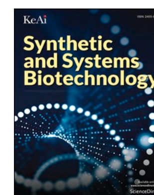
Synthetic and Systems Biotechnology 6 (2021) 192–199

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.sciencedirect.com/journal/synthetic-and-systems-biotechnology



A linear DNA template-based framework for site-specific unnatural amino acid incorporation

Xinjie Chen [1], Yingying Liu [1], Jiaqi Hou [1], Yuan Lu [*]

Key Laboratory of Industrial Biocatalysis, Ministry of Education, Department of Chemical Engineering, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Cell-free protein synthesis
Linear PCR product
Rapid screening
Unnatural amino acids

ABSTRACT

Site-specific incorporation of unnatural amino acids (UNAAs) into proteins using an orthogonal translation system (OTS) has expanded the scope of protein-coding chemistry. The key factor affecting UNAA embedding efficiency is the orthogonality of the OTS. Compared to traditional cell systems, cell-free systems are more convenient to control the reaction process and improve the utilization rate of UNAA. In this study, a linear DNA template-based cell-free unnatural protein synthesis system for rapid high-throughput screening and evolution was proposed. A total of 14 cell extracts were selected for screening out cell extract with high expression level. The result showed that EcAR7 ΔA ΔSer cell extract was optimal for the cell-free system. In addition, the screening results of four UNAAs, p-propargyloxy-l-phenylalanine (pPaF), p-azyl-phenylalanine (pAzF), p-acetyl-l-phenylalanine (pAcF), and p-benzoyl-l-phenylalanine (pBpF), showed that o-aaRS and o-tRNA of pPaF had good orthogonality. A new pair of corresponding o-aaRS and o-tRNA for pBpF was screened out. These results proved that this method could speed up the screening of optimal OTS components for UNAAs with versatile functions.

1. Introduction

Unnatural amino acids (UNAAs) embedded in protein synthesis is a powerful tool for biosynthesis. Using UNAA containing a special side chain to replace natural amino acid can provide a variety of new physical and chemical properties and biological functions of the protein [1,2]. UNAAs contain multiple reactive side chains that site-specifically bind to proteins or peptides by orthogonal tRNA (o-tRNA) and aminoacyl tRNA synthase to enrich protein structure and function [3,4]. At present, there are more than 150 UNAAs embedded into proteins through orthogonal translation systems (OTSs), which are widely used in protein labeling, enzyme activity improvement, biotherapeutics, and biocatalysis [5]. Site-specific binding of probes such as biotin and fluorescent groups can contribute to high-throughput analysis of proteins, protein-protein interactions, and protein-nucleic acid interactions [6–10]. Genetically encoded UNAAs have been used to synthesize antibody-drug conjugates (ADCs), which can preferentially deliver cytotoxic drugs to cells presenting tumor-associated antigens to achieve improved drug efficacy and safety. In addition, protein combined with UNAAs can also be used as a new biological macromolecule material, protein drugs, and large molecule covalent inhibitors [11,12]. The UNAA incorporation plays an important role in the development of protein engineering.

UNAAs embedding methods mainly include global suppression based on natural translation system [2], termination codon suppression based on OTSs [13], code shift suppression [14], meaningful codon redistribution, and unnatural base pair [15]. Currently, the most commonly used method is the amber suppression method based on UAG termination codon. This method is often used to embed UNAAs based on the OTS, including UNAA and its orthogonal o-tRNA/o-aaRS (orthogonal aminoacyl-tRNA synthetase) pairs. The orthogonality of OTS is the key to limit the embedding efficiency of UNAAs [16,17], which require the development of o-aaRSs with higher catalytic rates and stronger affinity for o-tRNA. At present, high-quality libraries of different aaRS mutants are critical for obtaining and screening highly active specific aaRSs. In the *E. coli* system, aaRS library can be generated and screened through the multiplex automated genome engineering (MAGE) and fluorescence activated cell sorting (FACS) [18]. However, some UNAAs are hard to pass through the cell membrane, generate toxicity to cells, and inhibit cell growth, and therefore, it is not easy to screen out effective OTS

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author.
E-mail address: yuanlu@tsinghua.edu.cn (Y. Lu).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.synbio.2021.07.003
Received 12 May 2021; Received in revised form 16 July 2021; Accepted 27 July 2021
2405-805X/© 2021 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

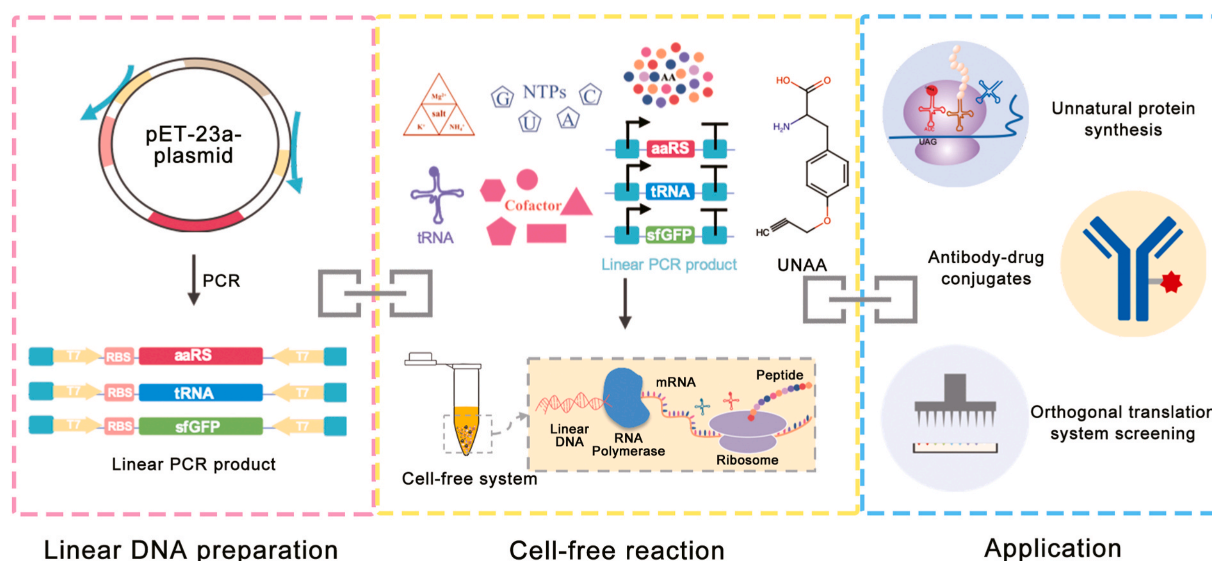


Fig. 1. Schematic diagram of experimental design. The design of linear sequences. These sequences included aaRS, tRNA and sfGFP. Optimal aaRS and tRNA were designed for four UNAAs, including P-propargyloxy-l-phenylalanine (pPaF), p-azyl-phenylalanine (pAzF), p-acetyl-l-phenylalanine (pAcF), and p-benzoyl-l-phenylalanine (pBpF). The same base modification was added at both ends of the upstream and downstream primers. The upstream primers were designed before the T7 promoter sequence, and the downstream primers were designed after the T7 terminator. The linear DNA template-based cell-free unnatural protein synthesis system was further performed. This system could be applied in unnatural protein synthesis, antibody-drug conjugates, and orthogonal translation system screening.

