the amino acids are chemically modified after tRNAs were charged with canonical amino acids by the AARS. One can consider this method similar to post-translational modification of proteins; instead, here it is the charged tRNA that is modified by a chemical reaction. Diverse functional groups such as N-methyl amino acid (Merryman and Green, 2004), glycosyl amino acid (Fahmi et al., 2007), and fluorescently-labeled amino acids (Iijima and Hohsaka, 2009) were generated with this approach. However, similar to the chemoenzymatic method, this method is laborious, technically demanding, and suffers from poor efficiency due to cyclic-tRNA by-product which inhibits protein synthesis, and also the short lifetime of aminoacylated tRNA (Yamanaka et al., 2004).

Another approach for incorporating NC-AA that avoids the use of AARS altogether is performed by ribozymes called flexizymes pioneered by Prof. Hiroaki Suga. A flexizyme is a small artificial ribozyme (44–46 nucleotides) capable of generating NC-AA tRNAs (Murakami et al., 2006). Flexizyme is derived from acyl-transferase ribozyme through directed evolution and sequence optimization. Flexizyme specifically recognizes amino acids with an activated carboxyl group and charge the 3′-CCA end of tRNA irrespective of tRNA body and anticodon sequence (Xiao et al., 2008). Flexizyme classes were expanded lately such that they can accept amino acids with different active groups. The complete set of flexizymes includes dinitro flexizyme (dFx) recognizing amino acid with activated dinitrobenzyl ester, enhanced flexizyme (eFx) recognizing amino acid with chlorobenzyl ester, and amino flexizyme (aFx) recognizing amino acid with the amino derivatized benzyl thioester group (Morimoto et al., 2011). Flexizymes can accept almost all amino acids as acyl-donor substrates and expand the diversity of amino acids that can be incorporated. For instance, flexizyme-mediated incorporation of NC-AA with D-α-amino acids (Katoh, Tajima, and Suga, 2017), β-amino acids (Katoh and Suga, 2018), γ-amino acids (Ohshiro et al., 2011), N-alkyl-L- α-amino acids (Kawakami, Ishizawa, and Murakami, 2013), benzoic acids (Kawakami et al., 2016), and exotic peptides (Goto and Suga, 2009) to highlight a few. Aminoacylation is performed by incubating the desired tRNA and amino acid together to yield charged tRNA (Goto, Katoh, and Suga, 2011). The only limitation that exists from the amino acid is the ability to activate the carboxyl group with a certain group and the chemical stability of amino acids during aminoacylation. This technique has opened up the possibility to charge any tRNA with nearly any amino acid and thereby reassigning any codon with

NC-AAs. The *in vitro* translation system based on flexizyme and translation apparatus is called the flexible *in vitro* translation system (FIT) and has been used for genetic code reprogramming (Goto, Katoh, and Suga, 2011).

5 CHALLENGES AND FUTURE DIRECTIONS

Protein synthesis is a highly complex process with many players involved that ensure fidelity. The process of translation has evolved over billions of years to attain the current state of efficiency and quality control. Attempts to modify, alter, or improve the efficiency of translation to suit a novel application require efforts at multiple levels *in vivo* to reach the desired function. CFPS systems eliminate some of the stringency associated with *in vivo* translation and simplified engineering of the protein synthesis process to some extent. For genetic code expansion, CFPS has opened new avenues for incorporating a diverse range of NC-AAs and helped achieve proteins with novel functional groups. The development of an orthogonal translation system requires a coordinated effect at every level of protein expression. Apart from tRNAs and AARSs engineering, engineering efforts on other translational elements such as elongation factors (Fan, Ip, and Söll, 2016; Gan et al., 2017), initiation factors (Goto et al., 2008; Goto and Suga, 2009), ribosomes (Semrad and Green, 2002; Jewett et al., 2013; Li et al., 2017a), and termination factors (Hong et al., 2015; Korkmaz and Sanyal, 2017) helped in the development of a better translational system. One key challenge lies in bringing together these individually engineered components. CFPS research will continue to increase our understanding and the capability to engineer the improved translational system. Failures will reveal the gaps in our understanding and guide our scientific research. Despite ongoing effects, many areas require advancements. For instance, the efficiency of NC-AA is low due to the non-compatibility of structurally diverse NC-AA with ribosomes and elongation factors (Fujino et al., 2013). This requires better engineered translation factors to successfully incorporate NC-AA with high efficiency. The advent of high-throughput methods and computational techniques like MAGE and PACE quickened the pace of library generation and screening, leading to faster protein engineering. GCE has increased amino acid repertoires with diverse functional and structural groups and has resulted in the creation of novel proteins that could not be synthesized before. The use of such novel proteins could be in creating better drugs, and functional biomaterials. Improvements and better characterization of translation elements, such as AARS and tRNA, using the latest techniques will help to standardize translation elements and aid in creating a predictable biological system. From a synthetic biology perspective, engineering life from scratch remains a grand challenge in the field. Toward that goal, different aspects of living systems have been reconstituted *in vitro* including ATP synthesis (Berhanu, Ueda, and Kuruma, 2019), DNA replication (Kurihara et al., 2011), PURE protein component self-regeneration (Lavickova, Laohakunakorn, and Maerkl, 2020), and ribosomal components (Jewett et al., 2013). With respect to the PURE system, the efficiency is improved by addition of external components into the system such as EF-P, EF4, and ArfA (Li et al., 2017b). Cell-free transcription and translation systems have enormous potential to overcome the limits of cell-based protein synthesis and could become the next generation platform for protein engineering that can go well beyond the scope of what could be accomplished in a cellular environment.

AUTHOR CONTRIBUTIONS

RBG and SJM wrote this review.

FUNDING

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program Grant 723106 and a Swiss National Science Foundation grant (182019). Open access funding provided by École Polytechnique Fédérale de Lausanne.

REFERENCES

Airas, R. K. (2006). Analysis of the Kinetic Mechanism of Arginyl-tRNA Synthetase. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1764 (2), 307–319. doi:10.1016/j.bbapap.2005.11.020

Amiram, M., Haimovich, A. D., Fan, C., Wang, Y.-S., Aerni, H.-R., Ntai, I., et al. (2015). Evolution of Translation Machinery in Recoded Bacteria Enables Multi-Site Incorporation of Nonstandard Amino Acids. *Nat. Biotechnol.* 33 (12), 1272–1279. doi:10.1038/nbt.3372

Asahara, H., Himeno, H., Tamura, K., Hasegawa, T., Watanabe, K., and Shimizu, M. (1993). Recognition Nucleotides of *Escherichia coli* tRNALeu and its Elements Facilitating Discrimination from tRNASer and tRNATyr. *J. Mol. Biol.* 231 (2), 219–229. doi:10.1006/jmbi.1993.1277

Asahara, H., Himeno, H., Tamura, K., Nameki, N., Hasegawa, T., and Shimizu, M. (1994). *Escherichia coli* Seryl-tRNA Synthetase Recognizes tRNASer by its Characteristics Tertiary Structure. *J. Mol. Biol.* 236 (3), 738–748. doi:10.1006/jmbi.1994.1186

Augustine, J., and Francklyn, C. (1997). Design of an Active Fragment of a Class II Aminoacyl-tRNA Synthetase and its Significance for Synthetase Evolution. *Biochemistry* 36 (12), 3473–3482. doi:10.1021/bi962395y

Avery, O. T., Macleod, C. M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *J. Exp. Med.* 79 (2), 137–158. doi:10.1084/jem.79.2.137

Awai, T., Ichihashi, N., and Yomo, T. (2015). Activities of 20 Aminoacyl-tRNA Synthetases Expressed in a Reconstituted Translation System in *Escherichia coli*. *Biochem. Biophysics Rep.* 3, 140–143. doi:10.1016/j.bbrep.2015.08.006

Baumann, T., Hauf, M., Richter, F., Albers, S., Möglich, A., Ignatova, Z., et al. (2019). Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine. *Ijms* 20 (9), 2343. doi:10.3390/ijms20092343

Bedouelle, H. (1990). Recognition of tRNATyr by Tyrosyl-tRNA Synthetase. *Biochimie* 72 (8), 589–598. doi:10.1016/0300-9084(90)90122-W

Berhanu, S., Ueda, T., and Kuruma, Y. (2019). Artificial Photosynthetic Cell Producing Energy for Protein Synthesis. *Nat. Commun.* 10 (1), 1325. doi:10.1038/s41467-019-09147-4

Beuning, P. J., and Musier-Forsyth, K. (2001). Species-specific Differences in Amino Acid Editing by Class II Prolyl-tRNA Synthetase. *J. Biol. Chem.* 276 (33), 30779–30785. doi:10.1074/jbc.M104761200

Borel, F., Vincent, C., Leberman, R., and Härtlein, M. (1994). Seryl-tRNA Synthetase fromEscherichia Coli: Implication of its N-Terminal Domain in Aminoacylation Activity and Specificity. *Nucl. Acids Res.* 22 (15), 2963–2969. doi:10.1093/nar/22.15.2963

Bryson, D. I., Fan, C., Guo, L.-T., Miller, C., Söll, D., and Liu, D. R. (2017). Continuous Directed Evolution of Aminoacyl-tRNA Synthetases. *Nat. Chem. Biol.* 13 (12), 1253–1260. doi:10.1038/nchembio.2474

Buchner, E. (1897). Alkoholische Gärung ohne Hefezellen. *Ber. Dtsch. Chem. Ges.* 30 (1), 1110–1113. doi:10.1002/cber.189703001215

Cader, M. Z., Ren, J., James, P. A., Bird, L. E., Talbot, K., and Stammers, D. K. (2007). Crystal Structure of Human Wildtype and S581L-Mutant Glycyl-tRNA Synthetase, an Enzyme Underlying Distal Spinal Muscular Atrophy. *FEBS Lett.* 581 (16), 2959–2964. doi:10.1016/j.febslet.2007.05.046

Cai, Q., Hanson, J. A., Steiner, A. R., Tran, C., Masikat, M. R., Chen, R., et al. (2015). A Simplified and Robust Protocol for Immunoglobulin Expression in E Scherichia Coli Cell-free Protein Synthesis Systems. *Biotechnol. Prog.* 31 (3), 823–831. doi:10.1002/btpr.2082

Campanacci, V., Dubois, D. Y., Becker, H. D., Kern, D., Spinelli, S., Valencia, C., et al. (2004). The *Escherichia coli* YadB Gene Product Reveals a Novel Aminoacyl-tRNA Synthetase like Activity. *J. Mol. Biol.* 337 (2), 273–283. doi:10.1016/j.jmb.2004.01.027

Chan, K. W., and Koeppe, R. E. (1994). Role of the TIGN Sequence in *E. coli* Tryptophanyl-tRNA Synthetase. *Biochimica Biophysica Acta (BBA) - Protein Struct. Mol. Enzym.* 1205 (2), 223–229. doi:10.1016/0167-4838(94)90237-2

Chargaff, E. (1950). Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation. *Experientia* 6 (6), 201–209. doi:10.1007/BF02173653

Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V., and Alfonta, L. (2015). Genetically Expanded Cell-free Protein Synthesis Using Endogenous Pyrrolysyl Orthogonal Translation System. *Biotechnol. Bioeng.* 112 (8), 1663–1672. doi:10.1002/bit.25587

Chen, J.-F., Guo, N.-N., Li, T., Wang, E.-D., and Wang, Y.-L. (2000). CP1 Domain in *Escherichia coli* Leucyl-tRNA Synthetase Is Crucial for its Editing Function. *Biochemistry* 39 (22), 6726–6731. doi:10.1021/bi000108r

Chu, W. C., and Horowitz, J. (1991). Recognition of *Escherichia coli* Valine Transfer RNA by its Cognate Synthetase: a Fluorine-19 NMR Study. *Biochemistry* 30 (6), 1655–1663. doi:10.1021/bi00220a031

Crick, F. (1955). *.'On Degenerate Templates and the Adaptor Hypothesis': RNA Tie Club*. Cambridge, United Kingdom: The Francis Crick Papers. Available at: https://wellcomecollection.org/works/j7qpsmdm.

Cui, Z., Stein, V., Tnimov, Z., Mureev, S., and Alexandrov, K. (2015). Semisynthetic tRNA Complement Mediates *In Vitro* Protein Synthesis. *J. Am. Chem. Soc.* 137 (13), 4404–4413. doi:10.1021/ja5131963

Doctor, B. P., and Mudd, J. A. (1963). Species Specificity of Amino Acid Acceptor Ribonucleic Acid and Aminoacyl Soluble Ribonucleic Acid Synthetases. *J. Biol. Chem.* 238 (11), 3677–3681. doi:10.1016/s0021-9258(19)75325-3

Doerr, A., Foschepoth, D., Forster, A. C., and Danelon, C. (2021). *In Vitro* synthesis of 32 Translation-Factor Proteins from a Single Template Reveals Impaired Ribosomal Processivity. *Sci. Rep.* 11 (1), 1–12. doi:10.1038/s41598-020-80827-8

Dondapati, S. K., Stech, M., Zemella, A., and Kubick, S. (2020). Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* 34 (3), 327–348. doi:10.1007/s40259-020-00417-y

Dopp, J. L., Jo, Y. R., and Reuel, N. F. (2019). Methods to Reduce Variability in E. Coli-Based Cell-free Protein Expression Experiments. *Synthetic Syst. Biotechnol.* 4 (4), 204–211. doi:10.1016/j.synbio.2019.10.003

Ebel, J. P., Giegé, R., Bonnet, J., Kern, D., Befort, N., Bollack, C., et al. (1973). Factors Determining the Specificity of the tRNA Aminoacylation Reaction. *Biochimie* 55 (5), 547–557. doi:10.1016/S0300-9084(73)80415-8

Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A System for the Continuous Directed Evolution of Biomolecules. *Nature* 472 (7344), 499–503. doi:10.1038/nature09929

Fahmi, N. E., Dedkova, L., Wang, B., Golovine, S., and Hecht, S. M. (2007). Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins. *J. Am. Chem. Soc.* 129 (12), 3586–3597. doi:10.1021/ja067466n

Fan, C., Ho, J. M. L., Chirathivat, N., Söll, D., and Wang, Y.-S. (2014). Exploring the Substrate Range of Wild-type Aminoacyl-tRNA Synthetases. *ChemBioChem* 15 (12), 1805–1809. doi:10.1002/cbic.201402083

Fan, C., Ip, K., and Söll, D. (2016). Expanding the Genetic Code of *Escherichia coli* with Phosphotyrosine. *FEBS Lett.* 590, 3040–3047. doi:10.1002/1873-3468.12333

Fauzi, H., Jack, K. D., and Hines, J. V. (2005). *In Vitro* selection to Identify Determinants in tRNA for Bacillus Subtilis tyrS T Box Antiterminator mRNA Binding. *Nucleic Acids Res.* 33 (8), 2595–2602. doi:10.1093/nar/gki546

Fersht, A. R. (1977). Editing Mechanisms in Protein Synthesis. Rejection of Valine by the Isoleucyl-tRNA Synthetase. *Biochemistry* 16 (5), 1025–1030. doi:10.1021/bi00624a034

Francklyn, C., and Schimmel, P. (1990). Enzymatic Aminoacylation of an Eight-Base-Pair Microhelix with Histidine. *Proc. Natl. Acad. Sci. U.S.A.* 87 (21), 8655–8659. doi:10.1073/pnas.87.21.8655

Francklyn, F., Shi, S., and Schimmel, S. (1992). Overlapping Nucleotide Determinants for Specific Aminoacylation of RNA Microhelices. *Science* 255 (5048), 1121–1125. doi:10.1126/science.1546312

Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2013). Reevaluation of the D-Amino Acid Compatibility with the Elongation Event in Translation. *J. Am. Chem. Soc.* 135 (5), 1830–1837. doi:10.1021/ja309570x

Gagoski, D., Polinkovsky, M. E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y., et al. (2016). Performance Benchmarking of Four Cell-free Protein Expression Systems. *Biotechnol. Bioeng.* 113 (2), 292–300. doi:10.1002/bit.25814

Gan, R., Perez, J. G., Carlson, E. D., Ntai, I., Isaacs, F. J., Kelleher, N. L., et al. (2017). Translation System Engineering in *Escherichia coli* Enhances Non-canonical Amino Acid Incorporation into Proteins. *Biotechnol. Bioeng.* 114 (5), 1074–1086. doi:10.1002/bit.26239

Ghosh, G., Pelka, H., Schulman, L. H., and Brunie, S. (1991). Activation of Methionine by Escherichia Coli Methionyl-tRNA Synthetase. *Biochemistry* 30 (40), 9569–9575. doi:10.1021/bi00104a002

Goto, Y., and Suga, H. (2009). Translation Initiation with Initiator tRNA Charged with Exotic Peptides. *J. Am. Chem. Soc.* 131 (14), 5040–5041. doi:10.1021/ja900597d

Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008). Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem. Biol.* 3 (2), 120–129. doi:10.1021/cb700233t

Goto, Y., Katoh, T., and Suga, H. (2011). Flexizymes for Genetic Code Reprogramming. *Nat. Protoc.* 6 (6), 779–790. doi:10.1038/nprot.2011.331

Gregory, S. T., and Dahlberg, A. E. (1995). Effects of Mutations at Position 36 of tRNAGluon Missense and Nonsense Suppression inEscherichia Coli. *FEBS Lett.* 361 (1), 25–28. doi:10.1016/0014-5793(95)00132-S

Gubbens, J., Kim, S. J., Yang, Z., Johnson, A. E., and Skach, W. R. (2010). *In Vitro* incorporation of Nonnatural Amino Acids into Protein Using tRNACys-Derived Opal, Ochre, and Amber Suppressor tRNAs. *RNA* 16 (8), 1660–1672. doi:10.1261/rna.2024810

Guo, M., and Schimmel, P. (2013). Essential Nontranslational Functions of tRNA Synthetases. *Nat. Chem. Biol.* 9 (3), 145–153. doi:10.1038/nchembio.1158

Hamano-Takaku, F., Iwama, T., Saito-Yano, S., Takaku, K., Monden, Y., Kitabatake, M., et al. (2000). A Mutant *Escherichia coli* Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently Than Tyrosine. *J. Biol. Chem.* 275 (51), 40324–40328. doi:10.1074/jbc.M003696200

Hamann, C. S., and Hou, Y.-M. (1997). An RNA Structural Determinant for tRNA Recognition. *Biochemistry* 36 (26), 7967–7972. doi:10.1021/bi970517w

Hartman, M. C. T., Josephson, K., and Szostak, J. W. (2006). Enzymatic Aminoacylation of tRNA with Unnatural Amino Acids. *Proc. Natl. Acad. Sci. U.S.A.* 103 (12), 4356–4361. doi:10.1073/pnas.0509219103

Hasegawa, T., Himeno, H., Ishikura, H., and Shimizu, M. (1989). Discriminator Base of tRNAAsp Is Involved in Amino Acid Acceptor Activity. *Biochem. Biophysical Res. Commun.* 163 (3), 1534–1538. doi:10.1016/0006291X(89)91154-610.1016/0006-291x(89)91154-6

Hasegawa, T., Miyano, M., Himeno, H., Sano, Y., Kimura, K., and Shimizu, M. (1992). Identity Determinants of *E. coli* Threonine tRNA. *Biochem. biophysical Res. Commun.* 184 (1), 478–484. doi:10.1016/0006-291x(92)91219-g

Hayase, Y., Jahn, M., Rogers, M. J., Sylvers, L. A., Koizumi, M., Inoue, H., et al. (1992). Recognition of Bases in *Escherichia coli* tRNA(Gln) by Glutaminyl-

tRNA Synthetase: a Complete Identity Set. *EMBO J.* 11 (11), 4159–4165. doi:10.1002/j.14602075.1992.tb05509.x10.1002/j.1460-2075.1992.tb05509.x

Hecht, S. M., Alford, B. L., Kuroda, Y., and Kitano, S. (1978). "Chemical Aminoacylation" of tRNA's. *J. Biol. Chem.* 253 (13), 4517–4520. doi:10.1016/s0021-9258(17)30417-9

Hentzen, D., Mandel, P., and Garel, J.-P. (1972). Relation between Aminoacyl-tRNA Stability and the Fixed Amino Acid. *Biochimica Biophysica Acta (BBA) - Nucleic Acids Protein Synthesis* 281 (2), 228–232. doi:10.1016/0005-2787(72)90174-8

Hershey, A. D., and Chase, M. (1952). Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *J. general physiology* 36 (1), 39–56. doi:10.1085/jgp.36.1.39

Hibi, K., Amikura, K., Sugiura, N., Masuda, K., Ohno, S., Yokogawa, T., et al. (2020). Reconstituted Cell-free Protein Synthesis Using *In Vitro* Transcribed tRNAs. *Commun. Biol.* 3 (1), 1–11. doi:10.1038/s42003-020-1074-2

Hill, K., and Schimmel, P. (1989). Evidence that the 3'-end of a Transfer RNA Binds to a Site in the Adenylate Synthesis Domain of an Aminoacyl-tRNA Synthetase. *Biochemistry* 28 (6), 2577–2586. doi:10.1021/bi00432a035

Himeno, H., Hasegawa, T., Asahara, H., Tamura, K., and Shimizu, M. (1991). Identity Determinants ofE.Colitryptophan tRNA. *Nucl. Acids Res.* 19 (23), 6379–6382. doi:10.1093/nar/19.23.6379

Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K., Miura, K.-i., and Shimizu, M. (1989). Role of the Extra G-C Pair at the End of the Acceptor Stem of tRNAHbin Aminoacylation. *Nucl. Acids Res.* 17 (19), 7855–7863. doi:10.1093/nar/17.19.7855

Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K., and Shimizu, M. (1990). Conversion of Aminoacylation Specificity from tRNATyrto tRNASerin Vitro. *Nucl. Acids Res.* 18 (23), 6815–6819. doi:10.1093/nar/18.23.6815

Hirsh, D. I. (1968). A Study of the Threonyl Adenylate Complex with Threonyl Transfer Ribonucleic Acid Synthetase and its Reaction with Hydroxylamine. *J. Biol. Chem.* 243 (21), 5731–5738. doi:10.1016/s0021-9258(18)91926-5

Hoagland, M. B., Keller, E. B., and Zamecnik, P. C. (1956). Enzymatic Carboxyl Activation of Amino Acids. *J. Biol. Chem.* 218 (1), 345–358. doi:10.1016/S0021-9258(18)65898-3

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., and Zamecnik, P. C. (1958). A Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *J. Biol. Chem.* 231 (1), 241–257. doi:10.1016/s0021-9258(19)77302-5

Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H., and Sisido, M. (2001). Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an *Escherichia coli In Vitro* Translation System. *Biochemistry* 40 (37), 11060–11064. doi:10.1021/bi0108204

Hong, S. H., Ntai, I., Haimovich, A. D., Kelleher, N. L., Isaacs, F. J., and Jewett, M. C. (2014). Cell-free Protein Synthesis from a Release Factor 1 Deficient *Escherichia coli* Activates Efficient and Multiple Site-specific Nonstandard Amino Acid Incorporation. *ACS Synth. Biol.* 3 (6), 398–409. doi:10.1021/sb400140t

Hong, S. H., Kwon, Y.-C., Martin, R. W., Des Soye, B. J., de Paz, A. M., Swonger, K. N., et al. (2015). Improving Cell-free Protein Synthesis through Genome Engineering ofEscherichia coliLacking Release Factor 1. *ChemBioChem* 16 (5), 844–853. doi:10.1002/cbic.201402708

Hoshika, S., Leal, N. A., Kim, M.-J., Kim, M.-S., Karalkar, N. B., Kim, H.-J., et al. (2019). Hachimoji DNA and RNA: A Genetic System with Eight Building Blocks. *Science* 363 (6429), 884–887. doi:10.1126/science.aat0971

Hou, Y.-M., and Schimmel, P. (1988). A Simple Structural Feature Is a Major Determinant of the Identity of a Transfer RNA. *Nature* 333 (6169), 140–145. doi:10.1038/333140a0

Hou, Y. M., and Schimmel, P. (1989). Modeling with In Vitro Kinetic Parameters for the Elaboration of Transfer RNA Identity In Vivo. *Biochemistry* 28 (12), 4942–4947. doi:10.1021/bi00438a005

Hou, Y. M., Westhof, E., and Giegé, R. (1993). An Unusual RNA Tertiary Interaction Has a Role for the Specific Aminoacylation of a Transfer RNA. *Proc. Natl. Acad. Sci. U.S.A.* 90 (14), 6776–6780. doi:10.1073/pnas.90.14.6776

Hunter, D. J. B., Bhumkar, A., Giles, N., Sierecki, E., and Gambin, Y. (2018). Unexpected Instabilities Explain Batch-to-batch Variability in Cell-free Protein Expression Systems. *Biotechnol. Bioeng.* 115 (8), 1904–1914. doi:10.1002/bit.26604

Ibba, M., Hong, K. W., Sherman, J. M., Sever, S., and Söll, D. (1996). Interactions between tRNA Identity Nucleotides and Their Recognition Sites in Glutaminyl-tRNA Synthetase Determine the Cognate Amino Acid Affinity of the Enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 93 (14), 6953–6958. doi:10.1073/pnas.93.14.6953

Iijima, I., and Hohsaka, T. (2009). Position-specific Incorporation of Fluorescent Non-natural Amino Acids into Maltose-Binding Protein for Detection of Ligand Binding by FRET and Fluorescence Quenching. *ChemBioChem* 10 (6), 999–1006. doi:10.1002/cbic.200800703

Ivanov, K. A., Lavrik, O. I., and Lavrik, O. I. (2000). Non-canonical Functions of Aminoacyl-tRNA Synthetases. *Biochem. (Mosc)* 65 (8), 888–897. doi:10.1134/S0006297912010026

Iwane, Y., Hitomi, A., Murakami, H., Katoh, T., Goto, Y., and Suga, H. (2016). Expanding the Amino Acid Repertoire of Ribosomal Polypeptide Synthesis via the Artificial Division of Codon Boxes. *Nat. Chem.* 8 (4), 317–325. doi:10.1038/nchem.2446

Jacob, F., and Monod, J. (1961). Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J. Mol. Biol.* 3 (3), 318–356. doi:10.1016/S0022-2836(61)80072-7

Jahn, M., Rogers, M. J., and Söll, D. (1991). Anticodon and Acceptor Stem Nucleotides in tRNAGln Are Major Recognition Elements for *E. coli* Glutaminyl-tRNA Synthetase. *Nature* 352 (6332), 258–260. doi:10.1038/352258a0

Jewett, M. C., Fritz, B. R., Timmerman, L. E., and Church, G. M. (2013). *In Vitro* integration of Ribosomal RNA Synthesis, Ribosome Assembly, and Translation. *Mol. Syst. Biol.* 9, 678. doi:10.1038/msb.2013.31

Josephson, K., Hartman, M. C. T., and Szostak, J. W. (2005). Ribosomal Synthesis of Unnatural Peptides. *J. Am. Chem. Soc.* 127 (33), 11727–11735. doi:10.1021/ja0515809

Katoh, T., and Suga, H. (2018). Ribosomal Incorporation of Consecutive β-Amino Acids. *J. Am. Chem. Soc.* 140 (38), 12159–12167. doi:10.1021/jacs.8b07247

Katoh, T., Tajima, K., and Suga, H. (2017). Consecutive Elongation of D-Amino Acids in Translation. *Cell Chem. Biol.* 24 (1), 46–54. doi:10.1016/j.chembiol.2016.11.012

Kawakami, T., Ishizawa, T., and Murakami, H. (2013). Extensive Reprogramming of the Genetic Code for Genetically Encoded Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics. *J. Am. Chem. Soc.* 135 (33), 12297–12304. doi:10.1021/ja405044k

Kawakami, T., Ogawa, K., Hatta, T., Goshima, N., and Natsume, T. (2016). Directed Evolution of a Cyclized Peptoid-Peptide Chimera against a Cell-free Expressed Protein and Proteomic Profiling of the Interacting Proteins to Create a Protein-Protein Interaction Inhibitor. *ACS Chem. Biol.* 11 (6), 1569–1577. doi:10.1021/acschembio.5b01014

Komatsoulis, G. A., and Abelson, J. (1993). Recognition of tRNACys by *Escherichia coli* Cysteinyl-tRNA Synthetase. *Biochemistry* 32 (29), 7435–7444. doi:10.1021/bi00080a014

Korencic, D., Söll, D., and Ambrogelly, A. (2002). A One-step Method for *In Vitro* Production of tRNA Transcripts. *Nucleic acids Res.* 30 (20), 105e–105. doi:10.1093/nar/gnf104

Korkmaz, G., and Sanyal, S. (2017). R213I Mutation in Release Factor 2 (RF2) Is One Step Forward for Engineering an Omnipotent Release Factor in Bacteria *Escherichia coli*. *J. Biol. Chem.* 292 (36), 15134–15142. doi:10.1074/jbc.M117.785238

Krüger, A., Mueller, A. P., Rybnicky, G. A., Engle, N. L., Yang, Z. K., Tschaplinski, T. J., et al. (2020). Development of a Clostridia-Based Cell-free System for Prototyping Genetic Parts and Metabolic Pathways. *Metab. Eng.* 62, 95–105. doi:10.1016/j.ymben.2020.06.004

Kurihara, K., Tamura, M., Shohda, K.-i., Toyota, T., Suzuki, K., and Sugawara, T. (2011). Self-reproduction of Supramolecular Giant Vesicles Combined with the Amplification of Encapsulated DNA. *Nat. Chem.* 3 (10), 775–781. doi:10.1038/nchem.1127

Laohakunakorn, N., Grasemann, L, Lavickova, B, Michielin, G, Shahein, A, Swank, Z, et al. (2020) 'Bottom-Up Construction of Complex Biomolecular Systems With Cell-Free Synthetic Biology', *Frontiers in Bioengineering and Biotechnology*. 8, 213. doi:10.3389/fbioe.2020.00213

Lapointe, J., and Söll, D. (1972). Glutamyl Transfer Ribonucleic Acid Synthetase of *Escherichia coli*. *J. Biol. Chem.* 247 (16), 4982–4985. doi:10.1016/s0021-9258(19)44927-2

Lavickova, B., and Maerkl, S. J. (2019). A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-free System. *ACS Synth. Biol.* 8, 455–462. doi:10.1021/acssynbio.8b00427

Lavickova, B., Laohakunakorn, N., and Maerkl, S. J. (2020). A Partially Self-Regenerating Synthetic Cell. *Nat. Commun.* 11 (1), 1–11. doi:10.1038/s41467-020-20180-6

Lee, C. P., Dyson, M. R., Mandal, N., Varshney, U., Bahramian, B., and RajBhandary, U. L. (1992). Striking Effects of Coupling Mutations in the Acceptor Stem on Recognition of tRNAs by *Escherichia coli* Met-tRNA Synthetase and Met-tRNA Transformylase. *Proc. Natl. Acad. Sci. U.S.A.* 89 (19), 9262–9266. doi:10.1073/pnas.89.19.9262

Levene, P. A. (1919). The Structure of Yeast Nucleic Acid. *J. Biol. Chem.* 40 (2), 415–424. doi:10.1016/S0021-9258(18)87254-4

Levi, O., and Arava, Y. (2019). mRNA Association by Aminoacyl tRNA Synthetase Occurs at a Putative Anticodon Mimic and Autoregulates Translation in Response to tRNA Levels. *PLoS Biol.* 17 (5), e3000274. doi:10.1371/journal.pbio.3000274

Li, J., Gu, L., Aach, J., and Church, G. M. (2014). Improved Cell-free RNA and Protein Synthesis System. *PLoS ONE* 9 (9), e106232. doi:10.1371/journal.pone.0106232

Li, J., Haas, W., Jackson, K., Kuru, E., Jewett, M. C., Fan, Z. H., et al. (2017a). Cogenerating Synthetic Parts toward a Self-Replicating System. *ACS Synth. Biol.* 6 (7), 1327–1336. doi:10.1021/acssynbio.6b00342

Li, J., Zhang, C., Huang, P., Kuru, E., Forster-Benson, E. T. C., Li, T., et al. (2017b). Dissecting Limiting Factors of the Protein Synthesis Using Recombinant Elements (PURE) System. *Translation* 5 (1), e1327006. doi:10.1080/21690731.2017.1327006

Li, S., Pelka, H., and Schulman, L. H. (1993). The Anticodon and Discriminator Base Are Important for Aminoacylation of *Escherichia coli* tRNA(Asn). *J. Biol. Chem.* 268 (24), 18335–18339. doi:10.1016/S0021-9258(17)46849-9

Lin, S. X., Wang, Q., and Wang, Y. L. (1988). Interactions between *Escherichia coli* Arginyl-tRNA Synthetase and its Substrates. *Biochemistry* 27 (17), 6348–6353. doi:10.1021/bi00417a023

Liu, H., Peterson, R., Kessler, J., and Musier-Forsyth, K. (1995). Molecular Recognition of tRNAProbyEscherichia Coliproline tRNA Synthetasein Vitro. *Nucl. Acids Res.* 23 (1), 165–169. doi:10.1093/nar/23.1.165

Liu, J., Ibba, M., Hong, K.-W., and Söll, D. (1998). The Terminal Adenosine of tRNAGln Mediates tRNA-Dependent Amino Acid Recognition by Glutaminyl-tRNA Synthetase. *Biochemistry* 37 (27), 9836–9842. doi:10.1021/bi980704+

Loftfield, R. B., and Vanderjagt, D. (1972). The Frequency of Errors in Protein Biosynthesis. *Biochem. J.* 128 (5), 1353–1356. doi:10.1042/bj1281353

Maddalena, L. L. d., Niederholtmeyer, H., Turtola, M., Swank, Z. N., Belogurov, G. A., and Maerkl, S. J. (2016). GreA and GreB Enhance Expression of *Escherichia coli* RNA Polymerase Promoters in a Reconstituted Transcription-Translation System. *ACS Synth. Biol.* 5 (9), 929–935. doi:10.1021/acssynbio.6b00017

Madern, D., Anselme, J., and Härtlein, M. (1992). Asparaginyl-tRNA Synthetase from theEscherichia Colitemperature-Sensitive Strain HO202 A Proline Replacement in Motif 2 Is Responsible for a Large Increase inKmfor Asparagine and ATP. *FEBS Lett.* 299 (1), 85–89. doi:10.1016/0014-5793(92)80106-Q

Martin, F., Sharples, G. J., Lloyd, R. G., Eiler, S., Moras, D., Gangloff, J., et al. (1997). Characterization of a Thermosensitive *Escherichia coli* Aspartyl-tRNA Synthetase Mutant. *J. Bacteriol.* 179 (11), 3691–3696. doi:10.1128/jb.179.11.3691-3696.1997

Martin, R. W., Des Soye, B. J., Kwon, Y.-C., Kay, J., Davis, R. G., Thomas, P. M., et al. (2018). Cell-free Protein Synthesis from Genomically Recoded Bacteria Enables Multisite Incorporation of Noncanonical Amino Acids. *Nat. Commun.* 9 (1), 1203. doi:10.1038/s41467-018-03469-5

McClain, W. H. (1993). Identity of *Escherichia coli* tRNA(Cys) Determined by Nucleotides in Three Regions of tRNA Tertiary Structure. *J. Biol. Chem.* 268 (26), 19398–19402. doi:10.1016/S00219258(19)36528-710.1016/s0021-9258(19)36528-7

McClain, W. H., and Foss, K. (1988a). Changing the Identity of a tRNA by Introducing a G-U Wobble Pair Near the 3′ Acceptor End. *Science* 240 (4853), 793–796. doi:10.1126/science.2452483

McClain, W. H., and Foss, K. (1988b). Changing the Acceptor Identity of a Transfer RNA by Altering Nucleotides in a "Variable Pocket". *Science* 241 (4874), 1804–1807. doi:10.1126/science.2459773

McClain, W. H., and Foss, K. (1988c). Nucleotides that Contribute to the Identity of *Escherichia coli* tRNAPhe. *J. Mol. Biol.* 202 (4), 697–709. doi:10.1016/0022-2836(88)90551-7

McClain, W. H., Foss, K., Jenkins, R. A., and Schneider, J. (1990). Nucleotides that Determine *Escherichia coli* tRNA(Arg) and tRNA(Lys) Acceptor Identities Revealed by Analyses of Mutant Opal and Amber Suppressor tRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 87 (23), 9260–9264. doi:10.1073/pnas.87.23.9260

McClain, W. H., Foss, K., Jenkins, R. A., and Schneider, J. (1991). Rapid Determination of Nucleotides that Define tRNA(Gly) Acceptor Identity. *Proc. Natl. Acad. Sci. U.S.A.* 88 (14), 6147–6151. doi:10.1073/pnas.88.14.6147

McClain, W. H., Schneider, J., and Gabriel, K. (1994). Distinctive Acceptor-End Structure and Other Determinants ofEscherichia colitRNAProidentity. *Nucl. Acids Res.* 22 (3), 522–529. doi:10.1093/nar/22.3.522

Meinnel, T., Mechulam, Y., Lazennec, C., Blanquet, S., and Fayat, G. (1993). Critical Role of the Acceptor Stem of tRNAsMet in Their Aminoacylation by *Escherichia coli* Methionyl-tRNA Synthetase. *J. Mol. Biol.* 229 (1), 26–36. doi:10.1006/jmbi.1993.1005

Merryman, C., and Green, R. (2004). Transformation of Aminoacyl tRNAs for the *In Vitro* Selection of "Drug-like" Molecules. *Chem. Biol.* 11 (4), 575–582. doi:10.1016/j.chembiol.2004.03.009

Miescher, F. (1869). Letter I; to Wilhelm His; Tübingen, February 26th, 1869. *Histochem. Physiol. Arb. Friedrich Miescher-aus dem Wiss. Briefwechsel Miescher* 1, 33–38.

Moor, N., Klipcan, L., and Safro, M. G. (2011). Bacterial and Eukaryotic Phenylalanyl-tRNA Synthetases Catalyze Misaminoacylation of tRNAPhe with 3,4-Dihydroxy-L-Phenylalanine. *Chem. Biol.* 18 (10), 1221–1229. doi:10.1016/j.chembiol.2011.08.008

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., et al. (2018). Rapid Acquisition and Model-Based Analysis of Cell-free Transcription-Translation Reactions from Nonmodel Bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 115 (19), E4340 LP–E4349. doi:10.1073/pnas.1715806115

Morimoto, J., Hayashi, Y., Iwasaki, K., and Suga, H. (2011). Flexizymes: Their Evolutionary History and the Origin of Catalytic Function. *Acc. Chem. Res.* 44 (12), 1359–1368. doi:10.1021/ar2000953

Mukba, S. A., Vlasov, P. K., Kolosov, P. M., Shuvalova, E. Y., Egorova, T. V., and Alkalaeva, E. Z. (2020). Expanding the Genetic Code: Unnatural Base Pairs in Biological Systems. *Mol. Biol. Mosk.* 54 (4), 531–541. doi:10.31857/S0026898420040126

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A Highly Flexible tRNA Acylation Method for Non-natural Polypeptide Synthesis. *Nat. Methods* 3 (5), 357–359. doi:10.1038/nmeth877

Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T., et al. (1988). Codon and Amino-Acid Specificities of a Transfer RNA Are Both Converted by a Single Post-transcriptional Modification. *Nature* 336 (6195), 179–181. doi:10.1038/336179a0

Nameki, N., Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992). *Escherichia coli* tRNAAsp Recognition Mechanism Differing from that of the Yeast System. *Biochem. biophysical Res. Commun.* 189 (2), 856–862. doi:10.1016/0006-291X(92)92282-3

Niederholtmeyer, H., Stepanova, V., and Maerkl, S. J. (2013). Implementation of Cell-free Biological Networks at Steady State. *Proc. Natl. Acad. Sci. U.S.A.* 110 (40), 15985–15990. doi:10.1073/pnas.1311166110

Niederholtmeyer, H., Sun, Z. Z., Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M., et al. (2015). Rapid Cell-free Forward Engineering of Novel Genetic Ring Oscillators. *eLife* 4, e09771. doi:10.7554/eLife.09771

Nierlich, D. P., Lamfrom, H., Sarabhai, A., and Abelson, J. (1973). Transfer RNA Synthesis *In Vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 70 (1), 179–182. doi:10.1073/pnas.70.1.179

Nirenberg, M. W., and Matthaei, J. H. (1961). The Dependence of Cell-free Protein Synthesis in *E. coli* upon Naturally Occurring or Synthetic Polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 47 (10), 1588–1602. doi:10.1073/pnas.47.10.1588

Normanly, J., Kleina, L. G., Masson, J.-M., Abelson, J., and Miller, J. H. (1990). Construction of *Escherichia coli* Amber Suppressor tRNA Genes. *J. Mol. Biol.* 213 (4), 719–726. doi:10.1016/S0022-2836(05)80258-X

Normanly, J., Ogden, R. C., Horvath, S. J., and Abelson, J. (1986). Changing the Identity of a Transfer RNA. *Nature* 321 (6067), 213–219. doi:10.1038/321213a0

Normanly, J., Ollick, T., and Abelson, J. (1992). Eight Base Changes Are Sufficient to Convert a Leucine-Inserting tRNA into a Serine-Inserting tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 89 (12), 5680–5684. doi:10.1073/pnas.89.12.5680

Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., et al. (1994). Molecular Recognition of the Identity-Determinant Set of Isoleucine

Transfer RNA from *Escherichia coli*. *J. Mol. Biol.* 236 (3), 710–724. doi:10.1006/jmbi.1994.1184

Nureki, O., Niimi, T., Muto, Y., Kanno, H., Kohno, T., Muramatsu, T., et al. (1993). "Conformational Change of tRNA upon Interaction of the Identity-Determinant Set with Aminoacyl-tRNA Synthetase," in *The Translational Apparatus: Structure, Function, Regulation, Evolution*. Editor K. H. Nierhaus (Boston, MA: Springer), 59–66. doi:10.1007/978-1-4615-2407-6_6

Ohshiro, Y., Nakajima, E., Goto, Y., Fuse, S., Takahashi, T., Doi, T., et al. (2011). Ribosomal Synthesis of Backbone-Macrocyclic Peptides Containing γ-Amino Acids. *ChemBioChem* 12 (8), 1183–1187. doi:10.1002/cbic.201100104

Ostrem, D. L., and Berg, P. (1974). Glycyl Transfer Ribonucleic Acid Synthetase from *Escherichia coli*. Purification, Properties, and Substrate Binding. *Biochemistry* 13 (7), 1338–1348. doi:10.1021/bi00704a006

Pak, M., Pallanck, L., and Schulman, L. H. (1992). Conversion of a Methionine Initiator tRNA into a Tryptophan-Inserting Elongator tRNA In Vivo. *Biochemistry* 31 (13), 3303–3309. doi:10.1021/bi00128a001

Palade, G. E. (1955). A Small Particulate Component of the Cytoplasm. *J. biophysical Biochem. Cytol.* 1 (1), 59–68. doi:10.1083/jcb.1.1.59

Pallanck, L., Li, S., and Schulman, L. H. (1992). The Anticodon and Discriminator Base Are Major Determinants of Cysteine tRNA Identity In Vivo. *J. Biol. Chem.* 267 (11), 7221–7223. doi:10.1016/s0021-9258(18)42508-2

Pallanck, L., and Schulman, L. H. (1991). Anticodon-dependent Aminoacylation of a Noncognate tRNA with Isoleucine, Valine, and Phenylalanine In Vivo. *Proc. Natl. Acad. Sci. U.S.A.* 88 (9), 3872–3876. doi:10.1073/pnas.88.9.3872

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., et al. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* 165 (5), 1255–1266. doi:10.1016/j.cell.2016.04.059

Park, S. J., and Schimmel, P. (1988). Evidence for Interaction of an Aminoacyl Transfer RNA Synthetase with a Region Important for the Identity of its Cognate Transfer RNA. *J. Biol. Chem.* 263 (32), 16527–16530. doi:10.1016/s0021-9258(18)37421-0

Pauling, L. (1958). *The Probability of Errors in the Process of Synthesis of Protein Molecules: Festschrift Fur Prof. Dr. Arthur Stoll*. Basel, Switzerland: Birkhäuser Verlag, 597–602.

