Characterizing and Improving pET Vectors for Cell-free Expression

Kara Jew¹, Philip E. J. Smith², Byungcheol So², Jillian Kasman², Javin P. Oza²* and Michael W. Black¹*

¹Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA, United States, ²Chemistry & 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

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.,...)
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; Kambhati 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 polyhistidine 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 MCI061 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-derived 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 MCI061 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 fmolues 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 MgCl₂, 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
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 #180755), pET30-T7-His/S-sfGFP (Addgene #180756), and pET30-T7-sfGFP (Addgene #180757). Graphic representations of the expression cassettes in each of the four plasmids are shown in Figure 1A. 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 T7-encoded N-terminal His & S tags (pET30-T7-His/S-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 was compared to the preferred sequences (pET30-T7-sfGFP). For the NEB Express system, the removal of lacO alone (pET30-T7-lacO-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

**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 | Tₘ | Annealing |
|------------------|----|----------|
| pET30-T7-sfGFP   |    |          |
| Insert: T7-Pro-Gib-F | cgcgaaattataataacgactcaactatagg | 59°C | 63°C |
| Insert: T7-Term-Gib-R | ctttacagcgaagaaacccccctcaag | 56°C | 63°C |
| Vector: T7-Pro-Gib-R | ctggagggtttttgtcgttaag | 56°C | 59°C |
| Vector: pET30-T7-lacO-sfGFP | ctatagtctatctcttatattcttgagg | 58°C |       |
| Insert: RBS-sfGFP-F | ctttgcagtaagggagatatacatag | 56°C | 59°C |
| Insert: T7-Term-Gib-R | ctttacagcgaagaaacccccctcaag | 56°C | 63°C |
| Vector: T7-Term-Gib-F | ctggagggtttttgtcgttaag | 56°C | 63°C |
| Vector: pET30-T7-His/S-sfGFP | ctttgaggggttttttgctgaaag | 56°C | 63°C |
| Insert: pET30-RBS-long-R | ctttgaggggttttttgctgaaag | 56°C | 63°C |
| Vector: pET30-T7-lacO-His/S-sfGFP | ctttgaggggttttttgctgaaag | 56°C | 63°C |
| Insert: N-tag-sfGFP-F | ctttgaggggttttttgctgaaag | 56°C | 63°C |
| Insert: T7-Term-Gib-R | ctttacagcgaagaaacccccctcaag | 56°C | 63°C |
| Vector: T7-Term-Gib-F | ctggagggtttttgtcgttaag | 56°C | 63°C |
| Vector: pET-No-Cut-DIC-R | ctggagggtttttgtcgttaag | 56°C | 63°C |
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 expression cassette included the deletion of lacO, 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.

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

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 expression cassette included the deletion of lacO, 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.

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 expression cassette included the deletion of lacO, 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.

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 expression cassette included the deletion of lacO, 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.

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 expression cassette included the deletion of lacO, 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.
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-Truettmannsdorff, T., et al. (2022). Differentially Optimized Cell-free Buffer Enables Robust Expression from Unprotected Linear DNA in Eucaryoic-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-Proparglyglyoxylphenylalanine 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 Polyster: A 100-Year Journey from Polyhydroxyalkanoates to Non-natural Microbial Polysters. 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 Reuzel, 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 of Start Codon, Nucleic Acids Res. 42, 2646–2659. doi:10.1093/nar/gkt1139

Espah Borujeni, A., and Salis, H. M. (2016). Translation Initiation Is Controlled by mRNA Structures Overlap with the Ribosomal Footprint in N-Terminal Coding Sequences. Nucleic Acids Res. 45, 5437–5448. doi:10.1093/nar/gkw061

Frontiers in Bioengineering and Biotechnology | www.frontiersin.org 5 June 2022 | Volume 10 | Article 895069

**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