components [4,19]. Therefore, more and more studies are using cell-free protein synthesis systems to embed UNAAs [20]. As the cell-free system is open, it is convenient to control the reaction process and improve the utilization rate of UNAA, which has a broad application prospect in ligand-protein interaction, biotherapy, and other aspects [21].

Since cell-free protein synthesis (CFPS) system has the advantages of short production cycle, no cell membrane barrier, and the toxin tolerance, it is conducive to the redistribution of the genetic code, fine regulation of the reaction process, and further expansion of protein functions [14,22], which is suitable for evaluating the catalytic efficiency of OTSs [23]. In addition, the cell-free system can directly use linear polymerase chain reaction (PCR) products as the template for gene expression, which can speed up the rate and make it more convenient for the synthesis of target proteins by CFPS system. In recent years, studies have been conducted to establish mutant libraries with linear PCR products as templates for rapid screening and evolution, such as promoters, genetic circuits [24], and ribosome binding site (RBS) [25]. The preparation, protein expression and activity screening of linear PCR products can be completed in one day, which has the potential for high-throughput screening. At present, some studies have succeed in embedding UNAA with expression template based on linear PCR products, which significantly reduces labor intensity [26], but these have not been thoroughly studied. A linear DNA template-based framework for site-specific unnatural amino acid incorporation has not yet been constructed.

Therefore, this paper proposed a linear DNA template-based cell-free unnatural protein synthesis system for rapid high-throughput screening and evolution (Fig. 1). The linear PCR products as the expression templates of target proteins and OTS fragments were combined to embed UNAAs into the target proteins. The embedding strategy of 4 UNAAs in cell-free systems, the fast screening method of o-aaRS, and the addition sequence of OTS components were explored. In this way, the potential of the linear DNA template-based framework for site-specific unnatural amino acid incorporation and orthogonal element screening had been proved.

2. Materials and methods

2.1. Strains

A total of 14 *E. coli* extracts were selected in this study, including K-12 series (K-12 [27] (Addgene #61440), K-12 ΔtnaA [28], K-12 ΔtnaAΔsdaB [28]), commercial series (BL21 (DE3) (Biomed, China), BL21 ΔserB [29] (Addgene #34929), Rosetta (DE3), Rosetta-gami B (DE3), Origami, Origami B) (Biomed, China), rEc series (rEc. 13 [30] (Addgene #69494), rEc. 13. ΔA [30] (Addgene #69495)), EcAR7 series (EcAR7 ΔA ΔSer [31]) (Addgene #52055), C321 series (C321 [32] (Addgene #48999), C321. ΔA [33] (Addgene #68306)). The information of bacteria used to prepare cell extracts was shown in Table S1.

2.2. UNAAs, plasmids and primers

P-propargyloxy-l-phenylalanine (pPaF), p-azyl-phenylalanine (pAzF), p-acetyl-l-phenylalanine (pAcF), and p-benzoyl-l-phenylalanine (pBpF) were purchased from Sigma. His-tag antibody was purchased from Sigma. DNA plasmids used in CFPS were obtained from cultures of *E. coli* DH10B strain (Biomed Biotechnology, China) using Plasmid Mini kits (Omega Bio-Tek, America). All plasmids used in this experiment were sequence-verified. All linear PCR products were amplified with pfu high fidelity DNA polymerase (Beyotime Biotechnology, China). Primers were purchased from GENEWIZ Biotechnology with no modifications (Table S2).

2.3. Cell extract preparation

First, single colonies were selected and inoculated in 10 mL 2 × YTP medium. The 1 mL overnight culture was then inoculated in 200 mL 2 × YTP medium. When the OD600 value reached 0.6–0.8. 50 mL of bacterial solution was inoculated in 1 L LB medium and cultured under the optimal culture conditions. Monitored the growth status during the cultivation. In the middle and late stages of the logarithmic growth phase (about 3–4 h), centrifuged the cell culture at 8000 rpm for 10 min to collect cells. The bacteria were washed with S30A buffer and centrifuged twice. Added 1 mL S30A to 1 g bacteria and suspended the cells. The cryogenic cells were broken twice with a high-pressure