Peterson, E. T., Blank, J., Sprinzl, M., and Uhlenbeck, O. C. (1993). Selection for Active *E. coli* tRNA(Phe) Variants from a Randomized Library Using Two Proteins. *EMBO J.* 12 (7), 2959–2967. doi:10.1002/j.1460-2075.1993.tb05958.x

Peterson, E. T., and Uhlenbeck, O. C. (1992). Determination of Recognition Nucleotides for *Escherichia coli* Phenylalanyl-tRNA Synthetase. *Biochemistry* 31 (42), 10380–10389. doi:10.1021/bi00157a028

Putney, S. D., and Schimmel, P. (1981). An Aminoacyl tRNA Synthetase Binds to a Specific DNA Sequence and Regulates its Gene Transcription. *Nature* 291 (5817), 632–635. doi:10.1038/291632a0

Reynolds, N. M., Ling, J., Roy, H., Banerjee, R., Repasky, S. E., Hamel, P., et al. (2010). Cell-specific Differences in the Requirements for Translation Quality Control. *Proc. Natl. Acad. Sci. U.S.A.* 107 (9), 4063–4068. doi:10.1073/pnas.0909640107

Rogers, M. J., Adachi, T., Inokuchi, H., and Söll, D. (1992). Switching tRNA(Gln) Identity from Glutamine to Tryptophan. *Proc. Natl. Acad. Sci. U.S.A.* 89 (8), 3463–3467. doi:10.1073/pnas.89.8.3463

Rogers, M. J., and Söll, D. (1988). Discrimination between Glutaminyl-tRNA Synthetase and Seryl-tRNA Synthetase Involves Nucleotides in the Acceptor Helix of tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 85 (18), 6627–6631. doi:10.1073/pnas.85.18.6627

Rubio Gomez, M. A., and Ibba, M. (2020). Aminoacyl-tRNA Synthetases. *Rna* 26 (8), 910–936. doi:10.1261/rna.071720.119

Saks, M. E., and Sampson, J. R. (1996). Variant Minihelix RNAs Reveal Sequence-specific Recognition of the Helical tRNA(Ser) Acceptor Stem by E.Coli Seryl-tRNA Synthetase. *EMBO J.* 15 (11), 2843–2849. doi:10.1002/j.1460-2075.1996.tb00645.x

Sampson, J. R., and Saks, M. E. (1993). Contributions of Discrete tRNASerdomains to Aminoacylation byE.Coliseryl-tRNA Synthetase: a Kinetic Analysis Using Model RNA Substrates. *Nucl. Acids Res.* 21 (19), 4467–4475. doi:10.1093/nar/21.19.4467

Sankaranarayanan, R., Dock-Bregeon, A.-C., Rees, B., Bovee, M., Caillet, J., Romby, P., et al. (2000). Zinc Ion Mediated Amino Acid Discrimination by Threonyl-tRNA Synthetase. *Nat. Struct. Mol. Biol.* 7 (6), 461–465. doi:10.1038/75856

Schulman, L. H., and Pelka, H. (1988). Anticodon Switching Changes the Identity of Methionine and Valine Transfer RNAs. *Science* 242 (4879), 765–768. doi:10.1126/science.3055296

Schulman, L. H., and Pelka, H. (1989). The Anticodon Contains a Major Element of the Identity of Arginine Transfer RNAs. *Science* 246 (4937), 1595–1597. doi:10.1126/science.2688091

Schulman, L. H., and Pelka, H. (1990). An Anticodon Change Switches the Identity of E.Coli tRNA mMet from Methionine to Threonine. *Nucleic acids Res.* 18 (2), 285–289. doi:10.1093/nar/18.2.285

Sekine, S. I., Nureki, O., Sakamoto, K., Niimi, T., Tateno, M., Gō, M., et al. (1996). Major Identity Determinants in the "Augmented D Helix" of tRNAGlu from *Escherichia coli*. *J. Mol. Biol.* 256 (4), 685–700. doi:10.1006/jmbi.1996.0118

Semrad, K., and Green, R. (2002). Osmolytes Stimulate the Reconstitution of Functional 50S Ribosomes from *In Vitro* Transcripts of *Escherichia coli* 23S rRNA. *Rna* 8 (4), 401–411. doi:10.1017/s1355838202029722

Serebrov, V., Vassilenko, K., Kholod, N., Gross, H. J., and Kisselev, L. (1998). Mg2+ Binding and Structural Stability of Mature and *In Vitro* Synthesized Unmodified *Escherichia coli* tRNAPhe. *Nucleic Acids Res.* 26 (11), 2723–2728. doi:10.1093/nar/26.11.2723

Shepherd, J., and Ibba, M. (2015). Bacterial Transfer RNAs. *FEMS Microbiol. Rev.* 39 (3), 280–300. doi:10.1093/femsre/fuv004

Shepherd, T. R., Du, L., Liljeruhm, J., Samudyata, fnm., Wang, J., Sjödin, M. O. D., et al. (2017). De Novo design and Synthesis of a 30-cistron Translation-Factor Module. *Nucleic acids Res.* 45 (18), 10895–10905. doi:10.1093/nar/gkx753

Sherman, J. M., Rogers, K., Rogers, M. J., and Söll, D. (1992). Synthetase Competition and tRNA Context Determine the In Vivo Identity of tRNA Discriminator Mutants. *J. Mol. Biol.* 228 (4), 1055–1062. doi:10.1016/00222836(92)90314-A10.1016/0022-2836(92)90314-a

Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T., and Himeno, H. (1992). The Role of Anticodon Bases and the Discriminator Nucleotide in the Recognition of Some *E. coli* tRNAs by Their Aminoacyl-tRNA Synthetases. *J. Mol. Evol.* 35 (5), 436–443. doi:10.1007/BF00171822

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., et al. (2001). Cell-free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* 19 (8), 751–755. doi:10.1038/90802

Swank, Z., Laohakunakorn, N., and Maerkl, S. J. (2019). Cell-free Gene-Regulatory Network Engineering with Synthetic Transcription Factors. *Proc. Natl. Acad. Sci. U.S.A.* 116 (13), 5892–5901. doi:10.1073/pnas.1816591116

Swartz, J. (2006). Developing Cell-free Biology for Industrial Applications. *J. Ind. Microbiol. Biotechnol.* 33 (7), 476–485. doi:10.1007/s10295-006-0127-y

Sylvers, L. A., Rogers, K. C., Shimizu, M., Ohtsuka, E., and Soll, D. (1993). A 2-thiouridine Derivative in tRNAGlu Is a Positive Determinant for Aminoacylation by *Escherichia coli* Glutamyl-tRNA Synthetase. *Biochemistry* 32 (15), 3836–3841. doi:10.1021/bi00066a002

Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1991). Identity Determinants of *E. coli* tRNAVal. *Biochem. Biophysical Res. Commun.* 177 (2), 619–623. doi:10.1016/0006-291X(91)91833-X

Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992). In Vitro Study of E.coli tRNA(Arg) and tRNA(Lys) Identity Elements. *Nucl. Acids Res.* 20 (9), 2335–2339. doi:10.1093/nar/20.9.2335

Tardif, K. D., and Horowitz, J. (2004). Functional Group Recognition at the Aminoacylation and Editing Sites of *E. coli* Valyl-tRNA Synthetase. *Rna* 10 (3), 493–503. doi:10.1261/rna.5166704

Uemura, H., Imai, M., Ohtsuka, E., Ikehara, M., and Söll, D. (1982). E.coliinitiator tRNA Analogs with Different Nucleotides in the Discriminator Base Position. *Nucl. Acids Res.* 10 (20), 6531–6539. doi:10.1093/nar/10.20.6531

Valencia-Sánchez, M. I., Rodríguez-Hernández, A., Ferreira, R., Santamaría-Suárez, H. A., Arciniega, M., Dock-Bregeon, A.-C., et al. (2016). Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases. *J. Biol. Chem.* 291 (28), 14430–14446. doi:10.1074/jbc.M116.730382

Villarreal, F., Contreras-Llano, L. E., Chavez, M., Ding, Y., Fan, J., Pan, T., et al. (2018). Synthetic Microbial Consortia Enable Rapid Assembly of Pure Translation Machinery. *Nat. Chem. Biol.* 14 (1), 29–35. doi:10.1038/nchembio.2514

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA Synthetase: an Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1844 (6), 1059–1070. doi:10.1016/j.bbapap.2014.03.002

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming Cells by Multiplex Genome Engineering and Accelerated Evolution. *Nature* 460 (7257), 894–898. doi:10.1038/nature08187

Wang, N., Ju, T., Niu, W., and Guo, J. (2015). Fine-tuning Interaction between Aminoacyl-tRNA Synthetase and tRNA for Efficient Synthesis of Proteins Containing Unnatural Amino Acids. *ACS Synth. Biol.* 4 (3), 207–212. doi:10.1021/sb500195w

Wang, S., Ibba, M. P., Ataide, S. F., Roy, H., and Michael Ibba*, fnm (2006). Discrimination of Cognate and Noncognate Substrates at the Active Site of Class I Lysyl-tRNA Synthetase. *Biochemistry* 45 (11), 3646–3652. doi:10.1021/bi0523005

Watson, J. D., and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171 (4356), 737–738. doi:10.1038/171737a0

Worst, E. G., Exner, M. P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N., et al. (2015). Cell-free Expression with the Toxic Amino Acid Canavanine. *Bioorg. Med. Chem. Lett.* 25 (17), 3658–3660. doi:10.1016/j.bmcl.2015.06.045

Xiao, H., Murakami, H., Suga, H., and Ferré-D'Amaré, A. R. (2008). Structural Basis of Specific tRNA Aminoacylation by a Small *In Vitro* Selected Ribozyme. *Nature* 454 (7202), 358–361. doi:10.1038/nature07033

Xu, B., Trawick, B., Krudy, G. A., Phillips, R. M., Zhou, L., and Rosevear, P. R. (1994). Probing the Metal Binding Sites of *Escherichia coli* Isoleucyl-tRNA Synthetase. *Biochemistry* 33 (2), 398–402. doi:10.1021/bi00168a002

Yamanaka, K., Nakata, H., Hohsaka, T., and Sisido, M. (2004). Efficient Synthesis of Nonnatural Mutants in *Escherichia coli* S30 *In Vitro* Protein Synthesizing System. *J. Biosci. Bioeng.* 97 (6), 395–399. doi:10.1016/S1389-1723(04)70225-X

Yan, W., Augustine, J., and Francklyn, C. (1996). A tRNA Identity Switch Mediated by the Binding Interaction between a tRNA Anticodon and the Accessory Domain of a Class II Aminoacyl-tRNA Synthetase. *Biochemistry* 35 (21), 6559–6568. doi:10.1021/bi952889f

Yan, W., and Francklyn, C. (1994). Cytosine 73 Is a Discriminator Nucleotide In Vivo for Histidyl-tRNA in *Escherichia coli*. *J. Biol. Chem.* 269 (13), 10022–10027. doi:10.1016/S0021-9258(17)36984-3

Zhang, C.-M., Christian, T., Newberry, K. J., Perona, J. J., and Hou, Y.-M. (2003). Zinc-mediated Amino Acid Discrimination in Cysteinyl-tRNA Synthetase. *J. Mol. Biol.* 327 (5), 911–917. doi:10.1016/s0022-2836(03)00241-9

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 | Frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 07 July 2022
doi: 10.3389/fbioe.2022.915035

Self-Assembling Protein Surfaces for *In Situ* Capture of Cell-Free-Synthesized Proteins

Ella Lucille Thornton[1], *Sarah Maria Paterson*[1], *Zoe Gidden*[1], *Mathew H. Horrocks*[2], *Nadanai Laohakunakorn*[1]* *and Lynne Regan*[1]*

[1]*Centre for Synthetic and Systems Biology, Institute of Quantitative Biology, Biochemistry and Biotechnology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom,* [2]*School of Chemistry, University of Edinburgh, Edinburgh, United Kingdom*

We present a new method for the surface capture of proteins in cell-free protein synthesis (CFPS). We demonstrate the spontaneous self-assembly of the protein BslA into functionalizable surfaces on the surface of a CFPS reaction chamber. We show that proteins can be covalently captured by such surfaces, using "Catcher/Tag" technology. Importantly, proteins of interest can be captured either when synthesised *in situ* by CFPS above the BslA surfaces, or when added as pure protein. The simplicity and cost efficiency of this method suggest that it will find many applications in cell-free-based methods.

Keywords: cell-free (CF) protein synthesis, surface immobilization, self-assembling, covalent attachment, protein-based tools

OPEN ACCESS

Edited by:
Simon J. Moore,
University of Kent, United Kingdom

Reviewed by:
Zhi-Gang Jeff Qian,
Shanghai Jiao Tong University, China
Amy J. Karlsson,
University of Maryland, College Park, United States

***Correspondence:**
Nadanai Laohakunakorn
nadanai.laohakunakorn@ed.ac.uk
Lynne Regan
lynne.regan@ed.ac.uk

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *07 April 2022*
Accepted: *16 June 2022*
Published: *07 July 2022*

Citation:
Thornton EL, Paterson SM, Gidden Z, Horrocks MH, Laohakunakorn N and Regan L (2022) Self-Assembling Protein Surfaces for In Situ Capture of Cell-Free-Synthesized Proteins. Front. Bioeng. Biotechnol. 10:915035. doi: 10.3389/fbioe.2022.915035

INTRODUCTION

The ability to immobilise proteins on surfaces facilitates the separation of the protein from all solution constituents, and therefore has a multitude of useful applications, including for protein purification, biosensing, and continuous flow enzymatic catalysis (Berrade et al., 2011).

Functional display of macromolecules on a surface, however, is difficult to achieve. Non-specific adherence of a molecule to the surface may cause it to denature, and non-specific orientation of presentation may make the active site of a molecule inaccessible (Kim and Herr, 2013; de Marco, 2018; Yang et al., 2018).

Cell-free protein synthesis (CFPS) is a powerful method to generate proteins *in vitro* using biochemical reactions. The process involves combining enzymatic machinery required for transcription and translation with an appropriate mix of small molecules and a DNA template encoding the protein of interest. The technique is now widely applied within synthetic biology for protein production, but also for diverse applications including biosensing, diagnostics, and materials production (Kelwick et al., 2020; Laohakunakorn et al., 2020). Much work has gone into improving the yield and lifetime of CFPS systems generated from a variety of organisms, with the best performing examples capable of producing more than 4 mg/ml of protein over 20 h in batch-mode reactions (Garenne et al., 2021).

Despite many advances in the optimisation of the CFPS reaction itself, there have been far fewer developments in surface capture technologies (Cole et al., 2020; Banks et al., 2022).

Surface immobilisation of proteins expressed in cell-free reactions was demonstrated in 2001, with a method named PISA (Protein *in situ* Array) (He and Taussig, 2001). In PISA, immobilisation of a His-tagged protein of interest, produced by CFPS, is achieved by attachment to a Ni-NTA coated surface. This method has the advantage of site-specific spontaneous immobilisation. It is limited by the relatively low strength of the Ni-NTA – His-tag interaction.

Subsequently, cell-free expression from immobilised DNA and capture of the protein product on the surface was reported in the method named NAPPA (Nucleic Acid Programmable Protein Array) (Ramachandran et al., 2004). In NAPPA, biotinylated DNA is immobilised on a surface coated with a mixture of avidin and polyclonal anti-GST antibodies. The latter are used to capture GST-tagged proteins, produced by CFPS. This method is powerful because both the DNA template and the protein product are attached to the surface. A limitation of this method is the heterogeneous orientation of the GST (Glutathione S-Transferase) antibodies on the surface, and the range of affinities of the polyclonal antibodies for the GST-tagged protein.

More recently, advanced protocols have been developed and implemented (Manzano-Román and Fuentes, 2019) including the MITOMI (mechanically-induced trapping of molecular interactions) method, which implements the NAPPA protocol in a microfluidic setting, at high throughput (Maerkl and Quake, 2007). In MITOMI, anti-His antibodies, biotinylated at multiple positions, are attached to a multi-layer surface of biotinylated BSA coated with neutravidin. This enables the pull-down of His-tagged protein products generated in CFPS. MITOMI has been used to characterise the binding energy landscape of transcription factors, as well as kinetic rate constants (Geertz et al., 2012).

The desirable features of a surface capture method for use with CFPS include: ease of assembly, homogeneity of surface presentation, retention of activity of the immobilised molecule, robust attachment, and cost effectiveness (Kilb et al., 2014). Here we describe a new system that we believe meets all desirable criteria.

We form a protein-based surface using the protein BslA, a small (15 kDa) amphiphilic protein that self-assembles at hydrophobic/hydrophilic interfaces (Hobley et al., 2013). BslA is capable of surface formation on the surface of hydrophobic glass, as shown previously by TEM (Bromley et al., 2015). We covalently link proteins to the BslA surface using Tag/Catcher technology, where a short peptide, SpyTag or SnoopTag, spontaneously forms a covalent bond with its cognate protein partner, SpyCatcher or SnoopCatcher, when mixed (Veggiani et al., 2016).

We have previously shown that fusing the hydrophilic end of BslA to a short peptide tag does not interfere with its ability to form surfaces. Thus, by fusing BslA to either SpyTag or SnoopTag peptides, we can capture a protein of interest (POI) that is fused to the cognate SpyCatcher or SnoopCatcher protein (Schloss et al., 2016).

SpyTag/SpyCatcher and SnoopTag/SnoopCatcher technology is especially well-suited for surface attachment applications because it is specific, rapid, and genetically encodable (Zakeri et al., 2012; Veggiani et al., 2016; Lange and Polizzi, 2021). The covalent linkage of the Tag/Catcher pairs is robust, and the availability of two orthogonal peptide/protein pairs enables different proteins to be attached to the surface. We show that specific surface capture occurs from a complex mixture of proteins. We also show that these surfaces can capture either proteins synthesised in the cell-free reaction, or purified proteins added to the cell-free reaction.

RESULTS

Spontaneously Self-Assembled BslA Protein Surfaces can be Functionalised by Covalent Attachment of Proteins

We have previously demonstrated BslA-Tag assembly and functionalisation, using Langmuir-Schaeffer deposition followed by incubation with a purified fluorescent protein (FP) fused to the appropriate Catcher protein (Williams et al., 2018). Here we show that a simpler, one-step, self-assembly method can also be used, and demonstrate that it is compatible with CFPS.

Specifically, we show that when an aqueous solution of wild-type (WT) BslA, or the fusion proteins BslA-SpT (BslA-SpyTag) or BslA-SnT (BslA-SnoopTag) are incubated with a hydrophobic glass slide, they spontaneously self-assemble into surfaces, which can be functionalised using the Tag/Catcher technology.

Figure 1 shows these data. After surface formation, the surface was incubated with fluorescent protein-Catcher fusion proteins (FP-Catcher), where FP is either mCherry or GFP. Unbound FP-Catcher was washed away, and the fluorescence retained on the surface was measured. We observe a significant fluorescent signal for both GFP-SpC incubated with BslA-SpT surfaces, and mCherry-SnC incubated with BslA-SnT surfaces indicating specific attachment to the surface via Tag/Catcher interaction.

By contrast, when either FP-Catcher protein is incubated with a BslA-WT surface, fluorescence is not retained, indicating that there is little non-specific binding of the fusion proteins to BslA-WT.

We also observed that when a mixture of mCherry-SnC and GFP-SC is incubated with a BslA-SnT surface, only mCherry-SnC is captured. Conversely, when a mixture of mCherry-SnC and GFP-SC is incubated with a BslA-SpT surface, only GFP-SC is captured (**Figure 1C**). In summary, the amount of either protein that is specifically captured is not diminished by the presence of the non-cognate protein. Additionally, we found that proteins specifically bound to the BslA surface were not affected by washing with additives such as BSA (Bovine Serum Albumin) or TWEEN20 (**Supplementary Figure S3**).

Following the observations described in **Figure 1**, we sought to determine if mCherry-SpC is covalently attached to the BslA-SnT surface.

Figure 2A shows purified mCherry-SnC added to BslA-WT, BslA-SnT or BslA-SpT in solution. After incubation, the mixture was analysed by SDS-PAGE. The presence of a new band corresponding to the mCherry-SnC-BslA-SnT fusion is evident only when the cognate pair are mixed, indicating specific and covalent attachment of mCherry-SnC to BslA-SnT.

We next sought to determine if mCherry-SnC is covalently linked to BslA-SnT after reaction with the BslA surface. To this end, we incubated BslA-WT, BslA-SnT, and BslA-SpT surfaces with mCherry-SnC. After washing, we extracted all proteins from the surface using SDS sample buffer and analysed the mixtures by SDS-PAGE. Because the amounts of the materials are low, we performed a Western blot, probing the gel with anti-mCherry

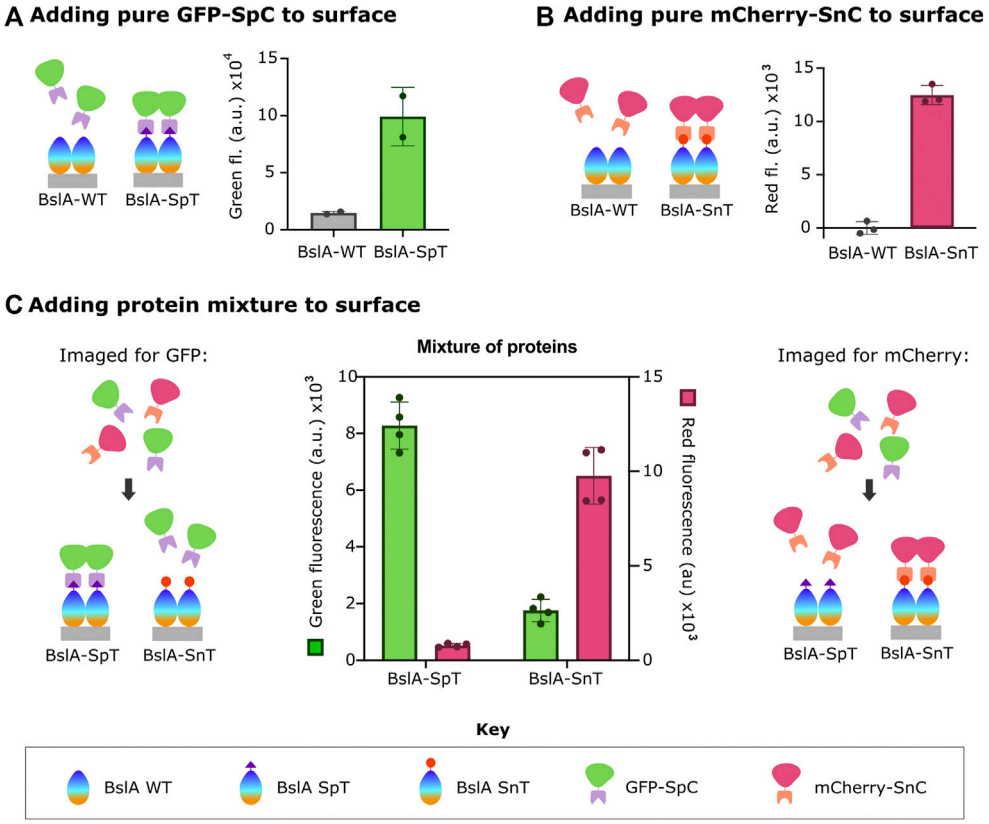


FIGURE 1 | Surfaces formed by BslA-SpT or BslA-SnT capture FP fused to SpC or SnC, respectively. The key at the bottom of the figure indicates the identity of the protein components in the cartoons. In A-C, the cartoons left illustrate the experiment in schematic form, with the corresponding data shown next to it. Purified **(A)** GFP-SpC or **(B)** mCherry-SnC proteins were incubated with BslA-WT, BslA-SpT or BslA-SnT surfaces. Fluorescence remaining in each well was measured after removal of non-bound protein and washing of the wells. In **(C)**, a similar protocol was followed, but here mixtures of GFP-SpC and mCherry-SnC were incubated (at equimolar concentration) with either BslA-SpT or BslA-SnT surfaces. Measured green fluorescence of either the BslA-SpT or BslA-SnT coated wells are shown with the green coloured bars, corresponding to the left *y*-axis. Measured red fluorescence of the wells is shown with the pink coloured bars, with values corresponding to the right *y*-axis. In all cases, the fluorescence of a well with BslA and buffer was subtracted as the blank. Original data is available in the **Supplementary Data Sheet**. Error bars show standard deviation.

antibodies. A band corresponding to a covalently joined complex is only present in the lane corresponding to mCherry-SnC + BslA-SnT.

This result supports the conclusion that BslA surfaces can be covalently functionalised with a protein of interest using the Tag/Catcher technology.

BslA Proteins do not Diffuse Laterally in the Surface Coating

Having established covalent attachment of fluorescent proteins, we sought to investigate the lateral mobility of BslA molecules in the surface. Such knowledge is important because it underlies the feasibility of localised attachment of different proteins to different areas of a BslA surface. We investigated BslA lateral mobility in the surface using Fluorescence Recovery After Photobleaching (FRAP).

We hypothesised that if the BslA-GFP fusions in the surface can move laterally, we would see recovery of fluorescence within a photobleached area of the surface. Conversely, if the BslA proteins do not move laterally in the surface, after photobleaching of a given area we would expect to see no recovery of fluorescence. A schematic illustration of these two scenarios is shown in (**Figure 3A**).

Figure 3B shows the experimentally observed behaviour. We observe no recovery after photo-bleaching an area of the BslA-SpT-SpC-GFP surface. This behaviour indicates that the BslA proteins do not diffuse laterally in the surface. This experiment was also repeated over longer time scales, and no recovery of fluorescence in the bleached area was observed even after 30 min (**Supplementary Figure S4**).

We therefore conclude that BslA proteins in a surface do not diffuse laterally. This result is important because it suggests that localised application of different Catcher-proteins to different areas of a BslA-Tag surface should be feasible.

The data shown here also supports our hypothesis that the fluorescent proteins are covalently attached to the BslA surface. If the interaction were non-covalent one might expect to observe the diffusion of fluorescent proteins back into the photobleached area as they unbind and rebind to the BslA surface.

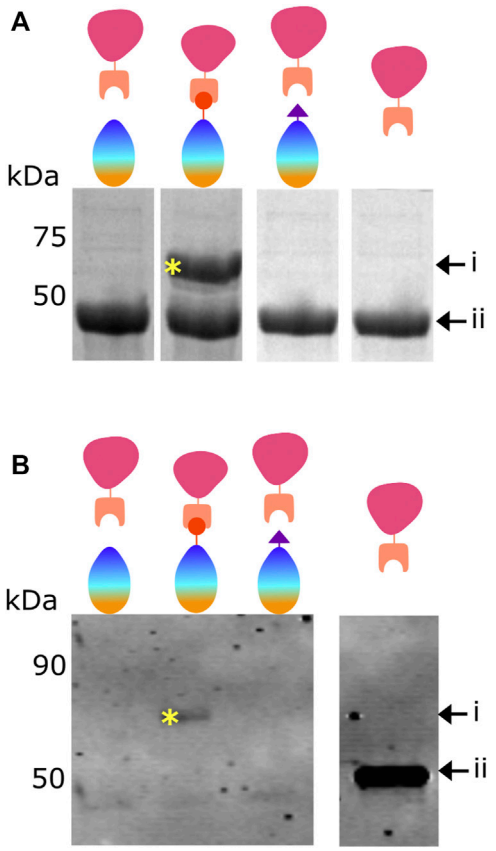


FIGURE 2 | Proteins bind to the BslA surface via a covalent bond. **(A)** Demonstration of covalent bond formation in solution. mCherry-SnC (44.8 kDa, labelled as ii) was incubated with BslA-WT, BslA-SnT, or BslA-SpT in solution and the products analysed by SDS-PAGE. A new protein band is seen at a molecular weight of approximately 60 kDa (labelled as i and with a yellow asterisk), only in the mCherry-SnC plus BslA-SnT reaction mix. Because it withstands the SDS-PAGE separation, the new band represents the covalently linked mCherry-SnC + BslA-SnT. Because mCherry-SnC protein is present in excess, bands corresponding to unreacted BslA proteins are not shown here for simplicity. The complete, uncropped SDS-PAGE gel is shown in **Supplementary Figure S5**. **(B)** Demonstration of covalent bond formation to a surface. Surfaces of BslA-WT, BslA-SnT, or BslA-SpT on a hydrophobic glass surface were reacted with mCherry-SnC. Excess protein was removed and the surface gently washed. The surfaces were then vigorously washed with SDS loading buffer, to extract the BslA surface and any associated proteins, and the samples analysed by SDS-PAGE. Because the amounts of protein were low, a Western blot was performed, and the membrane probed with an anti-mCherry primary antibody. A band at approximately 60 kDa is evident only with the BslA-SnT plus mCherry-SnC, labelled with a yellow asterisk. Because it withstands the SDS-PAGE separation, we interpret the new band as representing the covalently linked mCherry-SnC + Bsla-SnT. Uncropped versions of both SDS-PAGE gel and Western blot are supplied in **Supplementary Figure S5**.

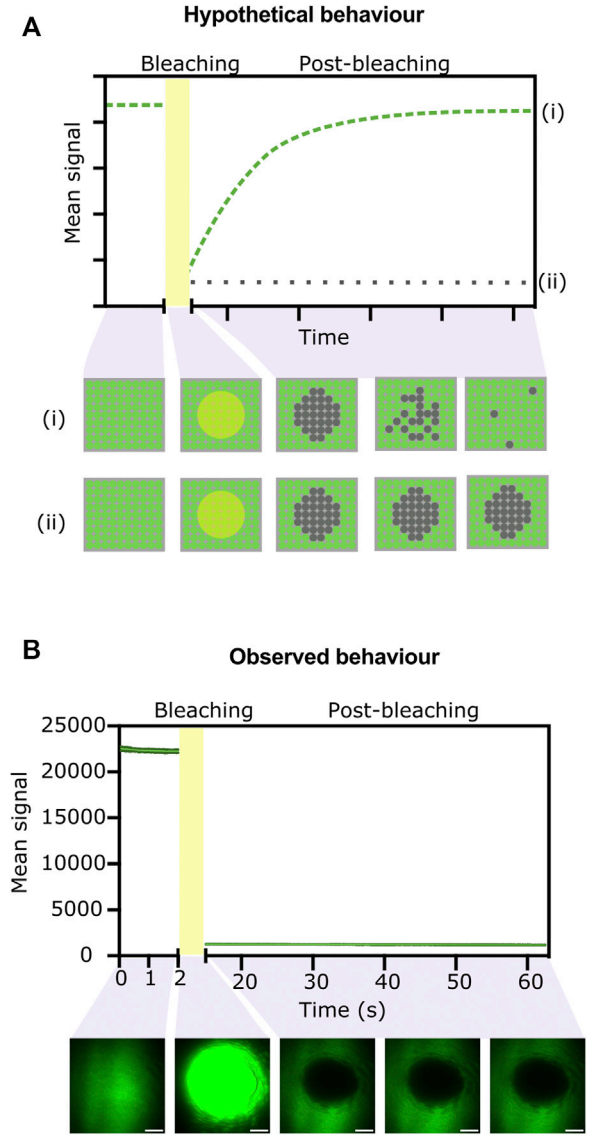


FIGURE 3 | FRAP experiments indicate that proteins do not move laterally in the BslA surface. **(A)** Hypothetical data showing the two extreme possibilities after a FRAP experiment. If the surface exhibits significant lateral mobility, the fluorescence in a bleached area will be recovered over time as non-bleached molecules move in. The expected data from this scenario is indicated by the dashed green line (i) on the graph. The behaviour is illustrated in cartoon form in the panel below, which shows a top-down view of the proteins on the surface. Conversely, if the fluorescent molecules do not diffuse laterally, fluorescence will not be recovered after photobleaching (dotted grey line on the graph, (ii) and matching cartoon). **(B)** Experimentally observed data for a BslA-SpT-SpC-GFP surface. No recovery of fluorescence after photobleaching is observed, indicative of no lateral movement. Error bars show standard deviation, but they are not clearly visible because the variability between replicates was very small (n = 4). Representative images of the surface at different stages of the experiment are shown below the graph. Images have been colored green using Fiji. Scale bar represents 10 μm.

BslA Surfaces can Capture Proteins Made by CFPS, *In Situ*

We investigated the ability of self-assembled BslA surfaces to capture proteins that are made in a CFPS reaction. **Figure 4A** shows a schematic of the experiment, where mCherry-SnC is produced by CFPS above a BslA-SnT or BslA-WT surface. When mCherry-SnC is synthesised, we expect that it will form a covalent bond with a BslA-SnT surface, but not with a BslA-WT surface. Therefore, after removing CFPS constituents and

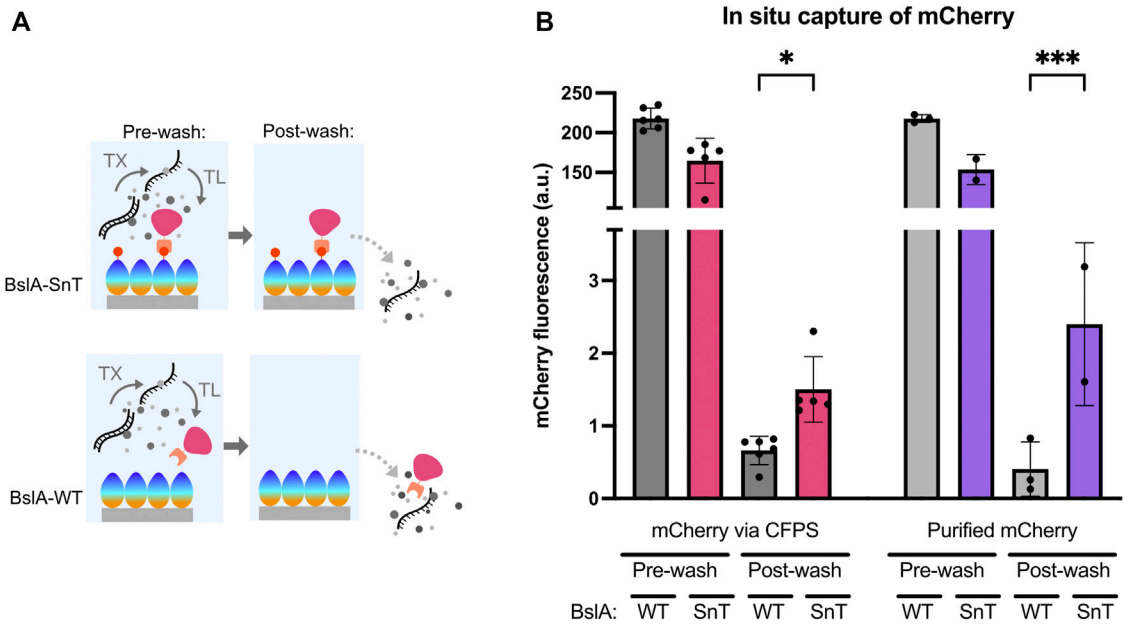


FIGURE 4 | BslA surfaces specifically capture protein produced in a cell-free reaction. **(A)** Schematic illustration of the experiment: A BslA-SnT or BslA-WT surface is assembled on the surface of a hydrophobically coated glass slide. mCherry-SnC is produced by CFPS above this surface. Because an equal amount of DNA is added to each reaction, we expect approximately equal amounts of fluorescent protein will be produced in both wells. After CFPS, the wells are washed. Only mCherry-SnC that is bound to the surface will be retained in the well after washing. **(B)** Experimental data resulting from the experiments illustrated in (A). A similar amount of total fluorescence is observed in the CFPS reactions performed above the BslA-WT and BslA-SnT surfaces. The wells are then washed, and the retained mCherry fluorescence is measured. Higher retention of fluorescence is observed for the CFPS performed above a BslA-SnT surface than above a BslA-WT surface, indicating capture of the mCherry protein synthesised *in situ*. When purified mCherry-SnC protein is added exogenously to the CFPS reaction mixture, rather than synthesised *in situ*, a similar result is obtained: with higher retention of fluorescence in BslA-SnT wells than in BslA-WT wells. Importantly, the fluorescence values of each well before and after washing does not correlate to efficiency of binding, as we expect this concentration of mCherry-SnC to be in excess of the possible binding sites on the surface of each well. **Supplementary Information (C)** provides estimates of maximum binding capacity of the given surface area. Post-wash signal was analysed for significance by one-way ANOVA using Šídák's multiple comparisons test. Significance is denoted by asterisk on the graph between fluorescence for WT and SnT surfaces post-wash, where * denotes $p \leq 0.05$ and *** denotes $p \leq 0.001$. Error bars show standard deviation.

washing the wells, we expect to see mCherry signal retained only in those wells coated with BslA-SnT.

In **Figure 4B**, we show the relative amounts of mCherry-SnC produced in the CFPS reactions. Post washing, mCherry signal is only retained in the wells coated with a BslA-SnT surface. The 4 bars on the right-hand side of **Figure 4B** visualise data where purified mCherry-SnC protein (5 μM) is added to a CFPS reaction with no DNA for protein synthesis. These samples serve as a control, and the results indicate that protein synthesised in CFPS and protein added exogenously to the CFPS mixture can bind to the surface specifically and with a similar efficiency.

The data we present here indicates not only the ability of the BslA surfaces to capture proteins produced by CFPS, but also their robustness over a 12 h incubation at 37°C during the CFPS reaction.

DISCUSSION

CFPS is a powerful approach with many applications. Here we demonstrate how CFPS can be expanded to include a new method of surface capture.

We demonstrate that the BslA protein spontaneously assembles into a surface, and that CFPS can be performed over such a surface. The method we present can be applied either to capture a protein produced in the cell-free reaction, or to bind purified protein. Importantly, the attachment of the protein to the BslA surface is covalent, and therefore robust to washing and other perturbations (**Supplementary Figure S3**). Binding proteins to the BslA surface can have differing background levels depending on the protein used and should be tested on a case-by-case basis if this is important for the desired application. To further reduce non-specific interactions between the target protein and BslA surface, washing conditions can be further optimised to ensure all proteins are covalently bound to the surface and not non-specifically stuck.

Our demonstration, by FRAP, that the BslA proteins do not move laterally on the hydrophobic surface suggests that it will also be possible to use this method to immobilise different proteins in different areas of the surface.

We envision that surface capture of a protein produced by CFPS could be used either to allow multiple rounds of screening of the activity of that protein against different substrates or to allow for consecutive rounds of CFPS to occur above the same surface. Also, it is a means by which an exogenously added protein could be present on the surface during a CFPS reaction.

The method we describe represents a simple yet powerful method to create functionalised surfaces that are compatible with CFPS. The surfaces are robust over the duration of a CFPS reaction, and the desired protein can be specifically captured from a heterogeneous mixture. By using proteins that are genetically encodable, and which can be fused to any protein of interest, at a unique position, this method is both cost-effective and versatile.

MATERIALS AND METHODS

E. coli General Methods

Standard overnight cell growth: *E. coli* cells were picked from a single colony on an LB agar plate with the appropriate antibiotic into 5 ml LB (10 g/L peptone, 5 g/L NaCl, 5 g/L yeast extract) with the appropriate antibiotic and grown overnight at 37°C with shaking (220 rpm). For DNA preparation, TOP10 cells were used [F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lacX*74 *nup*G *recA*1 *araD*139 Δ(*ara-leu*)7697 *gal*E15 *gal*K16 *rps*L(StrR) *end*A1 λ-]. For protein expression, BL21 Gold (DE3) cells were used [F- *omp*T *hsdSB*(*rB–mB–*) *dcm* (TetR) *gal* λ(DE3) *endA* The]. For CFPS lysate production, Rosetta-gami2 cells were used [Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC gale galK rpsL F'[lac + lacIq pro] gor522: Tn10 trxB pRARE2 (Cam^R, Str^R, Tet^R)]. Plasmids were transformed into competent *E. coli* cells following standard protocols (Green and Sambrook, 2012).

DNA purification and quantification: Plasmids were purified from *E. coli* following protocols described by the manufacturer using the QIAprep® Spin Miniprep Kit (Qiagen). Linear DNA was purified from PCR mixture or agarose gels using the Promega Wizard® SV Gel and PCR Clean-up System following protocols described by the manufacturer. Purified DNA solutions were quantified by A260 and stored at −20°C.

DNA sequencing: Plasmid sequences were verified by DNA sequencing performed either by DNA Sequencing & Services (www.dnaseq.co.uk) or Source Bioscience (www.sourcebioscience.com), using primers provided by the company.

Protein Expression and Purification

Cell growth and protein expression: Overnight cultures were diluted 100- fold into LB containing the appropriate antibiotic and grown at 37°C with shaking until OD600 reached 0.6–0.8. Protein expression was induced by addition of isopropyl β-d-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and growth continued for a further 20 h at 20°C with shaking. Cells were collected by centrifugation at 6,000 × *g* for 10 min and pellets stored at −20°C until needed.

Lysis and clarification: Cells were resuspended in lysis buffer (**Supplementary Table S1**) containing cOmplete Protease Inhibitor Cocktail (Sigma-Aldrich) according to manufacturer's instructions at a ratio of 1:50 volume of buffer to original cell culture volume. Resuspended cells were sonicated (Soniprep 150, MSE) on ice for 30 s, followed by a 30 s rest period. This sonication-rest cycle was repeated until cell lysis was achieved. Clarified cell lysates were prepared by centrifugation at 10,000 × *g* for 30 min at 4°C.

Affinity purification: mCherry-SnC and GFP-SpC were purified via hexahistidine tag using Ni-NTA agarose (Qiagen) according to the manufacturer's instructions. BslA-fusion proteins were purified via GST tag using GST agarose (ThermoFisher) according to the manufacturer's instructions. The GST tag was cleaved on column with PreScission Protease (GE Healthcare) according to the manufacturer's instructions. Purification was monitored using SDS-PAGE. Fractions containing protein at approximately 90% or greater purity, were pooled and dialysed against desired the storage buffer.

Note on BslA purification: During purification of BslA proteins, solution containing concentrated and pure protein could often look opaque and white in colour. When this happened, the solutions were left to stand at room temperature to allow for re-solubilisation of BslA protein, which could take between 5—30 min, depending on concentration of solution. Freeze thaw cycles were always kept to a minimum with pure protein solutions, but even more stringently with BslA: pure protein was aliquoted into small volumes before freezing at −20°C. For experiments involving functionalisation of surfaces, BslA solutions were only thawed once. These precautions were developed according to our own observations and in consultation with an author of (Bromley et al., 2015).

Size exclusion chromatography (SEC): Further purification of BslA proteins by size exclusion chromatography (SEC) with a Superdex-75 column was used when affinity chromatography did not provide sufficient purity.

Protein quantification: Protein concentration was determined by measuring absorbance at 280 nm using the extinction coefficient of each protein calculated from the amino acid sequence using the tools available on Benchling [Biology Software]. (2022) Retrieved from https://benchling.com.

Protein concentration: Buffer exchange and simultaneous concentration of protein solutions was performed using Amicon® Ultra Centrifugal Filters (Merck). Columns were selected based on the sample volume and POI size. Generally, the MW cut off was chosen to be at least 1.5 × smaller than that of the POI.

Dialysis: Buffer exchange of protein solutions was performed by dialysis using SnakeSkin Dialysis Tubing (ThermoFisher, various sizes depending on POI) according to manufacturer's instructions. Typically, samples were dialysed into 1 L buffer overnight at 4°C with stirring, followed by dialysis into 1 L fresh buffer twice more, for a period of 1 h each at 4°C with stirring.

SDS-PAGE: Protein expression, purification and reactivity was monitored using SDS-PAGE. For testing covalent bond formation between SpT/SpC and SnT/SnC fused proteins, each protein was incubated together at approximately equimolar concentrations for 1 hour at room temperature. Samples in 1x SDS loading buffer were heated at 100°C for 10 min before loading on SDS-PAGE gels alongside Precision Plus Protein Dual Xtra Prestained Protein Standards as molecular weight marker. Protein bands were visualised by staining using

InstantBlue® Coomassie Protein Stain according to manufacturer's instructions and were subsequently imaged using a Bio-Rad Gel Doc XR + system with white light filter in Coomassie mode.

Western blot: Protein samples were prepared and loaded onto an SDS-PAGE gel with Chameleon Duo Prestained Protein Ladder as a molecular weight marker. The Western blot procedure described in "Near-Infrared Western Blot Detection Document" (Doc. #988-13627, Li-cor) was followed. Antibodies recognising mCherry (Anti RFP-tag, pAb, Rabbit; A00682, GenScript) were used as a primary antibody at a dilution of 1: 3,000. IRDye 680RD Goat anti-Rabbit (925-68071, Li-cor) was used as a secondary antibody at a dilution of 1:20,000. Blotted membranes were visualised using an Odyssey CLx Infrared Imaging System (Li-cor) and analysed by Image Studio Lite software (Li-cor).

Surface Functionalisation

Preparation of glass slides: Ultra-clean, hydrophobic glass microscope slides were prepared following methods adapted from the literature (Cras et al., 1999). 150 ml HCl was slowly added into 150 ml MeOH in a glass beaker and gently swirled to mix. Slides (Thermo Scientific plain microscope slides, 12332098) were added to the solution and incubated at room temperature for 1 hour. Slides were removed and added to a fresh beaker with 300 ml distilled water, gently swirled to mix and this wash step was repeated an additional four times. Slides were tapped dry (avoiding touching the surface) and dried completely by incubation in a drying oven at 90°C for 2 hours. To add the hydrophobic coating to the cleaned sides, Dichloromethylsilane was dissolved in Trichloroethylene to a final concentration of 0.05% and incubated with the dried glass slides for 1 hour. Glass slides were transferred to a MeOH solution and incubated for 1 min, then slides were removed and incubated again with MeOH three more times. Slides were rinsed twice in distilled water and dried before packaging with lens paper in a sealed container. Slides were still sufficiently clean and hydrophobic for the experiments described after storage for up to 3 months. Hydrophobicity was assessed by measuring the contact angle of a water droplet on the glass surface.

Application of proteins to BslA surfaces: ProPlate Multi-Well Chambers (Grace Bio-Labs) allowed up to four slides with adaptors to be combined in one adaptor plate (ProPlate Multi-Array Slide System (Grace Bio-Labs) and imaged in a plate reader. 10 µL BslA protein (22 µM) was incubated in each well at room temperature for 1 h with high humidity to prevent evaporation. After surface formation, excess BslA was removed by pipette. 10 µL protein of interest was incubated with the well for 1 h to allow the SpyTag/SpyCatcher reaction to occur. Excess protein was removed from each well by pipetting, adaptors were removed and the whole slide was washed with 50 ml ultrapure water. Fluorescence was measured using a FLUOstar Omega plate reader with appropriate filters. For GFP = excitation 485 nm, emission 520 nm, for mCherry = excitation 584 nm, emission 620 nm, with gain set between 1500—2500.

Cell-Free Protein Synthesis

Lysate production: This protocol is adapted from (Kwon and Jewett, 2015), and was optimised as shown in **Supplementary Figure S1** and functionally demonstrated in **Supplementary Figure S2**. Typically, addition of 5 nM plasmid DNA to our homemade CFPS system yielded 5 µM protein in a 10 µL reaction. 2xYTPG (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L KH2PO4 3 g/L K2HPO4, 18 g/L glucose) was inoculated with 1/200 dilution of overnight cultures of Rosetta-gami2 *E. coli* cells. Cultures were grown for 2 h at 37°C with shaking, then induced with 0.4 mM IPTG and grown for a further 2 h in the same conditions, before growth arrest by placing on ice. Cells were harvested by centrifugation at 10,000 ×g for 10 min at 4°C, the supernatant was discarded, and cell pellets were resuspended with 80 ml Buffer A (**Supplementary Table S1**) per 400 ml cells harvested. Cells were collected by centrifugation at 4,500 rpm for 10 min at 4°C. The washing and cell harvesting process was repeated twice more, and cell pellets stored at −80°C for future downstream processing. Cell pellets were resuspended with 1 ml buffer A per 1 g wet cell mass and homogenised by vortexing. 1.5 ml aliquots were taken from the total mixture and sonicated (Fisher120 W sonicator with probe for 0.5–15 ml) with pulses of 10 s on, and 10 s off, until a total energy output of 556 J was achieved, while incubated on ice. Lysate was clarified by centrifugation at 12,000 × *g* at 4°C for 10 min. Supernatant was removed and re-spun under the same conditions. Clarified supernatant was placed in a clean 1.5 ml Eppendorf tube and incubated at 37°C for 1.5 h with shaking (220 rpm) in a "run-off" reaction (the impact of both run-off and dialysis on functionality of CFPS lysate is detailed further in **Supplementary Figure S1**). Samples were then centrifuged at 12,000 × g at 4°C for 10 min. Supernatant was removed and aliquoted into 25 µL samples, which were stored at −80°C until they were required for use.

Energy solution production: Energy solution was assembled from stock solutions of all constituents. Amino acid stock solution was made of final concentration 50 mM each of the following amino acids: Alanine, Arginine, Asparagine, Aspartate, Cysteine, Glutamate, Glutamine, Glycine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Proline, Serine, Threonine, Tryptophan, Valine. Tyrosine was prepared separately, in an acidic solution (pH ~5.2) also at a final concentration of 50 mM. Stock batches of energy solution were prepared in volumes of 3 ml, with the following recipe: HEPES (pH 8) 200 mM, ATP 6 mM, GTP 6 mM, CTP 3.6 mM, UTP 3.6 mM, tRNA 0.8 mM, CoA 1.04 mM, NAD 1.32 mM, cAMP 3 mM, Folinic acid 0.27 mM, Spermidine 4 mM, 3-PGA 120 mM, Amino acids 6 mM, Tyrosine 3 mM, PEG-8000 8%, Mg-glutamate 42 mM, K-glutamate 400 mM, DTT 1 mM.

Assembly of cell-free reaction: Cell free reactions were prepared with a final volume of 10 µL in wells of a 384 microplate (Greiner, 781906). Energy solution, lysate, DNA, and buffer A were combined in a 1:1:1:1 ratio by volume in each well. When required, master mixes were prepared which contained DNA, energy solution and buffer before dispensing into appropriate wells. Lysate was always added to each well individually and pipetted onto the side of the well wall. This protocol provided a useful visual note of progress through the

plate, but also allowed for the initiation of the cell-free reaction at the same time in every well, by centrifuging the plate at 500 rpm for 2 min at 4°C. This mixed all components simultaneously and removed any air bubbles from pipetting. After centrifugation, 35 µL BioRad chillout wax was added to each well to prevent evaporation of samples. Plates were sealed with non-breathable film and placed in the plate reader.

Preparation of DNA for cell-free reactions: Plasmid DNA was used for CFPS reactions, with expression from a pTrc promoter upstream of protein coding DNA sequence (full sequences available in **Supplementary Data (D)**). DNA used in cell-free reactions required a higher level of purity than those for standard molecular biology procedures. DNA was extracted from cells using the PureLink HiPure Plasmid Maxiprep Kit (Invitrogen) following described protocols. DNA was further purified by DNA Clean & Concentrator Kit (Zymo Research) following described protocols. Final DNA concentration was measured by absorbance at 260 nm on a Nanodrop (DeNovix DS-11) and DNA was stored at −20°C until required.

FRAP Microscopy

Samples for FRAP experiments were prepared in triplicate according to methods previously described, on glass coverslips with CultureWell Reusable Gaskets. BslA-WT or BslA-SpT was used at a concentration of 11 µM, and GFP-SpC at a concentration of 5 µM.

After incubation of proteins with the surface, the gasket was removed, and the coverslip washed thoroughly with ddH2O. Samples were then imaged using a custom-built TIRF (Total Internal Reflection Fluorescence) microscope (as previously described in (Pérez-Pi et al., 2019), using a 488 nm laser to excite and illuminate the GFP molecules throughout imaging. Photobleaching of one section of the surface was achieved using 405 nm laser excitation. Experiments were typically set up to collect data for 2 s pre-bleach, bleach for 10 s, and then collect data for the 30 s following. For some experiments, data was collected post-bleach for 5 min. The microscope was set up for each experiment as following. Laser light (Colbalt Diode Laser Systems, Cobalt, Sweden) were aligned and directed parallel to the optical axis at the edge of a 1.49 NA TIRF objective (CFI Apochromat TIRF 60XC Oil, Nikon, Japan), mounted on an inverted Nikon TI2 microscope (Nikon, Japan). The fluorescence was separated from the returning TIR beam by a dichroic mirror Di01-R405/488/561/635 (Semrock, Rochester, NY, United States), passed through the appropriate filters (Semrock, NY, United States) and then recorded on an EMCCD camera (Delta Evolve 512, Photometrics, Tuscon, AZ, United States). Each pixel was 103 nm in length. Data was analysed with Python code using Fiji to track mean fluorescence over time from the centre of the photobleached area. This numerical data was visualised in GraphPad Prism v8.1.2. To prepare the images shown in **Figure 3**, the raw data were processed using Fiji (Schindelin et al., 2012). The images were false coloured green by converting the greyscale image to RGB and splitting the image to only show the green channel.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the **Supplementary Material**. Further inquires can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

ET, LR, and NL led the study, designed the experiments, and wrote the manuscript, with input from SP, ZG, and MH. ET and ZG performed the FRAP experiments and analysed the data, with guidance from MH. ET and SP performed all other experiments and data analysis. ET and SP created the figures.

FUNDING

ET acknowledges the support of the DSTL during her PhD. ZG acknowledges support of the BBSRC via the EastBio DTP (BB/M010996/1). We also acknowledge the School of Biological Sciences and The Edinburgh Protein Production Facility at the University of Edinburgh. LR, ET, SP, and NL acknowledge support of the Leverhulme Trust (RPG-2021-230). NL is supported by a UKRI Future Leaders Fellowship (MR/V027107/1). MH acknowledges Jim Love and UCB Biopharma for funding the microscope used in the FRAP experiments.

ACKNOWLEDGMENTS

We thank Louise Holyoake, Curran Oi, Rossana Boni, Haresh Bhaskar, Fokhrul Islam, Megan Murdoch and Raef Shams for valuable suggestions on the figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.915035/full#supplementary-material

REFERENCES

Banks, A. M., Whitfield, C. J., Brown, S. R., Fulton, D. A., Goodchild, S. A., Grant, C., et al. (2022). Key Reaction Components Affect the Kinetics and Performance Robustness of Cell-free Protein Synthesis Reactions. *Comput. Struct. Biotechnol. J.* 20, 218–229. doi:10.1016/j.csbj.2021.12.013

Berrade, L., Garcia, A. E., and Camarero, J. A. (2011). Protein Microarrays: Novel Developments and Applications. *Pharm. Res.* 28 (7), 1480–1499. doi:10.1007/s11095-010-0325-1

Bromley, K. M., Morris, R. J., Hobley, L., Brandani, G., Gillespie, R. M. C., McCluskey, M., et al. (2015). Interfacial Self-Assembly of a Bacterial Hydrophobin. *Proc. Natl. Acad. Sci. U.S.A.* 112 (17), 5419–5424. doi:10.1073/pnas.1419016112

Cole, S. D., Miklos, A. E., Chiao, A. C., Sun, Z. Z., and Lux, M. W. (2020). Methodologies for Preparation of Prokaryotic Extracts for Cell-free Expression Systems. *Synthetic Syst. Biotechnol.* 5 (4), 252–267. doi:10.1016/j.synbio.2020.07.006

de Marco, A. (2018). Nanomaterial Bio-Activation and Macromolecules Functionalization: The Search for Reliable Protocols. *Protein Expr. Purif.* 147, 49–54. Academic Press Inc. doi:10.1016/j.pep.2018.02.010

Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A., and Noireaux, V. (2021). The All-E. coliTXTL Toolbox 3.0: New Capabilities of a Cell-free Synthetic Biology Platform. *Synth. Biol.* 6 (1), 1–8. doi:10.1093/synbio/ysab017

Geertz, M., Shore, D., and Maerkl, S. J. (2012). Massively Parallel Measurements of Molecular Interaction Kinetics on a Microfluidic Platform. *Proc. Natl. Acad. Sci. U.S.A.* 109 (41), 16540–16545. doi:10.1073/pnas.1206011109

Green, M., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual*. 4th ed. New York, NY: Cold Spring Harbor.