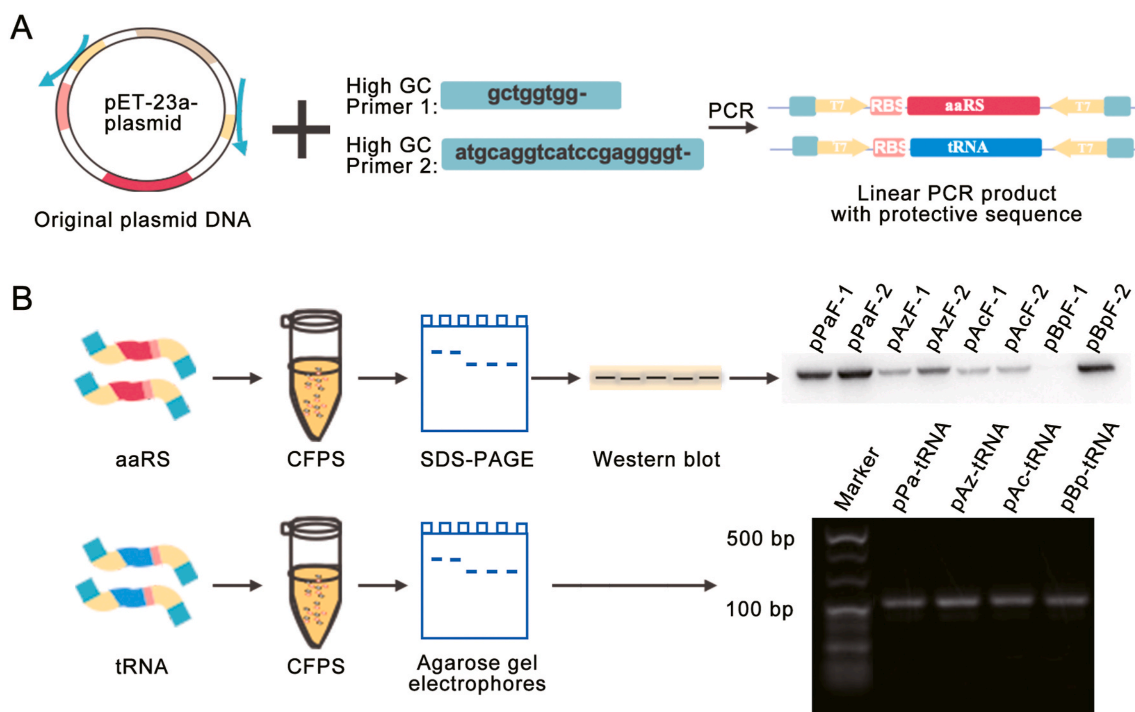


Fig. 2. The design of the aaRS fragment and tRNA fragment with sequence modification. (A) Schematic diagram of sequence design. High GC protective sequences were added to the ends of the o-tRNA and the o-aaRS linear DNA sequence to reduce the degradation of linear DNA in the cell-free system. Two different high GC protective sequences were added to both ends of the primer. The same base modification was added at both ends of the upstream and downstream primers. The linear product had a T7 promoter and RBS, ensuring that it was suitable for CFPS expression. (B) The expression of linear o-aaRS was detected by Western blotting. The expression of linear o-tRNA was detected by agarose gel electrophoresis. The number 1 and 2 meant high GC primer 1 and high GC primer 2. The letter M meant marker.

breaker. Centrifuged at 4 °C at 13,000×*g* for 30 min, and incubated the resuspended cells at 37 °C at 120 rpm in the dark for 80 min. After the dialysis at 4 °C overnight, they were centrifuged and stored at −85 °C in small aliquots [34].

2.4. Cell-free protein synthesis reactions

The CFPS reaction was performed in a 1.5 mL EP tube at 30 °C with a final volume of 20 μl. The reaction system included: 2 μl of 10 × salt, 20 mM magnesium glutamate, 0.1 mM PEP, 0.2 μl of T7 RNA polymerase, 0.2 μl of GSH, 0.8 μl of GSSH, 0.8 μl of 19AA, 0.8 μl of NTP, 5 μl of cell extract, 300 ng DNA template, 700 ng linear aaRS, 1500 ng t-DNA, 5 mM UNAA, 2% polyethylene glycol (PEG) 8000, and ddH$_2$O [34]. The sfGFP fluorescence was detected by a microplate reader.

2.5. Expression template design

The sfGFP, o-aaRS and o-tRNA sequences were located between the T7 promoter, RBS, ribozyme sequence and T7 termination on the vector pET-23a (Fig. S2), respectively. Two types of high GC-ratio protection sequences were added at both ends of the linear sequence. The amplified primers could be modified in two ways: GC sequence 1 (GCTGGTGG) and GC sequence 2 (ATGCAGGTCATCCGAGGGGT) [35].

2.6. Preparation of OTS components

Pfu high-fidelity DNA polymerase was used to perform PCR on o-aaRS and o-tRNA. Primer information was provided in the supplementary materials (Table S2). The reaction components included: 38 μl ddH$_2$O, 5 μl 10 × pfu buffer, 1 μl dNTPs (10 mM), 1.75 μl template, 2 μl forward primer, 2 μl reverse primer and 0.25 μl pfu polymerase. The program was run at 94 °C for 3 min, followed by 35 cycles of 94 °C for 30 s, 57 °C for 30 s and 72 °C for 2 min/kb. Final extension was running at 72 °C for 10 min and 4 °C forever. After PCR, the bands were confirmed by DNA agarose gel electrophoresis. The entire PCR product was then recovered, and DNA was recovered using ethanol precipitation. The specific method was to add 1/10 volume of sodium acetate and 1/3 of absolute ethanol to the product, and place it at −20 °C overnight. Collected the pellet after centrifugation, washed the pellet twice with 70% ethanol, dried the pellet and added an appropriate amount of water to dissolve the pellet and test the concentration.

2.7. Detection of expression products

First, the expression of o-aaRS was detected in a cell-free system by Western blotting using a His-Tag antibody. Second, the fluorescence value of sfGFP was detected by a microplate reader at a wavelength of 484 nm. Finally, it was necessary to detect the UNAA incorporation by mass spectrometry. Prior to mass spectrometry, the product was purified using affinity chromatography.

2.8. Mass spectrometric detection and analysis

After the detection of expression products, the samples were put into EP tubes and labeled. Then, the samples were sent to the protein platform at Tsinghua University for mass spectrometry analysis. *Protein-Prospector* were used to analyze the mass spectrum results (*https://prospector2.ucsf.edu/prospector/mshome.htm*). The software could analyze the reliability of the protein and provide information such as the coverage rate of amino acid sequence, the number of peptide segments, the abundance of protein, physicochemical properties, and so on.

2.9. Variance analysis of the orthogonal test

After obtaining the experimental data, the sum of test indexes of the

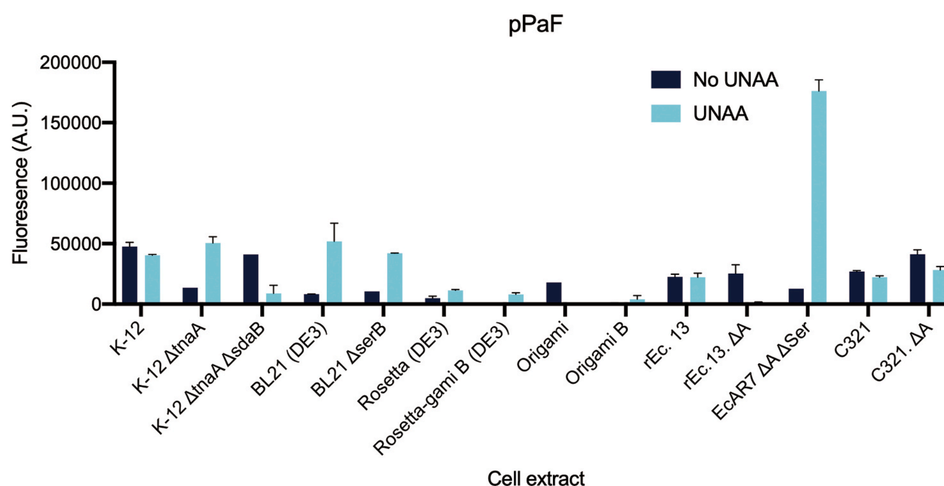


Fig. 3. Screening of 14 cell extracts. In this experiment, pPaF was selected for the UNAA embedding screening of 14 kinds of cell extracts. A total of 14 extracts were selected, including K-12 series (K-12, K-12 ΔtnaA, K-12 ΔtnaAΔsdaB), commercial series (BL21 (DE3), BL21 ΔserB, Rosetta (DE3), Rosetta-gami B (DE3), Origami, Origami B), rEc series (rEc. 13, rEc. 13. ΔA), EcAR7 series (EcAR7 ΔA ΔSer), C321 series (C321, C321. ΔA).

same level of each factor and the sum of test indexes of all test numbers were calculated. The average of the same level test index of each factor was also calculated. Then the sum of square and degree freedom of each factor and error variation were calculated to obtain the mean square. Then F-test was used to analyze the results to evaluate the significance of the factors.

3. Results and discussion

3.1. Design of linear o-tRNA and o-aaRS in CFPS

Linear DNA template-based cell-free systems could greatly improve the screening efficiency. However, linear DNA fragments were easily degraded by endonuclease I in the cell-free system [36]. Studies had found that adding high GC base sequences at both ends of the DNA sequence protected the DNA from degradation [35]. Using this GC-rich SP3 primer with the long sequence, Isokomosefe et al. amplified the PCR product with a high recovery rate, and proved that the PCR product with sequence modification has transcriptional activity [35]. Therefore, high GC protective sequences were added to the ends of the o-tRNA and the o-aaRS linear DNA sequence to reduce the degradation of linear DNA in the cell-free system (Fig. 2A). The expression of linear o-aaRS was detected by Western blotting, and the expression of linear o-tRNA was detected by agarose gel electrophoresis (Fig. 2B). Two different modified sequence were added to both ends of the primer, and OTS element was successfully prepared after PCR and electrophoresis recovery. The linear product had a T7 promoter and RBS, ensuring that it was suitable for CFPS expression. Optimal aaRS and tRNA, for four UNAAs, P-propargyloxy-l-phenylalanine (pPaF) [37], p-azyl-phenylalanine (pAzF) [38], p-acetyl-l-phenylalanine (pAcF) [18], and p-benzoyl-l-phenylalanine (pBpF) [39] were selected (Fig. S1, Table S3, and Table S4). The four UNAAs were designed for the target OTS, and could be used as crosslinking amino acids. 23TAG was selected as the embedding site. The expression verification results of PCR products in CFPS showed that both aaRS were expressed, but the expression level of aaRS modified by the sequence 1 was lower than that modified by the sequence 2, which was more obvious in the difference in the expression level of aaRS for pPaF and pAzF incorporation. Previous research found that when the GC content was between 60% and 65%, the protective efficiency of protective sequences was the best [40]. In this experiment, the GC content of sequence 1 was 75%, and the GC content of sequence 2 was 60%, which lead to a better protection efficiency of sequence 2. Besides, the length of sequence 2 was longer than sequence 1 and was closer to 20 bp, which could reduce the degradation of native nucleases [40]. These results indicated that PCR product-based o-aaRS had a higher efficiency on the embedding of UNAAs.

3.2. Screening of cell extracts

Different cell extracts in CFPS contained different specific factors that promoted protein expression, folding, or post-translational modification [41]. Cell extracts with high expression levels could better distinguish the differences in expression levels of the system and help with screening. Therefore, 14 cell extracts were screened for more effective follow-up experiments and proving the potential of this framework on the rapid screening. The basic cell-free system and pPaF were used to screen a variety of extracts with different genome modifications. The 14 cell extracts included K-12 series (K-12, K-12 ΔtnaA, K-12 ΔtnaAΔsdaB), commercial series (BL21 (DE3), BL21 ΔserB, Rosetta (DE3), Rosetta-gami B (DE3), Origami, Origami B), rEc series (rEc. 13, rEc. 13. ΔA), EcAR7 series (EcAR7 ΔA ΔSer), C321 series (C321, C321. ΔA) (Table S1). If the fluorescent value of the cell-free system after embedding UNAA was higher than that without the embedding of UNAA, it indicated that the embedding efficiency of UNAA was better, and the protein expression level was increased. Based on this point, after embedding UNAA into the target protein sfGFP, it could be found that K-12 ΔtnaA, BL21 (DE3), BL21 ΔserB, Rosetta (DE3), Rosetta-gami B (DE3), Origami B and EcAR7 ΔA ΔSer had better embedding efficiency. Among these cell extracts, only the protein expression level of the group with EcAR7 ΔA ΔSer cell extract increased obviously (Fig. 3), indicating that EcAR7 ΔA ΔSer was more suitable for embedding UNAA. This phenomenon resulted from that seven UAGs in the genome of EcAR7 strain were mutated into UAA codons, and current amber suppression method was based on UAG termination codon, which resulted in a higher embedding accuracy. In addition, this strain also removed the release factor 1 gene from the genome to reduce the ability of recognizing the stop codon UAG, which could help o-tRNA recognize the target codon. Besides, ΔSer meant phosphoserine aminotransferase gene was removed from the genome to avoid the hydrolysis of phosphorylated proteins. Based on the above results, EcAR7 ΔA ΔSer cell extract was selected for the following research.