He, M., and Taussig, M. J. (2001). Single Step Generation of Protein Arrays from DNA by Cell-free Expression and *In Situ* Immobilisation (PISA Method). *Nucleic Acids Res.* 29 (15), 73e–73. doi:10.1093/nar/29.15.e73

Hobley, L., Ostrowski, A., Rao, F. V., Bromley, K. M., Porter, M., Prescott, A. R., et al. (2013). BslA Is a Self-Assembling Bacterial Hydrophobin that Coats the Bacillus Subtilis Biofilm. *Proc. Natl. Acad. Sci. U.S.A.* 110 (33), 13600–13605. doi:10.1073/pnas.1306390110

Kelwick, R. J. R., Webb, A. J., and Freemont, P. S. (2020). Biological Materials: The Next Frontier for Cell-free Synthetic Biology. *Front. Bioeng. Biotechnol.* 8 (May), 399. doi:10.3389/fbioe.2020.00399

Kilb, N., Burger, J., and Roth, G. (2014). Protein Microarray Generation by *In Situ* Protein Expression from Template DNA. *Eng. Life Sci.* 14 (4), 352–364. doi:10.1002/elsc.201300052

Kim, D., and Herr, A. E. (2013). Protein Immobilization Techniques for Microfluidic Assays. *Biomicrofluidics* 7 (4), 041501–041547. doi:10.1063/1.4816934

Kwon, Y.-C., and Jewett, M. C. (2015). High-throughput Preparation Methods of Crude Extract for Robust Cell-free Protein Synthesis. *Sci. Rep.* 5, 1–8. doi:10.1038/srep08663

Lange, O. J., and Polizzi, K. M. (2021). Click it or Stick it: Covalent and Non-covalent Methods for Protein-self Assembly. *Curr. Opin. Syst. Biol.* 28, 100374. doi:10.1016/j.coisb.2021.100374

Laohakunakorn, N., Grasemann, L., Lavickova, B., Michielin, G., Shahein, A., Swank, Z., et al. (2020). Bottom-Up Construction of Complex Biomolecular Systems with Cell-free Synthetic Biology. *Front. Bioeng. Biotechnol.* 8 (March), 213. doi:10.3389/fbioe.2020.00213

Maerkl, S. J., and Quake, S. R. (2007). A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science* 315 (5809), 233–237. doi:10.1126/science.1131007

Manzano-Román, R., and Fuentes, M. (2019). A Decade of Nucleic Acid Programmable Protein Arrays (NAPPA) Availability: News, Actors, Progress, Prospects and Access. *J. Proteomics* 198 (December 2018), 27–35. doi:10.1016/j.jprot.2018.12.007

Pérez-Pi, I., Evans, D. A., Horrocks, M. H., Pham, N. T., Dolt, K. S., Koszela, J., et al. (2019). α-Synuclein-Confocal Nanoscanning (ASYN-CONA), a Bead-Based Assay for Detecting Early-Stage α-Synuclein Aggregation. *Anal. Chem.* 91 (9), 5582–5590. doi:10.1021/acs.analchem.8b03842

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A. Y., et al. (2004). Self-Assembling Protein Microarrays. *Science* 305 (July), 86–90. doi:10.1126/science.1097639

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9 (7), 676–682. doi:10.1038/nmeth.2019

Schloss, A. C., Liu, W., Williams, D. M., Kaufman, G., Hendrickson, H. P., Rudshteyn, B., et al. (2016). Fabrication of Modularly Functionalizable Microcapsules Using Protein-Based Technologies. *ACS Biomater. Sci. Eng.* 2 (11), 1856–1861. doi:10.1021/acsbiomaterials.6b00447

Veggiani, G., Nakamura, T., Brenner, M. D., Gayet, R. V., Yan, J., Robinson, C. V., et al. (2016). Programmable Polyproteams Built Using Twin Peptide Superglues. *Proc. Natl. Acad. Sci. U.S.A.* 113 (5), 1202–1207. doi:10.1073/pnas.1519214113

Williams, D. M., Kaufman, G., Izadi, H., Gahm, A. E., Prophet, S. M., Vanderlick, K. T., et al. (2018). Facile Protein Immobilization Using Engineered Surface-Active Biofilm Proteins. *ACS Appl. Nano Mat.* 1 (6), 2483–2488. doi:10.1021/acsanm.8b00520

Yang, F., Zuo, X., Fan, C., and Zhang, X.-E. (2018). Biomacromolecular Nanostructures-Based Interfacial Engineering: From Precise Assembly to Precision Biosensing. *Natl. Sci. Rev.* 5 (5), 740–755. doi:10.1093/nsr/nwx134

Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., et al. (2012). Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U.S.A.* 109 (12), E690–E697. doi:10.1073/pnas.1115485109

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 | Frontiers in Bioengineering and Biotechnology

OPINION
published: 14 July 2022
doi: 10.3389/fbioe.2022.927156

Exploring Information and Communication Theories for Synthetic Cell Research

Pasquale Stano *

Department of Biological and Environmental Sciences and Technologies (DiSTeBA), University of Salento, Lecce, Italy

Keywords: synthetic cells, artificial cells, bottom-up synthetic biology, information, communication

1 COMMUNICATING SYNTHETIC CELLS

The now-consolidated interest toward the bottom-up construction of cell-like systems, called "artificial" or "synthetic" cells (SCs), is probably one of the most innovative trends in synthetic biology. SCs are micro-compartmentalized chemical systems built from scratch, which are capable of mimicking cell properties and functions at various degrees of complexity. A quite relevant direction refers to SCs that exchange chemical signals with biological cells or with other SCs present in their surroundings (**Figures 1A,B**) [Cronin et al. (2006); Gardner et al. (2009); Stano et al. (2012); Lentini et al. (2017); Adamala et al. (2017); Rampioni et al. (2018); Mukwaya et al. (2021); Smith et al. (2022)]. The biotechnological relevance of communicating SCs points, for example, to their potential applications in nanomedicine, as a sort of "smart" programmable agents interfacing biological cells [Chang (1972); Leduc et al. (2007); Krinsky et al. (2018); Ding et al. (2018); Sato et al. (2021)].

The experimental articles reporting advancements in this field constantly increase, as it does the repertoire of available mechanisms for information processing and SC control, suggesting potential near-the-corner applications. We actually assume that impactful studies will be available soon. This opinion paper, however, does not deal with technical advancements. Instead here we will focus on certain theoretical aspects related to how information and communication concepts impinge on SC research, with possible original returns in modeling, interpretations, understandings. These aspects are crucial when the role of SCs (and synthetic biology, in general) is recognized as one approach—the wetware one—to the "Sciences of the Artificial" [Simon (1996); Cordeschi (2002)], the sciences devoted to modelling life and cognition by mean of man-made artifacts. When devoted to these goals, synthetic biology, then, complements well-known fields such as robotics and artificial intelligence, which instead play a role in the hardware and software domains [Damiano et al. (2011); Damiano and Stano (2018, 2021)].

1.1 Preludes and Current Descriptions

A preliminary remark is due. Despite several relevant advancements, current built-in-the-lab SCs are not yet alive (in particular, they miss key properties such as autonomy and autopoiesis). Looking at their organization, structure, function, current SCs most resemble machines, although endowed of very peculiar features. This machine-likeness, however, makes it possible to discuss and adapt some established concepts of information and communication theory (ICTs) to SCs by exploiting the rich

OPEN ACCESS

Edited by:
Yuan Lu,
Tsinghua University, China

Reviewed by:
Martín Gutiérrez,
Diego Portales University, Chile

***Correspondence:**
Pasquale Stano
pasquale.stano@unisalento.it

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *23 April 2022*
Accepted: *24 June 2022*
Published: *14 July 2022*

Citation:
Stano P (2022) Exploring Information and Communication Theories for Synthetic Cell Research. Front. Bioeng. Biotechnol. 10:927156. doi: 10.3389/fbioe.2022.927156

theoretical framework developed so far, since the age of the first cybernetics[1]. The new tone will be evident, as well as the call to familiarize with new languages[2].

When experiments about communicating SCs are presented, the discussions generally focus on molecular mechanisms, highlighting the nature of components that successfully concurred to establish a chemical communication, and showing whether the system succeeded or not to perceive signals or to communicate. Although the usual vocabulary of molecular biology is explicitly borrowed from ICTs (signal, receptor, encoding, transcription, translation, etc.), minor relevance is given, to date, to information theoretical aspects. The latter, instead, are the main focus in the recently started field of "molecular communications" (MCs) [Nakano et al. (2013); Nakano (2017)]. Here, communication engineers aim at extending the classical Shannon communication theory to the world of chemical signals, to generate and later exploit a sort of molecular communication protocol. The approach is quite general and can be adapted to any kind of chemical system, including the SCs [Magarini and Stano (2021)].

In this article, instead, we will look to still different descriptions of information and communication, pointing to system dynamics from the viewpoints of heteronomy (or machine-likeness, or "computer Gestalt" perspective—as Varela called it [Varela (1979)], or autonomy (or organism-like). The latter would certainly be more adequate to describe living systems, and it will be shortly commented in **Section 4** (as a sort of anticipation to future developments in the field). On the other hand, because current SCs essentially are non-living chemical machines, we maintain that descriptions based on heteronomy are also acceptable and workable for the moment.

The following four topics seem particularly interesting in relation to the above-mentioned goals: 1) the concepts of information and meaning according to Donald M. MacKay [MacKay (1969)]; 2) the cybernetic semiotics discussed by Doede Nauta [Nauta (1972)]; 3) the "in-formation" view of Francisco J. Varela in relation with biological autonomy [Varela (1979)]; and 4) the recent quantitative approach to semantic information proposed by Artemy Kolchinsky and David H. Wolpert [Kolchinsky and Wolpert (2018)].

2 MEANING AND "DESCRIPTIVE" INFORMATION ACCORDING TO DONALD M. MACKAY

The information theory developed in the 1950s by Donald M. MacKay (the main representative of the British information tradition [Bar-Hillel (1964); Nauta (1972)]) is strongly linked to cybernetic concepts, and it offers several interesting cues. We will touch here only a couple of them.

The first refers to the technical usage of the word "meaning" when referred to information (semantic aspects of information). The Shannon theory of communication of information explicitly splits the concepts of meaning and information. The theory only refers to the best and faithful way to transfer information (or better, its representation) from a sender to a receiver. The transmitted information is quantified (by the Shannon entropy function) with reference to the difficulty, or improbability, of selecting a specific signal/message from the set of all possible pre-defined signals/messages. What is, instead, the meaning of a signal? MacKay evidences that the meaning is related to what happens in the receiver, in its internal states, once that the signal/message has been received. The meaning should not be confused with the system response[3], but it refers to how the systems has changed of its propensity to behave in a specific manner. We can add that the system becomes "in-formed" because the signal has caused some changes of internal constraints. MacKay, at this purpose, introduced the concept of conditional "states of readiness". These are system's states, defined operatively, to which the system switches according to a matrix which describes the transition probabilities after interacting with signal from the environment. Consequently, MacKay defines the meaning as the selective function that a signal exerts on the set of the states of readiness. The meaning of a signal admits only a relational definition and is context-dependent.

The second cue instead refers to the difference between the "selective" nature of information (when a transmission context is considered) and the quantification of "descriptive" information, the latter being applied to representations of knowledge, with no reference to unexpectedness (which is at the basis of Shannon entropy). MacKay's descriptive information-content makes use of logons and metrons, which are, respectively, the number of logical classes needed to describe a representation, and the metrical measure of the contribution of each class to the representation. They can be thought as the number of bars and their heights in a histogram representation of a phenomenon, or as the measure of the projections on orthogonal axes of a representational vector in the information space (**Figure 1C**). It can be noted that when the metron measure

[1]There is no doubt, in our view, that if SCs had been available at the time, they would have attracted the interest of early cyberneticians. For example, Rosenblueth, Wiener and Bigelow concluded their 1943 seminal article on *Behavior*, *Purpose* and *Teleology* with a visionary remark: "[. . .] In future years, as the knowledge of colloids and proteins increases, future engineers may attempt the design of robots not only with a behavior, but also with a structure similar to that of a mammal. The ultimate model of a cat is of course another cat, whether it be born of still another cat or synthesized in a laboratory" [Rosenblueth et al. (1943)]. Dealing with cats is premature, yet synthetic biology can partially face the complexity of individual synthetic cells.

[2]Quite probably, most experimentally-oriented synthetic biologists have a background in chemistry and/or biology, but soon become interested also in engineering approaches, like those referring to control theory, or in some mathematical/systems biology aspects related to biological system modeling. Expanding the vocabulary to grasp the concepts described in this article is not a different activity than the above-mentioned examples; we believe it will reward interested readers with fascinating insights and inspiration for potential innovative research lines.

[3]Simple stimulus-response systems do not have sufficient complexity for a semantic information analysis.

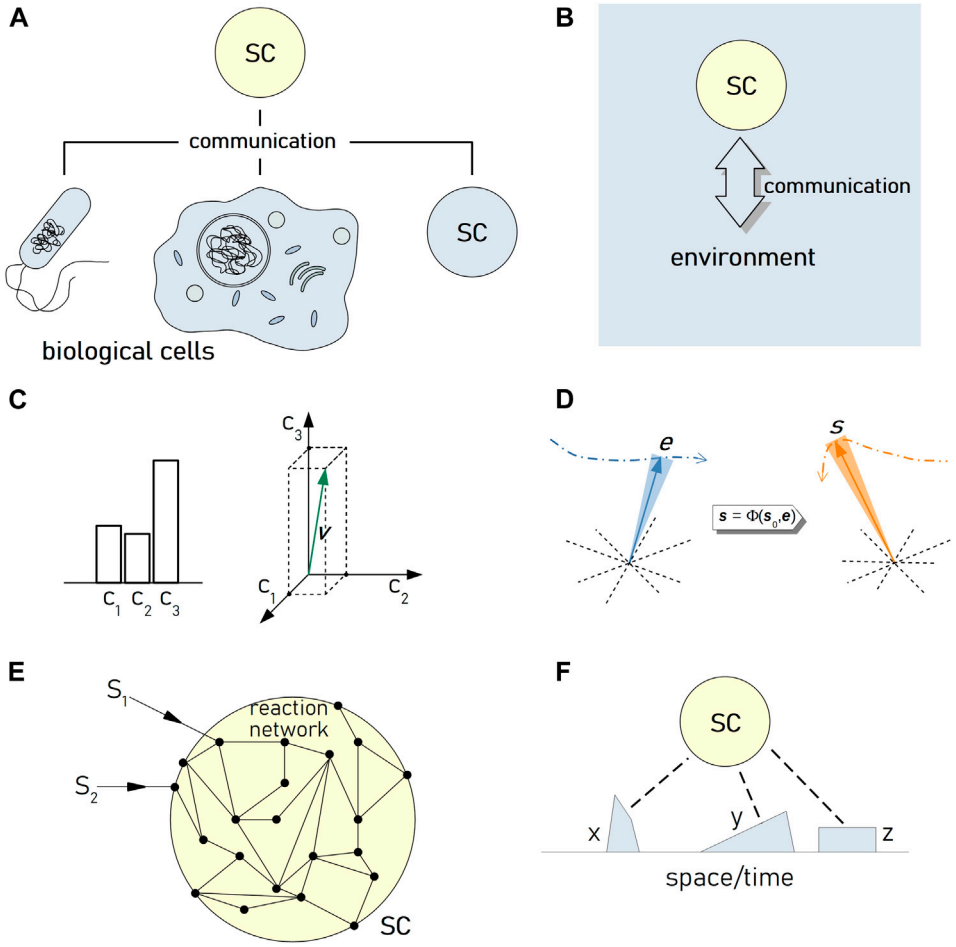


FIGURE 1 | Information and communication in SC research. **(A)** Several reports have demonstrated that SCs can send or receive chemical signals from biological cells or from other SCs. **(B)** Systems based on SCs that receive signals from other cells are equivalent to a generic scenario where a SC (a "system", an "agent") exchange signals with its environment. **(C)** Two ways of representing "descriptive" information. On the left, a histogram made of three bars. Bar identity represents the variable that carries information, bar height represents its measure. On the right, the same information is represented as a vector in an informational space; the variables are described by the axes, while the vector components represent the variables' measures. **(D)** A simplified pictograph representing the trajectories of the environmental vector ***e*** and of the system (the SC) vector ***s*** in their respective multidimensional spaces. The trajectory of ***s*** depends on the initial state of the system ($\boldsymbol{s_0}$) and on the environmental changes (described by ***e***) through the function Φ, which ultimately refers to the "organization" of the system (the set of relations between internal and external variables that determine the system dynamics). **(E)** A SC intended as a reaction network, which can be affected by external stimuli. **(F)** If a SC is *situated* in an environment that can "in-form" it, the resulting dynamics (including the achievement of a predetermined goal) is determined by the environment spatio/temporal pattern.

is not single-valued, the representation becomes a population of vectors, i.e., it becomes less defined, or fuzzy.

Possible applications of the MacKay theory in SCs research can be considered from the modeling perspective. In contrary to biological cells, SCs modeling does not suffer of lack-of-knowledge about conditions and chemical composition. For example, SCs can be built in order to display state transition dynamics with computable transition probability matrices. Once exposed to a set of signals, meaning can be assigned probabilistically, and then experimentally verified on the basis of predicted vs. observed system trajectories. Moreover, SCs could be ranked based on the descriptive information-amount they are able to perceive. The correlation between the environment information vector and the resulting SC state vector (and their changes) (**Figure 1D**) can provide a quantitative measure about how SCs cope with changes in their environment, exploiting probabilistic descriptions.

3 CYBERNETIC SEMIOTICS

Based on the combination of Peirce semiotics with systems theory and cybernetics, Doede Nauta "cybernetic semiotics" [Nauta (1972)] is another possible framework for discussing the dynamics of systems (SCs in our case) perceiving their environment. The theory introduces a classification of signs (signal/sign/symbols) and the domains in which they operate (syntactic, semantic, pragmatic). Emphasis is placed on how a system (in the semiotic language, the "interpreter") that receives information ("information vehicles") from its environment,

copes with it in terms of internal states and operations: the "interpretant" is generated (the effect of information vehicles on the interpreter). The approach somehow resonates with MacKay theories. Nauta also makes a clear distinction between the non-semiotic aspects of transmissional information (à la Shannon) and semiotic processes. The latter refer to the relation between information-vehicles and the cognitive map in the interpreter. Designing SCs that perceive signals or signs of their environment is equivalent to fix what Nauta calls "informational representations"[4] (e.g., interiorized constraints, patterns, correlations) in the molecular network that constitutes the SC organization (**Figure 1E**). This analysis makes it clear that SCs whose behavior is not plastic cannot self-generate meanings of the signals. Meanings—in these cases—are imposed by the designer, as it happens with machines. However, thanks to the modularity of synthetic biology molecular sub-systems and devices, meaning can be *de facto* engineered.

4 IN-FORMATION

We have seen that the two previous theories ground on the high-level symbolic description of informational processes, often applied when systems are interpreted within a machine paradigm. Maturana and Varela, on the other hand, have presented autopoiesis [Maturana and Varela (1980)] as a theoretical framework for understanding and describing life (and in particular cellular life) in terms of mere interactions between the components of a specific type of network, the one which accomplishes its own production. Because the whole phenomenology of living beings can be reduced to, and explained in terms of, interaction and transformation relations between the network constituents, autopoiesis also provides a self-standing operative description of what life (and cognition) is. No additional "symbolic" explanations are required. In this new perspective, how is communication interpreted? Varela maintains that in contrary to the "computer Gestalt" view, where signals from the environment are considered instructive inputs to the system, the "autonomy" view foresees that the environment and the system are engaged in a coupled dynamics, whereby the system adapts to perturbations (originated in the environment) [Varela (1979)]. The perturbation-adaptation dynamics is constrained by the need of maintaining the autopoietic organization, the perturbations playing a constructive (not instructive) role. Strictly speaking, there are no inputs for an autopoietic network. Autonomous systems deal with their environment in a cognitive ("in-formative") manner, due to their plasticity. This feature, together with the self-distinction, provides autonomous systems of mind-like character [Varela (1979)]. While the cybernetic interpretation of information and communication remains useful and workable for systems with restricted autonomy, Varela suggested that discussions on information and living systems require a different perspective. This co-dependent, constructive, correlational sense of information {dubbed: "in-formation", that which is formed within [Bateson (1972)]} should be intended as the structural adaptations of the system to environment perturbations, without the need of symbolic representations and mappings of any sort. The autonomy perspective, whereby a system engages with the environment a co-constructive dynamics can be applied in SC research as a guideline or a framework to for the long-term goal of constructing autopoietic and cognitive, and therefore living, SCs.

5 MEASURING SEMANTIC INFORMATION: AN APPROACH BASED ON SYSTEM DYNAMICS

Kolchinsky and Wolpert (2018) have proposed an operational definition of semantic information based on the coupled dynamics of an environment/system whole [Kolchinsky and Wolpert (2018)]. In particular, they suggested that "semantic information is the syntactic information that a physical system has about its environment which is causally necessary for the system to maintain its own existence". The approach is quite interesting and may constitutes an original framework in SC modeling. In order to adapt it to current SCs, the condition of "existence", which recalls the concepts of being alive or at least maintaining a structural or dynamical organization, can be provisionally substituted with less demanding properties, such as performing significant operations (i.e., significant for the observer). Imagine, then, responsive SCs that are situated in an environment where signals have certain spatio-temporal distributions (**Figure 1F**). The resulting SC behavior, even if determined by the internal chemical network, will depend from the signal patterns too. It is expected that only a sub-set of all possible environmental distributions will best "matches" the constraints *embodied* in the SC organization, leading to successful operations. The amount of semantic information will correspond to the threshold value of environment/SC mutual information, which must be overcome to transition from fail to success[5].

6 WHY SHOULD THESE TOPICS BE OF INTEREST TO SYNTHETIC BIOLOGISTS?

The recent growth of interest in SCs has led to sophisticated systems, constructed from scratch, which are able to exchange chemical signals with the environment and with other cells. One well-known remarkable example is the establishment of bidirectional communication between SCs and living cells [Lentini et al. (2017)], but more recent reports are similarly

[4]They are classified as 1) "implicit" when refer to simple signal/response dynamics; and 2) "concursive", when the interpreter deploys its cognitive network to orient itself according to received signals or signs.

[5]As a brief note connected with **Section 3**, we mention here a recently published report dealing with a potential "physiosemiotic" interpretation of the Kolchinsky-Wolpert approach [Herrmann-Pillath (2021)].

exciting. These developments mean that SC technology can become a reliable, versatile, powerful and pivotal platform to make one step further, and face either fundamental questions (e.g., what does it mean to be cognitive? Can meaning emerge in artificial systems? Can we engineer it?), either practical applications, such as developing cognitive artificial systems in the chemical domain and interface them with biological systems for whatever purpose—to parallel the impressive developments of artificial intelligence and robotics in the software and hardware domains, aimed at interfacing with humans. To progress the synthetic biology field, we believe that experiments and modeling should be firstly devised within well-thought-out theoretical frameworks, which in simpler cases have been just tacitly understood. For example, it can be agreed that targeting stimulus-response dynamics, even if experimentally challenging, does not require theoretical analyses like the one presented in this article. In contrary, devising an intelligent, or cognitive, or adaptive, or autonomous, or plastic artificial system (just to make some examples) presupposes a preliminary understanding of what these terms mean and under which framework should be understood. Information and communication theories (facing both syntactic and semantic approaches), and other theories as well, can guide more complex implementations, spark novel ideas, be used to confirm or reject working hypotheses.

7 CONCLUDING REMARKS

A fecund and unexplored area emerges clearly when one considers theoretical implications of SC research in the information and communication arenas, namely the construction of systems that interact with their environment [Damiano and Stano (2018; 2020)]. As mentioned, this research area is already under bright development, but the approaches, the results, and the descriptions have been generally reported and interpreted under the canonical lens of biochemical/molecular biology. This is fine for most of the scopes. In this article we propose that the same research can be also developed into new directions, under the guidance of some theoretical approaches, and try to convince that new concepts, languages, theories can further enrich the synthetic cell research. We hope that this short article will stimulate several scholars to start looking to these opportunities. Indeed, we are convinced that proper developments will definitely contribute not only to reach experimentally valuable goals, but also to advance in the sciences of artificial, with new synthetic approaches to model and understand life and cognition. By means of SC technology, synthetic biology participates to this fundamental challenge with one of the most innovative, multifaceted and versatile instruments.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The Author is grateful to Luisa Damiano (IULM, Milan, Italy) for stimulating discussions on Autopoiesis and on the Sciences of the Artificial. Investigations on the application of the semantic information of Kolchinsky-Wolpert to SCs are underway in collaboration with Maurizio Magarini (Politecnico di Milano, Milan, Italy).

REFERENCES

Adamala, K. P., Martin-Alarcon, D. A., Guthrie-Honea, K. R., and Boyden, E. S. (2017). Engineering Genetic Circuit Interactions within and between Synthetic Minimal Cells. *Nat. Chem.* 9, 431–439. doi:10.1038/nchem.2644

Bar-Hillel, Y. (1964). *Language and Information*. Reading, MA: Addison-Wesley.

Bateson, G. (1972). *Steps to an Ecology of Mind*. Northvale, New Jersey London: Jason Aronson.

Chang, T. M. S. (1972). *Artificial Cells*. Springfield, IL: Charles C Thomas.

Cordeschi, R. (2002). "The Discovery of the Artificial," in *Behavior, Mind and Machines Before and Beyond Cybernetics*. Studies in Cognitive Systems (Netherlands: Springer Netherlands).

Cronin, L., Krasnogor, N., Davis, B. G., Alexander, C., Robertson, N., Steinke, J. H. G., et al. (2006). The Imitation Game-A Computational Chemical Approach to Recognizing Life. *Nat. Biotechnol.* 24, 1203–1206. doi:10.1038/nbt1006-1203

Damiano, L., Hiolle, A., and Cañamero, L. (2011). "Grounding Synthetic Knowledge," in *Advances in Artificial Life, ECAL 2011*. Editors T. Lenaerts, M. Giacobini, H. Bersini, P. Bourgine, M. Dorigo, and R. Doursat (Boston: MIT Press), 200–207.

Damiano, L., and Stano, P. (2021). A Wetware Embodied AI? towards an Autopoietic Organizational Approach Grounded in Synthetic Biology. *Front. Bioeng. Biotechnol.* 9, 724023. doi:10.3389/fbioe.2021.724023

Damiano, L., and Stano, P. (2020). On the "Life-Likeness" of Synthetic Cells. *Front. Bioeng. Biotechnol.* 8, 953. doi:10.3389/fbioe.2020.00953

Damiano, L., Stano, P., and Stano, P. (2018). Synthetic Biology and Artificial Intelligence: Grounding a Cross-Disciplinary Approach to the Synthetic Exploration of (Embodied) Cognition. *Complex Systems* 27, 199–228. doi:10.25088/ComplexSystems.27.3.199

Ding, Y., Contreras-Llano, L. E., Morris, E., Mao, M., and Tan, C. (2018). Minimizing Context Dependency of Gene Networks Using Artificial Cells. *ACS Appl. Mat. Interfaces* 10, 30137–30146. doi:10.1021/acsami.8b10029

Gardner, P. M., Winzer, K., and Davis, B. G. (2009). Sugar Synthesis in a Protocellular Model Leads to a Cell Signalling Response in Bacteria. *Nat. Chem.* 1, 377–383. doi:10.1038/nchem.296

Herrmann-Pillath, C. (2021). The Natural Philosophy of Economic Information: Autonomous Agents and Physiosemiosis. *Entropy (Basel)* 23 (3), 277. doi:10.3390/e23030277

Kolchinsky, A., and Wolpert, D. H. (2018). Semantic Information, Autonomous Agency and Non-equilibrium Statistical Physics. *Interface Focus* 8, 20180041. doi:10.1098/rsfs.2018.0041

Krinsky, N., Kaduri, M., Zinger, A., Shainsky-Roitman, J., Goldfeder, M., Benhar, I., et al. (2018). Synthetic Cells Synthesize Therapeutic Proteins inside Tumors. *Adv. Healthc. Mat.* 7, 1701163. doi:10.1002/adhm.201701163

Leduc, P. R., Wong, M. S., Ferreira, P. M., Groff, R. E., Haslinger, K., Koonce, M. P., et al. (2007). Towards an *In Vivo* Biologically Inspired Nanofactory. *Nat. Nanotech* 2, 3–7. doi:10.1038/nnano.2006.180

Lentini, R., Martín, N. Y., Forlin, M., Belmonte, L., Fontana, J., Cornella, M., et al. (2017). Two-Way Chemical Communication between Artificial and Natural Cells. *ACS Cent. Sci.* 3, 117–123. doi:10.1021/acscentsci.6b00330

MacKay, D. M. (1969). *Information, Mechanism and Meaning*. Cambridge MA: MIT Press.

Magarini, M., and Stano, P. (2021). Synthetic Cells Engaged in Molecular Communication: An Opportunity for Modelling Shannon- and Semantic-Information in the Chemical Domain. *Front. Comms. Net.* 2, 48. doi:10.3389/frcmn.2021.724597

Maturana, H. R., and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. 1st edn. D. Reidel Publishing Company.

Mukwaya, V., Mann, S., and Dou, H. (2021). Chemical Communication at the Synthetic Cell/living Cell Interface. *Commun. Chem.* 4, 1–12. doi:10.1038/s42004-021-00597-w

Nakano, T., Eckford, A. W., and Haraguchi, T. (2013). *Molecular Communications*. Cambridge UK: Cambridge University Press.

Nakano, T. (2017). Molecular Communication: A 10 Year Retrospective. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* 3, 71–78. doi:10.1109/TMBMC.2017.2750148

Nauta, D. (1972). *The Meaning of Information. Approaches to Semiotics [AS]*. The Hague: Mouton (De Grutyer).

Rampioni, G., D'Angelo, F., Messina, M., Zennaro, A., Kuruma, Y., Tofani, D., et al. (2018). Synthetic Cells Produce a Quorum Sensing Chemical Signal Perceived byPseudomonas Aeruginosa. *Chem. Commun.* 54, 2090–2093. doi:10.1039/C7CC09678J

Rosenblueth, A., Wiener, N., and Bigelow, J. (1943). Behavior, Purpose and Teleology. *Philos. Sci.* 10, 18–24. doi:10.1086/286788

Sato, W., Zajkowski, T., Moser, F., and Adamala, K. P. (2021). Synthetic Cells in Biomedical Applications. *WIREs Nanomed Nanobiotechnol* 14, e1761. doi:10.1002/wnan.1761

Simon, H. A. (1996). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Smith, J. M., Chowdhry, R., and Booth, M. J. (2021). Controlling Synthetic Cell-Cell Communication. *Front. Mol. Biosci.* 8, 809945. doi:10.3389/fmolb.2021.809945

Stano, P., Rampioni, G., Carrara, P., Damiano, L., Leoni, L., and Luisi, P. L. (2012). Semi-synthetic Minimal Cells as a Tool for Biochemical ICT. *BioSystems* 109, 24–34. doi:10.1016/j.biosystems.2012.01.002

Varela, F. J. (1979). *Principles of Biological Autonomy. The North-Holland Series in General Systems Research*. New York: Elsevier/North Holland.

Conflict of Interest: The author declares 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 | Frontiers in Bioengineering and Biotechnology

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fbioe.2022.906295

OPEN ACCESS

EDITED BY
Yuan Lu,
Tsinghua University, China

REVIEWED BY
Cesare Orlandi,
University of Rochester, United States
Paul Raymond Gooley,
University of Melbourne, Australia

*CORRESPONDENCE
Frank Bernhard,
fbern@bpc.uni-frankfurt.de

SPECIALTY SECTION
This article was submitted to Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and Biotechnology

RECEIVED 28 March 2022
ACCEPTED 27 June 2022
PUBLISHED 22 July 2022

CITATION
Umbach S, Levin R, Neumann S, Steinmetzer T, Dötsch V and Bernhard F (2022), Transfer mechanism of cell-free synthesized membrane proteins into mammalian cells.
Front. Bioeng. Biotechnol. 10:906295.
doi: 10.3389/fbioe.2022.906295

Transfer mechanism of cell-free synthesized membrane proteins into mammalian cells

Simon Umbach[1], Roman Levin[1], Sebastian Neumann[2], Torsten Steinmetzer[2], Volker Dötsch[1] and Frank Bernhard[1]*

[1]Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University, Frankfurt am Main, Germany, [2]Institute for Pharmaceutical Chemistry, Philipps University, Marburg, Germany

Nanodiscs are emerging to serve as transfer vectors for the insertion of recombinant membrane proteins into membranes of living cells. In combination with cell-free expression technologies, this novel process opens new perspectives to analyze the effects of even problematic targets such as toxic, hard-to-express, or artificially modified membrane proteins in complex cellular environments of different cell lines. Furthermore, transferred cells must not be genetically engineered and primary cell lines or cancer cells could be implemented as well. We have systematically analyzed the basic parameters of the nanotransfer approach and compared the transfer efficiencies from nanodiscs with that from Salipro particles. The transfer of five membrane proteins was analyzed: the prokaryotic proton pump proteorhodopsin, the human class A family G-protein coupled receptors for endothelin type B, prostacyclin, free fatty acids type 2, and the orphan GPRC5B receptor as a class C family member. The membrane proteins were cell-free synthesized with a detergent-free strategy by their cotranslational insertion into preformed nanoparticles containing defined lipid environments. The purified membrane protein/nanoparticles were then incubated with mammalian cells. We demonstrate that nanodiscs disassemble and only lipids and membrane proteins, not the scaffold protein, are transferred into cell membranes. The process is detectable within minutes, independent of the nanoparticle lipid composition, and the transfer efficiency directly correlates with the membrane protein concentration in the transfer mixture and with the incubation time. Transferred membrane proteins insert in both orientations, N-terminus in and N-terminus out, in the cell membrane, and the ratio can be modulated by engineering. The viability of cells is not notably affected by the transfer procedure, and transferred membrane proteins stay detectable in the cell membrane for up to 3 days. Transferred G-protein coupled receptors retained their functionality in the cell environment as shown by ligand binding, induction of internalization, and specific protein interactions. In comparison to transfection, the cellular membrane protein concentration is better controllable and more uniformly distributed within the analyzed cell population. A further notable difference to transfection is the accumulation of transferred membrane proteins in clusters, presumably determined by microdomain structures in the cell membranes.

KEYWORDS

G-protein coupled receptors, cell-free expression, nanodiscs, Salipro nanoparticles, transfection, protein transfer, GPCR function, HEK 293 cell

1 Introduction

The *in vivo* analysis of membrane proteins (MPs) in cellular systems frequently suffers from low endogenous concentrations, unknown regulation mechanisms, or heterogeneous expression levels in a cell population. Recombinant MP production by genetic tools such as DNA or RNA transfection can address these issues, but still, a variety of problems such as MP toxicity by blocking translocon systems, misfolding, wrong trafficking, or failure to insert into cellular membranes may occur (Kim and Eberwine, 2010). Furthermore, artificial modifications of MPs by directed engineering, by attachment of probes for monitoring, or by modulation of MP complexes can be difficult or even hardly possible. The direct transfer of purified MPs into live cells could eliminate many problems and would be a straightforward approach to analyzing the *in vivo* effects of designed MPs by potentially having better control on protein quality, modifications, and copy number per cell.

The cotranslational insertion of MPs into nanodisc (ND) membranes by cell-free (CF) expression is an efficient method to produce a variety of functional MPs including G-protein coupled receptors (GPCRs), transporters, channels, or enzymes (Henrich et al., 2016; Henrich et al., 2017b; He et al., 2017; Rues et al., 2018; Keller et al., 2019; Kuroiwa et al., 2022). The NDs are first pre-assembled with the lipid or lipid mixture that is best suited for the stability and folding of the selected MP. CF expression reactions of the MP are then performed in presence of suitable concentrations of the supplied NDs (Henrich et al., 2016; Peetz et al., 2017). The technique gives fast access to even difficult proteins and MP/ND samples can be generated within 24 h. As no live cells are involved, any potentially toxic effects during MP production are eliminated. In addition, the MPs do not have contacts with destabilizing detergents during the whole synthesis and purification process. The open nature of CF systems further enables a high potential to modify the synthesized MPs, e.g., by a variety of engineering approaches, by attachment of probes, or by complex formation with artificial ligands.

A continuously growing toolkit for the transfer of soluble proteins into live cells by electroporation, direct injection, inclusion into liposomes, or by using cell-penetrating peptides is available (Chau et al., 2020). In contrast, the transfer of MPs is much less explored despite their high importance for pharmaceutical research (Overington et al., 2006). A problematic issue is the requirement of a suitable membrane mimetic to keep the MP soluble and stable while also being compatible with the transfer approach (Majeed et al., 2021). A significant advantage for MP transfer was the implementation of NDs as vectors (Denisov and Sligar, 2017). If NDs are mixed with other lipid bilayers such as micelles, supported bilayers, or even lipid cubic phases, lipids and inserted MPs can be transferred (Lai et al., 2015; Henrich et al., 2017a; Nikolaev et al., 2017; Dang et al., 2019). Furthermore, NDs have been used to transfer a variety of other biomolecules such as RNA, optical probes, or drugs into living cells (Murakami, 2012; Carney et al., 2015; He et al., 2020).

The successful nanotransfer of CF synthesized MPs, such as the human ß2-adrenergic receptor and the Her2 receptor tyrosine kinase, from NDs into live cells was already demonstrated (Patriarchi et al., 2018; He et al., 2021). This new process has a high potential to efficiently close a gap between the *in vitro* and *in vivo* analysis of MPs. Aliquots of CF-generated MP/ND samples can simultaneously be analyzed *in vitro* as well as in cell-based assays. During the complete procedure, the MPs stay in lipid environments and *in vitro* functions in defined lipid bilayers can be compared with those in the much more complex environment of living cells.

Mechanism and parameters important for the modulation of an efficient MP transfer from NDs into cells are still mostly unknown. Furthermore, details about the membrane topology and stability of transferred MPs are lacking. We have studied kinetic details of the MP transfer mechanism and analyzed the effect of key compounds such as the type of nanoparticle, lipid composition, or cell line on the transfer efficiency. As examples for human class A GPCRs, we have selected the endothelin B receptor (ETB), the prostacyclin receptor (IP), and the free fatty acid receptor 2 (FFAR2). In addition, the human orphan receptor GPRC5B as a member of the class C family of GPCRs and the seven-transmembrane domain-containing proton pump proteorhodopsin (PR) were selected as model MPs. We demonstrate that the functional nanotransfer of CF-synthesized MPs from NDs and also from Salipro particles into cellular membranes is a rather universal mechanism. We show the functionality of transferred GPCRs in mammalian cells and compare their localization with that of GPCR controls synthesized by transfection. The orientation of transferred MPs in the cell membrane was analyzed and modulated by engineering. As a particular feature of the nanotransfer approach, we first demonstrate that in contrast to MP production in transfected cells, a significant fraction of the transferred MPs is organized in membrane-located clusters.

2 Methods

DNA constructs and ligands: sequences encoding for human GPCRs (Supplementary Table S1) were codon-optimized for *E. coli* and flanked by an N-terminal Htag for improved CF expression yields and a C-terminal Strep-tag for purification and

detection (Haberstock et al., 2012). ETB constructs were based on the thermostabilized Y5 mutant and synthesized as published in some studies (Okuta et al., 2016; Rues et al., 2018). To monitor the nanotransfer, GPCR-mNeonGreen (mNG) fusion constructs were created by subcloning the mNG sequence downstream of the GPCR sequences via KpnI and XhoI restriction sites. For immunostaining experiments, a Myc-tag was inserted either between the Htag and the GPCR or between the GPCR and the Strep-tag using a standard quick-change site-directed mutagenesis approach. Flag-tagged MSP1E3FN was created by inserting the Flag-sequence between the N-terminal His6-TEV sequence and the MSP gene via quick change. This enables the removal of the His-tag after protein expression and purification, resulting in an N-terminally Flag-tagged MSP.

The circular ligand cET-1 for the endothelin type B receptor was obtained as described in Proverbio et al. (2013). The synthesis of biotinylated 4-Ala-ET-1 and DY647-4-Ala-Arg9-ET-1 is described in detail in the supplementary materials section.

Nanoparticles: NDs were performed with purified membrane scaffold protein (MSP) as described in previous studies (Rues et al., 2018; Köck et al., 2021). If not indicated otherwise, MSP1E3D1 was used for ND preparation, resulting in NDs with a diameter of 10–12 nm, which was a suitable compromise between lipid bilayer size and stability, since NDs with larger MSPs tend to be less stable. MSP1E3D1:lipid ratios were 1:80 (DEPG), 1:80 (DOPG), 1:90 (POPG) and 1:115 (DMPC). Rho-PE containing NDs was prepared by adding 2% (lissamine rhodamine B)-PE to DEPG before ND formation.

For SapA nanoparticle (SapNP) preparation, DOPG liposomes were prepared by first solubilization of DOPG in chloroform followed by vacuum evaporation. The dried lipid bilayers were then suspended in 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, and adjusted to a 30 mg/ml suspension. SapA was purified as described elsewhere (Frauenfeld et al., 2016) and concentrated to 20–30 mg/ml (1.7–2.6 mM). Subsequently, SapA solution and lipid suspension were mixed to yield a stoichiometric ratio of 1:24 SapA:DOPG. To induce particle formation, pH was shifted to 4.8–5.0 using 5% (v/v) acetic acid. Upon acidification, the turbid suspension was clarified and the solution was then rapidly diluted 1:10 with 50 mM Tris-HCl, pH 8.0, 50 mM NaCl to neutralize pH. The solution was then centrifuged at 30,000 x g for 20 min to remove insoluble material and the final concentration of formed particles was estimated via SapA concentration by UV-spectroscopy with ε = 10.345 M^{-1} cm^{-1} for non-TEV treated SapA. Total protein concentration should be 20–30 mg/ml (1.7–2.6 mM) to yield 400 µM total SapA in the CF mix which corresponds to 100–200 µM total SapNPs, considering that one particle is assembled with 2-4 SapA conformers.

CF expression: Two-compartment CF expression was based on *E. coli* A19 S30 lysate and performed as described in previous studies (Schwarz et al., 2007; Köck et al., 2021). Briefly, a semipermeable membrane (12–14 kDa cut-off) is used to separate a reaction mixture (RM) from a feeding mixture (FM). The RM contains all the necessary components for transcription and translation. The FM provides additional amino acids and energy sources and helps to dilute inhibitory by-products from the RM. Typically, reaction containers with 55 µl RM and 800 µl FM were used. Larger reaction volumes (1–3 ml RM) were incubated in dialysis cassettes. The final Mg^{2+} concentration was adjusted to 18 mM. Depending on the receptor, either DTT (final concentration 2 mM) or a 3:1 mixture of reduced and oxidized glutathione (GSH/GSSG; final concentration 3 and 1 mM) was used as a redox system. If not indicated otherwise, 60 µM preformed NDs were used in the RM containing lipids found to be suitable for the GPCR integration (Rues et al., 2018): DEPG (ETB and IP derivatives) and POPG (FFAR2 and GPRC5B derivatives). Different lipids were used for PR as indicated in the results section. Expression was performed overnight at 30°C under slight shaking. Afterward, precipitates were removed via centrifugation for 10 min at 18,000 x g and the synthesized MP/nanoparticle complexes were purified via the Strep-tag of the MP from the supernatant.

Protein purification and analysis: receptors were purified from the RM by Strep-Tactin affinity chromatography using a gravity flow column equilibrated in purification buffer (100 mM Tris-HCl, pH 8.0, 100 mM NaCl). Elution was done with *d*-desthiobiotin (25 mM, in purification buffer), which was subsequently removed through washing in Amicon centrifugal filters (50 kDa MWCO).

Size exclusion chromatography (SEC) was used for characterization and purification of the synthesized MPs. SEC was performed using a Superose 6 increase 3.2/300 column. For preparative SEC, samples were fractionated and fractions of interest were pooled and concentrated. For ligand binding and activity studies of ETB-mNG, purification was done via ligand affinity chromatography (LAC) using immobilized biotinylated 4Ala-ET-1. Therefore, biotinylated 4Ala-ET-1 in binding buffer was incubated with magnetic Strep-Tactin beads (MagStrep "type3" XT, IBA) overnight. ETB-mNG was bound for 1 h at 4°C while gently mixing by inversion. The beads were washed with binding buffer and the complex was eluted using a 10-fold excess of 4Ala-ET-1. Receptor-bound 4Ala-ET-1 was then removed by extensive washing in centrifugal filters. After purification, samples were concentrated in centrifugal filters and protein concentration was determined either via mNG fluorescence or via nanodrop measurements for receptors without fluorophore. Concentrations were calculated assuming a single receptor per ND.

For western blotting, samples were loaded onto a 4–15% continuous Mini-PROTEAN TGX gel (BioRad) and separated for 30 min at 200 V. Blotting was done using the Trans-Blot Turbo System (BioRad). Membranes were blocked with 4% skim milk powder for 1 h at room temperature. Primary antibodies

(anti-Myc (1:2000, 4A6, Millipore); anti-HA (1:1,000, A190-138A, Bethyl); anti-His (1:2,000, H1029)) were applied for 1 hour. After washing, secondary antibodies (anti-mouse-HRP (1:5,000, A9917, Sigma)) were applied for another hour at RT. For the detection of Strep-tagged proteins, an anti-Strep HRP conjugate was used (1:5,000; #1610381, BioRad). All antibody incubation and washing steps were done in PBS-T (0.05% Tween-20). Chemiluminescence was analyzed using a Lumi-Imager F1 (Roche).

Cell culture: CHO-K1, HEK293T, and H1299 cells were grown in Ham's F-12 medium, DMEM and RPMI, respectively. All media were supplemented with 10% fetal calf serum and 1% penicillin/streptomycin. All cell lines were grown at 37°C and 5% CO_2 and tested regularly negative for *mycoplasma* contamination. Transfections were performed using the Lipofectamine 2000 kit (ThermoFisher). For fluorescence microscopy, cells were seeded at a density of 2×10^5 cells/well onto glass coverslips in a 12-well plate. To prevent detachment of cells during washing steps, the medium for HEK293T cells was supplemented with 0.2 µg/ml poly-L-lysine.

Nanotransfer and confocal fluorescence microscopy: for nanotransfer experiments, cells seeded on coverslips in 12-well plates were washed once with 1 mL Dulbecco's phosphate-buffered saline (DPBS) and then incubated with 500 µL standard media supplemented with a defined concentration of nanoparticles. Cells were then kept at 37°C and 5% CO_2. After incubation, cells were washed five times with 1 mL DPBS and fixed for 10 min in RotiHistofix. After three additional washing steps to remove the residual fixation agent, coverslips were placed on microscope slides with a DAPI-containing mounting medium. If not stated otherwise, 0.5 µM nanoparticles and 4 h of incubation were used as standard conditions for all nanotransfer experiments.

Immunostainings were carried out by permeabilizing cells after fixation with 0.2% Triton X-100 in DPBS for 20 min. After three washing steps, cells were blocked for 1 h with DPBS containing 1% BSA. Primary antibodies (anti-Myc (1:100, 4A6, Millipore); anti-Flag (1:100, F3165, Sigma)) were added for 1 h. Cells were washed again and the secondary antibody (anti-mouse Alexa647 conjugate (1:500, A31571, life technologies)) was added. After washing, cells were mounted. All antibody incubation and washing steps were carried out in DPBS with 1% BSA at RT.

To evaluate ligand binding of ETB derivatives, HEK293T cells were washed five times with DPBS after transfer of LAC-purified ETB-mNG, and fresh media containing 100 nM DY647-ET-1 was added. Cells were incubated for 1 h at 4°C to prevent internalization. Afterward, they were washed with DPBS and mounted. For the internalization assay, cells were incubated with cET-1 for 1 h at 4°C. After ligand incubation, cells were transferred back to 37°C for 1 h before mounting.

Images were acquired using an inverted Zeiss Observer Z1 with a Yokogawa CSU-X1A 5000 spinning disc unit and an EMCCD camera. Imaging was done through a ×63 oil objective. Four lasers of 405, 488, 561, and 638 nm with corresponding 450/50, 485/30, 562/45, 690/50 nm bandpass filters for blue-, green-, orange-, and red-emitting fluorophores, respectively. Images were quantified using ImageJ. Membrane fluorescence intensity was calculated by tracing the plasma membrane of a number of randomly picked cells.

Co-IP of transfected and delivered GPCRs: for Co-IP experiments, HEK293T cells were seeded into 6-well plates at a density of 1.5×10^6 cells/well. After 24 h, cells were transfected with 2 µg per well plasmid DNA encoding for either GPRC5B-Myc or HA-IP. 6 h after transfection, the culture medium was removed and a fresh medium containing 0.5 µM nanoparticles with or without GPCR was added. After 16 h, cells were washed five times with ice-cold DPBS and lysed in lysis buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 5% glycerol) supplemented with complete protease inhibitor for 30 min on ice. The lysate was centrifuged at 22,000 × g for 15 min to remove cellular debris. The cleared lysate was then incubated with magnetic Strep-Tactin beads (MagStrep "type3" XT, IBA) for 2 h at 4°C under constant rotation. Afterward, beads were washed three times with ice-cold lysis buffer. SDS loading buffer was added and the sample was incubated at 95°C for 10 min before SDS-PAGE and western blot.