3.3. Orthogonal screening of linear OTS elements

After optimizing the key components of the cell-free system suitable for UNAA embedding, the linear DNA template-based framework was applied on the rapid screening of OTS translation components. In this framework, adding OTS as a linear PCR product could eliminate time-

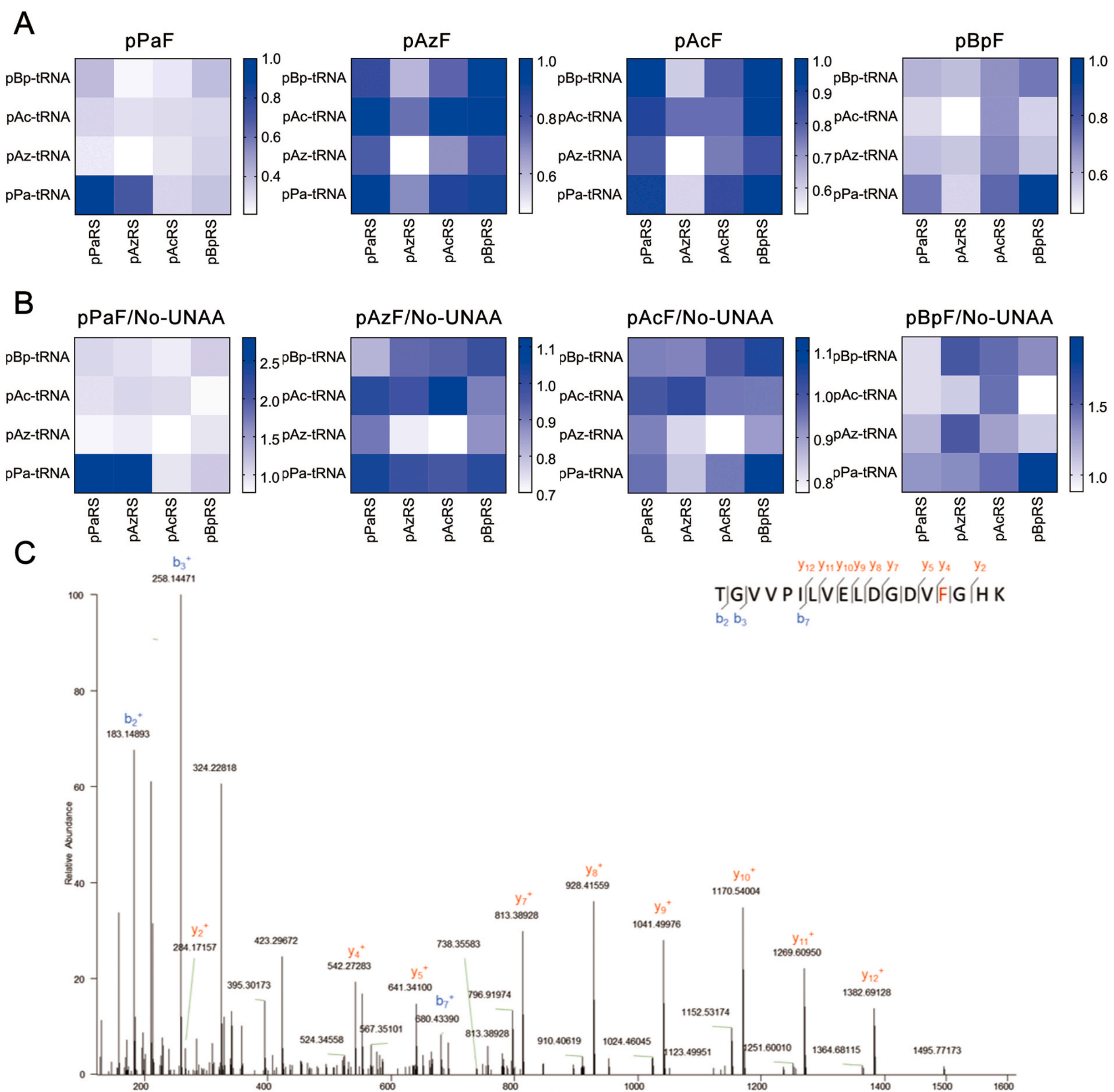
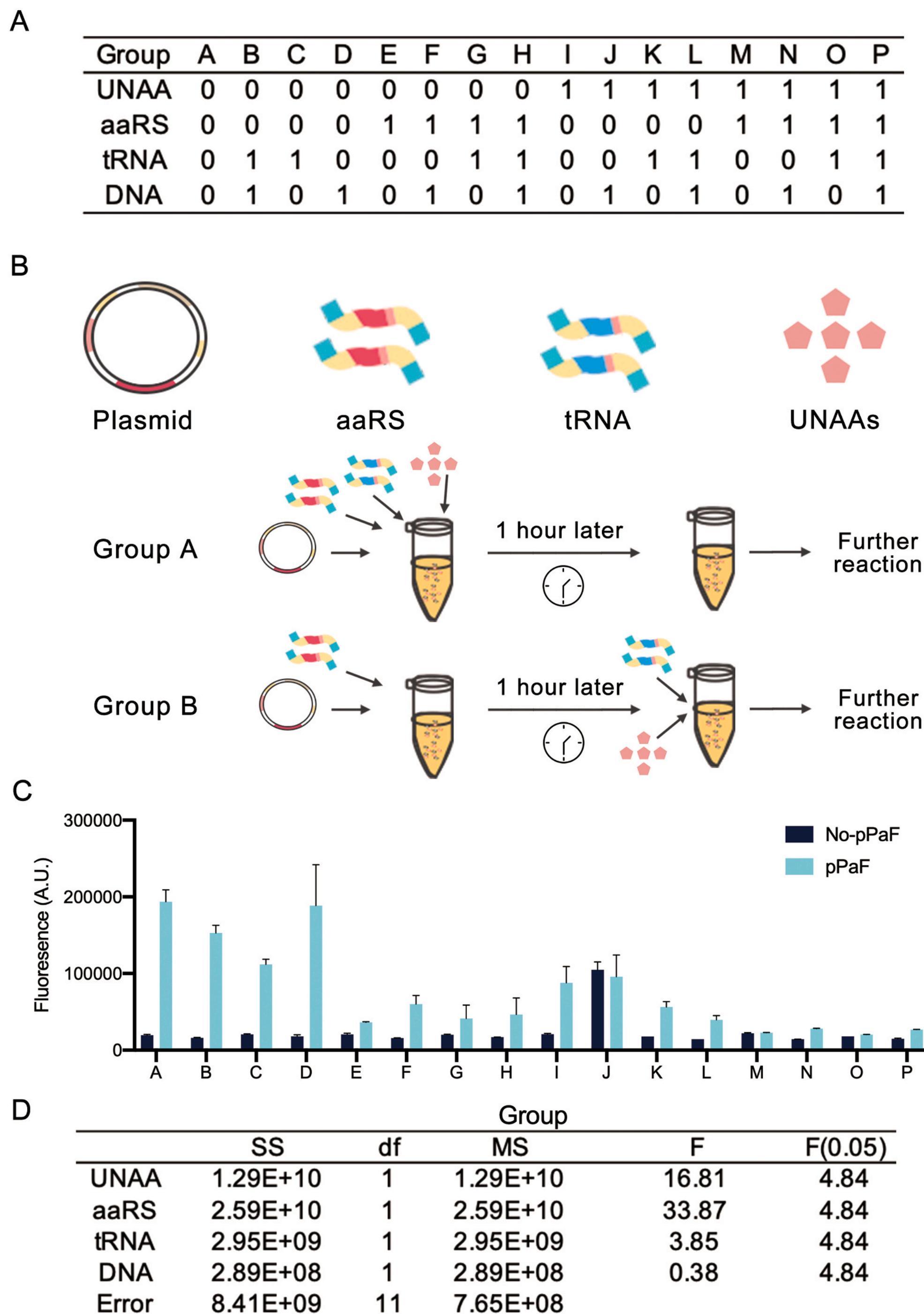


Fig. 4. Linear OTS was used to embed UNAA to select the orthogonal translation elements with the strongest orthogonality in the cell-free system. (A) The sfGFP proteins were embedded with UNAAs in cell-free systems. 23TAG was selected as the embedding site. Different UNAAs and different linear OTSs were added, respectively. The abscissa was different o-aaRS, and the ordinate was different o-tRNA. The darker the color in the hot spot map, the larger the fluorescent value. All the data in the hot spot map were normalized. The largest fluorescent value was set to 1. (B) The embedding efficiency of UNAAs. The ratio of the fluorescent value of the group with UNAAs to that of the group without UNAAs could be approximately regarded as the embedding efficiency of UNAAs. 23TAG was selected as the embedding site. The abscissa was different o-aaRS, and the ordinate was different o-tRNA. The darker the color in the hot spot map, the better embedding efficiency. (C) Mass spectrometry results of sfGFP with pPaF embedded at the 23TAG site. The results of mass spectrometry showed that pPaF was successfully embedded in the 23TAG site. According to the secondary mass spectrogram, blue was the matched b ion, and red was the matched y ion. In this picture, there was a lot of matching y ions, so the reliability was very good.

consuming aspects such as protein purification. To verify our hypothesis, UNAA was quickly embedded in sfGFP using a linear OTS fragment. Orthogonal translation modules expressed from linear DNAs were tested in CFPS, which could screen out highly efficient OTS for UNAAs. In this study, the screening system included four sets of UNAA, o-aaRS and o-tRNA. The OTS was added as a linear PCR product (Fig. 4A). From the orthogonal screening results of pPaF, it could be clearly found that the protein expression level of the corresponding o-aaRS and o-tRNA (pPaRS and pPa-tRNA) group was the highest, which was significantly higher than that of other groups, indicating that the pair of o-aaRS and o-tRNA had good orthogonality. As for the orthogonal screening results of pAzF and pAcF, it was found that there was no obvious orthogonality between the corresponding o-aaRS and o-tRNA. In addition, the orthogonal screening results of pBpF were also different from what was expected. The protein expression level of the pBpRS and pPa-tRNA was the highest, which was significantly different from the reported corresponding o-aaRS and o-tRNA (pBpRS and pBp-tRNA). This result indicated that there might be a pair of o-aaRS and o-tRNA that had better orthogonality. The results above showed that the linear DNA template-based framework had the potential on the rapid screening of OTS translation components.

Since the protein expression level did not directly represent the embedding efficiency of UNAA, the rough embedding efficiency was evaluated by the ratio of the protein expression level after the addition



Group	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
UNAA	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
aaRS	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
tRNA	0	1	1	0	0	0	1	1	0	0	1	1	0	0	1	1
DNA	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

	SS	df	MS	F	F(0.05)
UNAA	1.29E+10	1	1.29E+10	16.81	4.84
aaRS	2.59E+10	1	2.59E+10	33.87	4.84
tRNA	2.95E+09	1	2.95E+09	3.85	4.84
DNA	2.89E+08	1	2.89E+08	0.38	4.84
Error	8.41E+09	11	7.65E+08		

Fig. 5. The effect of OTS element addition order on the embedding efficiency of unnatural amino acids. (A) The order in which elements were added. The number 0 meant the addition of the component was at the start of the reaction, and the number 1 meant the addition of the component was 1 h after the start of the reaction. (B) Schematic diagram of the experiment with groups A and B as examples. (C) Fluorescence values of each group were measured after the reaction. All components added at the same time had the highest fluorescence value, and the fluorescence value of four groups was higher than that of the other groups. (D) Analysis of variance was used to analyze the key factors affecting the expression of the system. SS meant sum of square. Df meant degree freedom. MS meant mean square. F meant F statistic. F(0.05) was the F statistic at 95% confidence coefficient.

of UNAA to the protein expression level of the corresponding group without the addition of UNAA (Fig. 4B). The embedding efficiency of the four types of UNAAs was roughly the same as the orthogonal screening results above. The screening result of pPaF showed that the embedding efficiency of the group with the corresponding o-aaRS and o-tRNA (pPaRS and pPa-tRNA) was the highest, which was the same as the reported result. This result verified the screening efficiency of the linear DNA template-based framework. As for the screening result of pAzF, it was found that the corresponding group of pAcRS and pAc-tRNA had the highest embedding efficiency. However, for the embedding of pAzF, the orthogonality of these pairs of aaRS and tRNA was not obvious. For the screening result of pAcF, it was found that the corresponding group of pBpRS and pPa-tRNA had the highest embedding efficiency, which was different from the reported corresponding o-aaRS and o-tRNA (pAcRS and pAc-tRNA). The two results above indicated that the orthogonality of the reported corresponding o-aaRS and o-tRNA was not good. and there was still room for OTS translation components optimization. As for the screening result of pBpF, the embedding efficiency of the group with pBpRS and pPa-tRNA was the highest, which was the same as the result above. This result indicated that new corresponding o-aaRS and o-tRNA could be screened out by the linear DNA template-based framework.

In addition, mass spectrometry was also used to detect the specific embedding of UNAAs. The corresponding o-aaRS and o-tRNA, pPaRS and pPa-tRNA, were selected to embed pPaF into the target protein sfGFP. The results of the second-level mass spectrometry verified by mass spectrometry showed that pPaF was successfully embedded at the

23TAG site (Fig. 4C). The results showed that UNAAs could be embedded by adding linear DNA template directly *in vitro*.