3 Results

3.1 CF production and preparation of MP/ND samples

The GPCRs and PR were synthesized in an *E. coli*-based CF system and cotranslationally inserted into preformed NDs assembled with lipids found to be most suitable for the stability of the selected MPs. The production efficiency of the MP/ND complexes was between 0.7 and 1.1 mg/ml of RM for all analyzed targets. MP/ND complexes were purified via the C-terminal Strep-tag of the MPs. SEC profiling of the purified MP/ND complexes was used to evaluate the sample homogeneity (Figure 1). With all GPCR samples, two major SEC fractions were detectable. While fraction 2 is assumed to contain homogenous MP/ND particles, fraction 1 may include MP/ND aggregates. With PR, the most homogeneous profile, showing only fraction 2 was observed. GPCRs contain at least one disulfide bridge essential for their functional folding and thus the redox conditions during CF expression were optimized. ETB and ETB-tc were synthesized in presence of the redox system GSH/GSSG according to previously established protocols (Dong et al., 2018). GPRC5B/ND, FFAR2/ND, and IP/ND particles have not been CF synthesized before and their

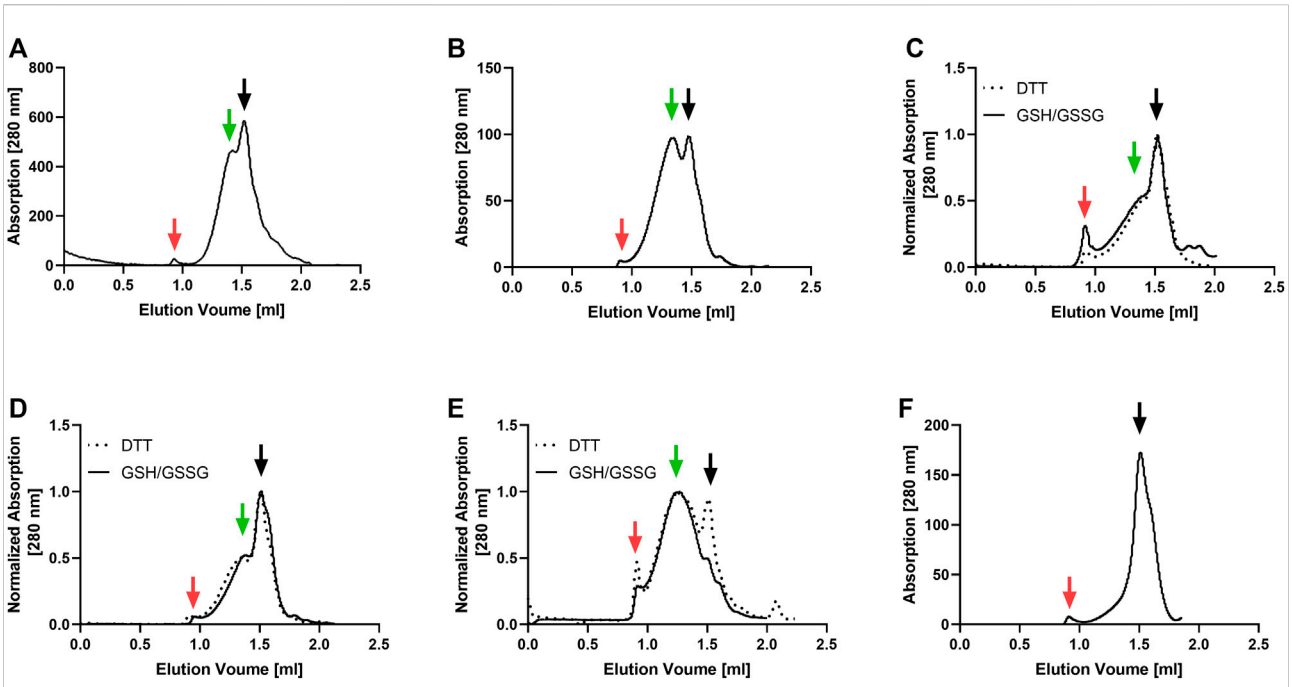


FIGURE 1
SEC profiling of MP/ND samples used for nanotransfer into mammalian cells. The receptors were cotranslationally inserted into preformed NDs, Strep-purified and analyzed with a Superose 6 increase 3.2/300 column. Solid and dotted lines correspond to elution profiles of GPCR/ND samples CF synthesized in presence of either a GSH/GSSG or a DTT redox system. **(A)** ETB **(B)** ETB-tc; **(C)** FFAR2; **(D)** GPRC5B; **(E)** IP; **(F)** PR. Void peak (red), fraction 2 (black) and fraction 1 (green) are indicated with arrows.

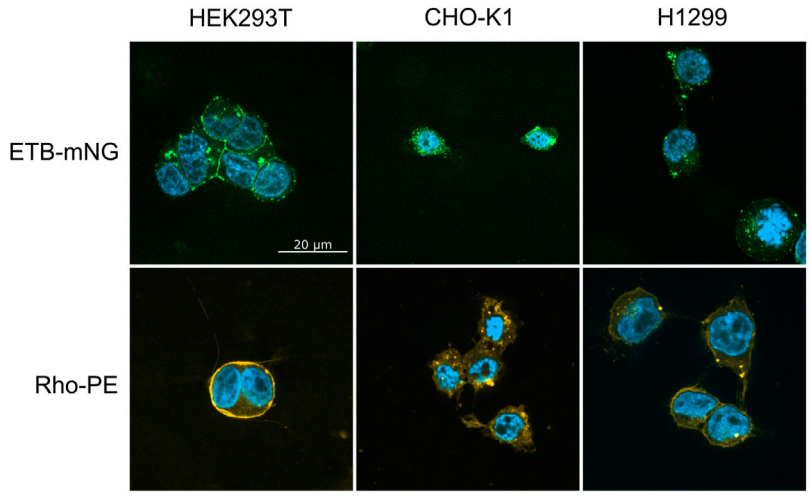


FIGURE 2
Transfer of ETB-mNG and Rho-PE from NDs into different mammalian cell lines. 0.5 µM of NDs (DEPG) containing either ETB-mNG or 2% Rho-PE were incubated for 4 h with HEK293T, CHO-K1, or H1299 cells. After washing and fixation, cells were analyzed using a spinning disk fluorescence microscope. Representative images are shown.

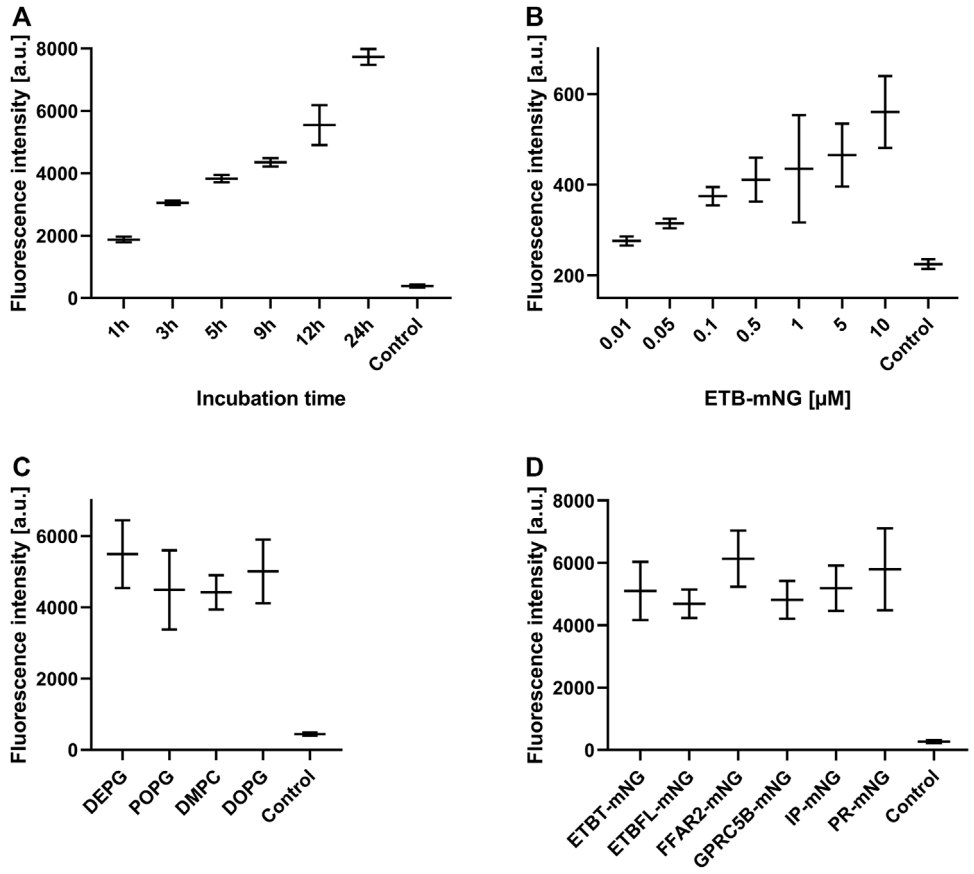


FIGURE 3
Evaluation of MP transfer efficiency into HEK293T cells. HEK293T cells were incubated with Strep-purified MP/ND particles at standard conditions (0.5 µM MP/ND particles, 4 h transfer) with the indicated modifications. After transfer, cells were fixed, and membrane fluorescence was quantified. **(A)** Effect of incubation times on ETB-mNG transfer. **(B)** Effect of ETB-mNG/ND particle concentration. **(C)** Transfer of PR-mNG from NDs containing either DEPG, POPG, DMPC, or DOPG lipids. **(D)** Transfer of GPCR-mNG and PR-mNG fusions from NDs containing customized membrane compositions (ETB-mNG/ND (DEPG), ETB-tc-mNG/ND (DEPG), FFAR2-mNG/NDs (POPG), GPRC5B-mNG/NDs (POPG), IP-mNG/NDs (DEPG) and PR-mNG/NDs (DEPG)). Mean ± SEM is shown. N = 20 cells. Controls are fluorescence of HEK293T cells after incubation with NDs without MPs.

quality was therefore analyzed after CF expression in presence of either DTT or GSH/GSSG as a redox system (Figure 1). A significant improvement of IP/ND homogeneity toward fraction 2 was monitored after synthesis in presence of DTT. With GPRC5B/NDs and FFAR2/NDs, no differences in the SEC profiles after production in the two redox systems could be detected.

3.2 Basic parameters of ND-mediated MP transfer into mammalian cells

In order to better understand the fundamental principles and mechanisms of the nanotransfer approach, a number of potentially relevant parameters including different cell lines and culture conditions as well as modified MPs and various ND lipid compositions were analyzed. First, 0.5 µM NDs (DEPG) containing either ETB-mNG or 2% Rho-PE lipids were incubated with HEK293T, CHO-K1, or H1299 cells for 4 h at 37°C. After washing and fixation, the cells were analyzed by fluorescence microscopy (Figure 2). In HEK293T cells, the transferred ETB-mNG and Rho-PE were the most evenly distributed in the plasma membrane, whereas in H1299 and in particular in CHO-K1 cells fluorescent spots appear to be more localized within the cytoplasm. Fluorescence quantification of representative ETB-mNG transferred cells further confirmed the primary localization of ETB-mNG in the HEK293T plasma membrane (Supplementary Figure S1). Notably, some cluster formation of the transferred ETB-mNG in the HEK293T membrane can be observed, while the distribution of the Rho-PE lipid appears to be more evenly. Based on the more preferential localization of the transferred compounds in the cell membrane, HEK293T cells

were selected for further experiments. The incubation with 0.5 μM MP/ND particles for 4 h at 37°C were taken as standard conditions. The transfer of ETB-mNG into HEK293T cells was unaffected by the presence of either antibiotics or FCS in the medium (data not shown).

Next, the effect of MP/ND concentration and incubation time on the transfer efficiency was analyzed (Figure 3, Supplementary Figure S2). Transfer by incubation of 0.5 μM ETB-mNG/ND with HEK293T cells for a period from 1 to 24 h was analyzed by quantification of membrane fluorescence (Figure 3A). The detection of ETB-mNG in the cell membrane increased linearly with the incubation time for the whole period. The concentration of ETB-mNG/ND particles in the incubation mixture was next screened from 10 nM to 10 μM by incubation for 4 h at 37°C (Figure 3B). Already with the lowest concentration of 10 nM, a transfer of the receptor into the cell membranes was detectable (Figure 3B). In a complementary experiment, the transfer of lipids from NDs (DEPG +2% Rho-PE) into HEK293T cells was analyzed within a concentration range of 50 nM to 5 μM NDs and incubation times from 1 to 24 h (Supplementary Figure S3). Similar to the MP transfer, the transfer of lipids was fast and detectable already with 50 nM NDs. The lipid transfer also increased linearly with ND concentration and incubation time.

The membrane composition of NDs can be crucial for the folding and stability of inserted GPCRs (Rues et al., 2018) and might also affect the transfer efficiency. PR-mNG/ND particles were prepared with the lipids—DEPG, POPG, DMPC, or DOPG—and subsequently transferred at standard conditions. Fluorescence analysis revealed no major differences in the transfer efficiency of PR-mNG from the different lipid environments into HEK293T cells (Figure 3C). This would allow to transfer individual MPs inserted into their most suitable lipid environment. A variety of GPCRs inserted into ND membranes of specific lipid compositions were then analyzed for their transfer efficiency into HEK293T cell membranes. Samples of ETB-mNG/ND (DEPG), ETB-tc-mNG/ND (DEPG), FFAR2-mNG/NDs (POPG), GPRC5B-mNG/NDs (POPG), IP-mNG/NDs (DEPG) and PR-mNG/NDs (DEPG) were incubated with HEK293T cells at standard conditions and the membrane fluorescence was quantified (Figure 3D). The overall transfer efficiency of all analyzed MPs was found to be similar, indicating that a wider range of MPs inserted in customized lipid environments can be used as a target for the nanotransfer approach.

3.3 NDs disintegrate upon MP transfer

The NDs may disintegrate upon transfer or at least some fraction may stay intact and attaches to the cellular membrane or may even become incorporated into the cell. To address this question, ETB-mNG was CF inserted into NDs (DEPG) containing 2% Rho-PE lipids. The resulting purified ETB-mNG/Rho-PE/NDs were then incubated with HEK293T cells under standard conditions and analyzed by fluorescence microscopy (Figure 4). A substantial fraction of transferred ETB-mNG again localized in clusters, while the Rho-PE lipids distributed homogenously across the cellular membrane. The absence of a clear co-localization of transferred ETB-mNG and Rho-PE already gives evidence that lipids and MP separate during the transfer procedure, indicating the disintegration of the NDs. This assumption was further analyzed by immunodetection of MSP. ETB-mNG/NDs (DEPG) and NDs (DEPG +2% Rho-PE) were prepared with Flag-tagged MSP (MSP1E3FN) and incubated with HEK293T cells under standard conditions. After incubation, the cells were permeabilized and stained with anti-Flag antibodies. As a staining control, transfected HEK293T cells expressing GPRC5B-Flag were used. Despite the successful transfer of ETB-mNG and Rho-PE, no MSP1E3FN was detected in the cells by immunostaining (Figure 5A). To further support this result, HEK293T cells after ETB transfer were washed, lysed, and analyzed by western blotting for the presence of ETB and MSP (Figure 5B). The transfer of ETB into the cells was verified, but MSP was only detectable in the transfer mixture and absent in the cell lysate.

3.4 MP transfer into HEK293T cells with SapNPs

SapNPs based on the scaffold protein SapA are an alternative option to solubilize MPs in nanomembranes (Frauenfeld et al., 2016). SapA is a sphingolipid activator protein in lysosomes and in contrast to NDs, SapNPs are adaptable to the size of the inserted MP (Chien et al., 2017). This specific characteristic of SapNPs might become interesting for the insertion and solubilization of larger MPs or MP complexes. Although similar in their function to solubilize lipids and MPs, SapA and MSP do not share extensive sequence homologies. We, therefore, analyzed whether SapNPs can be used similarly to NDs as vectors to transfer inserted MPs into living cells. PR-mNG was co-translationally integrated into preformed SapNPs (DOPG) and after Strep-purification incubated with HEK293T cells under standard conditions. After washing and fixation, membrane fluorescence was quantified. As a control, PR-mNG/NDs were incubated with HEK293T cells as well. PR-mNG was transferred from both particles into HEK293T cells (Figure 6A). The PR-mNG transfer from NDs showed a higher efficiency, giving evidence that potentially higher concentrations of MP/SapNPs need to be used in the transfer mixtures in order to obtain comparable MP transfer efficiencies as with NDs. In accordance, a notable transfer of PR-mNG from SapNPs was observed with PR-mNG/SapNP concentrations in the transfer mixture above 0.1 μM (Figure 6B). With NDs, already MP/ND concentrations starting with 0.01 μM resulted in detectable MP transfer (Figure 3B). Similar to the transfer of MPs from NDs, the

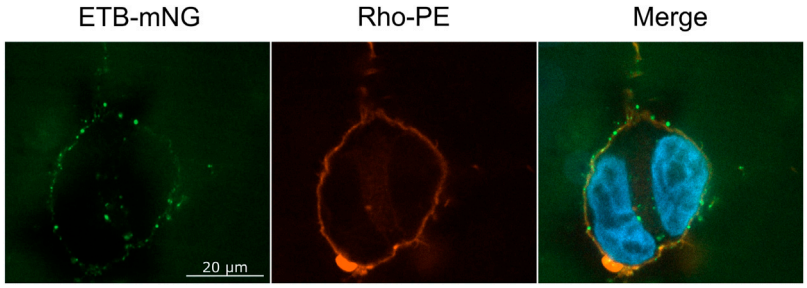


FIGURE 4
Localization of ETB-mNG and Rho-PE lipids after co-transfer into HEK293T cells. Purified ETB-mNG/NDs (DEPG +2% Rho-PE) were incubated with HEK293T cells at standard conditions. After fixation, the localization of transferred ETB-mNG and Rho-PE in the cells was analyzed via fluorescence microscopy. Representative images are shown.

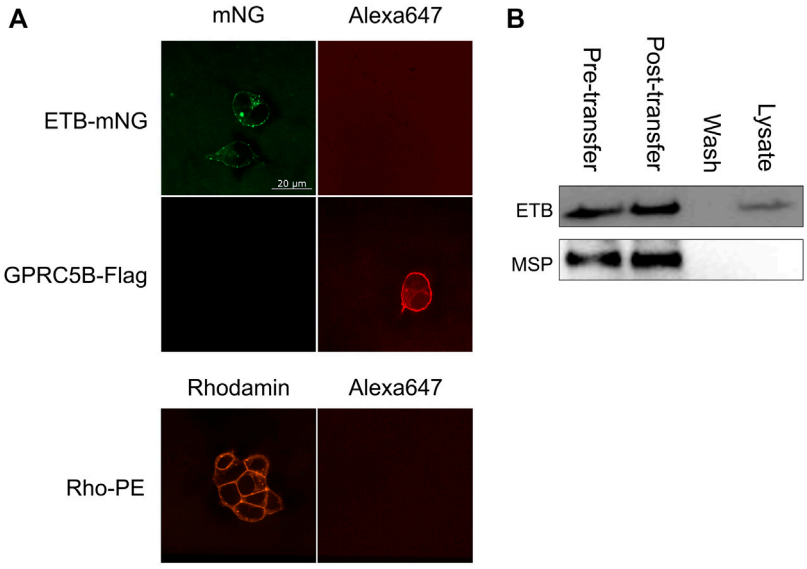


FIGURE 5
Analysis of MSP transfer into HEK293T cells. **(A)** ETB-mNG/NDs and NDs (DEPG +2% Rho-PE) performed with Flag-tagged MSP1E3FN were incubated with HEK293T cells at standard conditions. Cells were then washed, fixed, permeabilized, and incubated with anti-Flag primary antibodies and Alexa647 coupled secondary antibodies. Transfected HEK293T cells synthesizing GPRC5B-Flag were used as control. **(B)** ETB/NDs were incubated with HEK293T cells at standard conditions and transfer of ETB and MSP was monitored by western blotting with anti-Strep (ETB) and anti-His (MSP) antibodies, respectively. Pre-transfer: Samples of the transfer mixture before incubation; post-transfer: Samples of the transfer mixture after incubation; wash: DPBS after the last washing step; lysate: Cell lysate after incubation and washing.

transfer of PR-mNG from SapNPs continuously increased with prolonged incubation times (Figure 6C).

3.5 Membrane topology of transferred MPs

The membrane topology of transferred MPs is of general importance as it strongly affects the MP function. To this point, it was unclear whether MPs insert with a preferred orientation into the cellular membrane or if the insertion is rather random. It seems reasonable that larger soluble domains of an MP might affect its transfer and final membrane topology. Therefore, PR constructs containing a Myc-tag and with or without the large soluble mNG domain either at the N-terminus or at the C-terminus were created. All four constructs were cotranslationally inserted into NDs (POPG), purified, and transferred into HEK293T cells at standard conditions. The

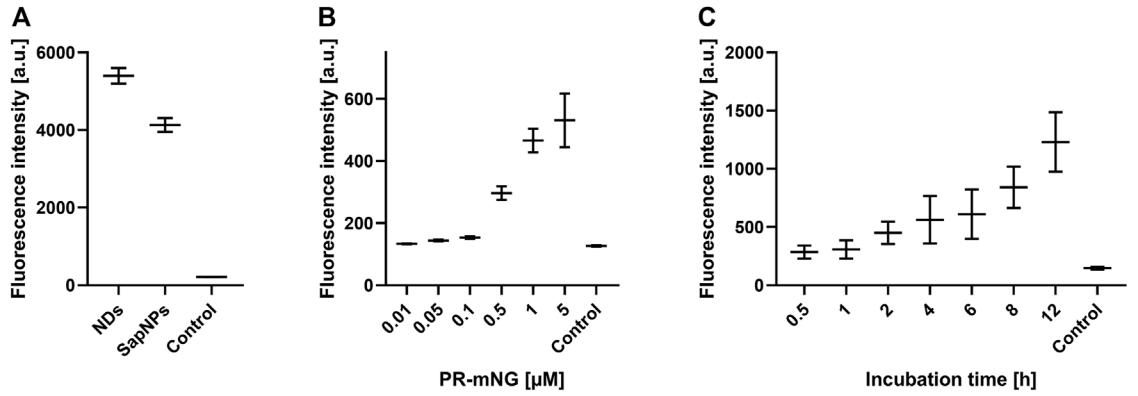


FIGURE 6
SapNPs as vector for MP transfer into HEK293T cells. PR-mNG/SapNP particles were prepared by CF expression of PR into preformed SapNPs (DOPG). The Strep-purified PR-mNG/SapNPs were incubated with HEK293T cells and the mNG fluorescence in the cellular membrane was quantified. **(A)** Comparison of PR-mNG transfer from NDs (DOPG) and SapNPs (DOPG) at standard conditions. **(B)** Effect of PR-mNG/SapNP concentration on PR-mNG transfer. **(C)** Effect of incubation time on PR-mNG transfer. Mean ± SEM is shown. N = 20 cells. MP-free NDs (DOPG) were used as a negative control.

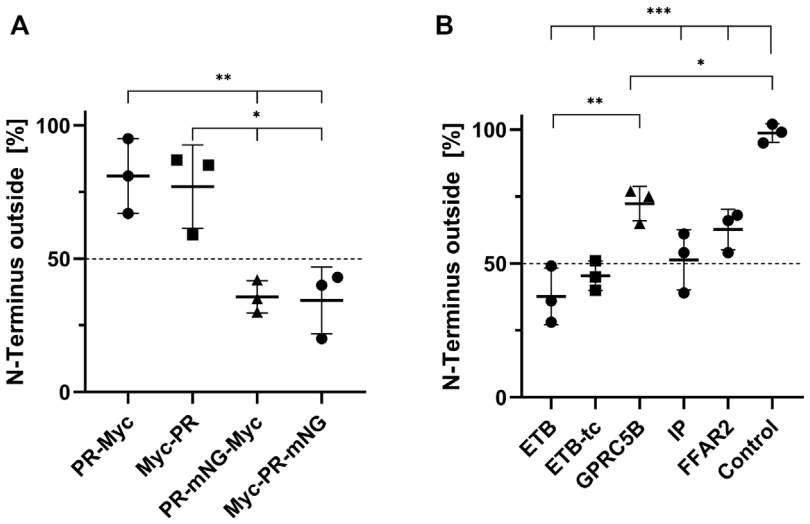


FIGURE 7
Orientation of transferred MPs in HEK293T cell membranes. NDs containing Myc-tagged MPs were transferred into HEK293T cells at standard conditions. The cells were fixed and either permeabilized with 0.2% Triton X-100 or left with intact cellular membranes. Anti-Myc immunostaining of permeabilized cells resulted in staining of all transferred MPs, while in non-permeabilized cells only MPs with an extracellularly accessibly Myc-tag were stained. The membrane fluorescence was quantified accordingly and the fraction of N-terminally accessible MPs was calculated by dividing the intensity of permeabilized by the intensity of non-permeabilized cells. **(A)** Membrane topology of transferred PR derivatives from NDs (POPG). **(B)** Membrane topology of transferred N-terminally Myc-tagged GPCRs. The transfer mixtures contained Myc-ETB/NDs, Myc-ETB-tc/NDs, Myc-FFAR2/NDs, Myc-GPRC5B/NDs or Myc-IP/NDs. GPRC5B-Myc synthesized in transfected cells was used as a positive control. Mean ± SEM is shown. N = 3. N = 20 cells (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; one-way ANOVA with Tukey's test).

position of the Myc-tags was then determined by immunostaining of permeabilized and non-permeabilized cells (Figure 7). The total amount of transferred PR derivatives in the membrane of permeabilized cells was then correlated with the number of transferred PR derivatives in non-permeabilized cells having their Myc-tag only extracellular accessible. The terminal position of the small Myc-tag did not affect the orientation of transferred PR in the HEK293T cell membrane. Approximately

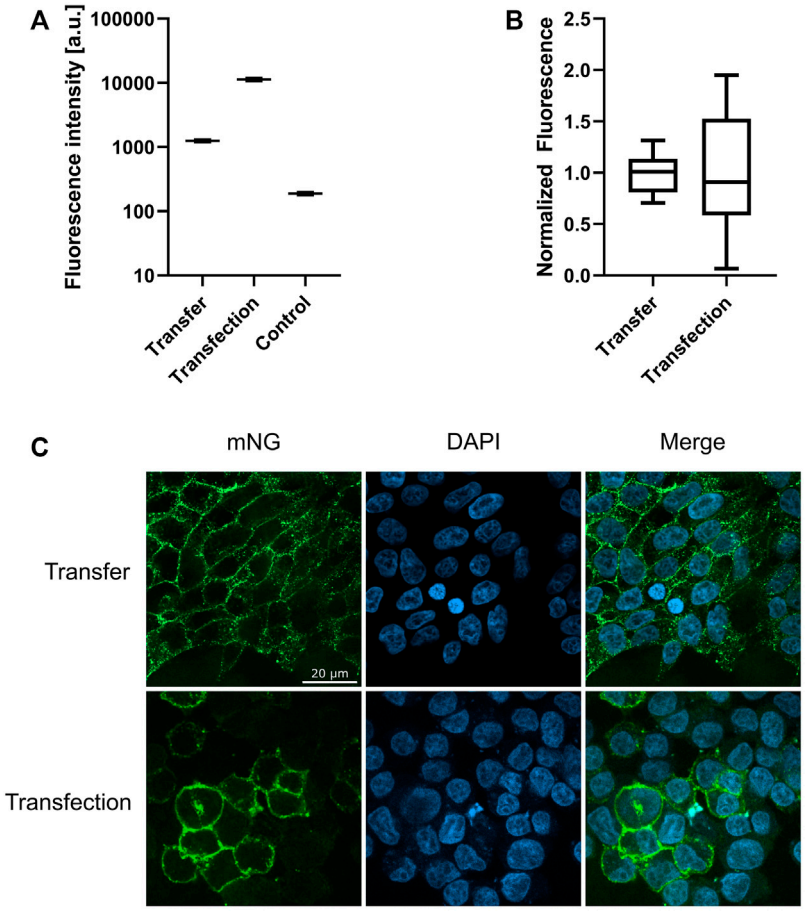


FIGURE 8
Membrane localization of transferred and transfected ETB-mNG. 0.5 µM ETB-mNG/NDs were incubated with HEK293T cells for 24 h. ETB-mNG was further synthesized in transfected HEK293T cells. Fluorescence microscopy images of fixed cells were taken after 24 h and fluorescence in the cellular membrane was quantified. **(A)** Comparison of membrane fluorescence in transfected and transferred cells. Mean ± SEM is shown. *n* = 20 cells. **(B)** Variation of mean-normalized membrane fluorescence. **(C)** Representative images of HEK293T cells with transferred ETB-mNG or with synthesized ETB-mNG after transfection.

78% of transferred PR-Myc as well as of Myc-PR were oriented with their N-terminus outside, i.e., in the correct and native orientation (Figure 7A). In contrast, fusions of PR with the larger mNG moiety had a strong effect on the insertion orientation and only approx. About 35% of both, PR-mNG-Myc or Myc-mNG-PR, remained with their N-terminal end outside of the cell membrane.

We further analyzed the membrane topology of a number of GPCRs after their transfer from NDs into HEK293T cells. CF prepared NDs containing Myc-ETB, Myc-ETB-tc, Myc-GPRC5B, Myc-FFAR2 or Myc-IP were incubated with HEK293T cells at standard conditions. The membrane topology of the transferred GPCRs in permeabilized and non-permeabilized cells was then determined by immunostaining (Figure 7B). The orientation of Myc-ETB-tc and Myc-IP appears to be rather balanced with approx. 50% having their N-terminal end outside in the correct topology. The truncations of the larger soluble IL3 and C-terminal domains in Myc-ETB-tc seem to have a slight positive effect on its final membrane orientation as only 40% of the full-length Myc-ETB had the correct topology. For Myc-FFAR2 and Myc-GPRC5B a small bias toward the correct membrane insertion with the N-terminus outside was observed. As a control, the membrane topology of GPRC5B-Myc synthesized in transfected HEK293T cells showed a nearly 100% correct N-terminus outside insertion.

3.6 Membrane distribution and cluster formation of transferred MPs

Transfer of ETB-mNG into HEK293T cells as well as the synthesis of ETB-mNG in transfected HEK293T cells results in

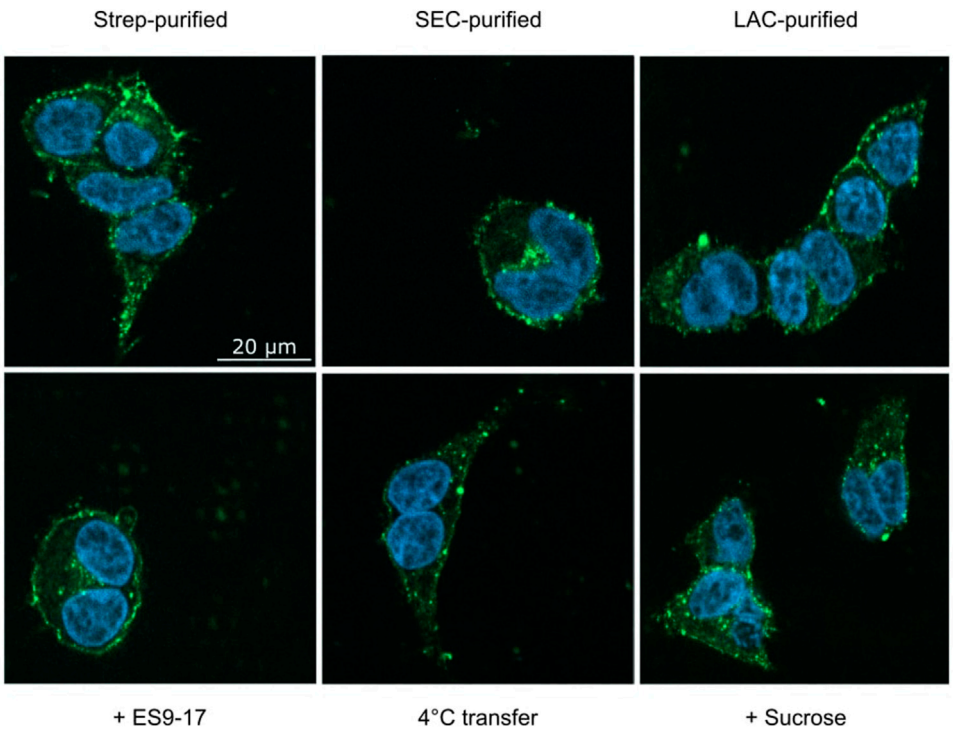


FIGURE 9
Analysis of cluster formation of transferred ETB-mNG. ETB-mNG/ND particles were purified either (i) via the C-terminal Strep-tag of the receptor, (ii) via the Strep-tag and subsequent SEC, or (iii) via LAC by using immobilized 4Ala-ET-1-biotin. ETB-mNG was then transferred into HEK293T cells at standard conditions and representative fluorescence images of the cells were taken. Lower panel: Effects of addition of the endocytosis inhibitors ES9-17 (30 µM) or sucrose (0.45 M) and of performing the transfer at 4°C on the cluster formation of Strep-purified ETB-mNG.

membrane localization of the GPCR (Figure 8). The total amount of ETB-mNG in transfected cells is approx. 10-times higher if compared with transferred cells (Figure 8A). However, the transferred ETB-mNG is more evenly distributed amongst the transferred cells and all cells in the population have a relatively similar ETB-mNG concentration (Figures 8B,C). In contrast, the ETB-mNG concentration in the transfected HEK293T cell population showed a considerable variation, including cells having very high expression as well as many cells with no detectable ETB-mNG production.

A striking difference in transfection is the prevalent clustering of transferred MPs in the cell membrane (Figures 2, 4, 8C). While in transfected HEK293T cells the synthesized MP is homogeneously distributed in the cell membranes, with all analyzed MPs a significant fraction of the transferred proteins appeared to be present in clusters. A possible reason for the observed cluster formation could be the transfer of already misfolded and aggregated MPs present in the MP/ND samples. The heterogeneous SEC profiles of all GPCR/ND samples could support this assumption, although even transfer of the homogeneous PR-mNG/ND samples resulted into similar cluster formation. Nevertheless, the effect of sample heterogeneity on the cluster formation was analyzed by transferring ETB-mNG samples of different purity. ETB-mNG/ND particles were either Strep-, SEC-, or LAC-purified. While the Strep-purified sample was the most heterogeneous, containing even larger aggregates, the SEC-purified sample contained mostly fraction 2 with presumably homogeneously inserted ETB-mNG, whereas the LAC-purified sample contained only correctly folded and ligand binding active ETB-mNG. However, the transfer of the differently purified ETB-mNG samples into HEK293T cells did not reveal any difference in membrane localization or cluster formation (Figure 9).

The observed MP cluster may result from endocytosis processes of the cell. To address this, effects by the addition of the endocytosis inhibitors ES9-17 (Dejonghe et al., 2019) or sucrose (Guo et al., 2015) have been tested. Furthermore, incubation at lower temperatures should also inhibit or retard endocytosis. ETB-mNG was transferred from NDs into HEK cells at standard conditions by pre- and co-incubation with either 30 µM ES9-17 or 0.45 M sucrose, or the transfer mixture was incubated at 4°C. However, none of the modified transfer conditions had a detectable effect on membrane localization or cluster formation of the transferred ETB-mNG (Figure 9).

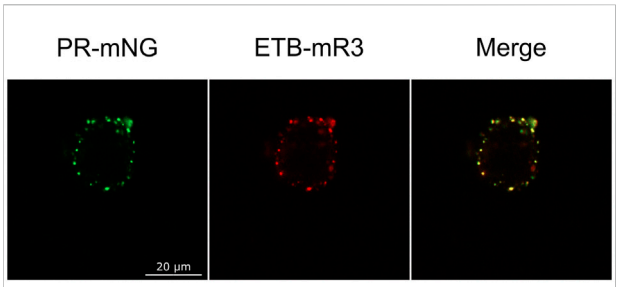


FIGURE 10
Co-localization of different MP clusters in transferred HEK293T cells. HEK293T cells were co-incubated with PR-mNG/NDs and ETB-mR3/NDs at standard conditions. After washing and fixation, cells were analyzed by fluorescence microscopy. Images of a representative cell are shown.

Furthermore, comparable cluster formation in HEK293T cells was independent of the transferred MP type, of the ND lipid composition and it was also visible after MP transfer from SapNPs. Moreover, the cluster formation is fast and already visible after 5 min of incubation (Supplementary Figure S4). The localization of the MP cluster in the cell membrane appears not to be random. HEK293T cells were incubated with PR-mNG/NDs and ETB-mR3/NDs together at standard conditions and analyzed (Figure 10). Cluster formation of both transferred MPs could be monitored individually, and the merged picture showed that most clusters for the two different MPs localize at identical positions in the cell membrane. The results give strong evidence that membrane structures or cellular mechanisms are responsible for the cluster formation, rather than the quality or type of the transferred MP.

3.7 Stability and activity of transferred GPCRs

For stability measurements, ETB-mNG transferred HEK293T cells were washed and further incubated in a fresh medium. Fluorescence was measured over a period of 3 days and then gradual decrease of ETB-mNG fluorescence followed a standard protein degradation kinetics (Chen et al., 2016) with a half-life of approx. 15.8 h (Supplementary Figure S5). The transferred cells remained viable and did not show decreased growth rates. The ETB-mNG clusters remained detectable throughout the whole observation period.

The functionality of transferred MPs was analyzed by 1) ligand binding, 2) ligand-induced receptor internalization and 3) by specific protein interactions. LAC-purified ETB-mNG was transferred into HEK293T cells at standard conditions. After washing, DY647 labeled ET-1 was added and the cells were further incubated at 4°C to inhibit the internalization of the ligand-receptor complex. Subsequent fluorescence microscopy revealed the co-localization of the transferred receptor with DY647-ET-1, indicating that the functional conformation of ETB-mNG remains unaltered after the transfer (Figure 11). It should be noted that, as shown in Figure 7B, less than 50% of the transferred receptor is inserted with the correct orientation in the membrane. Thus only a fraction of the transferred receptor should be able to bind the ligand. Nevertheless, almost all visible cluster of transferred ETB-mNG show interaction with DY647-ET-1. This gives evidence that the cluster contains mixtures of correct and non-correct inserted receptor and further indicates that cluster formation is not determined by a particular MP structure. As a positive control, DY647-ET-1 was added to HEK293T cells transfected with the corresponding ETB-mNG plasmid, and uniform labeling of membrane-localized ETB-mNG was monitored (Figure 11).

Internalization of transferred LAC-purified ETB-mNG was monitored after incubation with circular ET-1 (cET-1). A primary function of the ETB receptor in human tissue is to efficiently remove endothelin agonists by undergoing rapid internalization upon the formation of an ETB/agonist complex (Wu-Wong et al., 1995; Zrein et al., 2020). ETB-mNG transferred cells were incubated with 500 nM of the native agonist cET-1 for 1 hour and then further incubated at 37°C for 1 hour to allow for internalization. The amount of ETB-mNG in the cytosol of cET-1 stimulated and non-stimulated cells were then quantified after fixation (Figure 12). A significant increase of ETB-mNG fluorescence in the cytosol fraction after cET-1 stimulation gives further evidence of its functional integration into the cellular membrane environment after the transfer process.

The oligomerization of GPCRs and their interaction with other binding partners requires their functional conformation. The homo-oligomerization of GPRC5B and IP receptor as well as the hetero-oligomerization of GPRC5B with the IP receptor has been described before (Giguère et al., 2004; Carvalho et al., 2020). Strep-tagged GPRC5B or IP receptors were transferred into HEK293T cells previously transfected with constructs encoding for HA-IP or GPRC5B-Myc. After transfer, pulldown experiments based on the Strep-tag of the transferred receptors were performed. Both transferred GPRC5B and transferred IP were interacted with the synthesized GPRC5B-Myc (Figure 13). Furthermore, an interaction of transferred GPRC5B with the HA-IP receptor was detectable. However, no pulldown of transferred IP receptor with HA-IP receptor synthesized after transfection was observed.

4 Discussion

The nanotransfer of MPs into live mammalian cells is an emerging tool that complements established techniques for the

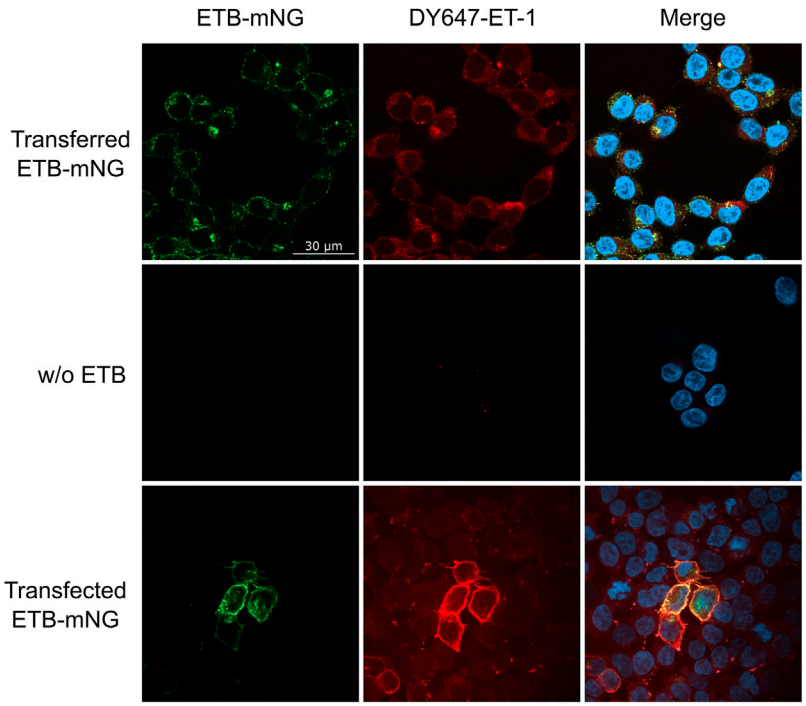


FIGURE 11
Ligand binding of transferred ETB-mNG in HEK293T cells. LAC-purified ETB-mNG was transferred into HEK293T cells at standard conditions. After washing, cells were incubated with 100 nM DY647-ET-1 for 1 h at 4°C. After additional washing, fluorescence microscopy was performed and representative pictures were taken. Transfected HEK293T cells synthesizing ETB-mNG were used as a positive control.

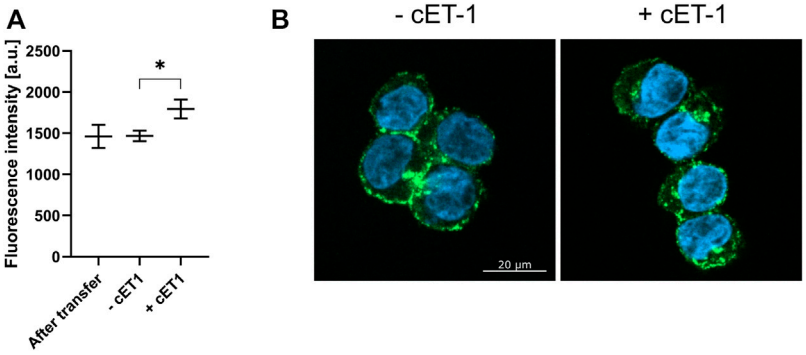


FIGURE 12
Induced internalization of transferred ETB-mNG by cET-1. LAC-purified ETB-mNG was transferred into HEK293T cells at standard conditions. Cells were washed and subsequently incubated with 500 nM cET-1 for 1 h at 4°C. Afterward, they were further incubated at 37°C for 1 h, fixed and ETB-mNG fluorescence in the cytoplasm was quantified via ImageJ. **(A)** Quantification of ETB-mNG in the cytoplasm of transferred HEK293T cells. Mean ± SEM is shown. $n = 20$ cells ($p < 0.05$; paired t-test). **(B)** Representative images of HEK293T cells treated with or without cET-1 after ETB-mNG transfer.

transfer of soluble proteins (Patriarchi et al., 2018; Chau et al., 2020; He et al., 2021). In combination with CF expression, MP/ND samples ready for transfer can be generated within 24 h. We have analyzed the nanotransfer of four human GPCRs and of PR also shows a seven-transmembrane domain structure. The CF synthesis of the ETB receptor and its truncated derivative ETB-tc

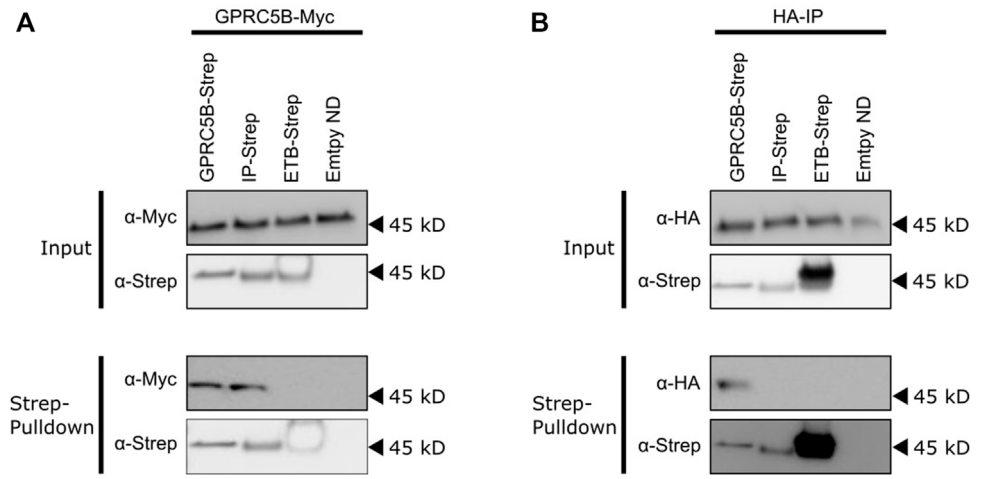


FIGURE 13
Interactions of transferred and transfected GPCRs. HEK293T cells were transfected with constructs expressing either GPRC5B-Myc **(A)** or HA-IP **(B)**. After 24 h, the transfected cells were incubated with 0.5 μM of either GPRC5B/NDs, IP/NDs, ETB/NDs or empty NDs for 16 h to allow efficient MP transfer. The cells were then washed, lysed and pulldowns based on the Strep-tags of the transferred GPCRs were performed. The MPs were subsequently analyzed for co-purification by SDS-PAGE and immunoblotting. **(A)** Analysis of transfected cells synthesizing GPRC5B-Myc. Upper panel input: Immunoblotting of cell lysates showing successful expression of GPRC5B-Myc (α-Myc) and the successful transfer of either GPRC5B, IP or ETB (α-Strep). Lower panel Strep-pulldown: Immunoblotting of the eluate from the Strep-column detecting transferred GPCRs (α-Strep) and co-purified transfected GPRC5B-Myc (α-Myc). No pulldown of GPRC5B-Myc was observed with transferred ETB or empty NDs. **(B)** Analysis of transfected cells synthesizing HA-IP. Upper panel input: Immunoblotting of cell lysates with anti-HA antibody shows a successful expression of HA-IP (α-HA) and the successful transfer of either GPRC5B, IP, or ETB into the cells (α-Strep). Lower panel Strep-pulldown: Immunoblotting of the eluate from the Strep-column detecting transferred GPCRs (α-Strep) and co-purified transfected HA-IP (α-HA).

into NDs was previously optimized. This work is the first description of the cotranslational insertion of the FFAR2, IP, and GPRC5B receptors into NDs. After optimization, all GPCR/ND particles were synthesized in μM concentrations, but still showed heterogeneous SEC profiles including fractions of functionally folded protein as well as potentially unfolded and aggregated fractions. The best quality was obtained with PR/ND samples showing a homogenous SEC profile.

MP transfer from nanoparticles into the cellular plasma membrane is a rapid process and is already detectable after a few minutes of incubation. The transfer is even much faster than previously reported and different detection techniques might account for this discrepancy (Patriarchi et al., 2018). Transcription, translation, and trafficking processes necessary for MP production in transfected cells are skipped and the recombinant MP appears almost instantly in the cellular plasma membrane. Furthermore, the MP transfer efficiency increases linearly with the incubation times of MP/NPs and cells for a period of several hours. Transfer efficiency further depends on the MP particle concentration in the transfer mixture. This offers an interesting tool to modulate the final concentration of the transferred MP in the cell. Effects of increasing MP concentrations could be determined and potential interaction partners in the cells could be identified and quantified by pulldown assays. The nanoparticle lipid composition appears to be of less importance for the MP transfer efficiency. This is an important finding as it allows us to integrate the synthesized MPs always into their most preferred lipid environment. The nature of the nanoparticle scaffold protein is also of less importance for the MP transfer efficiency. ND as well as SapA particles are suitable vectors for MP transfer. However, MP transfer from SapNPs appears to be a little less efficient than from NDs. This could be explained by the smaller membrane area in SapNPs potentially resulting in a weaker alignment of the SapNPs with the cell membrane (Lai et al., 2015). We could further demonstrate that NDs dissociate during transfer and that the transferred lipids and MPs separate from each other in the cell membrane. The scaffold protein MSP stays outside of the cells and we did not find any evidence of MSP uptake. This disagrees with previous reports showing uptake of fluorescent labeled MSP from HeLa cells after their incubation with NDs (Carney et al., 2015; Petrache et al., 2016). The labeled MSP was localized as punctuate spots inside the cells. However, labeled lipids from NDs are dispersed throughout the cell membrane, indicating a dynamic exchange of lipids between NDs and cells similar to our observations. It can therefore be concluded that NDs disintegrate upon transfer and release the lipids into the cell membranes, whereas a partial uptake of the MSP might be cell type-specific.

While the MP transfer appears to be possible with a variety of cell lines, we provide the first evidence that differences may exist

with regard to MP localization and stability after transfer. In our hands, the plasma membrane localization of transferred MPs was most prominent in HEK293T cells and somehow less pronounced in H1299 cells. With CHO cells, most of the transferred MPs are localized in the cytoplasm. It has been shown that the composition of supported lipid bilayers affects the transfer of MPs from NDs (Dang et al., 2019). Different membrane compositions of the analyzed cell lines might therefore contribute to the observed preference in MP localization.

The transfer mechanism appears to be more general as similar efficiencies were obtained with all five analyzed MPs. According to the fusion pore model for the transfer of cargo from NDs to bicelles, the NDs are suggested to align in a planar orientation on top of the target bilayer (Lai et al., 2015). The MPs may then spontaneously transfer between nanoparticle and cell membranes (Dang et al., 2019). Two contact sites of the MP/ND particles with the cell membrane are possible. Our results indicate that the membrane insertion of MPs with comparable topology is rather stochastic without a notable preferred orientation. However, steric hindrance by the presence of larger soluble domains could have a strong effect on the resulting MP membrane topology by causing an orientation bias of the aligned MP/ND particles. While a small soluble Myc-tag terminally attached to PR did not affect its orientation after membrane insertion, fusion of the larger mNG moiety to the PR C-terminus reduced the amount of "correctly" inserted MP (= N-terminus outside) from approx. 75% to less than 30%. It is important to note that the insertion of MPs into cellular plasma membranes with the wrong orientation is a distinct feature of the nanotransfer technique and might be interesting for particular experimental approaches. MPs produced from transfected cells are targeted to the plasma membrane exclusively in their correct orientation.

After prolonged incubation of 4 hours, the amount of transferred MP in HEK293T cells was still significantly lower if compared with transfected cells. However, the recombinant production of MPs in DNA transfected cells is hard to control and high variation exists with regard to the level of synthesized protein in individual cells (Kim and Eberwine, 2010). In accordance, a fraction of the HEK293T cells transfected with a DNA template of ETB-mNG showed high production of the recombinant MP, whereas a significant number of cells had no detectable ETB-mNG production at all. In contrast, the distribution of transferred ETB-mNG in HEK293T cells was much more homogenous within the cell population and cells without any transferred MP were hardly detectable. More homogenous MP production in a cell population may be obtained by transfection with RNA-based templates, while the requirement for specific RNA regulatory structures and difficult handling techniques still can cause problems (Kim and Eberwine, 2010). The nanotransfer approach could thus become interesting for approaches, where homogenous distribution and comparable MP concentrations within a cell population are more relevant than a total high protein concentration.

A striking difference in MP production in transfected cells is the accumulation of transferred MPs in clusters, while transferred lipids distribute homogeneously among the cellular plasma membrane. The cluster formation was generally visible with all analyzed MPs to similar extents. It was not due to the transfer of already aggregated MPs in the case of the GPCR samples, as control experiments with LAC-purified ETB-mNG gave similar results. In addition, also transfer of homogeneous PR samples resulted in cluster formation. Most interesting is the co-localization of cluster formed with different transferred MPs. Since clusters are visible already after a few minutes of transfer, they might show "hot spots" where the transfer of the MP from the ND takes place, indicating that the transfer might be dependent on certain microdomains in the cellular membrane. Alternatively, transferred MPs could diffuse to defined membrane areas where they become immobilized by certain mechanisms. Ligand binding indicated that the cluster appears to contain functionally folded MPs rather than unfolded aggregates. The cluster did not resolve during incubation and we could not find evidence for an involvement in endocytosis processes. It will be interesting to compare the cluster formation of transferred MPs in other cell lines in order to further reveal mechanistic details.

The transferred GPCRs show standard degradation kinetics and were detectable in the cells for approx. 48 h. The functional transfer of ETB was shown by binding to its ligand ET-1 and by the stimulation of ETB internalization after binding to the ligand cET-1. This is in accordance with previous reports showing signaling activation by transferred human ß2-adrenergic receptors (Patriarchi et al., 2018). We further demonstrated the interaction of transferred GPCRs with their cognate binding partners synthesized after transfection of the same cells. The GPRC5B receptor synthesized in transfected cells interacted with both, CF synthesized and transferred GPRC5B or IP receptor. This is in agreement with the previously observed homodimerization of GPRC5B and heterodimerization of GPRC5B with the IP receptor (Carvalho et al., 2020). Vice versa, the IP receptor synthesized in transfected cells interacted with CF synthesized and transferred GPRC5B, while interaction with the transferred IP receptor was not observed. Interestingly, although the IP homo-oligomerization has been described before, it is speculated that this interaction is based on intermolecular disulfide bonds formed already in intracellular compartments (Giguère et al., 2004). The failure to observe an interaction of transferred with transfected IP may therefore result from the inability of the

transferred MP to form these intermolecular covalent bonds in the cellular membrane. In addition to the mentioned examples of transferred functional GPCRs, a recent report demonstrated the alteration of the cellular transcriptome and induction of a tumor-like phenotype in originally non-malignant breast cells transferred with the receptor tyrosine kinase Her2 (He et al., 2021).

In conclusion, CF expression, nanoparticle technology and nanotransfer combine in excellent synergy to generally allow the fast insertion of recombinant MPs into membranes of live cells. The final MP amount in the transferred cells can be fine-tuned by incubation time and initial MP nanoparticle concentration, allowing for a more precise adjustment of recombinant MP in a cell population. The process is detergent-free and does not require potentially toxic chemicals, hazardous vectors or physical manipulation of the cells. The technique has a lower impact on cell physiology and might therefore in particular be suitable for sensitive, fragile or otherwise difficult to transfect cells such as primary cells, stem cells, cancer cell lines or stationary phase cells (Patriarchi et al., 2018; He et al., 2021). The speediness can induce immediate cellular reactions allowing the simultaneous *in vitro* and *in vivo* analysis of aliquots of identical samples. In addition to basic research applications, the nanodelivery approach might have potential for therapeutic studies as well. NDs were shown to be non-toxic and non-immunogenic in mice and displayed high stability in biological fluids (Fischer et al., 2014). Cell-type specific targeting might be addressed in the future by using chimeras of MSP and cell-specific antibody fragments (Crosby et al., 2015). It needs to be considered that, in contrast to MPs synthesized after transfection, the transferred MPs insert in more or less random orientation in the cell membranes, while engineering strategies may generate a bias in the orientation. Furthermore, the transferred MPs will be devoid of posttranslational modifications and are predominantly organized in membrane-located clusters. A variety of functions including ligand binding, protein interactions, and signaling are documented from transferred MPs, but the effects of this cluster formation on MP functionality can currently not be excluded.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material; further inquiries can be directed to the corresponding author.

Author contributions

SU and RL performed experiments and designed the study. SN and TS developed and performed the chemical synthesis and labeling of ligands. VD contributed to experimental design and provided infrastructure. SU and FB made the experimental concept and wrote the manuscript.

Funding

This work was funded by the Center for Biomolecular Magnetic Resonance and by the LOWE project GLUE of the state of Hessen. Financial support was further provided by the DFG project BE 1911/8-1.

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

Acknowledgments

The authors are grateful to Nina Wettschureck and Tobias Meckel for providing DNA templates for transfection. They thank Birgit Schäfer and Zoe Köck for helpful discussions and technical assistance. They further thank Jens Frauenfeld for helping to establish the Salipro technology.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.906295/full#supplementary-material

References

Carney, C. E., Lenov, I. L., Baker, C. J., MacRenaris, K. W., Eckermann, A. L., Sligar, S. G., et al. (2015). Nanodiscs as a modular platform for multimodal MR-optical imaging. *Bioconjug Chem.* 26, 899–905. doi:10.1021/acs.bioconjchem.5b00107

Carvalho, J., Chennupati, R., Li, R., Günther, S., Kaur, H., Zhao, W., et al. (2020). Orphan G protein-coupled receptor GPRC5B controls smooth muscle contractility and differentiation by inhibiting prostacyclin receptor signaling. *Circulation* 141, 1168–1183. doi:10.1161/circulationaha.119.043703

Chau, C., Actis, P., and Hewitt, E. (2020). Methods for protein delivery into cells: From current approaches to future perspectives. *Biochem. Soc. Trans.* 48, 357–365. doi:10.1042/bst20190039

Chen, W., Smeekens, J. M., and Wu, R. (2016). Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem. Sci.* 7, 1393–1400. doi:10.1039/c5sc03826j

Chien, C.-T. H., Helfinger, L. R., Bostock, M. J., Solt, A., Tan, Y. L., and Nietlispach, D. (2017). An adaptable phospholipid membrane mimetic system for solution NMR studies of membrane proteins. *J. Am. Chem. Soc.* 139, 14829–14832. doi:10.1021/jacs.7b06730

Crosby, N. M., Ghosh, M., Su, B., Beckstead, J. A., Kamei, A., Simonsen, J. B., et al. (2015). Anti-CD20 single chain variable antibody fragment-apolipoprotein A-I chimera containing nanodiscs promote targeted bioactive agent delivery to CD20-positive lymphomas. *Biochem. Cell Biol.* 93, 343–350. doi:10.1139/bcb-2015-0009

Dang, A. T., He, W., Ivey, D. B., Coleman, M. A., and Kuhl, T. L. (2019). Lipid and protein transfer between nanolipoprotein particles and supported lipid bilayers. *Langmuir* 35, 12071–12078. doi:10.1021/acs.langmuir.9b01288

Dejonghe, W., Sharma, I., Denoo, B., Munck, S. de, Lu, Q., Mishev, K., et al. (2019). Disruption of endocytosis through chemical inhibition of clathrin heavy chain function. *Nat. Chem. Biol.* 15, 641–649. doi:10.1038/s41589-019-0262-1

Denisov, I. G., and Sligar, S. G. (2017). Nanodiscs in membrane biochemistry and biophysics. *Chem. Rev.* 117, 4669–4713. doi:10.1021/acs.chemrev.6b00690

Dong, F., Rues, R. B., Kazemi, S., Dötsch, V., and Bernhard, F. (2018). Molecular determinants for ligand selectivity of the cell-free synthesized human endothelin B receptor. *J. Mol. Biol.* 430, 5105–5119. doi:10.1016/j.jmb.2018.10.006

Fischer, N. O., Weilhammer, D. R., Dunkle, A., Thomas, C., Hwang, M., Corzett, M., et al. (2014). Evaluation of nanolipoprotein particles (NLPs) as an *in-vivo* delivery platform. *PLoS One* 9, e93342. doi:10.1371/journal.pone.0093342

Frauenfeld, J., Löving, R., Armache, J.-P., Sonnen, A. F.-P., Guettou, F., Moberg, P., et al. (2016). A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* 13, 345–351. doi:10.1038/nmeth.3801

Giguère, V., Gallant, M. A., Brum-Fernandes, A. J. de, and Parent, J.-L. (2004). Role of extracellular cysteine residues in dimerization/oligomerization of the human prostacyclin receptor. *Eur. J. Pharmacol.* 494, 11–22. doi:10.1016/j.ejphar.2004.04.041

Guo, S., Zhang, X., Zheng, M., Zhang, X., Min, C., Wang, Z., et al. (2015). Selectivity of commonly used inhibitors of clathrin-mediated and caveolae-dependent endocytosis of G protein-coupled receptors. *Biochim. Biophys. Acta* 1848, 2101–2110. doi:10.1016/j.bbamem.2015.05.024

Haberstock, S., Roos, C., Hoevels, Y., Dötsch, V., Schnapp, G., Pautsch, A., et al. (2012). A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein Expr. Purif.* 82, 308–316. doi:10.1016/j.pep.2012.01.018

He, W., Evans, A. C., Hynes, W. F., Coleman, M. A., and Robertson, C. (2021). Nanolipoprotein-mediated Her2 protein transfection induces malignant transformation in human breast acinar cultures. *ACS Omega* 6, 29416–29423. doi:10.1021/acsomega.1c03086

He, W., Evans, A. C., Rasley, A., Bourguet, F., Peters, S., Kamrud, K. I., et al. (2020). Cationic HDL mimetics enhance *in vivo* delivery of self-replicating mRNA. *Nanomedicine* 24, 102154. doi:10.1016/j.nano.2020.102154

He, W., Felderman, M., Evans, A. C., Geng, J., Homan, D., Bourguet, F., et al. (2017). Cell-free production of a functional oligomeric form of a Chlamydia major outer-membrane protein (MOMP) for vaccine development. *J. Biol. Chem.* 292, 15121–15132. doi:10.1074/jbc.m117.784561

Henrich, E., Ma, Y., Engels, I., Münch, D., Otten, C., Schneider, T., et al. (2016). Lipid requirements for the enzymatic activity of MraY translocases and *in vitro* reconstitution of the lipid II synthesis pathway. *J. Biol. Chem.* 291, 2535–2546. doi:10.1074/jbc.m115.664292

Henrich, E., Peetz, O., Hein, C., Laguerre, A., Hoffmann, B., Hoffmann, J., et al. (2017a). Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic Biology. *Elife* 6. doi:10.7554/eLife.20954

Henrich, E., Sörmann, J., Eberhardt, P., Peetz, O., Mezhyrova, J., Morgner, N., et al. (2017b). From gene to function: Cell-free electrophysiological and optical analysis of ion pumps in nanodiscs. *Biophys. J.* 113, 1331–1341. doi:10.1016/j.bpj.2017.03.026

Keller, T., Gorboulev, V., Mueller, T. D., Dötsch, V., Bernhard, F., and Koepsell, H. (2019). Rat organic cation transporter 1 contains three binding sites for substrate 1-Methyl-4-phenylpyridinium per monomer. *Mol. Pharmacol.* 95, 169–182. doi:10.1124/mol.118.113498

Kim, T. K., and Eberwine, J. H. (2010). Mammalian cell transfection: The present and the future. *Anal. Bioanal. Chem.* 397, 3173–3178. doi:10.1007/s00216-010-3821-6

Köck, Z., Dötsch, V., and Bernhard, F. (2021). Screening methods for cell-free synthesized GPCR/nanoparticle samples. *Methods Mol. Biol.* 2268, 97–117. doi:10.1007/978-1-0716-1221-7_7

Kuroiwa, F., Nishino, A., Mandal, Y., Honzawa, M., Suenaga-Hiromori, M., Suzuki, K., et al. (2022). Reconstitution of prenyltransferase activity on nanodiscs by components of the rubber synthesis machinery of the para rubber tree and guayule. *Sci. Rep.* 12, 3734. doi:10.1038/s41598-022-07564-y

Lai, G., Forti, K. M., and Renthal, R. (2015). Kinetics of lipid mixing between bicelles and nanolipoprotein particles. *Biophys. Chem.* 197, 47–52. doi:10.1016/j.bpc.2015.01.006

Majeed, S., Ahmad, A. B., Sehar, U., and Georgieva, E. R. (2021). Lipid membrane mimetics in functional and structural studies of integral membrane proteins. *Membr. (Basel)* 11. doi:10.3390/membranes11090685

Murakami, T. (2012). Phospholipid nanodisc engineering for drug delivery systems. *Biotechnol. J.* 7, 762–767. doi:10.1002/biot.201100508

Nikolaev, M., Round, E., Gushchin, I., Polovinkin, V., Balandin, T., Kuzmichev, P., et al. (2017). Integral membrane proteins can be crystallized directly from nanodiscs. *Cryst. Growth & Des.* 17, 945–948. doi:10.1021/acs.cgd.6b01631

Okuta, A., Tani, K., Nishimura, S., Fujiyoshi, Y., and Doi, T. (2016). Thermostabilization of the human endothelin type B receptor. *J. Mol. Biol.* 428, 2265–2274. doi:10.1016/j.jmb.2016.03.024

Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996. doi:10.1038/nrd2199

Patriarchi, T., Shen, A., He, W., Baikoghli, M., Cheng, R. H., Xiang, Y. K., et al. (2018). Nanodelivery of a functional membrane receptor to manipulate cellular phenotype. *Sci. Rep.* 8, 3556. doi:10.1038/s41598-018-21863-3

Peetz, O., Henrich, E., Laguerre, A., Löhr, F., Hein, C., Dötsch, V., et al. (2017). Insights into cotranslational membrane protein insertion by combined LILBID-mass spectrometry and NMR spectroscopy. *Anal. Chem.* 89, 12314–12318. doi:10.1021/acs.analchem.7b03309

Petrache, A. I., Machin, D. C., Williamson, D. J., Webb, M. E., and Beales, P. A. (2016). Sortase-mediated labelling of lipid nanodiscs for cellular tracing. *Mol. Biosyst.* 12, 1760–1763. doi:10.1039/c6mb00126b

Proverbio, D., Roos, C., Beyermann, M., Orbán, E., Dötsch, V., and Bernhard, F. (2013). Functional properties of cell-free expressed human endothelin A and endothelin B receptors in artificial membrane environments. *Biochim. Biophys. Acta* 1828, 2182–2192. doi:10.1016/j.bbamem.2013.05.031

Rues, R.-B., Dong, F., Dötsch, V., and Bernhard, F. (2018). Systematic optimization of cell-free synthesized human endothelin B receptor folding. *Methods* 147, 73–83. doi:10.1016/j.ymeth.2018.01.012

Schwarz, D., Junge, F., Durst, F., Frölich, N., Schneider, B., Reckel, S., et al. (2007). Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat. Protoc.* 2, 2945–2957. doi:10.1038/nprot.2007.426

Wu-Wong, J. R., Chiou, W. J., Magnuson, S. R., and Opgenorth, T. J. (1995). Endothelin receptor in human astrocytoma U373MG cells: binding, dissociation, receptor internalization. *J. Pharmacol. Exp. Ther.* 274, 499–507.

Zrein, A., Bagher, A. M., Young, A. P., Denovan-Wright, E. M., and Kelly, M. E. M. (2020). Endothelin receptor heteromerization inhibits β-arrestin function in HEK293 cells. *Can. J. Physiol. Pharmacol.* 98, 531–540. doi:10.1139/cjpp-2019-0620

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Brief Research Report
PUBLISHED 16 September 2022
DOI 10.3389/fbioe.2022.992708

OPEN ACCESS

EDITED BY
Pasquale Stano,
University of Salento, Italy

REVIEWED BY
Cheemeng Tan,
University of California, Davis, United States
Kei Fujiwara,
Keio University, Japan

*CORRESPONDENCE
Simon J. Moore,
s.j.r.moore@kent.ac.uk

SPECIALTY SECTION
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 12 July 2022
ACCEPTED 30 August 2022
PUBLISHED 16 September 2022

CITATION
Nagappa LK, Sato W, Alam F, Chengan K, Smales CM, Von Der Haar T, Polizzi KM, Adamala KP and Moore SJ (2022), A ubiquitous amino acid source for prokaryotic and eukaryotic cell-free transcription-translation systems. *Front. Bioeng. Biotechnol.* 10:992708. doi: 10.3389/fbioe.2022.992708

A ubiquitous amino acid source for prokaryotic and eukaryotic cell-free transcription-translation systems

Lakshmeesha K. Nagappa[1], Wakana Sato[2], Farzana Alam[3], Kameshwari Chengan[1], Christopher M. Smales[1], Tobias Von Der Haar[1], Karen M. Polizzi[3,4], Katarzyna P. Adamala[2] and Simon J. Moore[1]*

[1]School of Biosciences, University of Kent, Canterbury, United Kingdom, [2]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, United States, [3]Centre for Synthetic Biology, Imperial College London, London, United Kingdom, [4]Department of Chemical Engineering, Imperial College London, London, United Kingdom

Cell-free gene expression (CFE) systems are an attractive tool for engineering within synthetic biology and for industrial production of high-value recombinant proteins. CFE reactions require a cell extract, energy system, amino acids, and DNA, to catalyse mRNA transcription and protein synthesis. To provide an amino acid source, CFE systems typically use a commercial standard, which is often proprietary. Herein we show that a range of common microbiology rich media (i.e., tryptone, peptone, yeast extract and casamino acids) unexpectedly provide an effective and low-cost amino acid source. We show that this approach is generalisable, by comparing batch variability and protein production in the following range of CFE systems: *Escherichia coli* (Rosetta™ 2 (DE3), BL21(DE3)), *Streptomyces venezuelae* and *Pichia pastoris*. In all CFE systems, we show equivalent or increased protein synthesis capacity upon replacement of the commercial amino acid source. In conclusion, we suggest rich microbiology media provides a new amino acid source for CFE systems with potential broad use in synthetic biology and industrial biotechnology applications.

KEYWORDS

cell-free gene expression, cell-free protein synthesis, TX-TL, protein production, pichia pastoris cell-free, industrial biotechnology

Introduction

There is a rising interest in cell-free gene expression (CFE) systems as an enabling technology for prototyping applications and bottom-up approaches to synthetic biology (Sun et al., 2014; Moore et al., 2017b; Garenne et al., 2021). CFE reactions require DNA, crude cell extract, an amino acid (AA) source, and an energy solution, to catalyse mRNA transcription and translation. Interest and potential in CFE systems for industrial

biotechnology is growing, with current yields and rates of protein synthesis approaching economic viability for industrial scale-up (Zawada et al., 2011). For recombinant protein production, CFE systems offer a distinct advantage for the synthesis of specialty proteins, peptides or small molecules that are difficult to make in a cell (Zimmerman et al., 2014; Dudley et al., 2020; Meyer et al., 2021; Tian et al., 2022). For example, non-canonical AAs or post-translation modifications (Cui et al., 2020; Charna et al., 2022).

Escherichia coli is the dominant and most productive CFE system (Garenne et al., 2021). In addition, several alternative prokaryotic and eukaryotic CFE systems platforms are emerging for distinct applications. The composition of the energy solution, AAs, and other additives is essential for optimising all CFE systems. The energy solution is a complex mixture comprising a primary (e.g., ATP/GTP/CTP/UTP nucleotides) and a secondary energy source, which is typically a high-energy glycolytic pathway intermediate. In addition, cofactors, AAs, macromolecular crowding agents and a range of additives also support the reaction (Gregorio et al., 2019). The primary energy source provides the initial nucleotides and energy to drive mRNA transcription and translation, respectively. In complement, native catabolic enzymes regenerate ATP equivalents from the secondary energy source, which prolongs the CFE time course–up to several hours in batch reactions (Caschera and Noireaux, 2015a; Garenne et al., 2021). If this regeneration cycle is inefficient or absent, some CFE systems use synthetic ATP regeneration systems–e.g., creatine kinase (Anderson et al., 2015). Other limiting factors for CFE systems include non-specific ATP phosphatases, which contribute to energy loss (Calhoun and Swartz, 2007), while other metabolic products (e.g., lactate) can alter the pH and/or inhibit individual enzyme activity (Caschera and Noireaux, 2015a). Therefore, to enhance the reaction length and productivity of CFE systems, the optimisation of the energy solution and energy/product recycling, has been a key focus in recent CFE studies (Shin and Noireaux, 2010; Caschera and Noireaux, 2015a; Anderson et al., 2015; Karim et al., 2018; Garenne et al., 2021). For example, maltodextrin (and maltose) provides a slow-release mechanism to generate glucose, and recycle inhibitory free inorganic phosphate (Wang and Zhang, 2009; Caschera and Noireaux, 2015a). Indeed, several CFE studies highlight the need to optimise L-glutamate (potassium or magnesium salt) as a key variable for achieving maximal CFE (Sun et al., 2013; Cai et al., 2015; Wiegand et al., 2019; Garenne et al., 2021; Moore et al., 2021). This is because L-glutamate is the main nitrogen donor in cells for AA biogenesis, and therefore is the most abundant metabolite in *E. coli* cells, close to 100 mM in concentration (Bennett et al., 2008). In addition, L-glutamate is converted into α-ketoglutarate by L-glutamate transaminase, which leads to ATP production through the Krebs cycle. Therefore, potassium glutamate can provide a sole energy source for industrial scale *E. coli* CFE systems (Zawada et al., 2011). By optimising the various components that contribute to energy recycling, the productivity of CFE systems is enhanced. So far, the maximum *E. coli* CFE batch yields of recombinant protein is 4 mg/ml for the model green fluorescence protein (GFP), using a combination of secondary energy sources including maltodextrin, L-glutamate and d-ribose (Garenne et al., 2021). To increase yields further, metabolite replenishment and removal of waste products via artificial cells increases *E. coli* CFE recombinant protein yields up to 8 mg/ml (Garenne et al., 2021). In summary, energy regeneration and overall metabolism is an important but understudied area of CFE systems.

AAs are a key component of the energy solution, essential for polypeptide synthesis, via tRNA aminoacylation. To provide the 20 canonical AAs, CFE systems rely on expensive commercial AA kits, available as solid or liquid form. While CFE systems can operate without the addition of AAs, this is a limiting factor (Zawada et al., 2011; Moore et al., 2021). For the preparation of AAs there are also some key limitations. First, AAs cost approximately 8% for *E. coli* CFE batch reactions (Sun et al., 2013). Second, most commercial AA solutions are proprietary and contain unknown additives to help solubilise the AAs, which is problematic for some CFE systems (Moore et al., 2018). Third, the manual preparation of single AA stocks from solid powder is time-consuming. Fourth, many AAs have low aqueous solubility (e.g., L-cysteine, L-leucine). To overcome these limitations, Caschera *et al* previously developed a method to solubilise all 20 AAs in concentrated potassium hydroxide (Caschera and Noireaux, 2015b). This method provides an advantage to customise the reaction, by varying the pH and individual AA composition (Caschera and Noireaux, 2015b).

Herein, we show generalisable and yet untapped AA sources to replace commercial AA mixtures in CFE systems. Specifically, we find tryptone, yeast extract, casamino acids and peptone provide effective AA sources for a wide range of CFE systems (e.g., prokaryotic, and eukaryotic). Rich media contain mixtures of AAs, peptides, proteolytic fragments, and sometimes metabolites (e.g., vitamins, primary metabolites), although there is remarkable variability between batches because of the diverse range of sources and extraction methods. In summary, we tested a selection of CFE systems (*E. coli*, *Streptomyces venezuelae* and *Pichia pastoris*) across three cell-free synthetic biology labs (in the United Kingdom and USA), to show the ease of use for rich media as an amino source for general use in CFE systems. We demonstrate specific advantages of this method and identify key research areas that our findings enhance.

Methods

Strains, cell-lines, and plasmids

S. venezuelae ATCC 10712, *E. coli* Rosetta™ 2 (DE3) (Millipore Sigma, 71400-3), and BL21 (DE3) pLysS, *P. pastoris*

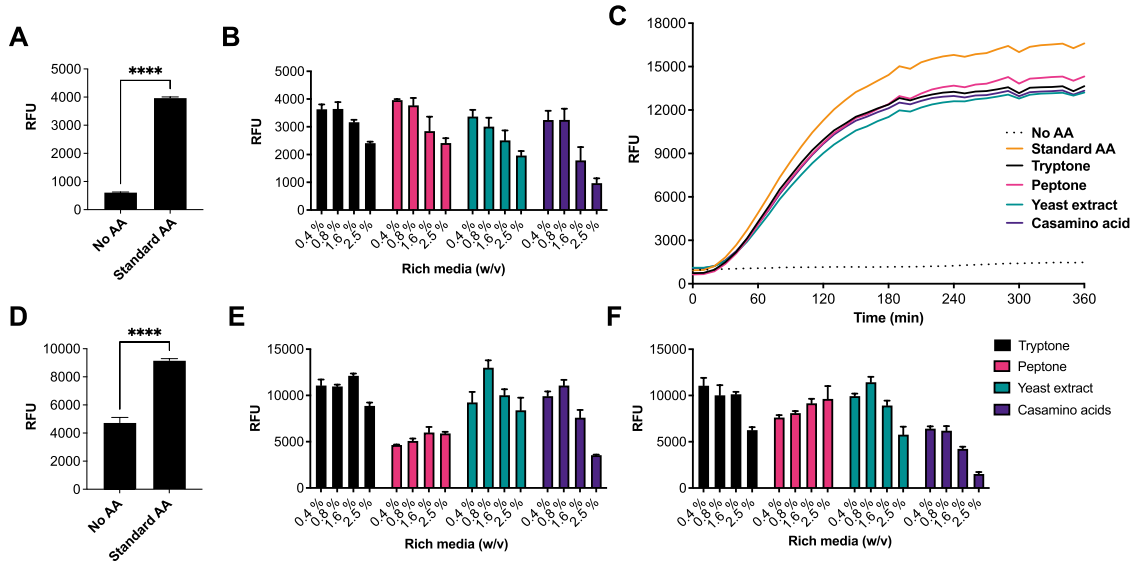


FIGURE 1
CFE activity of *E. coli* Rosetta™ 2(DE3) and BL21(DE3) systems with AA substitutes. **(A)** End-point fluorescence measurements of *E. coli* Rosetta™ (DE3) 2 control reactions with no AAs and 1 mM of AA standard (See methods). **(B)** End-point fluorescence measurements of *E. coli* Rosetta™ (DE3) 2 reactions. **(C)** *E. coli* Rosetta™ (DE3) 2 time-course reaction. **(D)** End-point fluorescence measurements of *E. coli* BL21 (DE3) control reactions with no AAs and 1 mM of commercial standard (Sigma). **(E)** End-point fluorescence measurements of *E. coli* BL21 (DE3) reactions with rich media commercial batch A, see Supplementary Table S1. **(F)** End-point fluorescence measurements of *E. coli* BL21 (DE3) reactions with rich media commercial batch B, see Supplementary Table S1. Cell-free reactions were set-up as described in the methods and results text. Experiments were performed on two independent days to ensure reproducibility. Error bars (removed in panel C for clarity) in standard deviation represent three technical measurements. End-point samples were collected after overnight incubation (16 h) at 30°C.

(also known as *Komagataella phaffi*) X33 overexpressing the *FHL1* gene (Aw and Polizzi, 2019) were used to prepare cell extracts.

Plasmids

pTU1A-SP44-*mScarlet-I* (AddGene - #163756) plasmid was used at a final concentration of 20 nM in *S. venezuelae* and 10 nM in *E. coli* BL21 (DE3) pLysS systems. pCI-T7Max-UTR1-*deGFP*-8xHis-T500 (AddGene–#178422) plasmid was used for *E. coli* Rosetta™ 2 (DE3) CFE reactions, at a final concentration of 10 nM.

CFE reactions

Original and unmodified CFE protocols were followed for *E. coli* (Sun et al., 2013; Sato et al., 2022), *S. venezuelae* (Moore et al., 2021; Toh et al., 2021) and *P. pastoris* (Spice et al., 2020; Spice et al., 2022). Luminescence measurements were performed as previously described (Aw et al., 2020). Commercial AAs (RTS Sampler kit–Biotech Rabbit, Germany) were used following the manufacturer's instructions, or individual standards were used as stated within the figure legends. For *E. coli* BL21 (DE3) CFE, the AAs were purchased from Sigma, United Kingdom (LAA21) and prepared as previously described (Moore et al., 2018). For *E. coli* Rosetta™ 2 (DE3) CFE, a 20 mM AA stock solution was prepared for the 20 AAs by dissolving in 400 mM potassium hydroxide solution pH 6.5. AAs prepared by this potassium hydroxide method were purchased from MP Biochemicals, with the exception that L-alanine and L-glycine were obtained from Santa Cruz Biotechnology, L-threonine from Sigma-Aldrich and L-arginine from Gold Biotechnology. All cell-free experiments were performed on two independent days to ensure reproducibility, where data is presented as a mean and standard deviation of three technical measurements. Data analysis was performed by GraphPad Prism 9.

Preparation of the microbiology rich media components for CFE reactions

All microbiology rich media components were prepared as 10% stock solutions in 60 mM HEPES-KOH buffer pH 8, except for *P. pastoris* CFE, where distilled water was used. Four different final concentrations of these components were used in the CFE reaction, i.e., 0.4, 0.8, 1.67 and 2.5% (w/v). The list of different commercial sources and batch numbers of the component used in this research are listed in the supplementary file (Supplementary Table S1).

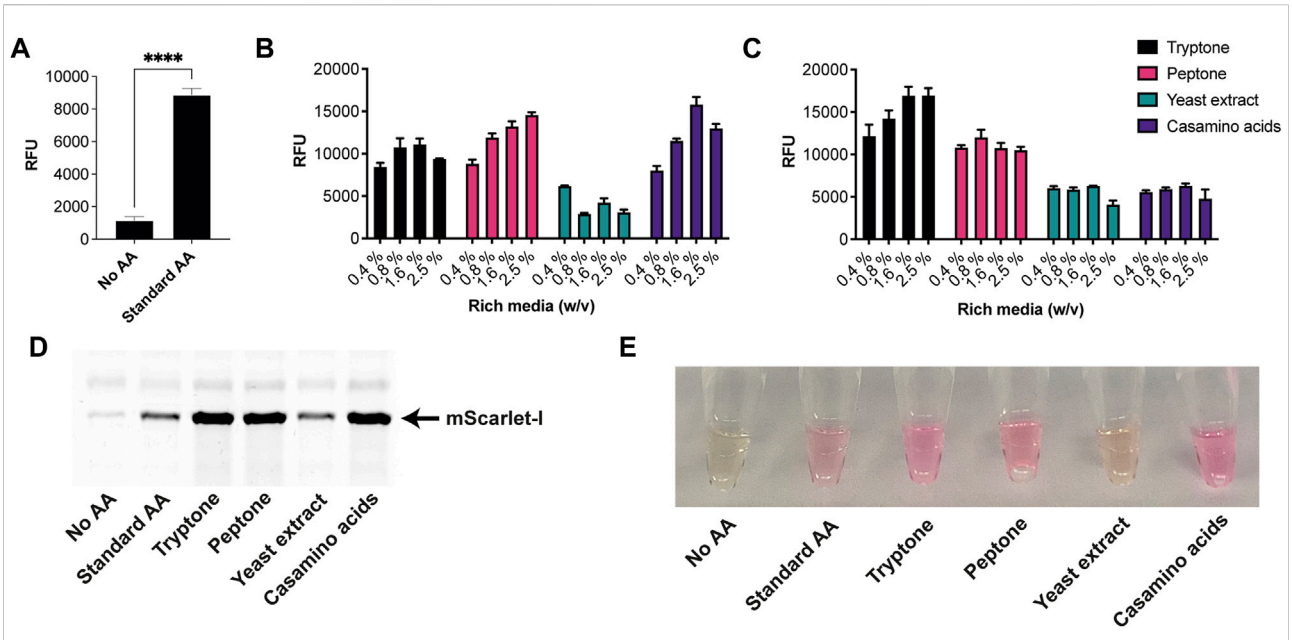


FIGURE 2
S. venezuelae CFE activity with AA substitutes. **(A)** End-point fluorescence measurements of *S. venezuelae* CFE control reactions with no AAs and 1 mM of commercial standard (Sigma). **(B)** End-point fluorescence measurements of *S. venezuelae* CFE reactions with rich media commercial batch A, see Supplementary Table S1. **(C)** End-point fluorescence measurements of *S. venezuelae* TX-TL reactions with rich media commercial batch B, see Supplementary Table S1. **(D)** In-gel fluorescence stain of C-terminal tetracysteine tagged mScarlet-I. **(E)** Visual image of mScarlet-I at endpoint. Cell-free reactions were set up as described in the methods and results text. Error bars in standard deviation represent three technical measurements. End-point samples were collected after overnight incubation (16 h) at 30°C.

Denaturing PAGE and in-gel FlAsH staining

For a 33 µL CFE sample, 1 ml of ice-cold acetone was added and incubated at -20 °C for 1 h. The sample was centrifuged at 16,000 × *g*, 10 min and the pellet was washed twice with 70% ice-cold acetone. The supernatant was removed, and the pellet was dried. To the dried pellet, 22 µL of ddH_2O, 10 µL of 4x SDS-PAGE loading buffer, and 4 µL of 0.5 M TCEP were added and mixed by vortexing. The samples were boiled for 5 min followed by the addition of 4 µL of 1 mM FlAsH-EDT_2 reagent. The samples were incubated at room temperature for 20 min, followed by centrifugation for 10 min, 16,000 × *g*. The supernatant was analyzed by SDS-PAGE and the gels were visualised under blue light as well as Coomasie Blue staining.

RT-PCR

The DNA in 2 µL of CFE reaction was degraded with 0.5 µL of TURBO DNase (2 U/µL, Invitrogen, AM2238) at 37°C for 30 min. The CFE reaction was quenched by addition of 15 mM EDTA and incubated at 75°C for 15 min. The denatured proteins were pelleted through centrifugation at 3,200 x *g* for 2 minutes. To prepare a 20 µL reverse transcription reaction, 2 µL of DNase-treated sample was mixed with 1 µM reverse primer (GATCCCGGCGGC), 10 mM DTT, 0.5 mM dNTP (Denville, CB4430-2), 5 U/µl protoscript II reverse transcriptase (NEB, M0368X), 1x protoscript II reverse transcriptase buffer, and 0.4 U/µL Murine RNase Inhibitor (NEB, M0314S). The reverse transcription was performed at 42°C for 1 hour, followed by the inactivation at 65°C for 20 min. A 25 µL qPCR reaction was performed by mixing the following: 1 µL of the reverse transcribed DNA, 0.8 µM forward (AAGTTCATCTGCACC ACC) and reverse (TTGAAGTCGATGCCCTTC) primers, 1x OneTaq Hot Start 2X Master Mix with Standard Buffer (NEB, M0484L), and 1x Chai Green Dye (CHAI, R01200S). The qPCR was performed on CFX96 Touch Real-Time PCR Detection System (BioRad). The thermocycling program was set up as follows: one cycle of 30 s denaturation at 95°C, 30 cycles of 15 s denaturation at 95°C, 15 s annealing at 54°C, 1 minute extension at 68°C, and one cycle of 5 minutes final extension at 68°C. The amplification curves plotted through CFX Maestro Software to determine Cq values and averaged across three replicates of each sample were calculated separately.

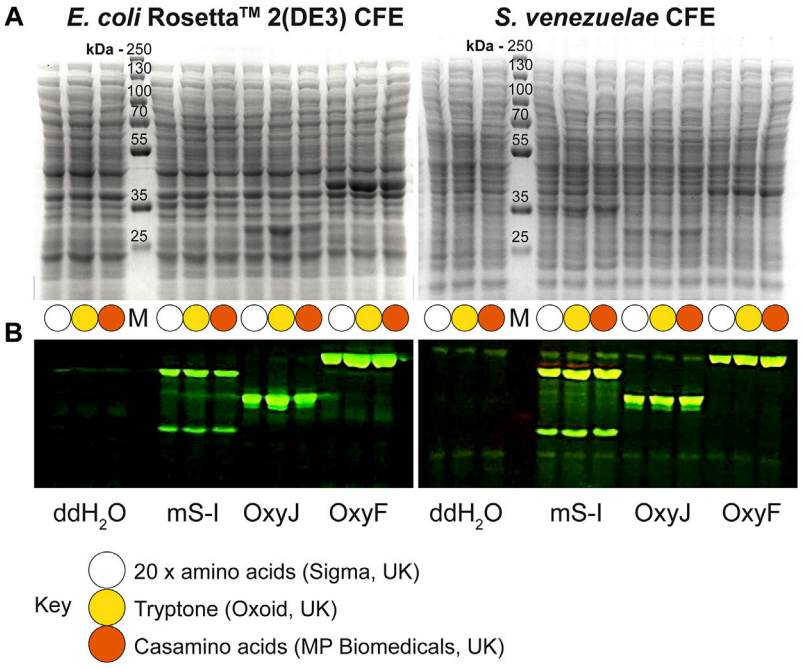


FIGURE 3
Denaturing PAGE analysis of *E. coli* Rosetta™ 2(DE3) and *S. venezuelae* CFE reactions making mScarlet-I, OxyJ and OxyF proteins using different AA sources. **(A)** Coomassie blue staining and **(B)** In-gel fluorescence stain of C-terminal tetracysteine tagged proteins. The best-performing batches for tryptone (Batch B) and casamino acids (Batch A) at 1.6% (w/v) were used to compare with the standard AAs source. Abbreviations: Pre-stained PageRuler™ (Thermo Scientific) protein marker (M); mScarlet-I (mS-I).

Results

AAs are a rate-limiting factor in CFE reactions, although some residual levels are present in most cell extracts, because of the processing of cells typically at mid-exponential growth phase. The addition of AAs to CFE makes up about 8% of the overall cost (Sun et al., 2013), although this figure depends on the AA concentration (variable) and scale. Conventionally, CFE systems use commercial AAs, which are sometimes proprietary, to mitigate for solubility and pH issues. To reduce cost and simplify the reaction mixture, we considered rich media components as an alternative AA source. This is plausible since most CFE protocols require the prior growth of microbial cells on rich media, prior to cell extract harvesting. While there are free AAs present in some rich media components, microbes also secrete proteases and peptidases to catabolise these complex mixtures.

E. coli CFE is active with a range of rich media sources

We began by investigating different rich media sources to support an *E. coli* CFE system. This is because *E. coli* is the dominant CFE system used widely throughout the international cell-free synthetic biology research community and within industrial biotechnology. Therefore, we performed these experiments in both *E. coli* Rosetta™ 2 (DE3) (Figures 1A–C) and BL21 (DE3) strains (Figures 1D–F), since these strains are the most widely used *E. coli* CFE hosts (Sun et al., 2013; Kwon and Jewett, 2015). For the *E. coli* Rosetta™ 2 (DE3) CFE reactions, we used an eGFP reporter coupled to a strong T7max promoter (Sato et al., 2022). First, there was a 6.5-fold increase in relative fluorescence when comparing no AAs with the 1 mM commercial standard (p = 0.0001). Interestingly, all four rich media components (tryptone, peptone, yeast extract and casamino acids), at a concentration of 0.4–0.8% (w/v), produced equivalent amounts of eGFP compared to the 1 mM commercial AA mixture. In contrast, concentrations above 1.67% (w/v) decreased eGFP production by approximately 40–60% (Figure 1A). In addition, we monitored fluorescent protein production over a time course. This determined that the rate of protein synthesis across assays carried out with different amino acid sources was broadly unchanged (Figure 1C). To verify whether there were any changes in mRNA abundance in the *E. coli* Rosetta™ 2 (DE3) CFE experiments, quantitative RT-PCR was performed. This experiment confirmed that none of the alternative AA sources altered mRNA expression (Supplementary Figure S1). We also repeated these experiments in a *E. coli* BL21 (DE3) CFE system using a mScarlet-I reporter coupled to a strong constitutive

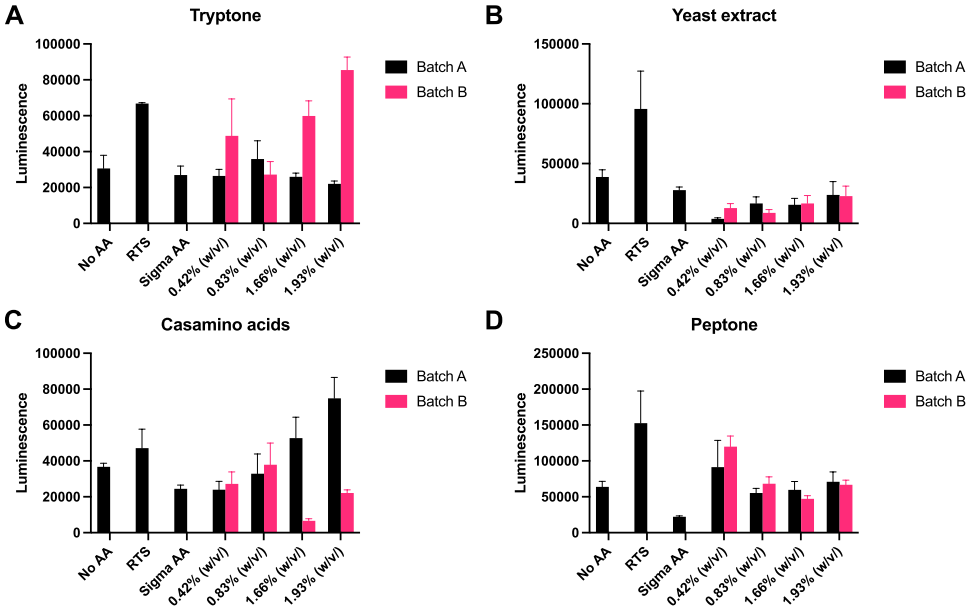


FIGURE 4
P. pastoris CFE with AA substitutes. **(A)** End-point luminescence measurements of *P. pastoris* CFE with **(A)** tryptone, **(B)** yeast extract, **(C)** casamino acids and **(D)** peptone. Error bars represent standard deviation of three technical measurements. End-point samples were collected after 2 h of incubation at 30°C.

synthetic promoter 44 (SP44) (Bai et al., 2015), which is highly active in both *E. coli* and *Streptomyces* CFE. For this, we observed broadly similar findings between the two different approaches and expression constructs (Figure 1C). Protein yields of CFE reactions with 0.4–0.8% (w/v) of alternative AA source were equivalent to that of CFE with 1 mM conventional AA mix. However, while there was significant variability between different companies/batches (Figures 1E,F), at least one batch (codes listed in supplementary information), gave equivalent activity to the commercial AA source for both the Rosetta™ 2 (DE3) and BL21 (DE3) CFE systems.

Tryptone and casamino acids provide strong *Streptomyces* CFE activity

Next, we tested the *S. venezuelae* CFE system, which we have specifically optimised to study high G + C genes from actinomycetes genomes. Here, we also used the SP44-*mScarlet-I* expression plasmid, optimised for *S. venezuelae* CFE (Bai et al., 2015). Remarkably, we observed strong production of mScarlet-I with tryptone, at up to 17 μM (0.45 mg/ml). This simple protocol modification provided up to 2-fold increase in activity (p = 0.0002) compared to the commercial AA source (Figures 2A–C), which was also observed through in-gel fluorescence staining of the C-terminal tetracysteine tagged mScarlet-I (Figure 2D) and visually (Figure 2E). However, an equivalent batch of tryptone from a different commercial source did not provide a significant increase in CFE activity (Figure 2B). Therefore, we repeated these experiments across several rich media sources, such as peptone, yeast extract, and casamino acids. In summary, while there is variation between commercial batches, tryptone and casamino acid consistently produced the highest yields of mScarlet-I protein yield. For comparison, the commercial AA source produced 8 μM mScarlet-I, while peptone yielded 14 μM mScarlet-I (Figure 2B). Yeast extract was the least effective AA source for *Streptomyces* CFE. It was surprising to observe CFE activity for casamino acids, since free L-tryptophan (L-Trp) is depleted during the manufacturing process. To test this, we compared *Streptomyces* CFE reactions with standard AA, to reactions lacking L-Trp. This only reduced the CFE activity by 48%, suggesting significant levels of L-Trp are present in the cell extract. Alternatively, if 1 mM L-Trp was supplemented to CFE reactions run with casamino acids, there was no significant change in activity. This suggests L-Trp is not sufficiently limiting in the reaction. In addition, we performed high-performance liquid chromatography analysis on two batches of casamino acids in comparison with a L-Trp standard. Casamino acids batch A (MP Biomedicals, United Kingdom), the best performing batch for CFE activity (Figure 2B), contained approximately 1 mM of L-Trp in 2% (w/v). L-Trp was not detected in batch B (Oxoid, United Kingdom), although CFE activity was equivalent to standard amino acids. In addition, to verify whether our findings were compatible with other model proteins, we compared the relative CFE production of two

Streptomyces enzymes, OxyJ and OxyF (Moore et al., 2017a). Here, we tested the best performing casamino acids and tryptone batches with both *Streptomyces* and *E. coli* Rosetta™ (DE3) 2 CFE. In-gel fluorescence and Coomassie blue staining, showed strong production of both proteins, at least equivalent to levels observed with standard AA in both *E. coli* Rosetta™ (DE3) 2 and *Streptomyces* CFE systems (Figure 3). This data overall suggests casamino acids and tryptone are an effective AA replacement for recombinant protein production purposes.

Rich media is active in *P. pastoris* CFE

Recently, eukaryotic CFE models have been optimised to provide a strong homologous system to make challenging eukaryotic proteins (Spice et al., 2020; Zhang et al., 2020) and for the study of cell-free prototyping (Kopniczky et al., 2020). To extend our findings, we repeated our experiments with *P. pastoris*, an emerging eukaryotic CFE expression system (Thoring et al., 2016; Zhang et al., 2020). For this system, we followed previous published literature, and used a luciferase reporter for detection of protein synthesis (Spice et al., 2020; Spice et al., 2022). Due to the stability of luminescence assays, the controls (no AA, RTS and Sigma AA) were tested separately against each rich media component. Across the four rich media options (tryptone, yeast extract, casamino acids and peptone), we observed broadly similar findings to the prokaryotic CFE results. Interestingly, while luminescence was observed in the absence of added AA, in comparison, the RTS AA standard, tryptone or casamino acids, showed increased activity (two to four-fold). Like the *E. coli* and *Streptomyces* CFE systems, this effect was also batch dependent(Figure 4).

Discussion

We have previously tested a variety of AA sources and methods in different prokaryotic CFE systems, with variable findings. While *E. coli* CFE systems are robust, we found that the activity of alternative CFE systems (e.g., *Streptomyces*/*Bacillus*) were sensitive to the salts or proprietary reagents required for AA stock preparation (Moore et al., 2018; Moore et al., 2021). Therefore, we desired a simple AA source for general use in CFE systems. Rich media are ubiquitous AA sources used for general microbial growth, and contain free AAs, short peptides, proteolytic fragments, and, depending on the source (e.g., yeast extract) also vitamins, cofactors, and primary metabolites. In summary, specific batches (Supplementary Table S1) of tryptone and casamino acids gave up to a 2-fold increase in protein yield (up to 0.45 mg/ml) for the *S. venezuelae* CFE system (Moore et al., 2021), which is of specific interest for studying high G + C (%) genes from related microbial genomes. In the *E. coli* and *P. pastoris* CFE systems we tested, while there was no general preference for AA source, there was at least equivalent activity for when the commercial AA standard mixture was replaced with rich media. Our experiments also show that each CFE system has specific preferences for rich media source, while there is also variation in activity between commercial brands and batches. Batch variation is an important limitation of our study, likely due to different AA frequencies. However, *E. coli* CFE reactions were shown to be more productive with a fixed AA concentration, in comparison to a mixture adjusted close to the distribution of AAs in *E. coli* cells (Lobry and Gautier, 1994; Caschera and Noireaux, 2015b). Furthermore, our finding was reproducible in three separate laboratories (United Kingdom and USA) with different CFE systems and commercial rich media sources. Considering batch variation, we suggest that while rich media may not be suitable for all cell-free experiments, especially where modelling or metabolomics is important, our method provides a cost-effective AA source with potential application for high-value recombinant protein production. For example, a typical 0.5 kg unit of rich media will provide a long-lasting and low-cost AA source for CFE–approximately 1–2 g rich media (~$6) is required per litre of CFE reagent. Therefore, this method advancement has a clear cost benefit over commercial AA sources. Compared to the widely used RTS AA kit (Biotech Rabbit), our method provides up to a 300-fold reduction in cost for the AA source (Supplementary Table S2). Depending on scale and format, AAs cost approximately 8% of the overall CFE reaction. Finally, rich media components are available in most biological laboratories. Therefore, our method is easily accessible and provides an improvement that lowers the threshold for engagement with cell-free systems, which require multiple components and where access to some laboratory reagents is limiting for some research groups. Therefore, we believe our method advancement will widen interest in cell-free systems in both academic and educational settings (Huang et al., 2018), as well as provide a cost-benefit to those interested in using CFE systems for high-value recombinant protein production.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

LN, WS, FA, and KC performed all experiments and analysed the data. TV and CS were involved in discussions and supervision of

LN and KC, respectively. SM, KA, and KP designed the study. LN, WS, TV, KP, KA, and SM wrote the manuscript. SJM and LN conceived the study. All authors read and approved the manuscript.