3.4. Adding order of orthogonal translation components

Considering that the effective time and interaction relation were significantly different between different OTS components, it was necessary and meaningful to explore the influence of the adding order of the OTS components in the system. The pPaF with better embedding efficiency was used to be embedded into 23TAG-sfGFP to explore the effect of different OTS adding order on the protein expression level of the system (Fig. 5). The number 0 meant the addition of the component was at the start of the reaction, and the number 1 meant the addition of the component was 1 h after the start of the reaction. According to the fluorescent value of the reaction, the embedding efficiency of the group A was the highest, that was, all OTS components were added in the initial reaction at the same time. The result showed that, when performing UNAA embedding experiments, it was necessary to ensure that all OTS components participate in the reaction at the beginning of the reaction. In addition, analysis of variance was used to analyze the key factors affecting the expression of the system (Fig. 5D). Sum of square (SS), degree freedom (*df*), mean square (MS) and F statistic (F) were displayed. F(0.05) was the F statistic at 95% confidence coefficient. In this experiment, since each group had the same *df*, F statistic represented the disturbance of each factor to the system. The larger F statistic was, the larger the disturbance was. The results of F-test indicated that the adding order of UNAA ($F = 16.81$, 1 and 11 *df*) and aaRS ($F = 33.87$, 1 and 11 *df*) had a significant impact on the embedding efficiency of the cell-free system. The adding order of tRNA ($F = 3.85$, 1 and 11 *df*) and DNA ($F = 0.38$, 1 and 11 *df*) was not significant. Among the four OTS components, the adding order of aaRS had the largest effect on the cell-free system. This result also indicated that the method established in this study could quickly and easily screen out the optimal OTS corresponding to UNAA.

4. Conclusion

The most critical factor limiting the efficiency of UNAA embedding was the efficiency of OTS components. Therefore, it was important to establish a method to quickly, effectively and conveniently screen out the orthogonal OTS for an unknown UNAA in an acellular system, improve the embedding efficiency of UNAA, and reduce the heavy work of mass screening.

In this study, a linear DNA template-based cell-free unnatural protein synthesis system was proposed for rapid high-throughput screening out the orthogonal translation components and key factors affecting the efficiency of UNAA embedding. In this framework, linear PCR products were added as the template of o-aaRS and o-tRNA. Both ends of the linear PCR product were modified with GC-rich sequences to achieve higher expression rates. The effectiveness of the linear DNA template-based framework has been proved by several experiments. First, a total of 14 cell extracts were selected for screening out cell extract with high expression level. The result showed that EcAR7 ΔAΔSer cell extract was optimal for the cell-free system. After the optimization of the cell-free system, four types of UNAAs and corresponding o-aaRS and o-tRNA were orthogonally screened by the linear DNA template-based framework. From the difference between fluorescent value and rough embedding efficiency, the orthogonality of pPaRS and pPa-tRNA was proved. The corresponding o-aaRS and o-tRNA of pAzF and pAcF were found not to be orthogonal. In addition, a new pair of corresponding o-aaRS and o-tRNA for pBpF was screened out by this linear DNA template-based framework. This framework could also be used to explore the influence of the adding order of OTS components on the system. The results indicated that it was necessary to ensure that all OTS components participated in the reaction at the beginning of the reaction, and the adding order of aaRS had the largest effect on the cell-free system. Compared with other cell-free UNAA incorporation systems [26], this system innovated by using linear DNA templates with protective sequences. This system also successfully realized the screening of efficient OTS components and the optimization of component addition order. The results of these experiments showed that the method established in this study could quickly and easily screen out the optimal OTS components corresponding to the target UNAA. The original long experiment period was shortened to 1–2 days. Based on these advantages, the linear DNA template-based cell-free unnatural protein synthesis system could be applied in the optimization of cell extracts, the screening of OTS components, and the exploration of OTS components' addition order, which could greatly improve the embedding efficiency of UNAAs. This study laid a foundation for accelerating the synthesis and application development of unnatural proteins.

CRediT authorship contribution statement

Xinjie Chen: Methodology, Investigation, Writing – original draft, Writing – review & editing. **Yingying Liu:** Methodology, Investigation, Resources. **Jiaqi Hou:** Methodology, Resources. **Yuan Lu:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgment

This work was supported by the National Key R&D Program of China (2018YFA0901700), National Natural Science Foundation of China (21878173), Beijing Natural Science Foundation (2192023), a grant from the Institute Guo Qiang, Tsinghua University (2019GQG1016), and Laboratory Innovation Fund of Tsinghua University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2021.07.003.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Wang L. Genetically encoding new bioreactivity. N Biotech 2017;38:16–25. https://doi.org/10.1016/j.nbt.2016.10.003.
[2] Des Soye BJ, Patel JR, Isaacs FJ, Jewett MC. Repurposing the translation apparatus for synthetic biology. Curr Opin Chem Biol 2015;28:83–90. https://doi.org/10.1016/j.cbpa.2015.06.008.
[3] Albayrak C, Swartz JR. Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. Nucleic Acids Res 2013;41:5949–63. https://doi.org/10.1093/nar/gkt226.
[4] Hirao I, Kanamori T, Ueda T. Cell-free synthesis of proteins with unnatural amino acids. The PURE system and expansion of the genetic code. Protein eng., vol. 22. Springer Berlin Heidelberg; 2009. p. 271–90. https://doi.org/10.1007/978-3-540-70941-1_10.
[5] Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment-Expanding the chemistry in biology. Chem Sci 2015;6:50–69. https://doi.org/10.1039/c4sc01534g.
[6] Hohsaka T, Sisido M. Incorporation of non-natural amino acids into proteins. Curr Opin Chem Biol 2002;6:809–15. https://doi.org/10.1016/S1367-5931(02)00376-9.
[7] Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem 2010;79:413–44. https://doi.org/10.1146/annurev.biochem.052308.105824.
[8] Lang K, Davis L, Chin JW. Genetic encoding of unnatural amino acids for labeling proteins. Methods Mol Biol 2015;1266:217–28. https://doi.org/10.1007/978-1-4939-2272-7_15.
[9] Klippenstein V, Mony L, Paoletti P. Probing ion channel structure and function using light-sensitive amino acids. Trends Biochem Sci 2018;43:436–51. https://doi.org/10.1016/j.tibs.2018.02.012.