Funding

LN was supported by Leverhulme Trust award RPG-2021-018. WS was supported by the Funai Overseas Scholarship of The Funai Foundation for Information Technology and by the National Science Foundation award 2123465. KC was supported by a Global Challenges Research Fund studentship and University of Kent. We gratefully acknowledge funding from the EPSRC Future Targeted Healthcare Manufacturing Hub (Grant Reference: EP/P006485/1) for financial assistance with publication costs.

Acknowledgments

SM would also like to acknowledge the Future Targeted Healthcare Manufacturing Hub (EP/P006485/1) for funding support.

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/fbioe.2022.992708/full#supplementary-material

References

Anderson, M. J., Stark, J. C., Hodgman, C. E., and Jewett, M. C. (2015). Energizing eukaryotic cell-free protein synthesis with glucose metabolism. *FEBS Lett.* 589, 1723–1727. doi:10.1016/j.febslet.2015.05.045

Aw, R., and Polizzi, K. M. (2019). Biosensor-assisted engineering of a high-yield *Pichia pastoris* cell-free protein synthesis platform. *Biotechnol. Bioeng.* 116, 656–666. doi:10.1002/bit.26901

Aw, R., Spice, A. J., and Polizzi, K. M. (2020). Methods for expression of recombinant proteins using a *Pichia pastoris* cell-free system. *Curr. Protoc. Protein Sci.* 102, e115. doi:10.1002/cpps.115

Bai, C., Zhang, Y., Zhao, X., Hu, Y., Xiang, S., Miao, J., et al. (2015). Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 12181–12186. doi:10.1073/pnas.1511027112

Bennett, B. D., Yuan, J., Kimball, E. H., and Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat. Protoc.* 3, 1299–1311. doi:10.1038/nprot.2008.107

Cai, Q., Hanson, J. A., Steiner, A. R., Tran, C., Masikat, M. R., Chen, R., et al. (2015). A simplified and robust protocol for immunoglobulin expression in *Escherichia coli* cell-free protein synthesis systems. *Biotechnol. Prog.* 31, 823–831. doi:10.1002/btpr.2082

Calhoun, K. A., and Swartz, J. R. (2007). "Energy systems for ATP regeneration in cell-free protein synthesis reactions," in In vitro *transcription and translation protocols*. Editor G. Grandi (Totowa, NJ: Humana Press), 3–17.

Caschera, F., and Noireaux, V. (2015a). A cost-effective polyphosphate-based metabolism fuels an all *E. coli* cell-free expression system. *Metab. Eng.* 27, 29–37. doi:10.1016/j.ymben.2014.10.007

Caschera, F., and Noireaux, V. (2015b). Preparation of amino acid mixtures for cell-free expression systems. *Biotechniques* 58, 40–43. doi:10.2144/000114249

Charna, A. K. R., Soye, B. J. D., Ntai, I., Kelleher, N. L., and Jewett, M. C. (2022). An efficient cell-free protein synthesis platform for producing proteins with pyrrolysine-based non-canonical amino acids. *Biotechnol. J.* 17, e2200096. doi:10.1002/biot.202200096

Cui, Z., Johnston, W. A., and Alexandrov, K. (2020). Cell-free approach for non-canonical amino acids incorporation into polypeptides. *Front. Bioeng. Biotechnol.* 8, 1031. doi:10.3389/fbioe.2020.01031

Dudley, Q. M., Karim, A. S., Nash, C. J., and Jewett, M. C. (2020). *In vitro* prototyping of limonene biosynthesis using cell-free protein synthesis. *Metab. Eng.* 61, 251–260. doi:10.1016/j.ymben.2020.05.006

Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A., and Noireaux, V. (2021). The all-*E. coli* TXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. *Synth. Biol.* 6, ysab017. doi:10.1093/synbio/ysab017

Gregorio, N. E., Levine, M. Z., and Oza, J. P. (2019). A user's guide to cell-free protein synthesis. *Methods Protoc.* 2, 24. doi:10.3390/mps2010024

Huang, A., Nguyen, P. Q., Stark, J. C., Takahashi, M. K., Donghia, N., Ferrante, T., et al. (2018). BioBits™ explorer: A modular synthetic biology education kit. *Sci. Adv.* 4, eaat5105. doi:10.1126/sciadv.aat5105

Karim, A. S., Heggestad, J. T., Crowe, S. A., and Jewett, M. C. (2018). Controlling cell-free metabolism through physiochemical perturbations. *Metab. Eng.* 45, 86–94. doi:10.1016/j.ymben.2017.11.005

Kopniczky, M. B., Canavan, C., McClymont, D. W., Crone, M. A., Suckling, L., Goetzmann, B., et al. (2020). Cell-free protein synthesis as a prototyping platform for mammalian synthetic biology. *ACS Synth. Biol.* 9, 144–156. doi:10.1021/acssynbio.9b00437

Kwon, Y.-C., and Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* 5, 8663. doi:10.1038/srep08663

Lobry, J. R., and Gautier, C. (1994). Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* 22, 3174–3180. doi:10.1093/nar/22.15.3174

Meyer, C., Nakamura, Y., Rasor, B. J., Karim, A. S., Jewett, M. C., and Tan, C. (2021). Analysis of the innovation trend in cell-free synthetic biology. *Life* 11, 551. doi:10.3390/life11060551

Moore, S. J., Lai, H.-E., Chee, S.-M., Toh, M., Coode, S., Chengan, K., et al. (2021). A *Streptomyces venezuelae* cell-free toolkit for synthetic biology. *ACS Synth. Biol.* 10, 402–411. doi:10.1021/acssynbio.0c00581

Moore, S. J., Lai, H.-E., Needham, H., Polizzi, K. M., and Freemont, P. S. (2017a). *Streptomyces venezuelae* TX-TL - a next generation cell-free synthetic biology tool. *Biotechnol. J.* 12, 1600678. doi:10.1002/biot.201600678

Moore, S. J., MacDonald, J. T., and Freemont, P. S. (2017b). Cell-free synthetic biology for *in vitro* prototype engineering. *Biochem. Soc. Trans.* 45, 785–791. doi:10.1042/bst20170011

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., et al. (2018). Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4340–E4349. doi:10.1073/pnas.1715806115

Sato, W., Sharon, J., Deich, C., Gaut, N., Cash, B., Engelhart, A. E., et al. (2022). Akaby-Cell-free protein expression system for linear templates. *PLoS One* 17, e0266272. doi:10.1371/journal.pone.0266272

Shin, J., and Noireaux, V. (2010). Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.* 4, 8. doi:10.1186/1754-1611-4-8

Spice, A. J., Aw, R., Bracewell, D. G., and Polizzi, K. M. (2020). Improving the reaction mix of a Pichia pastoris cell-free system using a design of experiments approach to minimise experimental effort. *Synth. Syst. Biotechnol.* 5, 137–144. doi:10.1016/j.synbio.2020.06.003

Spice, A. J., Aw, R., and Polizzi, K. M. (2022). Cell-free protein synthesis using *Pichia pastoris*. *Methods Mol. Biol.* 2433, 75–88. doi:10.1007/978-1-0716-1998-8_4

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* 50762, e50762. doi:10.3791/50762

Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014). Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* 3, 387–397. doi:10.1021/sb400131a

Thoring, L., Wüstenhagen, D. A., Borowiak, M., Stech, M., Sonnabend, A., and Kubick, S. (2016). Cell-free systems based on CHO cell lysates: Optimization strategies, synthesis of "difficult-to-express" proteins and future perspectives. *PLoS One* 11, e0163670. doi:10.1371/journal.pone.0163670

Tian, X., Liu, W.-Q., Xu, H., Ji, X., Liu, Y., and Li, J. (2022). Cell-free expression of NO synthase and P450 enzyme for the biosynthesis of an unnatural amino acid L-4-nitrotryptophan. *Synth. Syst. Biotechnol.* 7, 775–783. doi:10.1016/j.synbio.2022.03.006

Toh, M., Chengan, K., Hanson, T., Freemont, P. S., and Moore, S. J. (2021). A high-yield *Streptomyces* transcription-translation toolkit for synthetic biology and natural product applications. *J. Vis. Exp.* 175, e63012 (1–18). doi:10.3791/63012

Wang, Y., and Zhang, Y.-H. P. (2009). Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnol.* 9, 58. doi:10.1186/1472-6750-9-58

Wiegand, D. J., Lee, H. H., Ostrov, N., and Church, G. M. (2019). Cell-free protein expression using the rapidly growing bacterium Vibrio natriegens. *J. Vis. Exp.*, 10.3791/59495. doi:10.3791/59495

Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., et al. (2011). Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* 108, 1570–1578. doi:10.1002/bit.23103

Zhang, L., Liu, W.-Q., and Li, J. (2020). Establishing a eukaryotic *Pichia pastoris* cell-free protein synthesis system. *Front. Bioeng. Biotechnol.* 8, 536. doi:10.3389/fbioe.2020.00536

Zimmerman, E. S., Heibeck, T. H., Gill, A., Li, X., Murray, C. J., Madlansacay, M. R., et al. (2014). Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system. *Bioconjug. Chem.* 25, 351–361. doi:10.1021/bc400490z

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Methods
PUBLISHED 28 September 2022
DOI 10.3389/fbioe.2022.947508

OPEN ACCESS

EDITED BY
Yuan Lu,
Tsinghua University, China

REVIEWED BY
Karthik Raman,
Indian Institute of Technology Madras, India
Jinyun Tang,
Berkeley Lab (DOE), United States
Mike Lyons,
Trinity College Dublin, Ireland

*CORRESPONDENCE
Rahul Sarpeshkar,
rahul.sarpeshkar@dartmouth.edu

SPECIALTY SECTION
This article was submitted to Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and Biotechnology

RECEIVED 18 May 2022
ACCEPTED 09 August 2022
PUBLISHED 28 September 2022

CITATION
Deng Y, Beahm DR, Ran X, Riley TG and Sarpeshkar R (2022), Rapid modeling of experimental molecular kinetics with simple electronic circuits instead of with complex differential equations.
Front. Bioeng. Biotechnol. 10:947508.
doi: 10.3389/fbioe.2022.947508

Rapid modeling of experimental molecular kinetics with simple electronic circuits instead of with complex differential equations

Yijie Deng[1], Douglas Raymond Beahm[1], Xinping Ran[1], Tanner G. Riley[2] and Rahul Sarpeshkar[1,3]*

[1]Thayer School of Engineering, Dartmouth College, Hanover, NH, United States, [2]School of Undergraduate Arts and Sciences, Dartmouth College, Hanover, NH, United States, [3]Departments of Engineering, Microbiology and Immunology, Physics, and Molecular and Systems Biology, Dartmouth College, Hanover, NH, United States

Kinetic modeling has relied on using a tedious number of mathematical equations to describe molecular kinetics in interacting reactions. The long list of differential equations with associated abstract variables and parameters inevitably hinders readers' easy understanding of the models. However, the mathematical equations describing the kinetics of biochemical reactions can be exactly mapped to the dynamics of voltages and currents in simple electronic circuits wherein voltages represent molecular concentrations and currents represent molecular fluxes. For example, we theoretically derive and experimentally verify accurate circuit models for Michaelis-Menten kinetics. Then, we show that such circuit models can be scaled via simple wiring among circuit motifs to represent more and arbitrarily complex reactions. Hence, we can directly map reaction networks to equivalent circuit schematics in a rapid, quantitatively accurate, and intuitive fashion without needing mathematical equations. We verify experimentally that these circuit models are quantitatively accurate. Examples include 1) different mechanisms of competitive, noncompetitive, uncompetitive, and mixed enzyme inhibition, important for understanding pharmacokinetics; 2) product-feedback inhibition, common in biochemistry; 3) reversible reactions; 4) multi-substrate enzymatic reactions, both important in many metabolic pathways; and 5) translation and transcription dynamics in a cell-free system, which brings insight into the functioning of all gene-protein networks. We envision that circuit modeling and simulation could become a powerful scientific communication language and tool for quantitative studies of kinetics in biology and related fields.

KEYWORDS

biological circuits, cell-free system, kinetic modeling, michaelis-menten equation, biochemical engineering, circuit modeling, enzyme kinetics, reversible reactions

1 Introduction

Kinetic modeling has been a powerful tool for studying biological systems from simple enzymatic reactions to metabolic pathways, drug kinetics in hosts, gene circuits in synthetic biology, and host-pathogen interactions (Alves et al., 2006; Resat et al., 2009; Bevc et al., 2011; Eshtewy and Scholz, 2020; Néant et al., 2021). Modeling molecular kinetics can provide quantitative insights and mechanistic understandings of biological systems. However, kinetic modeling of biological processes relies on a substantial number of mathematical equations to describe even simple biochemical reactions. The heavy dependence on long and tedious differential equations hinders many biologists from appreciating and taking advantage of kinetic modeling as a powerful tool for studying biological questions. In particular, for many biological researchers, the long list of parameters and abstract terms that are used during the process of mathematical derivation are exhausting and difficult to follow. In addition, it can be challenging to resolve complex, nonlinear, coupled differential equations that require sophisticated algorithms/programs including numerical approaches (Bevc et al., 2011) for simulating time-course kinetics.

However, ordinary differential equations (ODEs), commonly used to model biochemical reactions and processes, can be represented by simple electronic circuits (Sarpeshkar, 2010; Teo and Sarpeshkar, 2020) in a mathematically exact fashion. We can thus take advantage of electronic design software to design circuits *in silico* that represent the kinetics of the target system and then run simulations in software without the need to manufacture the physical circuits (Teo et al., 2019a; Teo and Sarpeshkar, 2020). Therefore, not only can we visualize all the math equations in one circuit but also solve them by just running simulations on the circuit. Using virtual electronic circuits enables one to do rapid kinetic modeling of biochemical reactions without deriving tedious differential equations. In addition, circuit simulation in electronic design software is able to provide accurate time-course dynamics, not just equilibrium solutions. Circuit software has built-in algorithms to automatically solve underlying equations represented by the circuits, which has evolved over 75 + years of circuit design for multiple forms of design and analysis (Sarpeshkar, 2010).

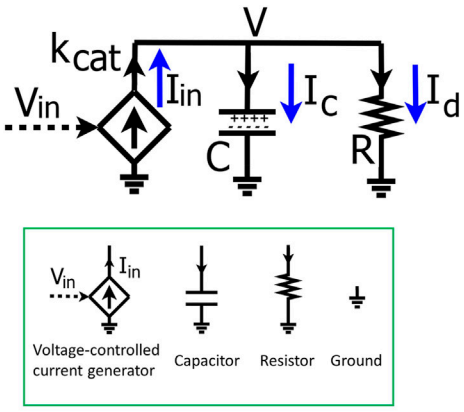
The overall mechanism of the simulation is that, given some preset parameters of a circuit, the voltage and current at any node of the circuit at any time are readily available upon simulation; these voltages and currents exactly represent the corresponding molecular concentrations and molecular reaction flux rates, respectively. For example, we have used electronic circuits to model and simulate complex biological processes including genetic circuits in synthetic biology (Daniel et al., 2013; Teo et al., 2015; Zeng et al., 2018); kinetics of microbial growth and energetics (Deng et al., 2021); tissue homeostasis (Teo et al., 2019b); and virus-host interactions (Beahm et al., 2021). However, there are gaps in biologists' understanding of electronic circuits and the underlying mathematics; and, in their understanding of the analogy of circuit variables to reaction kinetic parameters. These gaps have prevented many researchers from understanding circuit models and using circuits to do kinetic modeling in practice.

Therefore, an important goal of this work is to illustrate how the mathematics describing the kinetics of biochemical reactions can be exactly mapped to electronic circuits; and, to demonstrate how to use such circuits to do rapid kinetic modeling without deriving math equations. To demonstrate the circuit modeling approach, we start with the basics of a simple resistor-capacitor (RC) circuit and its use in representing the dynamics of a simple biochemical reaction. We then illustrate how to use circuits to model an enzyme-substrate reaction that is characterized by Michaelis-Menten kinetics, one of the most fundamental processes in biology. Next, we develop widely applicable circuit motifs for biochemical reactions including different types of enzyme inhibition, multi-molecular binding, multi-substrate reactions, reversible reactions, and DNA transcription and translation. Notably, these circuit models are validated by good fits to our experimental data. These fundamental circuit motifs can be easily used to construct large-scale circuit models for complicated biological networks/pathways without using cumbersome math equations; mature circuit-simulation software can then automatically provide accurate solutions including the time-course kinetics of molecules. Large circuit models are useful in analyzing the behavior of biological systems and to discover natural algorithms and architectures in biology. In addition, our approach provides mechanistic insights into fundamental biochemical reactions, such as the kinetics of enzyme inhibition and the kinetics of sequential binding reactions.

2 Results

2.1 Mapping a basic chemical reaction to a simple electronic circuit

To help biologists understand the basic mathematics of electronic circuits, we first derive, step-by-step, the basics of a simple RC circuit that is foundational for kinetic modeling in biological systems. The RC circuit consists of a resistor (R), a capacitor (C), and an input current (I_{in}) that is generated by a voltage-controlled current generator (a 'transconductor') that converts the input voltage (V_{in}) into the current (I_{in}) (Figure 1). In electronics, a current is denoted by I (measured in Amps, A), a voltage is denoted by V (measured in Volts, V), a resistor has a resistance of R (measured in Ohms, Ω), and a capacitor has a capacitance of C (measured in Farad, F = A*s/V). As shown in Figure 1, the input current (I_{in}) goes through the capacitor and the resistor; the voltage (V) on the capacitor and



By Ohm's law, the current (I_d) through a resistor (R):

$$I_d = \frac{V}{R} \quad (1)$$

The current (I_c) through a capacitor (C):

$$I_c = \frac{C*dV}{dt} \quad (2)$$

The input current by the current generator:

$$I_{in} = V_{in} * k_{cat} \quad (3)$$

By Kirchhoff's current law:

$$I_c = I_{in} - I_d \quad (4)$$

Thus, the dynamics of voltage (V) is described as:

$$\frac{C*dV}{dt} = V_{in} * k_{cat} - \frac{V}{R} \quad (5)$$

FIGURE 1
The basics of a resistor-capacitor (RC) circuit fed by a transconductor input. The input current is generated by the transconductor (diamond symbol), i.e., a voltage-controlled current generator that converts the input voltage (V_{in}) into the input current (I_{in}) with a conversion factor of k_{cat}. The dynamics of the voltage (V) over the capacitor (C) and the resistor (R) are determined by the input current (I_{in}) and the current (I_d) through the resistor. Electronic circuit symbols are shown in the green box.

the resistor keeps increasing until it reaches a steady state wherein the capacitor is fully charged; thereafter, all the input current (I_{in}) goes through the resistor. To describe how the voltage (V) changes over time, we first calculate three currents in the circuit as below:

According to Ohm's law, the current (I_d) going through the resistor is defined by Eq. 1:

$$I_d = \frac{V}{R} \quad (1)$$

The total charge on the capacitor is Q and thus the current (I_c) through the capacitor is described as:

$$I_c = \frac{dQ}{dt} = \frac{C^*dV}{dt} \quad (2)$$

As mentioned above, the input current (I_{in}) is generated by a voltage-controlled current source (the diamond-shaped device or 'transconductor' in Figure 1) that converts the input voltage (V_{in}) into the current with a conversion factor (k_{cat}), and thus we have:

$$I_{in} = V_{in}{}^* k_{cat} \quad (3)$$

The input current is split into I_c and I_d in the circuit. By Kirchhoff's current law, we have:

$$I_c = I_{in} - I_d \quad (4)$$

We substitute Eqs 1–3 into Eq. 4 and thus have Eq. 5 that describes the voltage dynamics in the RC circuit:

$$\frac{C^*dV}{dt} = I_{in} - I_d = V_{in}{}^* k_{cat} - \frac{V}{R} \quad (5)$$

The dynamics of the voltage (V) over the capacitor and resistor are determined by the input current (I_{in}) and the current through the resistor (I_d). A simple analogy for the circuit is that the product concentration in a reaction system is determined by the production rate and the degradation rate. Given constant C and k_{cat} in Eq. 5, the voltage dynamics are thus determined by V_{in} and R which can be translated into biological relevance, as we discuss later. Eqs 1–5 describe the basics of a simple RC circuit and are foundational for understanding the map between circuit modeling and mathematical modeling of kinetic processes. So, we have summarized their derivation in Figure 1.

With the voltage dynamics described by Eq. 5, we normalize the equation by C such that the change of V over time is described as:

$$\frac{dV}{dt} = \frac{V_{in}{}^* k_{cat}}{C} - \frac{V}{RC} \quad (6)$$

To simplify the equation, we normally set C = 1 (F) = 1 (A*s/V) in circuit modeling and the above equation becomes:

$$\frac{dV}{dt} = V_{in}{}^* k_{cat} - \frac{V}{R} \quad (7)$$

Eq. 7 describes the voltage kinetics in the RC circuit (Figure 1) when C = 1. Setting C = 1 in circuit models enables a direct map to the mathematics behind the kinetics of a chemical reaction, as we illustrate in Figure 2. In Figure 2A, substrate S is converted into a product P with a production rate constant of k_{cat} (1/s) and the product decays with a rate constant of 1/r (1/s). The kinetics of this reaction can be exactly described by an equivalent RC circuit fed by a transconductor (Figure 2B) which is identical to the RC circuit above (Figure 1). In the context of biological systems, we can consider the same reaction taking place in a container or a cell with a volume of C (liter, L) (Figure 2D). According to the law

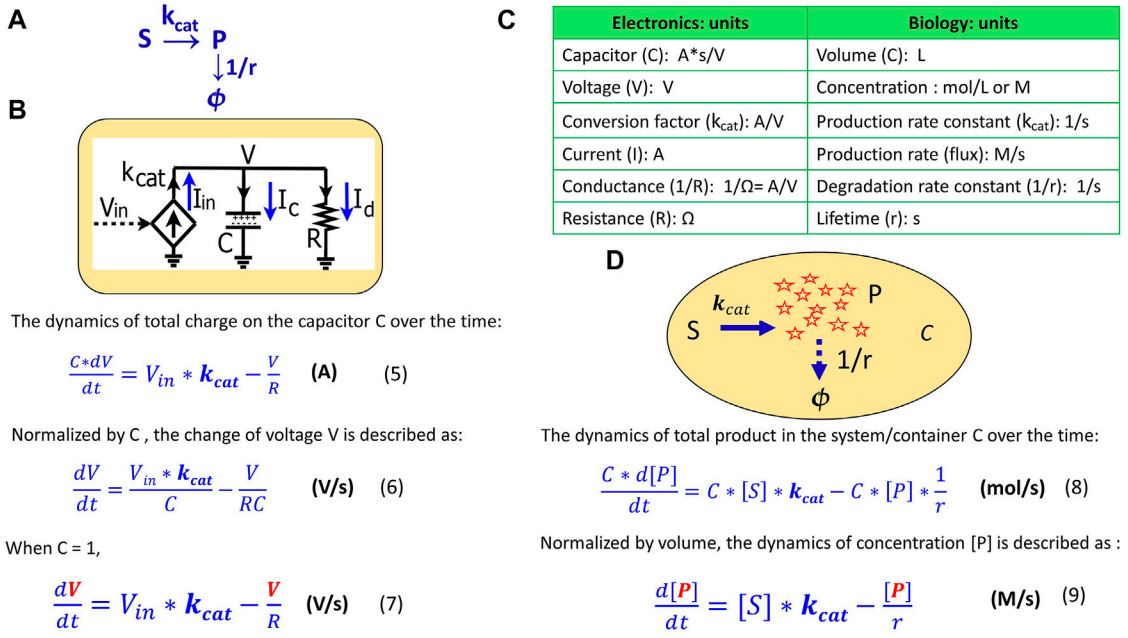


FIGURE 2
The mapping of an elementary biochemical reaction to an equivalent electronic circuit. **(A)** An example of a simple biochemical reaction wherein substrate S is converted to product P at a rate constant of k_{cat} (1/s) while the product also decays at a rate constant of 1/r (1/s). **(B)** A simple RC circuit in the context of the chemical reaction. **(C)** Translation of electronic variables into biochemical kinetics in a reaction. **(D)** The same biochemical reaction taking place in a container or a cell with a volume of C. The capacitance of a capacitor is normally set C = 1 A*s/V, which represents a volume-normalized container in a system (per L). Some important equations are summarized in this figure for comparison.

of mass action in chemistry, the production rate is proportional to the concentration of S and thus is S^*k_{cat} (M/s); similarly, the decay rate of the product is P * (1/r) (M/s). The total amount of product P changes over time (mol/s) in the container, and is thus described as below:

$$\frac{C^*d[P]}{dt} = C^*[S]^*k_{cat} - C^*[P]^*\frac{1}{r} \quad (8)$$

where [S] and [P] are the concentrations (M); C is the container volume (L); k_{cat} and 1/r are rate constants (1/s). Eq. 5 is physically parallel to Eq. 8, wherein the former describes the change of the total charge of the capacitor while the latter describes the change of the total amount of product in a container. Then, we normalize Eq. 8 by the container volume C and thus have the concentration kinetics:

$$\frac{d[P]}{dt} = [S]^*k_{cat} - \frac{[P]}{r} \quad (9)$$

Eq. 9 describes how the concentration of a product changes over time in a container, a cell, or in any reaction system. When comparing Eq. 9 to Eq. 7, we notice that they become mathematically identical with the input voltage V_{in} representing the substrate concentration [S]; the voltage V representing the product concentration [P]; k_{cat}, the conversion factor for the current-generating transconductor, representing a production rate constant; and, the resistance R defining the time constant (or lifetime, r) of the product. The side-to-side comparison between electronic dynamics and chemical dynamics for this foundational production reaction is summarized in Figure 2C.

As in this production-reaction example, the dynamics in electronic circuits can be translated into the kinetics of biochemical reactions in more complex systems as well: the voltage (V) corresponds to the concentration (M) of a reagent in a chemical reaction; the current (A) is analogous to a reaction flux (M/s); the resistor R (Ω) defines the degradation of a product, with 1/(RC) corresponding to the degradation rate constant and RC corresponding to the equivalent time constant (lifetime); the capacitor with capacitance C = 1 (A*s/V) corresponds to a volume-normalized container in a biochemical system (per L). Unless otherwise mentioned, all capacitors in our circuits have C = 1. The dynamics of P are determined by the production flux and the degradation flux (Figure 2D) while the dynamics of the voltage (V) in the RC circuit (Figure 2B) are determined by the input source current (I_{in}) and the sink current (I_d).

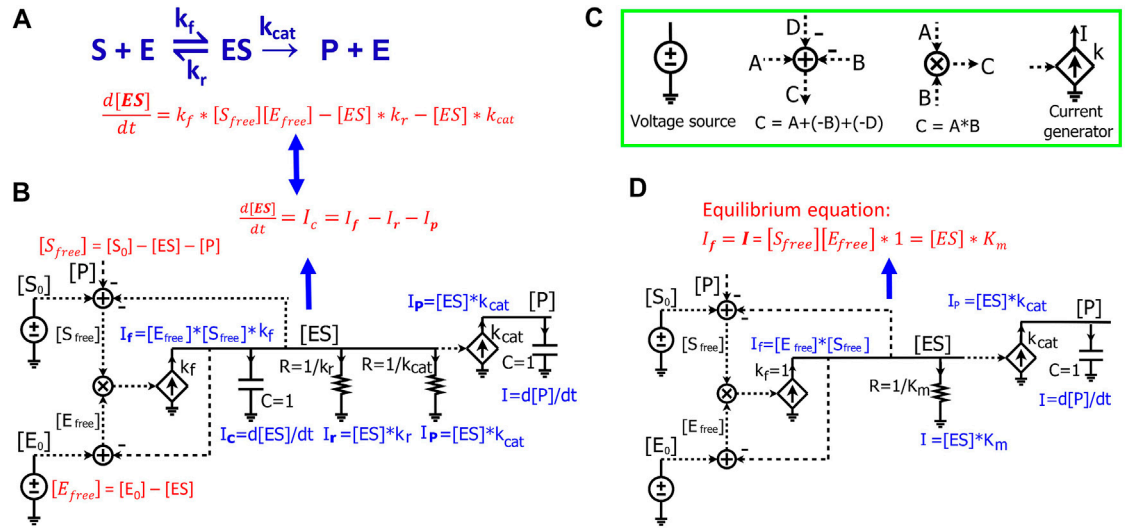


FIGURE 3
Modeling Michaelis-Menten kinetics of enzymatic reactions by simple electronic circuits. **(A)** A general enzymatic reaction wherein the enzyme E binds to the substrate S, forming an enzyme-substrate complex ES, which converts S to a product P. **(B)** The electronic circuit exactly describes the kinetics of the enzymatic reaction in **(A)**. All the math equations describing the voltages and/or currents of the circuit are indicated near the corresponding nodes. The dashed lines are wires connecting the same voltage between two nodes/components in the circuit and have no current running through them. The voltages labeled with the same names indicate that they have the same values. The voltages are mainly for math calculations such as calculating the mass conservation of a reagent via the adder/subtracter blocks, or multiplying two concentrations via a multiplier block. They are also used as inputs to voltage-dependent current generators (transconductors, the diamond symbols) to control their output currents. **(C)** Electronic symbols used in the circuits in addition to the symbols from Figure 1. **(D)** The Michaelis-Menten circuit of **(A)**, but with a steady-state approximation such that the [ES] capacitor is removed. Since the capacitor has been removed, resistors are directly related to steady-state Michaelis-Menten constants only and do not affect dynamic parameters like time constants. In this case, the resistor R = 1 /K_m (Ω) and K_m are in the standard molar concentration unit, M.

2.2 Circuit modeling of Michaelis-Menten kinetics

We next demonstrate how to use circuits to simulate Michaelis-Menten kinetics for enzyme-substrate interactions. For a general enzymatic reaction (Figure 3A), enzyme E binds to a substrate S to form an intermediate enzyme-substrate complex ES at a rate constant of k_f; ES will then either dissociate into E and S with a reverse rate constant of k_r or be converted to product P and free enzyme E_{free} with a rate constant of k_{cat}. From mass conservation, we have:

$$\left[E_{free}\right] = [E_0] - [ES] \tag{10}$$

$$\left[S_{free}\right] = [S_0] - [ES] - [P] \tag{11}$$

where $[E_0]$ and $[S_0]$ are the initial concentrations of enzyme and substrate, respectively. The enzyme-substrate complex is converted into product P with a rate proportional to its concentration [ES], such that:

$$\frac{d[P]}{dt} = k_{cat}{}^{*}[ES] \tag{12}$$

The kinetics of [ES] is determined by three fluxes: the forward reaction rate, $k_f{}^{*}[E_{free}][S_{free}]$, the reversed reaction rate, $[ES]^{*}k_r$, and the catalytic reaction rate $[ES]^{*}k_{cat}$. Therefore, the dynamics of [ES] are described as below:

$$\frac{d[ES]}{dt} = k_f{}^{*}\left[S_{free}\right]\left[E_{free}\right] - [ES]^{*}k_r - [ES]^{*}k_{cat} \tag{13}$$

The circuit (Figure 3B) exactly represents the enzymatic reaction (Figure 3A). In this circuit, voltages of the wires are labeled with names corresponding to components of the enzymatic reaction. The dashed lines are wires that don't have current going through them but still have the same voltage as the wires or nodes that they originate from. We first derive equations for currents and voltages in the circuit. Since there is no current running through any of the dashed lines/wires, this circuit (Figure 3B) is similar to the circuits we derived above (Figures 1, 2B), consisting of two RC blocks connected together, one with two resistors and the other with no resistors. The dynamics of the voltage [ES] are determined by three currents: I_f, I_r and I_p. Because the voltage across the resistors and capacitor is [ES], by Ohm's law, the current through the resistor (R = $1/k_r$) is: I_r = [ES]/R = $[ES]^{*}k_r$ which represents the reverse reaction rate/flux; similarly, given the other resistor (R = $1/k_{cat}$), the current through it is: $I_p = [ES]^{*}k_{cat}$ which represents the catalytic flux; given that the current generator has a conversion factor of k_f and an input voltage $[E_{free}]^{*}[S_{free}]$ that is calculated by the

multiplier, we have the input current $I_f = [E_{free}]^*[S_{free}]^*k_f$. We note that the latter three currents are mathematically exactly the same as the three reaction fluxes in the enzymatic reaction (Eq. 13). Finally, the current through the capacitor is $I_c = \frac{C^*d[ES]}{dt} = \frac{d[ES]}{dt}$ (when C = 1) (Figure 3B). According to Kirchhoff's current law, the current through the capacitor is given by:

$$\begin{aligned}\frac{d[ES]}{dt} &= I_c = I_f - I_r - I_p \\ &= k_f^*\left[S_{free}\right]\left[E_{free}\right] - [ES]^*k_r - [ES]^*k_{cat}\end{aligned} \tag{14}$$

Eq. 14 describes the dynamics of the voltage [ES] in the RC circuit (Figure 3B), which is the same as Eq. 13 that we derived from the enzymatic reaction (Figure 3A). Just as in the enzymatic reaction, the dynamics of [ES] are determined by one generation reaction flux and two consumption fluxes (Eq. 13); the dynamics of the voltage [ES] in the RC circuit are determined by one source generation current and two sink consumption currents through the resistors (Eq. 14).

Finally, the voltage [P] is determined by I_p which is generated by a current generator with an input voltage [ES] and a conversion factor of k_{cat}. Since all current generated flows into the capacitor (C = 1), we have:

$$I = \frac{C^*d[P]}{dt} = \frac{d[P]}{dt} = I_p = k_{cat}^*[ES] \tag{15}$$

Eq. 15 describes the dynamics of the voltage [P] in the RC circuit, which is exactly the same as Eq. 12 that describes the product kinetics in the enzymatic reaction.

In the circuit model (Figure 3B), two adders are used to calculate $[E_{free}]$ and $[S_{free}]$ based on the law of mass conservation. The binding of enzyme and substrate is represented by the multiplier symbol resulting in a signal, $[E_{free}]^*[S_{free}]$, which is the input voltage used to generate current in the first transconductor in Figure 3B. Given the conversion factor of k_f, the resulting current is $[E_{free}]^*[S_{free}]^*k_f$.

Since all the equations describing the electronic circuit and the enzymatic reaction (Figures 3A,B) are mathematically identical, we can directly use the electronic circuit to simulate the kinetics of enzymatic reactions without deriving the underlying equations. The changes of concentrations and reaction fluxes over time are directly mapped to the corresponding changes in voltages and currents, respectively. Therefore, electronic circuits enable a powerful and intuitive method for visualizing multiple math equations in one pictorial schematic. Using these circuits is especially advantageous when one wants to simulate complicated biological pathways/networks where hundreds of differential equations can be represented in a single circuit. To draw/construct and simulate such electronic circuits, multiple electronic software packages are widely and easily available, including Cadence (Cadence Design Systems, Inc.), CircuitLab (https://www.circuitlab.com/), or MATLAB Simulink/Simscape Electrical (The MathWorks, Inc.). Once the circuits are constructed, we can simply run simulations with these tools. The dynamics of the voltages and currents then directly represent real-time changes in the concentrations and reaction fluxes in biochemical reactions, respectively.

The circuit in Figure 3B is an exact circuit for representing biochemical reactions without using any mathematical assumptions/approximations; however, circuit simulation requires known values of k_f and k_r which are not normally available for most enzymatic reactions. To circumvent this requirement and make the circuit more useful in practice, we apply the same assumptions that Michaelis-Menten equation uses to simplify dynamics: Under the quasi steady-state assumption that enzyme-substrate binding is much faster than the substrate-to-product conversion output reaction, the ES concentration is assumed to reach a steady state almost instantaneously. Therefore,

$$\frac{d[ES]}{dt} = k_f^*\left[S_{free}\right]\left[E_{free}\right] - [ES]^*k_r - [ES]^*k_{cat} = 0 \tag{16}$$

Normalizing the equation by k_f and grouping the [ES] terms, we have:

$$\left[S_{free}\right]\left[E_{free}\right] - [ES]^*(k_r + k_{cat})/k_f = 0 \tag{17}$$

Letting $K_m = (k_r + k_{cat})/k_f$, we have the equilibrium equation below:

$$\left[S_{free}\right]\left[E_{free}\right] = [ES]^*K_m \tag{18}$$

where K_m (M) is the Michaelis-Menten constant. Accordingly, we can also modify the circuit of Figure 3B to reflect the equilibrium Eq. 18. Since it is assumed that [ES] reaches a steady state instantaneously, it means that the capacitor connected to the [ES] voltage node is zero (Figure 3B), architecting an effective RC time constant of zero. Accordingly, we remove the capacitor for [ES] from the circuit in Figure 3B. Next, to normalize both forward and reverse reaction rates by k_f, we set $k_f = 1$ in the circuit. Finally, since we have combined k_f and k_{cat} as in Eq. 17, we merge the two currents, I_r and I_p, into one current $([ES]^*(k_r + k_{cat})/k_f)$ through a resistor with resistance of $1/K_m$. Therefore, we have a simplified circuit (Figure 3D) containing only the parameters K_m, k_{cat}, S_0, and E_0, which are readily available from experiments.

We can easily confirm that the simplified circuit correctly reflects Eq. 18; the forward current generated by the current generator is $I_f = [S_{free}][E_{free}]$, and the current that goes into the resistor is: $I = [ES]^*K_m$. These two currents are equal since all generated current goes through the resistor. Therefore, we have $I_f = I = [S_{free}][E_{free}] = [ES]^*K_m$, which is Eq. 18 describing [ES] in the reaction. This simplified circuit (Figure 3D) exactly characterizes the kinetics of the enzymatic reaction (Figure 3A) under the steady-state approximation and is the Michaelis-Menten

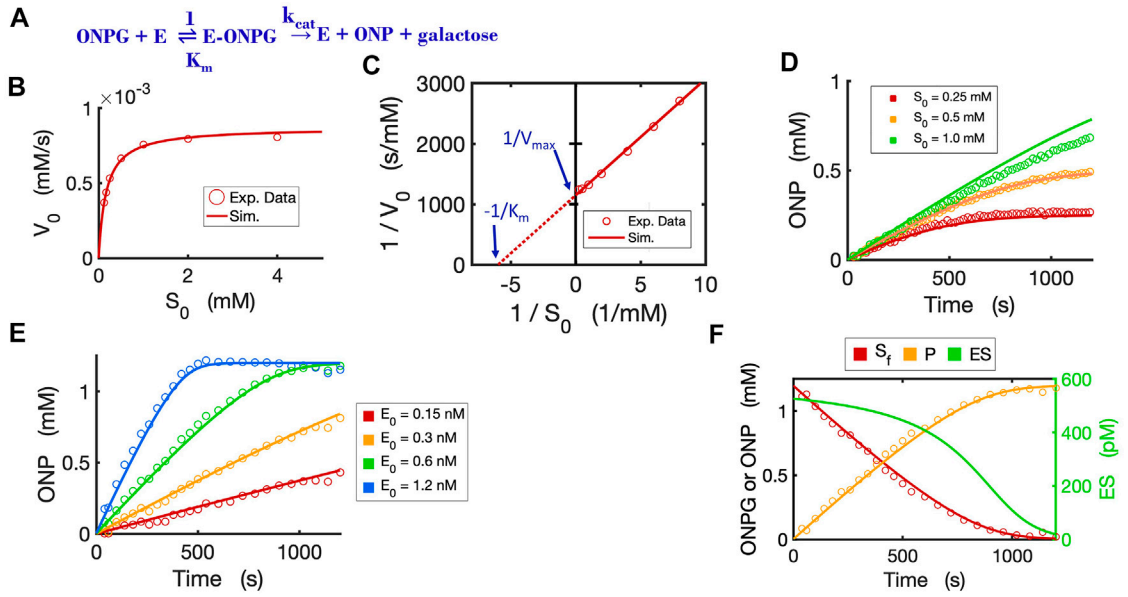


FIGURE 4
Circuit modeling of the kinetics of the beta-galactosidase reaction. **(A)** The enzymatic reaction scheme for E (beta-galactosidase) with its substrate ONPG. **(B)** Circuit simulation curve of initial reaction rate (V_0) versus initial substrate concentration [S_0] fitted to the experimental data (when E_0 = 0.3 nM). **(C)** Lineweaver-Burk plot showing linearized curves of $1/V_0$~$1/S_0$ with a *Y*-intercept of $1/V_{max}$ and *X*-intercept of -1 /K_m. **(D)** Circuit simulation curves of product dynamics over time with varying initial S concentrations fitted to experimental data with E_0 = 0.3 nM. **(E)** Circuit simulation curves of product dynamics over time with varying initial enzyme concentrations (0.15–1.2 nM) fitted to experimental data with S = 1.2 mM. **(F)** Model-predicted curves of the dynamics of [S_f], [ES] and [P] over time with E_0 = 0.6 nM and S_0 = 1.2 mM. The data points for [P] are experimental data while the data points for the free substrate [S_f] are calculated from stoichiometric conservation to be [S_f] = [S_0] - [P]. All simulation curves are obtained from the Michaelis-Menten circuit model with experimentally measured K_m = 0.167 mM and k_{cat} = 2903/s. Note that the resistor R = 1/K_m = 1/0.000167 = 5988 Ω. All data points are means of three independent replicates. The standard deviations are relatively small (less than 20% of the corresponding mean) and are not shown.

equation in the circuit form. This Michaelis-Menten circuit is the basic building block for enzymatic reactions and can be easily extended into more complicated circuits for different mechanisms of enzyme inhibition as we show below. It should be noted that besides ensuring enzyme conservation, i.e., that $[E_{free}]$ and $[ES]$ sum to $[E_0]$, the circuit of Figure 3D also ensures substrate conservation: $[S_{free}] = [S_0] - [ES]$. Ensuring the conservation of both the enzyme and substrate species allows the Michaelis-Menten circuit to be more robust and accurate, especially under scenarios where the enzyme concentration and substrate concentration are comparable (discussed later).

2.3 Circuit modeling of a hydrolytic reaction by beta-galactosidase

To validate the Michaelis-Menten circuit (Figure 3D), we fit our circuit model to experimental data that we collected from an enzyme-substrate reaction. We chose to use a hydrolytic reaction wherein beta-galactosidase is the enzyme and ONPG is the substrate (Figure 4A). We ran the circuit simulation with the four necessary parameters (K_m, k_{cat}, S_0 and E_0), which were all experimentally determined under our test conditions. As we mentioned above, the reaction flux is given by the current I_p in the circuit while the voltage [S_0] reflects the initial substrate concentration. As expected, the circuit model perfectly matches the measured initial reaction rates when we varied the initial substrate concentrations (Figure 4B). The predicted curve (V_0 ~ S_0) is a classic hyperbolic curve for an enzymatic reaction; as S_0 increases the initial reaction rate also increases until it reaches the maximal rate. For Lineweaver-Burk plotting, the circuit model accurately predicts a straight line ($1/V_0$ versus $1/S_0$) with K_m = 0.167 mM (derived from the *X*-intercept in Figure 4C) and V_{max} = 0.00087 mM/s (derived from the *Y*-intercept in Figure 4C) with an excellent fit to the experimental data.

The circuit model can also accurately simulate the time-course dynamics. As shown in Figure 4D, the simulation curves of [P] dynamics fit our experimental data closely under varying initial substrate concentrations. In addition, when we changed the initial concentration of the enzyme with a constant substrate concentration, as expected, the circuit model predicts product dynamics that are in good agreement with our experimental data (Figure 4E). As more enzyme is added, the reaction consumes substrate faster and reaches a plateau sooner. The circuit model

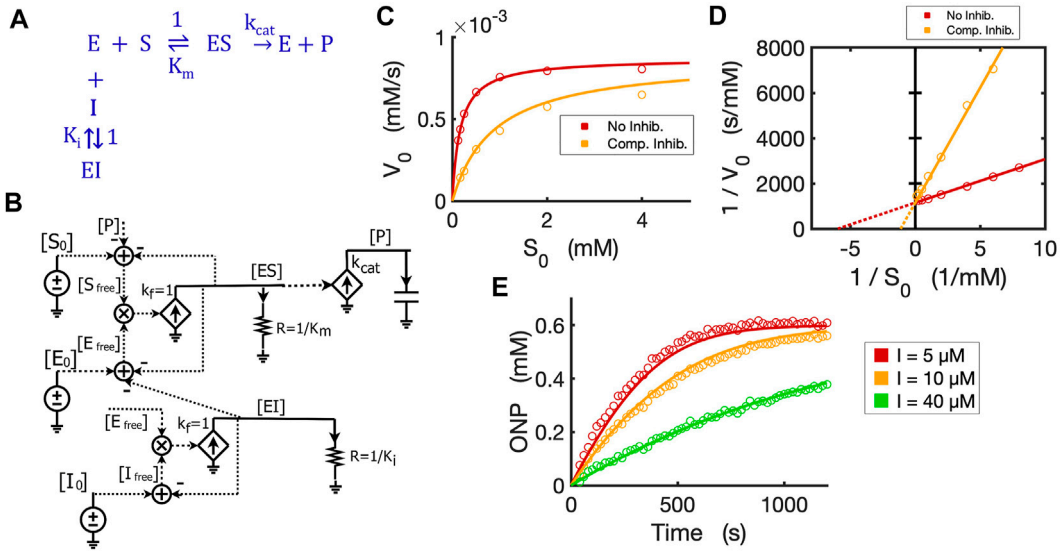


FIGURE 5
Circuit model of competitive inhibition. **(A)** The classic reaction scheme for competitive inhibition. **(B)** The circuit model for competitive inhibition. In this case, E is beta-galactosidase, S is ONPG, and the competitive inhibitor is phenylethyl beta-D-thiogalactopyranoside (PETG). **(C)** Simulation curves of initial reaction rate (V_0) versus initial substrate concentration [S_0] fitted to experimental data in the absence and presence of PETG (10 μM) when E_0 = 0.3 nM. **(D)** Lineweaver-Burk plot showing linearized curves of $1/V_0$~$1/S_0$ fitted to experimental data with or without the inhibitor, which have the same *Y*-intercept, $1/V_{max}$. **(E)** The model curves of product dynamics over time with varying inhibitor concentrations fitted to experimental data points (E_0 = 1.0 nM, S_0 = 0.6 mM). All simulation curves are obtained from the circuit model with experimentally measured K_i = 2.33 μM, and the same K_m and k_{cat} from Figure 4. All data points are means of three independent replicates. The standard deviations are relatively small (<20% of the corresponding mean) and are not shown.

also accurately predicts the dynamics of [S_{free}], [P], and [ES] over time (Figure 4F).

2.4 Circuit modeling of competitive inhibition and product-feedback inhibition

With the basic Michaelis-Menten circuit validated, we developed a circuit model for competitive inhibition. In the classic competitive inhibition model, an inhibitor binds to an enzyme at the substrate-binding site and competes with the substrate for the free enzyme, as shown in the reaction scheme (Figure 5A). To model the competitive inhibition in a circuit, we only need to add an enzyme-inhibitor binding circuit to the same Michaelis-Menten circuit (Figure 3D). Given the assumption that inhibitor binding is also much faster than the catalytic reaction and that the enzyme-inhibitor complex [*EI*] also, therefore, reaches steady-state instantaneously, we have the equilibrium condition:

$$\left[I_{free}\right]\left[E_{free}\right] = [EI]^{*}K_i \tag{19}$$

where K_i is the dissociation constant for the inhibitor. Eq. 19 is similar to Eq. 18, so we make a similar circuit (the lower circuit of Figure 5B) representing the equilibrium enzyme-inhibitor binding as described in Eq. 19. In this circuit block, all we need is to set R = $1/K_i$ to reflect the inhibitor binding constant. The voltage [*EI*] in the circuit is then wired to the [E_0] adder to account for the consumption of free enzyme that has been competitively bound by the inhibitor. Therefore, we have a competitive inhibition circuit model (Figure 5B). We used a competitive inhibitor of beta-galactosidase and experimentally demonstrated that this circuit is identical to the classic equations describing the kinetics of competitive inhibition. The circuit model accurately predicts the relationship between initial reaction rates and initial substrate concentrations (V_0 ~ S_0) in the presence and absence of the inhibitor (Figure 5C). As [S_0] increases, the initial reaction rate V_0 also increases and eventually will reach the same maximal rate V_{max} even in the presence of the inhibitor. The linearized curves (Lineweaver-Burk plot) show the expected behavior of competitive inhibition where the inhibitor increases the apparent K_m but not the maximal reaction rate (V_{max}) (Figure 5D). In addition, the circuit model can also exactly predict the product dynamics over time under different inhibitor concentrations (Figure 5E). Thus, using our experimental data, we have verified that the circuit model accurately describes the kinetics for competitive inhibition.

The circuit for competitive inhibition is a useful building block and can be used to construct circuits for complicated

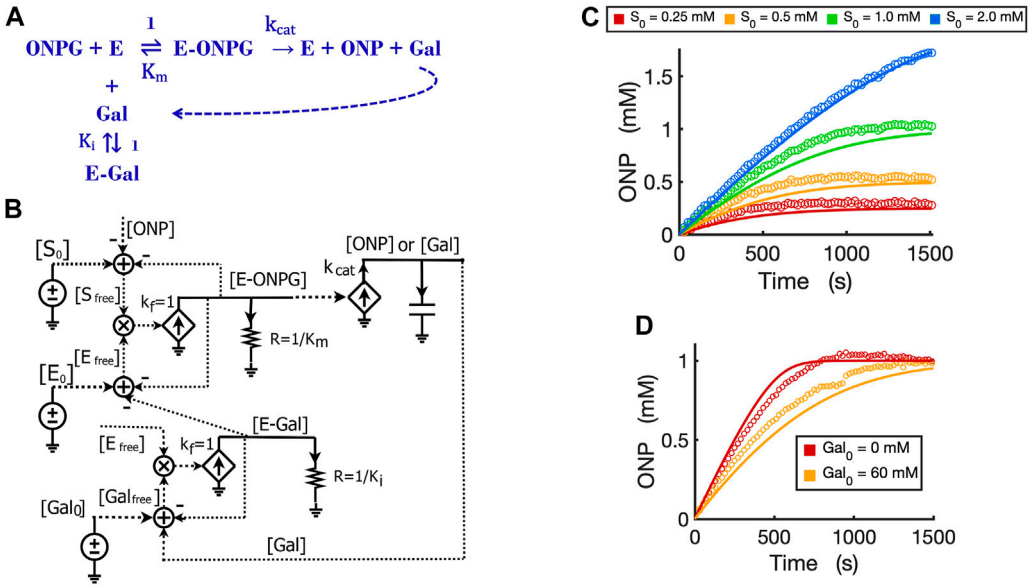


FIGURE 6
Circuit modeling of product feedback inhibition. **(A)** The scheme of product feedback inhibition based on competitive inhibition. Galactose [Gal] is one of the products and is also an inhibitor to the enzyme, beta-galactosidase. **(B)** The circuit model for product feedback inhibition: $[S_0]$ is the initial ONPG added to the reaction while $[Gal_0]$ is the initial galactose added to the reaction. **(C)** Model curves of product dynamics over time fitted to experimental data with varying initial substrate concentrations (when $[Gal_0]$ = 40 mM, E_0 = 0.7 nM). **(D)** Model curves of product dynamics fitted to experimental data with or without galactose added to the reaction (when ONPG = 1 mM, E_0 = 0.85 nM). All simulation curves are obtained from the circuit model **(B)** with experimentally measured K_i = 13.7 mM, and the same K_m and k_{cat} from Figure 4. All data points are means of three independent replicates with standard deviations less than 20% of the corresponding mean (not shown).

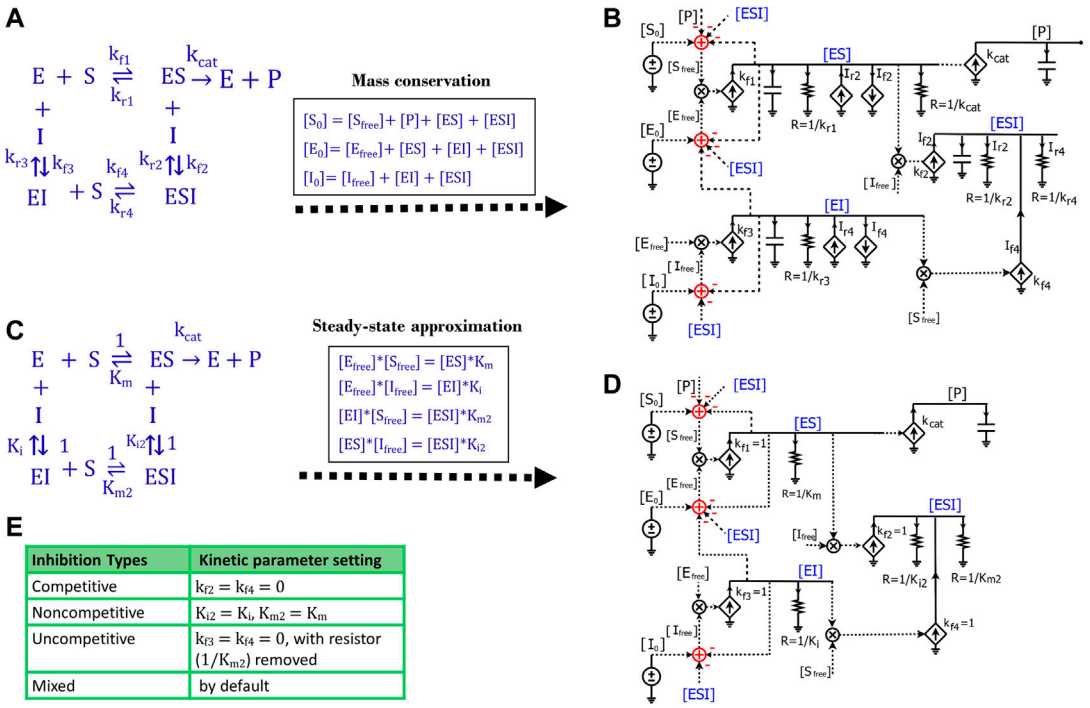
biological pathways when there are competitive inhibitors involved. As an example, we now use the competitive inhibition circuit (Figure 5B) to model the kinetics of product inhibition in an enzymatic reaction. Product inhibition is a common way to regulate reaction rates in metabolic pathways. Based on previous reports that beta-galactosidase can be competitively inhibited by a relatively high concentration of its own product galactose (Portaccio et al., 1998; Nguyen et al., 2006), we easily construct a reaction scheme for the product inhibition wherein the product galactose competes for the free enzyme (Figure 6A): In Figure 5B, we simply replace the inhibitor I_0 with product $[Gal_0]$ that can be externally added to the reaction and also wire newly produced [Gal] to the total product pool to architect the feedback inhibition; the resultant circuit is shown in Figure 6B.

It is worth noting that simple rewiring and reuse of circuit building blocks avoids the need for any math equations, and preserves physical and chemical intuition. We can directly and rapidly map the reaction mechanism of Figure 6A to a quantitatively accurate representation of its function and dynamics in Figure 6B. The implicit (caused by subtractive inputs from the "use-it-and-lose-it" mass conservation in Figures 4–6) and explicit (due to product inhibition) feedback loops are all evident and clearly represented.

We validated the product-inhibition circuit model by fitting it to experimental data. The circuit model shows good fits to the experimental data for product dynamics over time when varying substrate concentrations were added but with constant galactose concentration (Figure 6C). In another experiment, we compared the reactions with and without the product galactose added before starting the reaction. As expected, when the initial amount of galactose is added, the reaction is inhibited and takes a longer time to reach a plateau wherein all substrate has been consumed (Figure 6D).

2.5 A generalized molecular-binding circuit block for enzyme inhibition and two-substrate reactions

We next sought to develop a generalized circuit model for all types of enzyme inhibition including competitive, non-competitive, and uncompetitive inhibition. In the generalized reaction scheme (Figure 7A), the enzyme forms ES, EI, and ESI complexes with the substrate, inhibitor, or both, respectively; the specific reaction fluxes can be derived from the corresponding rate constants. We can directly translate the reaction scheme into an equivalent circuit (Figure 7B) that exactly describes all the dynamics of each species in the reaction and has identical math



Inhibition Types	Kinetic parameter setting
Competitive	$k_{f2} = k_{f4} = 0$
Noncompetitive	$K_{i2} = K_i$, $K_{m2} = K_m$
Uncompetitive	$k_{f3} = k_{f4} = 0$, with resistor $(1/K_{m2})$ removed
Mixed	by default

FIGURE 7
Generalized circuit models for enzyme inhibition. **(A)** A reaction scheme of general inhibition using rate constants without any approximation or assumption. **(B)** A circuit model translated from the reaction scheme in **(A)** describes reaction dynamics. All symbols used in this circuit are the same as the ones used in Figures 1, 3. The dependent current generators (diamond symbols) can provide input source currents (arrow up) or sink currents (arrow down) at nodes that they are wired to. Voltages or currents labeled with same names indicate that they have the same values: in accord with the reaction, the same current or voltage is appropriately re-used or regenerated at multiple locations with the use of implicit rather than explicit wiring to avoid clutter. **(C)** The reaction scheme of general inhibition in **(A)**, but with a steady-state approximation (all complexes reach equilibria instantaneously). Here, normalized k_f parameters are set to 1, such that all k_r parameters are mapped to their corresponding equilibria dissociation-constant (K_m or K_i) values. **(D)** The generalized circuit for enzyme inhibition translated from the reaction scheme in **(C)**. Note that, in accord with the steady-state approximation, the capacitors in **(B)** are removed in **(D)**; and kinetic parameters in the reaction scheme **(C)** are mapped to equivalent circuit parameters in **(D)**. The latter circuit can simulate all common enzyme-inhibition mechanisms including competitive, noncompetitive, uncompetitive, and mixed inhibition. **(E)** Kinetic parameter settings for different types of enzyme inhibition.

equations, as we derive below. Based on the mass conservation law, we have the following relationships for $[S_0]$, $[E_0]$ and $[I_0]$ (Figure 7A):

$$[S_0] = \left[S_{free}\right] + [P] + [ES] + [ESI] \tag{20}$$

$$[E_0] = \left[E_{free}\right] + [ES] + [EI] + [ESI] \tag{21}$$

$$[I_0] = \left[I_{free}\right] + [EI] + [ESI] \tag{22}$$

Such mass conservation is represented in the equivalent circuit (Figure 7B) by an adder block with a positive conserved total species input (S_0, E_0, or I_0); subtractive (negative) inputs are caused by the use of the species (for binding or product transformation) to generate other species (P, ES, EI, or ESI); finally, free variables (S_{free}, E_{free}, or I_{free}) are the resulting outputs of the adder blocks. The subtractive inputs always cause the 'use-it-and-lose-it' implicit negative-feedback loops in chemical reaction networks (Teo et al., 2015; Teo and Sarpeshkar, 2020). As shown in Figure 7A, the dynamics of $[ES]$ are determined by five reaction fluxes, including two generation fluxes and three consumption fluxes. Therefore, we have:

$$\frac{d[ES]}{dt} = k_{f1}{}^{*}\left[S_{free}\right]\left[E_{free}\right] + [ESI]{}^{*}k_{r2} - [ES]{}^{*}k_{r1} - [ES]\left[I_{free}\right]{}^{*}k_{f2} - [ES]{}^{*}k_{cat} \tag{23}$$

Likewise, the voltage dynamics of $[ES]$ in the circuit (Figure 7B) are also determined by five currents/fluxes. The two generation fluxes include the forward reaction flux, $k_{f1}{}^{*}[S_{free}][E_{free}]$ for enzyme-substrate binding indicated by the k_{f1} current source, and the reverse reaction flux (I_{r2}), $[ESI]{}^{*}k_{r2}$ from $[ESI]$, indicated by the I_{r2} current source; the three consumption fluxes include the dissociation reaction flux, $[ES]{}^{*}k_{r1}$, indicated by the $1/k_{r1}$ resistor, the reaction flux for ES and inhibitor binding (I_{f2}), $[ES][I_{free}]{}^{*}k_{f2}$, indicated by the I_{f2} sink current source, and the catalytic reaction flux, $[ES]{}^{*}k_{cat}$.

The dynamics of $[EI]$ are determined by two generation fluxes and two consumption fluxes in the reaction scheme (Figure 7A). Therefore, we have:

$$\frac{d[EI]}{dt} = k_{f3}{}^{*}\left[I_{free}\right]\left[E_{free}\right] + [ESI]^{*}k_{r4} - [EI]^{*}k_{r3} - [EI]^{*}\left[S_{free}\right]^{*}k_{f4} \quad (24)$$

Similarly, in the circuit (Figure 7B), the voltage dynamics of $[EI]$ are also determined by two source currents and two sink currents. Two supply currents include $k_{f3}{}^{*}[I_{free}][E_{free}]$, indicated by the k_{f3} current source, and the dissociation flux (I_{r4}), $[ESI]^{*}k_{r4}$, indicated by the I_{r4} current source. The two sink currents include the current through the resistor $(1/k_{r3})$, $[EI]^{*}k_{r3}$, and the current through the I_{f4} current source, which is $[EI]^{*}[S_{free}]^{*}k_{f4}$.

The dynamics of $[ESI]$ in the reaction scheme (Figure 7A) are quantified by two generation fluxes and two consumption fluxes, so we have:

$$\frac{d[ESI]}{dt} = k_{f2}{}^{*}[ES]\left[I_{free}\right] + k_{f4}{}^{*}[EI]\left[S_{free}\right] - [ESI]^{*}k_{r2} - [ESI]^{*}k_{r4} \quad (25)$$

This equation is also exactly reflected by the voltage dynamics of $[ESI]$ in the circuit (Figure 7B), which are determined by two source currents and two sink currents. As shown in the circuit, the two source currents include one from the k_{f2} current source, $k_{f2}{}^{*}[ES][I_{free}]$, and one from the k_{f4} current source, $k_{f4}{}^{*}[EI][S_{free}]$. The two sink currents are indicated by two resistors $1/k_{r2}$ and $1/k_{r4}$, with quantities of $[ESI]^{*}k_{r2}$, and $[ESI]^{*}k_{r4}$, respectively (Figure 7B).

Finally, the product [P] dynamics that are described by Eq. 12 are also described by the dynamics of voltage [P] in the circuit (Figure 7B). As illustrated above, Eqs 12 and 20–25 describing the reaction kinetics/dynamics for the reaction scheme (Figure 7A) are exactly mapped to a single circuit (Figure 7B). Given the rate constants and initial conditions, these equations (Eqs 12 and 20–25) can be solved by running simulations of the circuit. In short, this circuit not only accurately visualizes all differential equations in one diagram but can also easily provide solutions to these equations via simulations of the circuit.

Since we haven't applied any assumptions while deriving the equations or in developing the circuit (Figures 7A,B), this circuit is very general and can be used to model the kinetics of any reaction with the same topology. However, we need to know the values of all the rate constants to run the simulation of this circuit, which is not very convenient. To make the circuit more useful in practice, we simplified the reaction by applying the steady-state approximation (Figure 7C), which is also used in the derivation of enzyme inhibition kinetics from the Michaelis-Menten equation. Under this approximation, substrate binding and inhibitor binding to an enzyme are viewed as instantaneous (much faster than the catalytic reaction) and thus all intermediate complexes reach quasi-steady states. Therefore, we have the equilibrium equations below:

$$\left[E_{free}\right]^{*}\left[S_{free}\right] = [ES]^{*}K_m \quad (26)$$

$$\left[E_{free}\right]^{*}\left[I_{free}\right] = [EI]^{*}K_i \quad (27)$$

$$[EI]^{*}\left[S_{free}\right] = [ESI]^{*}K_{m2} \quad (28)$$

$$[ES]^{*}\left[I_{free}\right] = [ESI]^{*}K_{i2} \quad (29)$$

Similar to the tactics used to derive the Michaelis-Menten circuit (Figure 3D), these equilibrium equations are reflected in the circuit (Figure 7D) by removing all capacitors, setting all four k_f values equal to 1, and combining currents for each complex into one or two resistors. We thus achieve a parameter-reduced circuit generalized for enzymatic reactions with inhibition (Figure 7D). Note that ESI has two production fluxes, one from ES and the other from EI, and two consumption fluxes, dissociation from ESI to EI and to ES, respectively. We can sum Eqs 28, 29, and have Eq. 30 which determines the steady state of $[ESI]$:

$$[EI]^{*}\left[S_{free}\right] + [ES]^{*}\left[I_{free}\right] = [ESI]^{*}(K_{m2} + K_{i2}) \quad (30)$$

Accordingly, Eq. 30 is also reflected in the circuit (Figure 7D) for the voltage of $[ESI]$, which is determined by two source currents and two sink currents through the two resistors. The two source currents are provided through the k_{f2} current source and k_{f4} current source; the two sink currents are realized by the $1/K_{m2}$ resistor and $1/K_{i2}$ resistor, respectively.

Based on the mathematical derivations above, we have developed a parameter-reduced circuit (Figure 7D) generalized for all four kinds of enzyme inhibition including competitive, non-competitive, uncompetitive, and mixed inhibition. By default, the generalized circuit (Figure 7D) can be viewed as a circuit model for mixed inhibition while the other three types of inhibition are just special cases. For competitive inhibition, since there is only one $[EI]$ reaction (Figure 7C), setting the parameters $k_{f2} = k_{f4} = 0$ gives $[ESI] = 0$ and the circuit (Figure 7D) becomes the same circuit in Figure 5B for competitive inhibition. For non-competitive inhibition, the inhibitor binds to the enzyme at a different site from the catalytic site and this binding is independent of the substrate binding, resulting in $K_{m2} = K_m$ and $K_{i2} = K_i$ in the reaction scheme (Figure 7C). Consequently, using those same settings in the circuit (Figure 7D), we have the circuit for non-competitive inhibition. Since non-competitive inhibition is a special case of mixed inhibition, when there are no special requirements on K_{i2} and K_{m2}, the circuit (Figure 7D) by itself is a model for mixed inhibition. For uncompetitive inhibition, there is only one production flux for ESI (from ES), meaning that $k_{f3} = k_{f4} = 0$ in the circuit (Figure 7D). As a result, there is only one consumption flux for $[ESI]$ in the uncompetitive inhibition circuit, such that there is only current through the resistor $(R = 1/K_i)$, achieved by either removing the other resistor

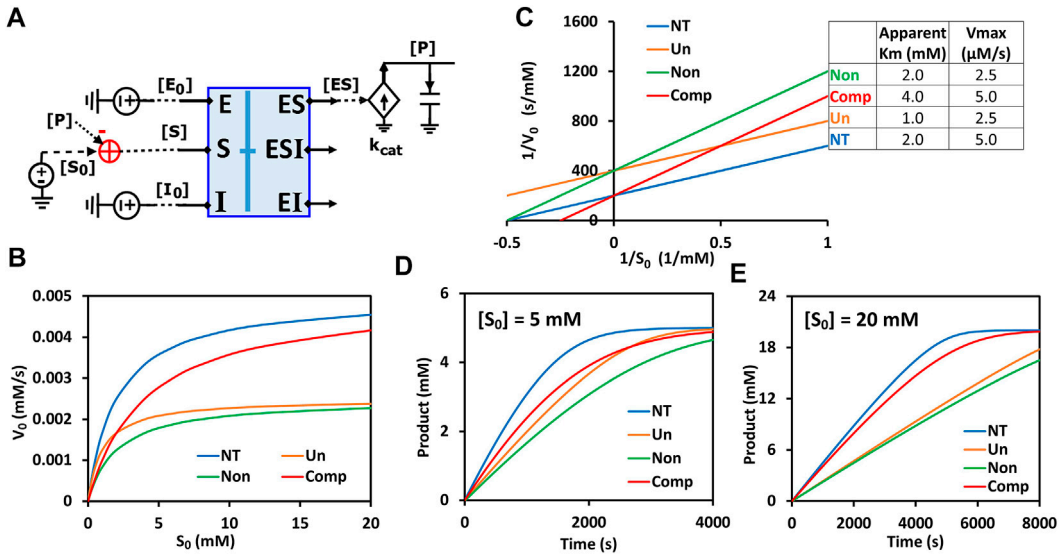


FIGURE 8
Generalized circuit simulations of enzymatic reactions with different inhibition types. The kinetic parameters used in the simulations are K_m = 2 mM, K_i = 3 mM, k_{cat} = 500/s, E_0 = 10 nM, I_0 = 3 mM, with varying amounts of substrate S. **(A)** The simplified circuit schematic for the generalized inhibition circuit from Figure 7D with a binding block (blue box) and a catalytic reaction. The binding block is the same circuit used for the binding of enzyme, inhibitor, and substrate in Figure 7D, with a separated catalytic reaction block. **(B)** Simulation curves of initial reaction rate V_0 versus initial substrate concentration S_0 for enzymatic reactions without inhibitor (NT), with an uncompetitive inhibitor (Un), with a noncompetitive inhibitor (Non), and a competitive inhibitor (Comp). **(C)** Lineweaver-Burk plot plotting $1/V_0$ versus $1/S_0$. The inserted table corresponds to parameters derived from the X- and Y-intercepts of the corresponding line. These derived parameters exactly match the input parameters. **(D,E)** Predicted product dynamics over time with S = 5 mM **(D)** and S = 20 mM **(E)** for all inhibition types listed.

R = $1/K_m$ or setting its resistance to a huge value such that the current running through it can be ignored. In summary, we have developed a generalized circuit for four types of enzyme inhibition. Its simulation requires five parameters (K_m, k_{cat}, K_i, K_{m2} and K_{i2}) for modeling mixed inhibition and only three kinetic parameters, K_m, k_{cat}, and K_i, for modeling competitive, non-competitive, or uncompetitive inhibition.

The generalized circuit (Figure 7D) consists of two parts: a binding block and a catalytic reaction block. The binding block is the circuit representing the binding of enzyme-substrate (ES), enzyme-inhibitor (EI), and enzyme-substrate-inhibitor (ESI) in Figure 7D. Since the binding circuit can fundamentally simulate different bindings among biomolecules, it is a useful building block for modeling more complicated interactions. Therefore, we extracted the binding block and rearranged the generalized circuit (Figure 7D) into a more concise schematic symbol (Figure 8A) for easy visualization. The new circuit consists of a binding block (blue box) and a catalytic reaction block (Figure 8A), describing exactly the same kinetics/dynamics as in the original circuit (Figure 7D). This binding block has three inputs ($[E_0]$, $[S_0]$, and $[I_0]$) and three outputs ([ES], [EI], and [ESI]), which are the same as those in Figure 7D. The reaction block is separated; hence there is product [P] feedback to the substrate input to maintain mass conservation of S as shown in Figure 8A. With all the underlying schematic circuits built within it (Figure 7D), our binding-block circuit motif symbol is convenient to use: As in highly complex electronic integrated-circuit design with hierarchical building blocks, it provides a useful visualization aid; it can be repeatedly instantiated to create ever-more complex biochemical reaction networks.

To verify that our generalized circuit can accurately simulate all enzyme inhibition types, we ran simulations with the circuit using the same kinetic parameters (K_m, k_{cat}, and K_i) and initial concentrations of enzyme, substrate, and inhibitor for all of its special cases, i.e., for competitive, noncompetitive, and uncompetitive inhibition. As expected, our circuit model predicted the characteristic curves for the relationship between $[S_0]$ and initial reaction rate V_0 for all three inhibition types (Figure 8B). Lineweaver-Burk plotting further confirmed that our circuit model is accurate (Figure 8C): For competitive inhibition, the apparent K_m increases but V_{max} remains the same upon the addition of the inhibitor; for non-competitive inhibition, the V_{max} decreases but apparent K_m remains the same; for uncompetitive inhibition the apparent K_m and V_{max} both decrease with the same proportion such that the curve's slope (K_m/V_{max}) without inhibitor (NT, the blue line in Figure 8C) is identical to that with inhibitor (Un, the orange line in Figure 8C). Furthermore, this circuit can also simulate mixed inhibition where K_m, K_{m2}, K_i, and K_{i2} are independent of each other, the most general case. As an example, when using the

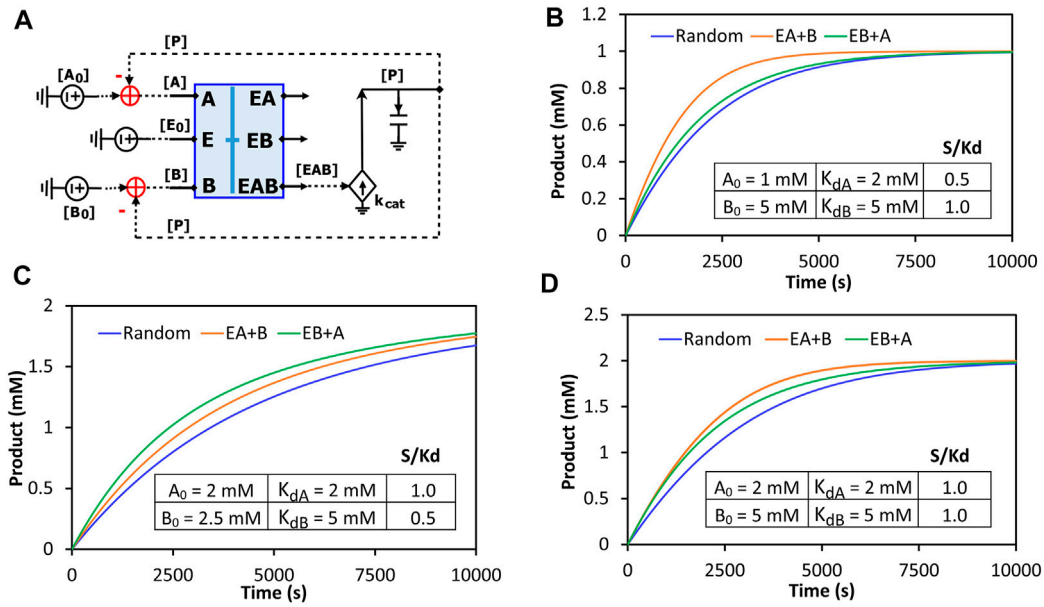


FIGURE 9
Comparison of the reaction kinetics of two-substrate reactions with random binding and ordered-binding. **(A)** The concise circuit schematic for the generalized reaction circuit from Supplementary Figure S2D, with a binding block (blue box) and a catalytic reaction. The binding block circuit motif (blue box) is the same circuit used in Figure 7D for ESI binding but with substrates A and B replacing S and I. Depending on parameter settings, this circuit can also simulate ordered-binding reactions as mentioned in Supplementary Figure S2E. **(B–D)** Product dynamics for the random- and ordered-binding reactions with different substrate-to-Kd (S/K_d) ratios. Random: random-binding reaction; EA + B: enzyme binds to A then B; EB + A: enzyme binds to B then A. The parameters used in the simulations are listed in the inserted table in each graph. The two common parameters used are: E_0 = 10 nM, k_{cat} = 500/s.

same K_m and K_i values as in Figure 8 but setting K_{i2} = 6 mM and K_{m2} = 4 mM, we obtain a typical curve for mixed inhibition with different X- and Y-intercepts on the Lineweaver-Burk plot (Supplementary Figure S1).

Remarkably, this circuit shows that the time-course dynamics of the product differ greatly amongst the three inhibition types (Figures 8D,E) and that the difference depends on substrate concentration: When S = 5 mM, the product dynamics of uncompetitive inhibition are similar to that of competitive inhibition (Figure 8D), while when S = 20 mM the product dynamics of uncompetitive inhibition are closer to that of noncompetitive inhibition (Figure 8E). Non-competitive inhibition usually has the strongest inhibition among the three types, with the slowest product production (Figures 8B,D,E). Our simulation results suggest that, to obtain accurate molecular kinetics, it is important to choose the right inhibition mechanism when constructing models. Choosing the right type of inhibition is especially critical for cascade reactions such as in metabolic pathways where there are many steps with feedback inhibition caused by the final product and/or intermediate metabolites.

The binding block circuit motif (Figures 7D, 8A) is very useful and versatile for different applications. As another example, we used the same binding block and made a circuit to simulate reactions with two substrates binding to the same enzyme. There are three possibilities of binding order: Enzyme binds to A and then to B; Enzyme binds to B and then to A; Enzyme binds to A and B randomly. We called the first two "ordered-binding reactions" and the last one "random-binding reaction". The reaction schemes and a circuit model for a generalized reaction are demonstrated in Supplementary Figure S2. Our circuit simulations clearly show that the binding order can cause different reaction dynamics. Normally, the random-binding reaction has the slowest rate while the rate for ordered-binding reactions depends on the substrate-to-K_d ratio (S/K_d) (Figure 9). Given the same concentration of substrate A and B, the reaction is faster when the enzyme first binds to B (greater K_d and thus smaller S/K_d) while given the same K_d for both substrates, the reaction is faster when the enzyme first binds to A (lower concentration and thus smaller S/K_d) (Supplementary Figure S3). Indeed, when we vary the S/K_d ratio for both substrates, the reaction is always faster when the enzyme first binds to the substrate with the smaller S/K_d (Figure 9). It is interesting to point out that given the same S/K_d ratio (1:1 in Figure 9D), the rates are initially similar for both ordered-binding reactions but as two substrates are consumed to cause one to have smaller S/K_d (substrate A in Figure 9D), the reaction (with A binding first) is eventually faster

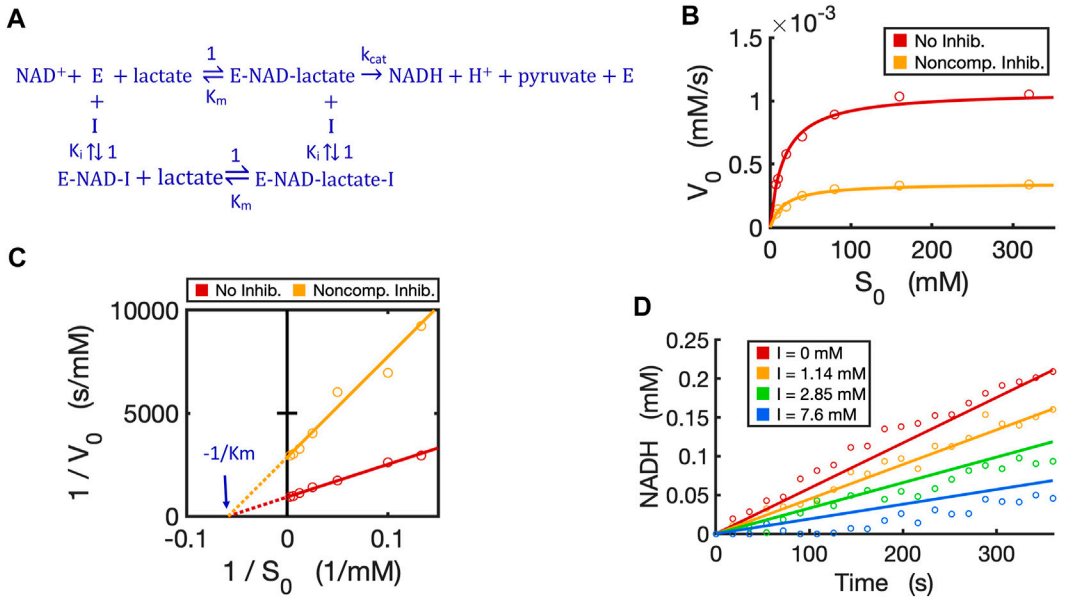


FIGURE 10
Circuit modeling of non-competitive inhibition of LDH by oxamate. **(A)** The reaction scheme of non-competitive inhibition against LDH, where the inhibitor I is oxamate. For simplification, the initial catalysis can be viewed as a directional reaction, though the LDH reaction is known to be reversible; the fast binding step of NAD^+ to enzyme is neglected in this reaction scheme. **(B)** Simulation curves of V_0 versus S_0 fitted to measured experimental data in the absence and presence of oxamate (7.6 mM), when E_0 = 5.0 nM. **(C)** Lineweaver-Burk plot showing linearized curves of $1/V_0$ versus $1/S_0$ with and without the inhibitor, which have the same *X*-intercept of -1 /K_m. **(D)** The model curves of product dynamics over time with varying inhibitor concentrations fitted to experimental data points (when S = 80 mM, E_0 = 3.3 nM). All simulation curves are obtained from the circuit model (Figures 7D, 8A) with experimentally measured K_m = 17.1 mM, k_{cat} = 215.5 (1/s), and K_i = 3.66 mM. All data points are means of three independent replicates with standard deviations less than 20% of the corresponding mean (not shown).

than the reaction with the opposite binding order. Given that the binding order matters for reaction kinetics, modelers may need to carefully select which binding order best approximates the reaction they are modeling.

2.6 Circuit modeling of non-competitive inhibition

We have now verified in theory that the generalized circuit (Figure 8A) is accurate in predicting kinetics for all enzyme inhibition types. This circuit is also convenient to use; with inputs of K_m, k_{cat} and K_i as well as experimental initial conditions, running simple simulations can provide all kinetic data, which can be compared to experimentally measured data. As an example, we now use our circuit to model the noncompetitive inhibition wherein lactate oxidation by lactate dehydrogenase is inhibited by a noncompetitive inhibitor, oxamate (Powers et al., 2007). In this redox reaction (Figure 10A), we varied the concentration of lactate while keeping a constant NAD^+ concentration. As expected, the circuit-predicted curves capture the experimental data nicely. The initial reaction rate increases as substrate concentration increases, but, in the presence of the noncompetitive inhibitor, the reaction rate never approaches the V_{max} of the reaction when without the inhibitor (Figure 10B). Linearized curves from Lineweaver-Burk plots show that noncompetitive inhibition has the same K_m as the reaction without the inhibitor, but a smaller V_{max} (i.e., greater $1/V_{max}$) (Figure 10C). The generalized circuit can also simulate the time-course kinetics of NADH production for the initial period (the first 360s) (Figure 10D) where the catalytic reaction can be largely viewed in the forward direction.

2.7 Circuit modeling of reversible reactions

We developed a circuit to simulate reversible reactions because they are very common in metabolic pathways. Since the kinetic mechanism for the reversible reaction of alcohol dehydrogenase (ADH) is well studied with all rate constants known (Dickinson and Dickenson, 1978; Ganzhorn et al., 1987; Plapp, 2010), we developed a circuit model for this reversible reaction with ethanol and NAD^+ both as substrates. The kinetics of this reaction by yeast ADH follow the ordered Bi-Bi mechanism, resulting in a five-step reversible reaction (Ganzhorn et al., 1987) as shown in Figure 11A. First, NAD^+ binds to the enzyme ADH and forms an E-NAD complex; the substrate ethanol then binds to E-NAD and forms an

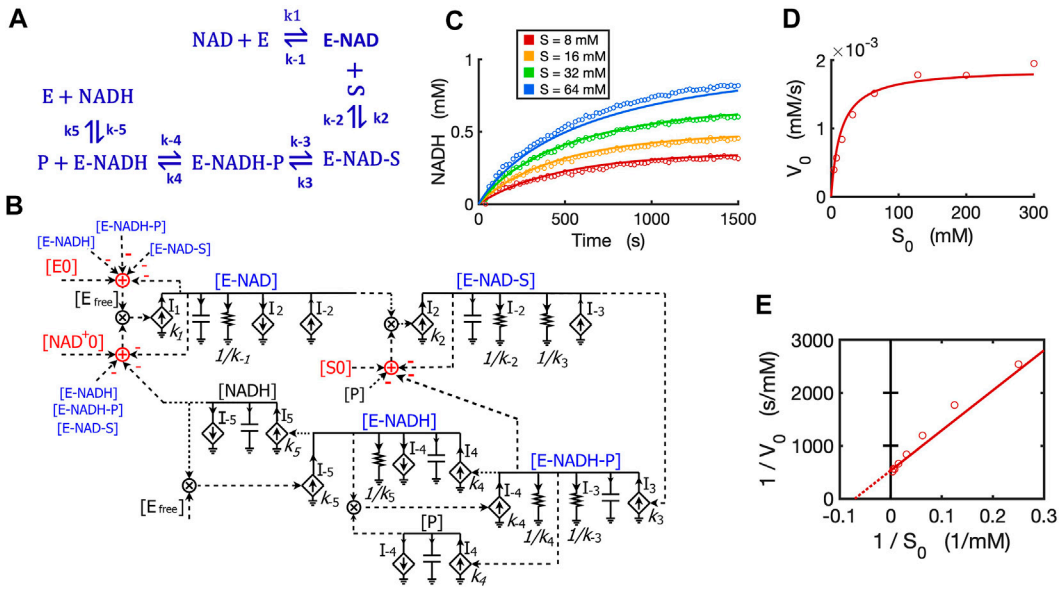


FIGURE 11
Circuit modeling of a reversible reaction catalyzed by yeast alcohol dehydrogenase (ADH). **(A)** The mechanistic scheme for a five-step reversible reaction for yeast ADH, where E is ADH enzyme, S is ethanol, and P is the acetaldehyde produced. **(B)** Circuit model that exactly matches the kinetics of the reaction scheme in **(A)**. $[E_0]$ is the initial ADH concentration (3.9 nM), $[NAD^+_0]$ is the initial NAD^+ concentration (4 mM), $[S_0]$ is the initial ethanol concentration. Arrows along solid lines or in current generators (diamond symbols) indicate the direction of the corresponding current (reaction flux). Voltages or currents labeled with the same names indicate the same values. All circuit symbols are shown in Figures 1, 3. **(C)** Model curves of product [NADH] dynamics over time fitted to experimental data points with varying ethanol concentrations. **(D)** Simulation curves of initial reaction rates V_0 versus initial alcohol concentrations S_0 fitted to experimental data. **(E)** Lineweaver-Burk plot showing a straight curve of $1/V_0$ versus $1/S_0$. All data points are means of three independent replicates with standard deviations less than 20% of the corresponding mean (not shown).

E-NAD-S complex; this complex turns into a new complex E-NADH-P and then the first product aldehyde (P) releases, resulting in an E-NADH complex; the last step is the dissociation of NADH from the enzyme. The total reaction consists of five reversible steps, each with forward and reverse rate constants as indicated in the reaction scheme (Figure 11A).

To mathematically simulate the dynamics/kinetics of all species in this multi-step reaction would require a long list of differential equations as well as many mass conservation equations. Instead, we can simply translate the reaction scheme into a single circuit (Figure 11B) without deriving differential equations. The circuit model is constructed by using a circuit block similar to the one for the enzyme-substrate circuit (Figure 3B). The difference is that each intermediate complex (blue text) including E-NAD, E-NAD-S, E-NADH-P, and E-NADH in steps 1–4 has two production fluxes and two consumption fluxes (Figure 11B), instead of only three fluxes for the [ES] complex in the circuit (Figure 3B) for classic Michaelis-Menten kinetics without a reversible catalytic reaction. It should be noted that the product aldehyde [P] is produced in proportion to [E-NADH-P] with a rate constant of k_4 while it is consumed by binding back to [E-NADH], forming [E-NADH-P] with a rate constant k_{-4}, resulting in a reaction rate of $I_{-4} = k_{-4}*[P]*[E\text{-}NADH]$. Similarly, another product [NADH] is produced in proportion to [E-NADH] with a reaction rate of $k_5*[E\text{-}NADH]$ while it is also consumed by binding back to the enzyme, forming [E-NADH] with a rate of $I_{-5} = k_{-5}*[NADH]*[E_{free}]$. Another important note is that the conservations of enzyme, NAD^+, and ethanol (S) during the reaction are indicated by adder and subtraction symbols in red in the circuit (Figure 11B).

To verify the accuracy of our circuit model, we fit experimental data for [NADH] produced from ethanol oxidation by ADH for varying amounts of alcohol and a fixed concentration of initial NAD^+. Given that the 10 rate constants for the whole reaction are known from previous studies (Dickinson and Dickenson, 1978; Ganzhorn et al., 1987), we used these kinetic parameters as input to our circuit model. We were able to get a good fit with only some adjustments (Supplementary Table S1) relevant to our specific experimental conditions. Our circuit model can accurately simulate the product dynamics over time with varying amounts of ethanol input (Figure 11C). Importantly, our circuit model also correctly predicts the classic relationship between initial reaction rates (V_0) and initial ethanol

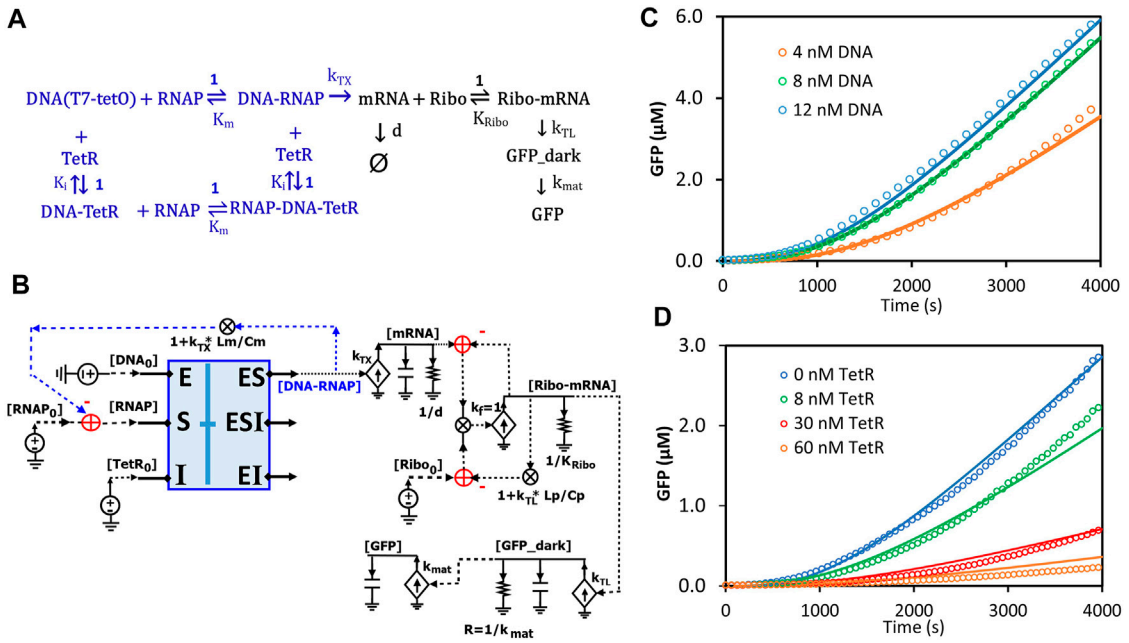


FIGURE 12
Circuit modeling of cell-free transcription and translation regulated by TetR. **(A)** The scheme for molecular binding and associated reactions of TXTL in the cell-free system with repression by TetR. The TXTL is turned on by a hybrid T7 promoter (DNA) and T7 RNA polymerase (RNAP) and repressed by TetR via non-competitive inhibition. The free RNAP-DNA complexes turn on transcription followed by the translation of GFP. The TetR used in the model is its homodimer. K_i represents the dissociation constant (K_d) of the homodimer binding to the DNA (tetO). **(B)** The circuit model exactly describes the kinetics of molecular binding and associated reactions in **(A)**. In the non-competitive binding block (blue box), DNA acts in a manner like an enzyme where both the substrate (RNAP) and the inhibitor (TetR) bind to it. Only free [DNA-RNAP] turns on transcription. **(C)** Model curves of GFP dynamics fitted to experimental data points with varying DNA concentrations when TetR = 0 nM. **(D)** Model curves of GFP dynamics fitted to experimental data points with varying TetR concentrations at constant DNA concentration (6 nM). All data points are means of three independent replicates with standard deviations less than 20% of the corresponding mean (not shown).

concentrations (S_0) based on modeling the initial 60 s of the reaction (Figure 11D). The Lineweaver-Burk plot shows a straight line for $1/V_0$ versus $1/S_0$, which captures our data closely (Figure 11E). From the X- and Y-intercepts, we calculated a K_m of 14.3 mM for ethanol and k_{cat} of 483/s, which agrees well with previous reports (Ganzhorn et al., 1987).

2.8 Circuit modeling of transcription and translation in a cell-free system

Circuits can be used to model any biological system, not just enzymatic reactions (Sarpeshkar, 2010; Teo and Sarpeshkar, 2020). Here, we sought to model transcription and translation (TXTL) kinetics regulated by TetR in an *E. coli*-based cell-free system. In this system, the DNA insert region (a hybrid T7 promoter regime) on a plasmid (DA313) has a T7 RNAP binding site and a TetR binding site (tetO) that is 5-bp downstream; given the relatively small sizes of T7RNAP and TetR, the DNA region acts like an enzyme capable of binding both the "substrate" (T7 RNAP) and the inhibitor protein (TetR) at two different sites. Since T7 RNAP and the inhibitor TetR bind to the DNA regime ("enzyme") at two different sites, we can assume their interactions to architect noncompetitive binding (Figure 12A). Such transcriptional binding is similar to the noncompetitive binding/reaction scheme (Figures 7C,D, 8A) but with the difference that none of the "substrate" (T7 RNAP) is consumed or converted into products. The free DNA-RNAP complex turns on the transcription followed by the translation of GFP (Figure 12A). The dynamics of mRNA are determined by production and degradation, with rate constants of k_{TX} (1/s) and d (1/s), respectively. Ribosomes then bind to the ribosome binding site (RBS) of the mRNA and turn on the translation of GFP_dark with a rate constant of k_{TL} (1/s). GFP_dark undergoes folding and maturation with a rate constant of k_{mat} (1/s), resulting in fluorescent GFP (Figure 12A). In this study, we focused only on the initial TXTL reactions in the first ~4000 s, wherein the reactions reach their maximal production rates (steady states) and are not limited by the available amino acids, energy sources, and/or other components. Since TetR forms stable homodimers (Krafft et al., 1998), TetR in this model represents its homodimer. With these simplifications and assumptions,

we thus have the reaction scheme for TXTL regulated by TetR in the cell-free system (Figure 12A).

Based on the binding and reaction scheme (Figure 12A), we used the general binding block (Figure 8A) and designed a circuit (Figure 12B) that exactly describes the kinetics of the binding interactions and reactions for the regulated TXTL in the cell-free system. In this circuit (Figure 12B), the binding block (blue box) has voltage inputs of initial [DNA_0] for total plasmid concentration, [$RNAP_0$] for total T7 RNAP concentration, and [$TetR_0$] for total TetR concentration, via the pins of E, S, and I, respectively; free [TetR] and [RNAP] bind to DNA noncompetitively, exactly described by the same circuit as Figures 7D, 8A; among the three outputs, only the [DNA-RNAP] turns on the transcription; the mRNA is then bound by ribosomes, which results in GFP production. Since several RNAP molecules can bind to one DNA molecule during transcriptional elongation, the average number of T7-RNAP bound in each DNA-RNAP complex would be 1 + k_{TX}*Lm/Cm where 1 is the one RNAP binding to the promoter, k_{TX} is the transcription rate (1/s), Lm is the length of mRNA (nt), and Cm is the average RNAP velocity during elongation (nt/s) (Bremer et al., 2003; Marshall and Noireaux, 2019). Therefore, for each DNA-RNAP complex the number of RNAP, 1 + k_{TX}*Lm/Cm, has to be subtracted from the total [$RNAP_0$] pool. Similar to the counting of RNAP copy numbers on each DNA-RNAP complex, for each Ribo-mRNA complex, there are 1 + k_{TL}*Lp/Cp ribosomes, where 1 is the one ribosome binding to the ribosome binding site, k_{TL} is the translation rate (1/s), Lp is the GFP coding length of mRNA (nt), and Cp is the ribosome velocity during elongation (nt/s). The last part of the circuit is the GFP production where GFP_dark is produced with a rate constant of k_{TL} and consumed to make fluorescent GFP with a rate constant of k_{mat}.

All the kinetic parameters (Supplementary Table S2) used in this model were obtained either from previous studies (Újvári and Martin, 1996; Kamionka et al., 2004; Skinner et al., 2004; Marshall and Noireaux, 2019) under similar conditions or from fitting to our experimental data. The simulation curves fitted our experimental data closely as shown in Figures 12C,D. As expected, as more plasmid was added to the cell-free system more GFP was produced until saturation was reached (up to ~12 nM DNA in our experiments) (Figure 12C), where adding more plasmid would not increase GFP production. In another experiment, when varying amounts of TetR were added to the cell-free system with a constant DNA concentration, GFP production was increasingly repressed by more TetR addition.

3 Discussion

We have shown that the ordinary differential equations used to describe the dynamics of voltages and currents in relatively simple electronic circuits are exactly the same as the ODEs describing molecular kinetics in chemical reactions and biological systems. Voltages faithfully represent molecular concentrations while currents faithfully represent reaction fluxes. Furthermore, we have built Michaelis-Menten circuits and validated them by using well-defined parameters and by fitting experimental data from enzymatic reactions. Based on the MM circuits, we then developed circuit models for more complicated reactions including various enzyme inhibition types, product feedback inhibition, reversible reactions, two-substrate reactions, and regulated TXTL in a cell-free system. These circuits provide foundational building blocks for kinetic modeling of complex chemical and biological networks in the fields of synthetic biology (such as different types of oscillators), metabolism, and cellular signaling.

Taking advantage of circuit models, we are able to translate reaction schematics directly into circuits and to directly perform rapid kinetic modeling for biochemical reactions/networks without the need for deriving math equations or writing code. Using electronic circuits as a new modeling language can thus enable faster and easier modeling of molecular kinetics than using ODEs and numerical solvers. Importantly, circuit modeling is also very effective for scientific communication. Circuit modeling is not only accurate but also very concise because all the underlying equations and parameters/terms describing molecular kinetics, and the detailed interconnectivity between species/components are visualized and clearly labeled in one circuit.

It is worth noting that our circuit models are made of active transconductor-resistor-capacitor circuits that are more general than the passive-only resistor-based circuits used previously in biogeochemical modeling (Domingo-Félez and Smets, 2020; Tang et al., 2021). Our circuits can simulate steady-state and nonlinear dynamic evolution including loading and non-modularity between arbitrary interacting reactions, stochastics at low molecular copy number, and directly map physical parameters in the molecular domain (e.g., concentrations and rates) to equivalent physical circuit parameters (e.g., voltages and currents) (Sarpeshkar, 2010; Teo and Sarpeshkar, 2020). Thus, our circuits are designed mechanistically even at the molecular level and are flexible for use in a variety of applications. For example, circuit modeling allows us to simulate fundamental biochemical mechanisms such as different types of molecular binding and inhibition between enzyme and substrate/inhibitor/activator (Figures 8, 9). Such models otherwise are challenging to obtain without using many assumptions and mathematical derivations.

It should also be noted that our circuit models are more general and accurate than classic MM kinetics because we don't need the free ligand assumption where [S_0] is assumed to be much greater than [E_0]. Instead, subtraction of [ES] and [P] from [S_0] in all of our circuits always ensures exact mass conservation. Similarly, if present, automatic subtraction of [EI] and [ESI] ensures exact mass conservation. These subtractions may not

necessarily improve the model accuracy for normal enzymatic reactions that are run under conditions of $[S_0] >> [E_0]$ and/or $[I_0]>>[E_0]$. However, under some circumstances where the concentrations of substrate and/or inhibitor are close to the enzyme concentration, such as the case of our TXTL modeling in the cell-free system (Figure 12), the concentrations of complexes including [ES], [EI] and/or [ESI] account for considerable portions of the total concentrations of $[S_0]$, $[I_0]$ and/or $[E_0]$. Thus, with built-in accounting for the concentrations of these complexes, we can apply circuits more accurately and conveniently than MM equations that may not be accurate under all conditions.

Our circuit modeling approach is unique because it allows for a biological system to be both intuitively visualized and quantitatively simulated within a single interactive and graphical software program (e.g., Cadence, CircuitLab). Professional circuit-design software directly enables model-order reduction, hierarchy, modular scalable design, and 25 + forms of sophisticated built-in mathematical analysis, including transient, steady state, frequency response, parametric variation, noise, Monte-Carlo, and other analyses automatically (Teo et al., 2019a; Teo and Sarpeshkar, 2020). Sophisticated adaptive Runge-Kutta and other numerical convergence tools are built into the circuit software preventing the user from having to focus on them. Although many useful tools exist for creating models in systems biology, such as SBML (Hucka et al., 2003) and BioCRNpyler (Poole et al., 2022), these tools require coding skills and are not as convenient to use as circuit design tools that fully take advantage of pictorial and architectural design. Therefore, we believe that our circuit approach will be significantly easier to use for experimental biologists. Synthetic biologists, who are typically more mathematically inclined, will also appreciate the quantitative design aspects and great flexibility that directly port from electronic circuit design to biological circuit design.