[10] Zheng S, Kwon I. Controlling enzyme inhibition using an expanded set of genetically encoded amino acids. Biotechnol Bioeng 2013;110:2361–70. https://doi.org/10.1002/bit.24911.
[11] Neumann-Staubitz P, Neumann H. The use of unnatural amino acids to study and engineer protein function. Curr Opin Struct Biol 2016;38:119–28. https://doi.org/10.1016/j.sbi.2016.06.006.
[12] Gao W, Cho E, Liu Y, Lu Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. Front Pharmacol 2019;10. https://doi.org/10.3389/fphar.2019.00611.
[13] Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. A general method for site-specific incorporation of unnatural amino acids into proteins. Science 1989;244(80):182–8. https://doi.org/10.1126/science.2649980.
[14] Quast RB, Mrusek D, Hoffmeister C, Sonnabend A, Kubick S. Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. FEBS Lett 2015;589:1703–12. https://doi.org/10.1016/j.febslet.2015.04.041.
[15] Hirao I, Ohtsuki T, Fujiwara T, Mitsui T, Yokogawa T, Okuni T, et al. An unnatural base pair for incorporating amino acid analogs into proteins. Nat Biotechnol 2002;20:177–82. https://doi.org/10.1038/nbt0202-177.
[16] Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia coli. Science 2001;292(80):498–500. https://doi.org/10.1126/science.1060077.
[17] Kleina LG, Masson JM, Normanly J, Abelson J, Miller JH. Construction of Escherichia coli amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. J Mol Biol 1990;213:705–17. https://doi.org/10.1016/S0022-2836(05)80257-8.
[18] Amiram M, Haimovich AD, Fan C, Wang YS, Aerni HR, Ntai I, et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat Biotechnol 2015;33:1272–9. https://doi.org/10.1038/nbt.3372.
[19] Gao W, Bu N, Lu Y. Efficient incorporation of unnatural amino acids into proteins with a robust cell-free system. Methods Protoc 2019;2:1–12. https://doi.org/10.3390/mps2010016.
[20] Seki E, Yanagisawa T, Yokoyama S. Cell-free protein synthesis for multiple site-specific incorporation of noncanonical amino acids using cell extracts from RF-1 deletion E. coli strains. In: Methods mol. Biol. vol. 1728. Humana Press Inc.; 2018. p. 49–65. https://doi.org/10.1007/978-1-4939-7574-7_3.
[21] Katzen F, Chang G, Kudlicki W. The past, present and future of cell-free protein synthesis. Trends Biotechnol 2005;23:150–6. https://doi.org/10.1016/j.tibtech.2005.01.003.
[22] Lu Y. Cell-free synthetic biology: engineering in an open world. Synth Syst Biotechnol 2017;2:23–7. https://doi.org/10.1016/j.synbio.2017.02.003.
[23] Hong SH, Kwon YC, Jewett MC. Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. Front Chem 2014;2. https://doi.org/10.3389/fchem.2014.00034.
[24] Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM. Linear DNA for rapid prototyping of synthetic biological circuits in an escherichia coli based TX-TL cell-free system. ACS Synth Biol 2014;3:387–97. https://doi.org/10.1021/sb400131a.
[25] Wang H, Li J, Jewett MC. Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. Synth Biol 2018;3. https://doi.org/10.1093/synbio/ysy003.
[26] Shrestha P, Smith MT, Bundy BC. Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates. N Biotech 2014;31:28–34. https://doi.org/10.1016/j.nbt.2013.09.002.
[27] Kunjapur AM, Tarasova Y, Prather KLJ. Synthesis and accumulation of aromatic aldehydes in an engineered strain of escherichia coli. J Am Chem Soc 2014;136:11644–54. https://doi.org/10.1021/ja506664a.
[28] Jiang N, Ding X, Lu Y. Development of a robust Escherichia coli-based cell-free protein synthesis application platform. Biochem Eng J 2021;165:107830. https://doi.org/10.1016/j.bej.2020.107830.
[29] Park HS, Hohn MJ, Umehara T, Guo LT, Osborne EM, Benner J, et al. Expanding the genetic code of Escherichia coli with phosphoserine. Science 2011;333(80):1151–4. https://doi.org/10.1126/science.1207203.
[30] Hong SH, Ntai I, Haimovich AD, Kelleher NL, Isaacs FJ, Jewett MC. 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 2014;3:398–409. https://doi.org/10.1021/sb400140t.
[31] Heinemann IU, Rovner AJ, Aerni HR, Rogulina S, Cheng L, Olds W, et al. Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion. FEBS Lett 2012;586:3716–22. https://doi.org/10.1016/j.febslet.2012.08.031.
[32] Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, et al. Genomically recoded organisms expand biological functions. Science 2013;342(80):357–60. https://doi.org/10.1126/science.1241459.
[33] Pirman NL, Barber KW, Aerni HR, Ma NJ, Haimovich AD, Rogulina S, et al. A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation. Nat Commun 2015;6. https://doi.org/10.1038/ncomms9130.
[34] Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. JoVE 2013. https://doi.org/10.3791/50762.
[35] Airen IO. Genome-wide functional genomic analysis for physiological investigation and improvement of cell-free protein synthesis. 2011.
[36] Lehman IR, Roussos GG, Pratt EA. The deoxyribonucleases of Escherichia coli. II. Purification and properties of a ribonucleic acid-inhibitable endonuclease. J Biol Chem 1962;237:819–28. https://doi.org/10.1016/S0021-9258(18)60378-3.
[37] Deiters A, Schultz PG. In vivo incorporation of an alkyne into proteins in Escherichia coli. Bioorg Med Chem Lett 2005;15:1521–4. https://doi.org/10.1016/j.bmcl.2004.12.065.
[38] Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc 2002;124:9026–7. https://doi.org/10.1021/ja027007w.
[39] Chin JW, Martin AB, King DS, Wang L, Schultz PG. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. Proc Natl Acad Sci U S A 2002;99:11020–4. https://doi.org/10.1073/pnas.172226299.
[40] Chen X, Lu Y. In silico design of linear DNA for robust cell-free gene expression. Front Bioeng Biotechnol 2021;9. https://doi.org/10.3389/fbioe.2021.670341.
[41] Sawyer N, Gassaway BM, Haimovich AD, Isaacs FJ, Rinehart J, Regan L. Designed phosphoprotein recognition in escherichia coli. ACS Chem Biol 2014;9:2502–7. https://doi.org/10.1021/cb500658w.