Our approach leverages the scalability and modularity of electronic circuit design with built-in hierarchy and circuit motifs. Circuit modeling enables modularity without sacrificing accuracy (e.g., due to the built-in use-it-and-lose-it feedback loops that automatically incorporate the 'loading' of one reaction module on another reaction module in our approach), which is also essential in scalable design. Circuit modeling also makes model-order reduction or expansion very easy such that the impact of a simple or complex model of any portion of a system on outputs can be evaluated with a few keystrokes. Circuit models are therefore scalable to the design of vast complex circuit networks. Furthermore, our circuit modeling approach directly maps physical parameters to physical parameters such that it is easily generalizable. Consequently, it is also more accurate and flexible even in extreme cases such as when concentrations of enzyme and substrate are comparable, enabling detailed models of molecular binding and enzymatic inhibition (Figures 8, 9). With such scalability and flexibility, we anticipate our circuit modeling approach will become a powerful tool to analyze the behaviors of large biological networks and to mine useful natural algorithms in such systems.

Computational time is an important issue for modeling complicated systems. Since our circuit modeling approach uses similar numerical algorithms (built into the circuit design software) as traditional ODE modeling, we do not expect advancement in computational time between the two modeling methods. Circuit software has some overhead in it to enable graphical design and multiple forms of analysis so it typically leads to a slight increase in simulation time, especially for small networks. For example, depending on the complexity and scale of circuits, the simulation takes a few seconds for small circuits (most circuits in this work) and a few minutes or longer for very large circuits. Running an ODE solver in MATLAB gave identical results to those obtained by circuit simulations in Cadence (Supplementary Figure S4). The MATLAB ODE solver was found to be faster than Cadence in that case (several tenths of a second vs. several seconds) due to the overhead being more significant for a small network. For large-scale models, hardware integrated circuits, namely, cytomorphic chips can run simulations of the "virtual" circuit models at speeds orders of magnitude faster than normal computers (Sarpeshkar, 2010; Woo et al., 2015; Woo et al., 2018). Cytomorphic chips integrate many fundamental biochemical reaction motifs, whose parameters and connectivity can be programmed to run simulations of arbitrary chemical networks. On such integrated circuit chips, the voltages (concentrations) and currents (rates) are emulated in a parallel fashion, even for highly stochastic simulations, rather than simulated. They are almost instantaneously accessible in a MATLAB interface that both programs and reads data from the chips. The computation time of the cytomorphic chips is independent of the reaction network's size and complexity even for stochastic simulations, which cannot be parallelized to represent Poisson noise in digital simulations (Kim et al., 2018). For example, models of computation time can be significantly reduced by several orders of magnitude for circuit models including oscillatory repressilators and more complex networks in cancer (Woo et al., 2015; Teo and Sarpeshkar, 2020) or in SARS-COV2 infection (Beahm et al., 2021).

4 Conclusion

Circuit modeling is advantageous for rapid design and accurate kinetic modeling of chemical and biological systems. Circuit models visualize all math equations and their relationships in one circuit, which is very concise and convenient for effective communication. This approach takes advantage of the high scalability in circuit design and has wide applicability. We envision that circuit modeling could be adopted

as a new language for kinetic modeling and promote scientific communication in the fields of biochemistry, systems biology, biogeochemistry, and synthetic biology.

5 Methods and materials

All enzymes and chemicals were purchased from Sigma Aldrich unless otherwise mentioned. Three enzymes used in this study were beta-galactosidase from *E. coli* (cat# G5635), rabbit muscle lactate dehydrogenase (cat# L2500), and yeast alcohol dehydrogenase (cat# A7011).

5.1 Enzymatic assays

All enzymatic assays were run in 96-well plates (Costar, cat#3595) with 100 µl of reaction mixture per well and reactions were monitored in real time by using a microplate reader (Molecular Devices Inc.).

Beta-galactosidase (beta-Gal) assays were performed using o-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate. All reagents were prepared in PBS buffer (pH7.2) with 5 mM Dithiothreitol (DTT) and 4 mM $MgCl_2$. The reactions (100 µl/well) with varying concentrations of ONPG were monitored by measuring the absorbance of the product, 2-nitrophenol, at 420 nm in real time for over 20 min. The concentrations of the product were determined by a calibration curve that was made by using a freshly prepared 2-nitrophenol solution under the same condition. Initial reaction rates were calculated from the slopes of each reaction curve normally within the first 6–8 min. For competitive inhibition, phenylethyl beta-D-thiogalactopyranoside (PETG), a known competitive inhibitor of beta-Gal (Xu and Ewing, 2004), was added to the reactions under the same condition. For the product feedback inhibition, galactose at varying concentrations was added to the reaction mixture before running the enzymatic assay.

The lactate dehydrogenase (LDH) assay was performed in 96-well plates using a constant concentration of 2 mM NAD^+ and varying the concentration of lactate as substrate. All reagents were prepared in 0.5 M glycine buffer (pH9.5). The end product NADH concentration was determined by measuring absorbance at 350 nm with a calibration curve pre-established under the same condition. All the reactions (100 µl/well) were monitored in real time in 96-well plates by a microplate reader. For modeling non-competitive inhibition, oxamate was used as a non-competitive inhibitor of LDH, in the lactate oxidation direction (Powers et al., 2007). The alcohol dehydrogenase (ADH) assay was performed in 40 mM Tris buffer (pH8.3) by using a constant concentration of 4 mM NAD^+ and varying the concentration of ethanol as substrate. The reactions were monitored similarly to the LDH assay.

All experiments were done with three replicates unless otherwise mentioned. Control reactions without any enzyme were also included in each experiment as blanks. The kinetic parameters including K_m, K_i, V_{max}, and k_{cat}, were derived from Lineweaver-Burk plotting. These experimentally obtained kinetic parameters and initial concentrations were put into the corresponding circuit models before running simulations.

5.2 Transcription and translation in the cell-free system

The cell-free *E. coli* protein synthesis system was purchased from New England BioLabs (cat# E5360). The T7 RNA Polymerase (NEB, cat# M0251S) and RNase Inhibitor (NEB, cat#M0314S) were added to the reaction mixture. The transcription and translation reactions (15 µl/each) with varying concentrations of plasmid DA313 and purified TetR were run in a 384-well plate (ThermoFisher Scientific, cat# 12-566-2). The reactions were monitored in real time by measuring GFP fluorescence at Ex485/Em528 in a microplate reader (Molecular Devices Inc.). The plasmid DA313 was made by fusing fragments of a hybrid T7 promoter with a tetO binding site (Jung et al., 2020), 5′UTR sequence, and a superfolder GFP (sfGFP) gene into the backbone of pJBL7010 (Silverman et al., 2019) by Gibson assembly (NEB, cat# E2621L). The 5′UTR sequence includes an mRNA stability hairpin (Carrier and Keasling, 1999; Silverman et al., 2019) and a ribosome binding site that was designed by the RBS calculator (Salis, 2011). The plasmid DA313 was purified by Monarch plasmid miniprep kit (New England BioLabs, cat# T1010S) and its concentration was determined by a fluorometric assay with a DNA-specific dye EvaGreen (Ihrig et al., 2006). The sfGFP concentration was estimated by a pre-established calibration curve using a pure EGFP (BioVision, cat#4999-100) with a conversion factor of 1.6 to account for different fluorescent brightness between sfGFP and EGFP (Fluorescent Protein Database, https://www.fpbase.org/). All experiments were performed with three independent replicates.

The recombinant TetR with 6xHis tag was over-expressed in *E. coli* JM109 (DE3) with plasmid DA303 and purified using Nickel-NTA agarose resin (ThermoFisher Scientific, cat# 88221). The purity of the recombinant TetR was analyzed by SDS-PAGE gel and the concentration was measured by BCA assay (ThermoFisher Scientific, cat#23252). Since TetR forms stable homodimers (Krafft et al., 1998) its concentration is then converted to the molar concentration of the homodimer. The plasmid DA303 was made by fusing TetR into the pQE80 backbone by Gibson assembly. All plasmids used were confirmed by Sanger sequencing and their maps are provided in the supplemental material (Supplementary Figure S5).

5.3 Circuit design and simulation

All circuits were designed and simulated in sophisticated circuit software including Cadence Virtuoso (version IC6.1.6–64b.500.6, Cadence Design Systems, Inc.) and an online tool, CircuitLab (https://www.circuitlab.com/). To design circuits, electric components such as capacitors, resistors, grounds, voltage sources, current sources/generators, and math functions (adders, subtractors, multipliers) were chosen directly from the library in the software and placed into the design window. All components were then wired together to make a circuit according to the system being modeled. Parameters and initial conditions were assigned to each component before running the simulations in the software. As examples, three circuits designed in this work are provided on the CircuitLab website via the links below:

Enzyme inhibitions: (https://www.circuitlab.com/circuit/5jsn32mbt42d/enzyme-inhibition/); Two-substrate enzymatic reactions: (https://www.circuitlab.com/circuit/fk2rya5d6s2p/eab-reaction/); Cell-free system: (https://www.circuitlab.com/circuit/5n6sug25t29j/cell-free-simulation/).

Parameters may need changes to fit the specific condition of the simulation. Other circuits were designed and simulated in Cadence. The design files (schematics) and simulation settings (cellview simulation states) for the Cadence circuits are included in the Supplementary Material. For comparison of ODE coding and circuit modeling, the MATLAB (R2020a, Mathworks) ODE solution code used to simulate the reversible reaction is also included in the Supplementary Material.

Data availability statement

All supporting data generated or analyzed in this study are included in this published article and the Supplementary Material or are available upon request to the corresponding author.

Author contributions

YD, DB, and RS conceived and designed the circuits and projects; YD and XR conducted the experiments and collected all the data; YD and DB analyzed the data, and performed circuit modeling and simulation; YD and TR made the circuit schematics; YD wrote the manuscript and all authors revised and approved the final manuscript.

Funding

This work was supported by AFOSR under grant number FA9550-18-1-0467 and by the NIH under grant number R01 GM 123032-01 to R. Sarpeshkar.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.947508/full#supplementary-material

References

Alves, R., Antunes, F., and Salvador, A. (2006). Tools for kinetic modeling of biochemical networks. *Nat. Biotechnol.* 24, 667–672. doi:10.1038/nbt0606-667

Beahm, D. R., Deng, Y., Riley, T. G., and Sarpeshkar, R. (2021). Cytomorphic electronic systems: A review and perspective. *IEEE Nanotechnol. Mag.* 15, 41–53. doi:10.1109/MNANO.2021.3113192

Bevc, S., Konc, J., Stojan, J., Hodošček, M., Penca, M., Praprotnik, M., et al. (2011). Enzo: A web tool for derivation and evaluation of kinetic models of enzyme catalyzed reactions. *PLoS One* 6, e22265. doi:10.1371/journal.pone.0022265

Bremer, H., Dennis, P., and Ehrenberg, M. (2003). Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie* 85, 597–609. doi:10.1016/S0300-9084(03)00105-6

Carrier, T. A., and Keasling, J. D. (1999). Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog.* 15, 58–64. doi:10.1021/bp9801143

Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature* 497, 619–623. doi:10.1038/nature12148

Deng, Y., Beahm, D. R., Ionov, S., and Sarpeshkar, R. (2021). Measuring and modeling energy and power consumption in living microbial cells with a synthetic ATP reporter. *BMC Biol.* 19, 101–121. doi:10.1186/s12915-021-01023-2

Dickinson, F. M., and Dickenson, C. J. (1978). Estimation of rate dissociation constants involving ternary complexes in reactions catalysed by yeast alcohol dehydrogenase. *Biochem. J.* 171, 629–637. doi:10.1042/bj1710629

Domingo-Félez, C., and Smets, B. F. (2020). Modeling denitrification as an electric circuit accurately captures electron competition between individual reductive steps: The activated sludge model-electron competition model. *Environ. Sci. Technol.* 54, 7330–7338. doi:10.1021/acs.est.0c01095

Eshtewy, N. A., and Scholz, L. (2020). Model reduction for kinetic models of biological systems. *Symmetry (Basel)* 12, 863–922. doi:10.3390/SYM12050863

Ganzhorn, A. J., Green, D. W., Hershey, A. D., Gould, R. M., and Plapp, B. V. (1987). Kinetic characterization of yeast alcohol dehydrogenases. Amino acid residue 294 and substrate specificity. *J. Biol. Chem.* 262, 3754–3761. doi:10.1016/s0021-9258(18)61419-x

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531. doi:10.1093/bioinformatics/btg015

Ihrig, J., Lill, R., and Mühlenhoff, U. (2006). Application of the DNA-specific dye EvaGreen for the routine quantification of DNA in microplates. *Anal. Biochem.* 359, 265–267. doi:10.1016/j.ab.2006.07.043

Jung, J. K., Alam, K. K., Verosloff, M. S., Capdevila, D. A., Desmau, M., Clauer, P. R., et al. (2020). Cell-free biosensors for rapid detection of water contaminants. *Nat. Biotechnol.* 38, 1451–1459. doi:10.1038/s41587-020-0571-7

Kamionka, A., Bogdanska-Urbaniak, J., Scholz, O., and Hillen, W. (2004). Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res.* 32, 842–847. doi:10.1093/nar/gkh200

Kim, J., Woo, S. S., and Sarpeshkar, R. (2018). Fast and precise emulation of stochastic biochemical reaction networks with amplified thermal noise in silicon chips. *IEEE Trans. Biomed. Circuits Syst.* 12, 379–389. doi:10.1109/TBCAS.2017.2786306

Krafft, C., Hinrichs, W., Orth, P., Saenger, W., and Welfle, H. (1998). Interaction of Tet repressor with operator DNA and with tetracycline studied by infrared and Raman spectroscopy. *Biophys. J.* 74, 63–71. doi:10.1016/S0006-3495(98)77767-7

Marshall, R., and Noireaux, V. (2019). Quantitative modeling of transcription and translation of an all-E. coli cell-free system. *Sci. Rep.* 9, 11980. doi:10.1038/s41598-019-48468-8

Néant, N., Lingas, G., Le Hingrat, Q., Ghosn, J., Engelmann, I., Lepiller, Q., et al. (2021). Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2017962118. doi:10.1073/pnas.2017962118

Nguyen, T.-H., Splechtna, B., Steinböck, M., Kneifel, W., Lettner, H. P., Kulbe, K. D., et al. (2006). Purification and characterization of two novel beta-galactosidases from Lactobacillus reuteri. *J. Agric. Food Chem.* 54, 4989–4998. doi:10.1021/jf053126u

Plapp, B. V. (2010). Conformational changes and catalysis by alcohol dehydrogenase. *Arch. Biochem. Biophys.* 493, 3–12. doi:10.1016/j.abb.2009.07.001

Poole, W., Pandey, A., Shur, A., Tuza, Z. A., and Murray, R. M. (2022). BioCRNpyler: Compiling chemical reaction networks from biomolecular parts in diverse contexts. *PLoS Comput. Biol.* 18, e1009987–19. doi:10.1371/journal.pcbi.1009987

Portaccio, M., Stellato, S., Rossi, S., Bencivenga, U., Mohy Eldin, M. S., Gaeta, F. S., et al. (1998). Galactose competitive inhibition of β-galactosidase (Aspergillus oryzae) immobilized on chitosan and nylon supports. *Enzyme Microb. Technol.* 23, 101–106. doi:10.1016/S0141-0229(98)00018-0

Powers, J. L., Kiesman, N. E., Tran, C. M., Brown, J. H., and Bevilacqua, V. L. H. (2007). Lactate dehydrogenase kinetics and inhibition using a microplate reader. *Biochem. Mol. Biol. Educ.* 35, 287–292. doi:10.1002/bmb.74

Resat, H., Petzold, L., and Pettigrew, M. F. (2009). "Kinetic modeling of biological systems," in.*Computational systems biology*, eds. R. Ireton, K. Montgomery, R. Bumgarner, R. Samudrala, and J. McDermott (Totowa, NJ: Humana Press), 311–335. doi:10.1007/978-1-59745-243-4_14

Salis, H. M. (2011). The ribosome binding site calculator. *Methods Enzymol.* 498, 19–42. doi:10.1016/B978-0-12-385120-8.00002-4

Sarpeshkar, R. (2010). *Ultra low power bioelectronics: Fundamentals, biomedical applications, and bio-inspired systems*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511841446

Silverman, A. D., Kelley-Loughnane, N., Lucks, J. B., and Jewett, M. C. (2019). Deconstructing cell-free extract preparation for *in vitro* activation of transcriptional genetic circuitry. *ACS Synth. Biol.* 8, 403–414. doi:10.1021/acssynbio.8b00430

Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E., and Hoggett, J. G. (2004). Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase: A single-molecule view of the transcription cycle. *J. Biol. Chem.* 279, 3239–3244. doi:10.1074/jbc.M310471200

Tang, J., Riley, W. J., Marschmann, G. L., and Brodie, E. L. (2021). Conceptualizing biogeochemical reactions with an Ohm's law analogy. *J. Adv. Model. Earth Syst.* 13. doi:10.1029/2021MS002469

Teo, J. J. Y., Kim, J., Woo, S. S., and Sarpeshkar, R. (2019a). "Bio-molecular circuit design with electronic circuit software and cytomorphic chips," in *2019* IEEE biomedical Circuits and systems conference (BioCAS) *(IEEE)*, 1–4. doi:10.1109/BIOCAS.2019.8918684

Teo, J. J. Y., and Sarpeshkar, R. (2020). The merging of biological and electronic circuits. *iScience* 23, 101688. doi:10.1016/j.isci.2020.101688

Teo, J. J. Y., Weiss, R., and Sarpeshkar, R. (2019b). An artificial tissue homeostasis circuit designed via analog circuit techniques. *IEEE Trans. Biomed. Circuits Syst.* 13, 540–553. doi:10.1109/TBCAS.2019.2907074

Teo, J. J. Y., Woo, S. S., and Sarpeshkar, R. (2015). Synthetic biology: A unifying view and review using analog circuits. *IEEE Trans. Biomed. Circuits Syst.* 9, 453–474. doi:10.1109/TBCAS.2015.2461446

Újvári, A., and Martin, C. T. (1996). Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase. *Biochemistry* 35, 14574–14582. doi:10.1021/bi961165g

Woo, S. S., Kim, J., and Sarpeshkar, R. (2015). A cytomorphic chip for quantitative modeling of fundamental bio-molecular circuits. *IEEE Trans. Biomed. Circuits Syst.* 9, 527–542. doi:10.1109/TBCAS.2015.2446431

Woo, S. S., Kim, J., and Sarpeshkar, R. (2018). A digitally programmable cytomorphic chip for simulation of arbitrary biochemical reaction networks. *IEEE Trans. Biomed. Circuits Syst.* 12, 360–378. doi:10.1109/TBCAS.2017.2781253

Xu, H., and Ewing, A. G. (2004). A rapid enzyme assay for β-galactosidase using optically gated sample introduction on a microfabricated chip. *Anal. Bioanal. Chem.* 378, 1710–1715. doi:10.1007/s00216-003-2317-z

Zeng, J., Teo, J., Banerjee, A., Chapman, T. W., Kim, J., and Sarpeshkar, R. (2018). A synthetic microbial operational amplifier. *ACS Synth. Biol.* 7, 2007–2013. doi:10.1021/acssynbio.8b00138

frontiers | Frontiers in Bioengineering and Biotechnology

TYPE Brief Research Report
PUBLISHED 09 March 2023
DOI 10.3389/fbioe.2023.1115159

OPEN ACCESS

EDITED BY
Simon J. Moore,
Queen Mary University of London,
United Kingdom

REVIEWED BY
Thomas Schoenfeld,
Tamarack Bioscience, Inc., United States
Hung-En Lai,
Victoria University of Wellington,
New Zealand

*CORRESPONDENCE
Shuichi Asakawa,
asakawa@g.ecc.u-tokyo.ac.jp

SPECIALTY SECTION
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

RECEIVED 03 December 2022
ACCEPTED 27 February 2023
PUBLISHED 09 March 2023

CITATION
Bhuiyan A and Asakawa S (2023), Synthesis and cloning of long repeat sequences using single-stranded circular DNA.
Front. Bioeng. Biotechnol. 11:1115159.
doi: 10.3389/fbioe.2023.1115159

Synthesis and cloning of long repeat sequences using single-stranded circular DNA

Afsana Bhuiyan and Shuichi Asakawa*

Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Non-coding repeat expansion causes several neurodegenerative diseases, such as fragile X syndrome, amyotrophic lateral sclerosis/frontotemporal dementia, and spinocerebellar ataxia (SCA31). Such repetitive sequences must be investigated to understand disease mechanisms and prevent them, using novel approaches. However, synthesizing repeat sequences from synthetic oligonucleotides is challenging as they are unstable, lack unique sequences, and exhibit propensity to make secondary structures. Synthesizing long repeat sequence using polymerase chain reaction is often difficult due to lack of unique sequence. Here, we employed a rolling circle amplification technique to obtain seamless long repeat sequences using tiny synthetic single-stranded circular DNA as template. We obtained 2.5–3 kbp uninterrupted TGGAA repeats, which is observed in SCA31, and confirmed it using restriction digestion, Sanger and Nanopore sequencing. This cell-free, *in vitro* cloning method may be applicable for other repeat expansion diseases and be used to produce animal and cell culture models to study repeat expansion diseases *in vivo* and *in vitro*.

KEYWORDS

repeat sequence, cell-free synthetic biology, in vitro cloning, neurodegenerative diseasaes, spinocerebellar ataxia (SCA) 31, benign adult familial myoclonic epilepsy (BAFME)

Introduction

The discovery that simple tandem repeats or microsatellites can cause neurological diseases was revolutionary in the field of neurodegenerative disorders. Nearly 50 neurological diseases have been identified so far, of which 26 diseases are related to the repeat expansion in coding, non-coding, intron and 5′and 3′UTR regions (Rohilla and Gagnon, 2017; Paulson, 2018; Chintalaphani et al., 2021). Among the repeat expansion-related diseases, spinocerebellar Ataxia type 31 (SCA31) is caused by the repeat expansion of 2.5 to 3.8 kbp pentanucleotide TGGAA, where a pure $(TGGAA)_n$ extended for at least 110 repeats in the intron region of the BEAN gene (Sato et al., 2009), whereas benign adult familial myoclonic epilepsy (BAFME1) is associated with the repeat expansion of 105 to 3,680 units of TTTCA (Cen et al., 2018; Ishiura et al., 2018). Such diseases can be classified based on the DNA sequences of repeat units (trinucleotide, tetranucleotide, pentanucleotide, or hexanucleotide). To develop transgenic models of these diseases, we need to obtain disease sequences from patients, except for some special cases (Mizielinska et al., 2014; Swinnen et al., 2018). However, this is often difficult due to ethical regulations and unavailability of patient-derived genomic DNA. In addition, the sizes of disease sequences are limited to that of the patients. Yet, longer DNAs are desirable for developing disease models because of the anticipation found in the repeat diseases (Wells, 1996). In the case of non-repeat mutation

TABLE 1 Sequences for oligonucleotides used in ssDNA template preparation, rolling circle amplification, PCR amplification check and Sanger sequencing.

Oligonucleotides	Sequence (5′–3′)
80nt TGGAA ssDNA	TGGAATGGAATGGAATGGAATGGAATG GAATGGAATGGAATGGAA TGGAATGGA ATGGAATGGAATGGAATGGAATGGAA
Sca31_RCA_F	TGGAATGGAATGGAA
Sca31_RCA_R	TTCCATTCCATTCCA
pIRES_F	GTAACAACTCCGCCCCATT
pIRES_R	GGTACCGTCGACTGCAGAA
M13_90mer ssDNA (Control for RCA)	GTAAAACGACGGCCAGTAAACAGTGAC CATGATAGTGGCCACCCTGCAACCGTG TTGTTTGTCAGGTTCATTTGTCATAGCTGT TTCCTG
M13_RCA_F	CTGCAACCGTGTTGTTT
M13_RCA_R	GGTGGCCACTATCAT
80 nt TTTCA ssDNA	TTTCATTTCATTTCATTTCATTTCATTTCA TTTCATTTCATTTCATTTC ATTTCATTT CATTTCATTTCATTTCATTTCA
BAFME1_RCA_F	TTTCATTTCATTTCA
BAFME1_RCA_R	TGAAATGAAATGAAA

disorders, disease genes can be synthesized and studied without patient sources. Moreover, methods to obtain uninterrupted long repeats are lacking owing to the technical obstacles associated with amplification and cloning. Recombinant vectors containing these synthetic repeats can have numerous applications in biological, medical, and bioengineering research. These vectors can be used for studying repeat associated-non-AUG translation (RAN polypeptides) as well as formation of RNA foci and its interaction with RNA binding proteins (Lee et al., 2013; Malik et al., 2021).

Several methods have been described for trinucleotide repeat synthesis using synthetic oligonucleotides. For example, 20 bp trinucleotide repeat has been used as template as well as primer for conventional polymerase chain reaction (PCR) and cloned in a vector (Ordway and Detloff, 1996). Synthesis of long iterative polynucleotide (SLIP) and non-template PCR method for trinucleotide repeat synthesis was developed based on the theory that filling gaps leads to repeat expansion (Takahashi et al., 1999). Another ligation-based method required iterative ligation reactions to obtain expanded repeats (Kim et al., 2005). Alternatively, concatenated DNA was obtained by random insertion of restriction sites using ligation and PCR (Jiang et al., 1996). PCR has also been used to obtain repeat sequences from DNA sources for polyglutamine expansions disease like Huntington's disease and spinocerebellar ataxia type 10 (Laccone et al., 1999; Peters and Ross, 1999; Michalik et al., 2001; Matsuura and Ashizawa, 2002). Amplification of dimerized expanded repeats (ADER) method was developed to obtain 2,000 CTG repeats using phi29 DNA polymerase in a cell free system (Osborne and Thornton, 2008). In rolling circle amplification or RCA (also known as hyperbranched amplification), small single stranded circular DNA is used as a template for amplification (Fire and Xu, 1995). This method has a low error rate, strong strand displacement activity, high processivity, and uses circular DNA for isothermal amplification (Hafner et al., 2001). Here we describe a cell-free synthetic method taking SCA31 and BAFME1 as examples for the synthesis and cloning of long repeat sequences using rolling circle amplification. As we were able to reproducibly obtain long repeat sequences in a cell-free manner, we regarded this method as a kind of "*in vitro* cloning".

Material and methods

Table 1 shows all oligonucleotide sequences used in the experiment. Repeat single-stranded DNA (ssDNA) with phosphorylated 5′ end (80 nt, $(TGGAA)_{16}$, HPLC purified) was obtained from Shanghai Generay Biotech Co., Ltd. (Generay, Shanghai, China). For RCA control, M13 DNA (90 nt) without repeat sequence, was obtained from Hokkaido System Science Co., Ltd. (Hokkaido, Japan). The 20 pmol ssDNA was incubated overnight at 60°C with CircLigase II (Lucigen CL9021K) according to the manual condition to obtain single-stranded circular DNA. The ligase was inactivated by heating at 80°C for 10 min. To remove excess ssDNA, exonuclease T (NEB M0265) was used at 25°C for 30 min and inactivated at 65°C for 20 min. Alternatively Exonuclease I (NEB M0293) can be used (See the results & discussion). This clean circular DNA (2 pmol) was used as a template for RCA, which was performed using Bst DNA polymerase, large fragment (NEB M0275) with 5′phospholated forward and reverse primers (Sca31_RCA_F, Sca31_RCA_R) at 60°C for 12 h. The amplified products were treated with 5 U mung bean nuclease (Takara Bio, 2420) and 0.5 U nuclease P1 (Wako 145-08221) at 37°C for 10 min. After digestion, the DNA was run on a 1.5% agarose gel and 2-3 kbp DNA samples were excised from the gel and purified with the Fastgene gel/PCR extraction kit (Nippon genetics co. FG-91302). The pIRES2 DsRed-Express2 vector (Takara Bio, 632540) was digested by AfeI (NEB R0652) and Shrimp Alkaline Phosphatase (rSAP) (NEB M0371) to prepare blunt ended vector. pIRES and the purified 2-3 kbp DNA insert were ligated for 2 h using T4 DNA Ligase (NEB M0202) and electroporated in Stable Competent *E. coli* (NEB C3040) cells. After 30 min recovery of competent cells, they were plated on kanamycin-supplemented culture plates and incubated at 30°C overnight. The insert lengths of selected clones were checked by PCR amplification using Phusion High-Fidelity DNA Polymerase (NEB M0530) and vector primers (pIRES_F pIRES_R) and by restriction digestion with NheI-HF & EcoRI-HF whose sites are located close to the cloning site, and by Sanger sequencing. The same process was applied to another repeat TTTCA (BAFME1) (80 nt, $(TTTCA)_{16}$, HPLC purified) to check the versatility of the method. The whole length of one SCA31 plasmid was sequenced using Oxford Nanopore Technology (ONT) (Stevanovski et al., 2022). Library preparation with the ligation sequencing kit (SQK-LSK110) was followed by sequencing on Flongle Flow Cell (FLO-FLG001). Multiple alignment using fast fourier transform (MAFFT) on raw sequence data and consensus sequence polishing by the Medaka tool (Lee et al., 2021) were employed to construct a complete plasmid map.

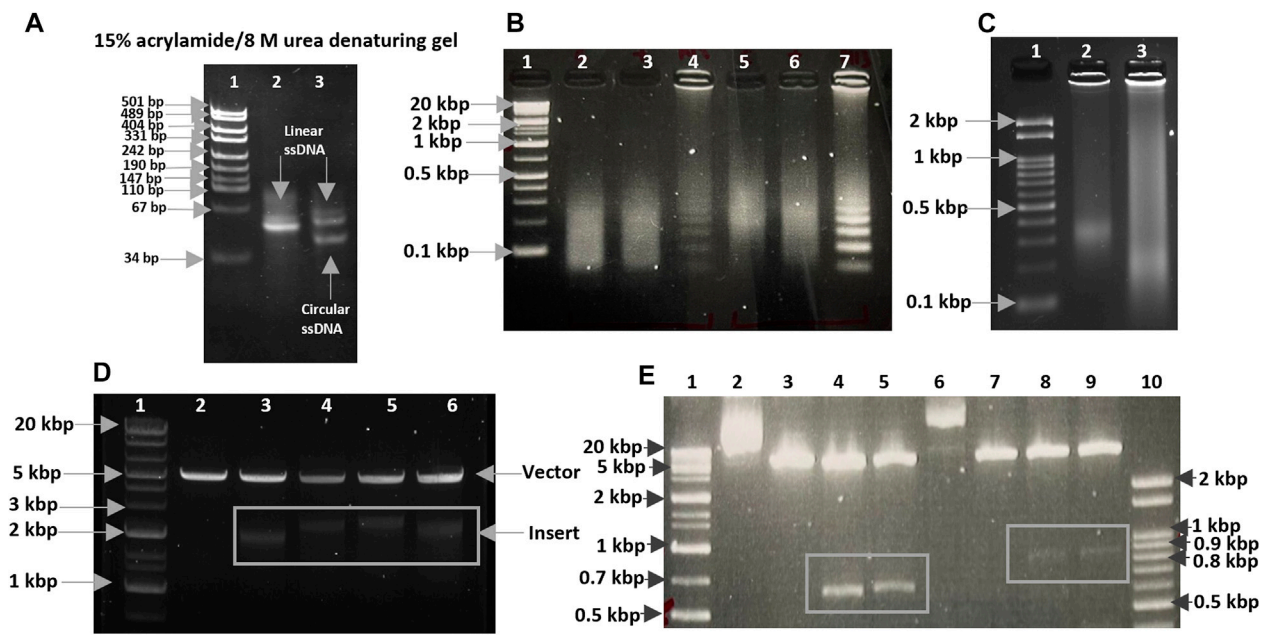


FIGURE 1
(A) CircLigase II ssDNA ligase converted 80 nt linear ssDNA oligos into circular ssDNA. Lane 1. pUC19 DNA/MspI (HpaII) Marker (Thermo Scientific); Lane 2. 80 nt linear ssDNA oligos (upper band); Lane 3. Unreacted 80 nt linear ssDNA oligos (upper band), and 80 nt circular ssDNA product (lower band). **(B)** RCA products after 1 h incubation. Lane 1. Size maker, Gene Ladder Wide 1 (0.1–20 kbp) (Nippon genetics co. 313-06961); Lane 2,3,5,6. RCA for SCA31; Lane 4,7. RCA for M13_90mer. **(C)** RCA products after 12 h incubation. Lane 1. Size maker, Gene Ladder 100 (0.1-2 kbp) (Nippon genetics co. 316-06951); Lane 2. RCA for SCA31 (12 h); Lane 3. RCA digested with 5 U mung bean nuclease and 0.5 U nuclease P1. **(D)** Size analysis of the TGGAA plasmids. Lane 1. Size maker, Gene Ladder Wide 1 (0.1–20 kbp) (Nippon genetics co. 313-06961); Lane 2. Restriction digestion of pIRES plasmid (control) using NheI-HF and EcoRI-HF. Lane 3-6. Restriction digestion of TGGAA plasmid obtained from four different clones using NheI-HF and EcoRI-HF. Inserts were shown in a white box. **(E)** Size analysis of the of TTTCA plasmids. Lane 1. Size maker, Gene Ladder Wide 1 (0.1-20 kbp) (Nippon genetics co. 313-06961); Lane 2,6. Undigested plasmids; Lane 3,7. Restriction digestion of pIRES plasmid (control) using NheI-HF and EcoRI-HF. Lane 4,5,8,9. Restriction digestion of TTTCA plasmids obtained from four different clones using NheI-HF and EcoRI-HF. Inserts were shown in white boxes. Lane 10. Size maker, Gene Ladder 100 (0.1–2 kbp) (Nippon genetics co. 316-06951).

Results & discussion

The initial oligonucleotide containing 80 nucleotides was selected for our experiment, as the circularized template exhibits maximum amplification efficiency for this size (Joffroy et al., 2018). The formation of circular DNA was confirmed using 15% acrylamide/8 M urea denaturing gel (Figure 1A; Supplementary Figure S1A, B). The size discrepancy between marker and ssDNA (Figure 1A; Supplementary Figure S1A) may be due to base composition bias of ssDNA causing differential mobility. The reaction mixture of circularization was treated with a single-strand specific nuclease, Exonuclease T. Unreacted linear ssDNA found in lane 3 of Supplementary Figure S1B was digested and not found in lane 5, which showed only circular DNA was left. Alternatively, Exonuclease I may be more effective for ssDNA containing C nucleotide(s). After removing linear ssDNA, RCA was employed for 1 h to SCA31 (repeat unit is TTGGA) and M13_90mer circular DNA (control). Ladder like pattern (Kuhn and Frank-Kamenetskii, 2005) was observed for control (M13_90mer) and smear was observed for concatenated repeat sequence showing the success of RCA (Figure 1B). In order to produce a longer repeat sequence, the RCA reaction was extended to 12 h; consequently, extremely large RCA products were obtained, which were concatemers of the repeat DNA (Figure 1C). DNA ranging from 2 to 3 kbp was cut out from the randomly elongated RCA products and cloned into pIRES vector (Figure 1D). Sanger sequence result for one clone confirmed at least 90-unit repeats of TGGAA from the 5′ end (Figure 2A) and 180-unit repeats of TTCCA from the 3’ end (Figure 2B). The same procedure was applied for the BAFME1 repeat (repeat unit is TTTCA). Figure 1E white boxes shows that inserts of about 600 bp (Lane 4-5) and about 900 bp (Lane 8-9) were cloned in the pIRES vector. We confirmed the sequence of one clone and found that the insert contained at least 80-unit repeats of TTTCA (Figure 2C). These results show the reproducibility of this method; repetitive sequences of different repeat sequences with various unit lengths could be generated by RCA and be cloned into plasmid vectors.

To characterize the whole sequence composition and repeat unit length, Oxford Nanopore Technology (ONT) was attempted on a SCA31 plasmid containing a 2.5 kbp repeat. From 8,000 reads (N50: 6,833bp), 3,831 reads were extracted with 6,000-8,000 bp length (almost full length of the plasmid), which were assembled to form the plasmid containing the repeat and vector (Supplementary Figure S2A). Supplementary Figure S2B shows 50 random reads aligned onto the consensus

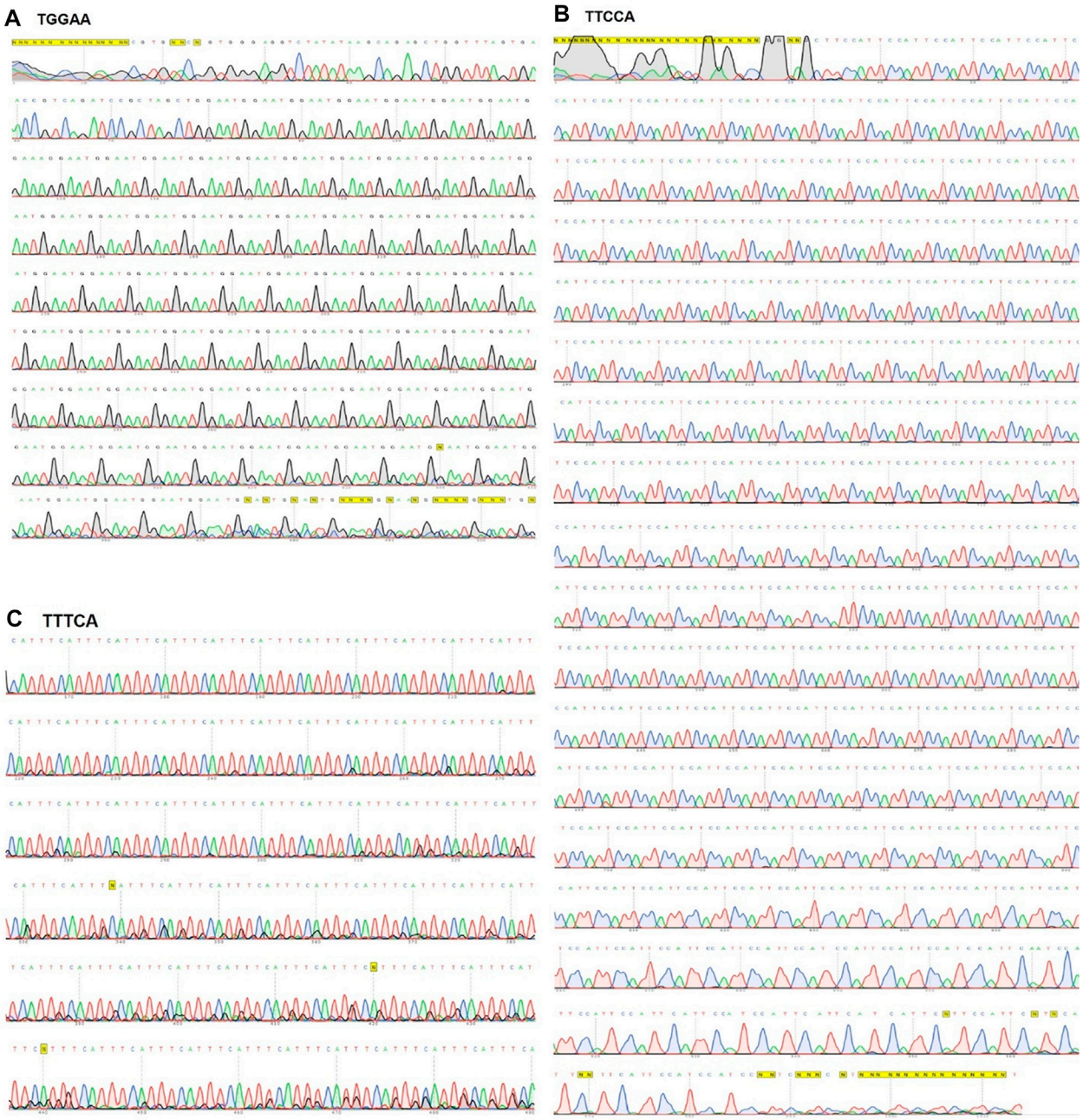


FIGURE 2
Sanger sequencing result for SCA31 and BAFME1 repeats. **(A)** TGGAA—Up to 500 bp sequence from the 5′side of a plasmid containing 2.5 kb SCA31 insert was confirmed by Sanger sequencing. **(B)** TTCCA—Up to 1 kbp sequence from the 3′side of the SCA31 plasmid was confirmed by Sanger sequencing. **(C)** TTTCA—About 180 units of the BAFME1 repeat of was inserted into the vector. Up to 400 bp sequence from the 5′side was confirmed by Sanger sequencing.

sequence. After polishing with the Medaka tool, full length sequence of the plasmid was obtained (Supplementary Figure S2C), containing 445 TGGAA repeats that included four mutated sequences—TGTAA, TTAAA, TGGAGG, TGGAT. Although these point mutations were found in the sequence, the longest stretch of a perfect TGGAA repeat was 1960 bp (392 TGGAA), which covered the pathogenic size. In summary, our method successfully obtained uninterrupted long repeat sequences using Bst DNA polymerase. However, our trial with phi29 polymerase was not successful despite trying with various concentrations and different reaction conditions (Supplementary Figure S3). RCA products did not migrate from the loading well into the gel and did not run after treating with various concentrations of mung bean nuclease. As a result, the DNA could not be extracted out from the gel. Our method can be used for generating long tandem repeats of tailored size in a cell-free manner and combined with

the ADER method (Osborne and Thornton, 2008) to stabilize the repeat units in *E. coli*. Moreover, this *in vitro* cloning method may be applicable for other types of repeat expansion, including those for neurodegenerative diseases, undiscovered disease-causative repeats or artificial repeats and may be used to produce animal and cell culture models to study repeat expansion diseases *in vivo* and *in vitro*. Lastly, we believe that this method can also be used for studying the function of repeats in the genome, such as alpha-satellite, centromeres and telomeres.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

SA and AB contributed to conception and design of the study. AB performed all experiments and analysis. AB wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the Japan Society for the Promotion of Science (grant no. JP20H00429).

Acknowledgments

We acknowledge Mariko Kondo for editing the manuscript. We acknowledge Kazutoshi Yoshitake, Ryo Yonezawa and Hideaki Mizobata for various suggestions regarding the Nanopore experiment.

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/fbioe.2023.1115159/full#supplementary-material

References

Cen, Z., Jiang, Z., Chen, Y., Zheng, X., Xie, F., Yang, X., et al. (2018). Intronic pentanucleotide TTTCA repeat insertion in the SAMD12 gene causes familial cortical myoclonic tremor with epilepsy type 1. *Brain* 141 (8), 2280–2288. doi:10.1093/brain/awy160

Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., and Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta neuropathol. Commun.* 9 (1), 98–20. doi:10.1186/s40478-021-01201-x

Fire, A., and Xu, S. Q. (1995). Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci.* 92 (10), 4641–4645. doi:10.1073/pnas.92.10.4641

Hafner, G. J., Yang, I. C., Wolter, L. C., Stafford, M. R., and Giffard, P. M. (2001). Isothermal amplification and multimerization of DNA by Bst DNA polymerase. *Biotechniques* 30 (4), 852–867. doi:10.2144/01304rr03

Ishiura, H., Mitsui, J., Yoshimura, J., Matsukawa, M. K., Fujiyama, A., Toyoshima, Y., et al. (2018). Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat. Genet.* 50 (4), 581–590. doi:10.1038/s41588-018-0067-2

Jiang, S. W., Trujillo, M. A., and Eberhardt, N. L. (1996). An efficient method for generation and subcloning of tandemly repeated DNA sequences with defined length, orientation and spacing. *Nucleic acids Res.* 24 (16), 3278–3279. doi:10.1093/nar/24.16.3278

Joffroy, B., Uca, Y. O., Prešern, D., Doye, J. P., and Schmidt, T. L. (2018). Rolling circle amplification shows a sinusoidal template length-dependent amplification bias. *Nucleic acids Res.* 46 (2), 538–545. doi:10.1093/nar/gkx1238

Kim, S. H., Cai, L., Pytlos, M. J., Edwards, S. F., and Sinden, R. R. (2005). Generation of long tracts of disease-associated DNA repeats. *BioTechniques* 38 (2), 247–253. doi:10.2144/05382rr01

Kuhn, H., and Frank-Kamenetskii, M. D. (2005). Template-independent ligation of single-stranded DNA by T4 DNA ligase. *FEBS J.* 272 (23), 5991–6000. doi:10.1111/j.1742-4658.2005.04954.x

Laccone, F., Maiwald, R., and Bingemann, S. (1999). A fast polymerase chain reaction-mediated strategy for introducing repeat expansions into CAG-repeat containing genes. *Hum. Mutat.* 13 (6), 497–502. doi:10.1002/(sici)1098-1004(1999)13:6<497:aid-humu10>3.0.co;2-6

Lee, J. Y., Kong, M., Oh, J., Lim, J., Chung, S. H., Kim, J. M., et al. (2021). Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. *Sci. Rep.* 11 (1), 20740. doi:10.1038/s41598-021-00178-w

Lee, Y. B., Chen, H. J., Peres, J. N., Gomez-Deza, J., Attig, J., Štalekar, M., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell. Rep.* 5 (5), 1178–1186. doi:10.1016/j.celrep.2013.10.049

Malik, I., Kelley, C. P., Wang, E. T., and Todd, P. K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell. Biol.* 22 (9), 589–607. doi:10.1038/s41580-021-00382-6

Matsuura, T., and Ashizawa, T. (2002). Polymerase chain reaction amplification of expanded ATTCT repeat in spinocerebellar ataxia type 10. *Ann. Neurol.* 51 (2), 271–272. doi:10.1002/ana.10049

Michalik, A., Kazantsev, A., and Broeckhoven, C. V. (2001). Method to introduce stable, expanded, polyglutamine-encoding CAG/CAA trinucleotide repeats into CAG repeat-containing genes. *Biotechniques* 31 (2), 250–254. doi:10.2144/01312bm02

Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* 345 (6201), 1192–1194. doi:10.1126/science.1256800

Ordway, J. M., and Detloff, P. J. (1996). *In vitro* synthesis and cloning of long CAG repeats. *Biotechniques* 21 (4), 609–612. doi:10.2144/96214bm08

Osborne, R. J., and Thornton, C. A. (2008). Cell-free cloning of highly expanded CTG repeats by amplification of dimerized expanded repeats. *Nucleic acids Res.* 36 (4), e24. doi:10.1093/nar/gkn025

Paulson, H. (2018). Repeat expansion diseases. *Handb. Clin. neurology* 147, 105–123. doi:10.1016/B978-0-444-63233-3.00009-9

Peters, M. F., and Ross, C. A. (1999). Preparation of human cDNas encoding expanded polyglutamine repeats. *Neurosci. Lett.* 275 (2), 129–132. doi:10.1016/s0304-3940(99)00758-2

Rohilla, K. J., and Gagnon, K. T. (2017). RNA biology of disease-associated microsatellite repeat expansions. *Acta neuropathol. Commun.* 5 (1), 63–22. doi:10.1186/s40478-017-0468-y

Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., et al. (2009). Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. *Am. J. Hum. Genet.* 85 (5), 544–557. doi:10.1016/j.ajhg.2009.09.019

Stevanovski, I., Chintalaphani, S. R., Gamaarachchi, H., Ferguson, J. M., Pineda, S. S., Scriba, C. K., et al. (2022). Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci. Adv.* 8 (9), eabm5386. doi:10.1126/sciadv.abm5386

Swinnen, B., Bento-Abreu, A., Gendron, T. F., Boeynaems, S., Bogaert, E., Nuyts, R., et al. (2018). A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta neuropathol.* 135 (3), 427–443. doi:10.1007/s00401-017-1796-5

Takahashi, N., Sasagawa, N., Suzuki, K., and Ishiura, S. (1999). Synthesis of long trinucleotide repeats *in vitro*. *Neurosci. Lett.* 262 (1), 45–48. doi:10.1016/s0304-3940(99)00031-2

Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats (*). *J. Biol. Chem.* 271 (6), 2875–2878. doi:10.1074/jbc.271.6.2875

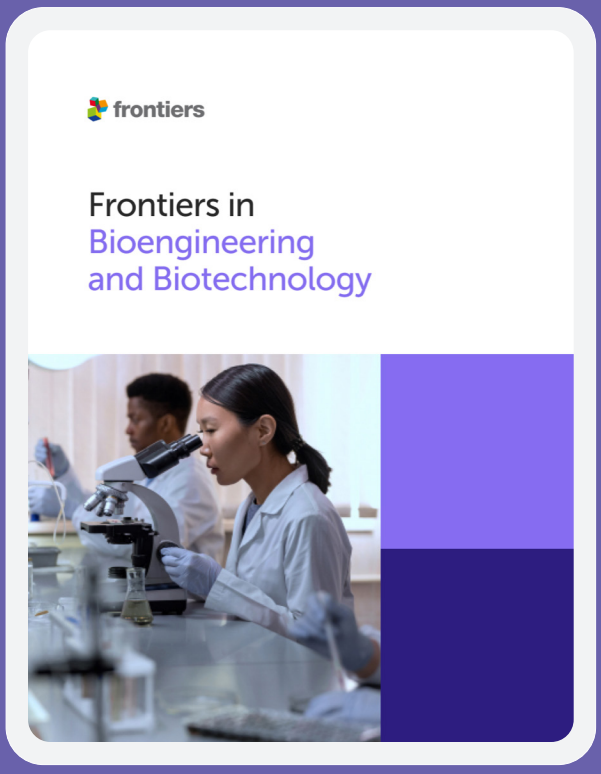
Frontiers in
Bioengineering and Biotechnology

Accelerates the development of therapies, devices, and technologies to improve our lives

A multidisciplinary journal that accelerates the development of biological therapies, devices, processes and technologies to improve our lives by bridging the gap between discoveries and their application.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact