Synthetic Biology Bicistronic Designs Support Gene Expression Equally Well *in vitro* and *in vivo*

Owen Koucky^a, Jacob Wagner^b, Sofia Aguilera^b, Benjamin Bashan^b, Queena Chen^a, Anthony Eckdahl^b, Elise Edmart, Paul Gomez^a, Nick Hanlan^b, Nick Kempf^d, Devin Mattoon^d, Sam McKlin^a, Christopher Mazariegos^a, Alex Morehead^d, Shi Qing Ong^b, Andy Peterson^e, Maria Rojas^a, Kyla Roland^b, Kaitlyn Schildknecht^d, Haley Seligmann^e, Kaden Slater^d, Ali Tauchen^b, Raechel Tittor^b, Tatianna Travieso^a, Dannie Urban^d, Caroline Willis^a, John Zhou^a, Nicole L. Snyder^e, Laurie J. Heyer^e, Jeffrey L. Poet^a, Todd T. Eckdahl^b, & A. Malcolm Campbell^{bat}

^aDepartment of Biology, Davidson College, Davidson, NC ^bDepartment of Biology, Missouri Western State University, St. Joseph, MO ^cDepartment of Mathematics and Computer Science, Davidson College, Davidson, NC ^dDepartment of Computer Science, Math, and Physics, Missouri Western State University, St. Joseph, MO ^cDepartment of Chemistry, Davidson College, Davidson, NC

https://doi.org/10.33697/ajur.2020.012

Students: owkoucky@davidson.edu, jwagner3@missouriwestern.edu, saguileratovar@missouriwestern.edu, bbasham@missouriwestern.edu, quchen@davidson.edu, aneckdahl@davidson.edu, eledman@davidson.edu, pagomez@davidson.edu, nhanlan@missouriwestern.edu, nkempf1@missouriwestern.edu, dmattoon@missouriwestern.edu, samcklin@davidson.edu, chmazariegos@davidson.edu, amorehead1@missouriwestern.edu, song@missouriwestern.edu, apeterson5@missouriwestern.edu, marojas@davidson.edu, kschildknecht@missouriwestern.edu, haseligmann@davidson.edu, kslater3@missouriwestern.edu, atauchen1@missouriwestern.edu, rtittor@missouriwestern.edu, tatravieso@davidson.edu, durban1@missouriwestern.edu, cawillis@davidson.edu, jzbou@davidson.edu Mentors: nisnyder@davidson.edu, laheyer@davidson.edu, poet@missouriwestern.edu, eckdahl@missouriwestern.edu, macampbell@davidson.edu*

ABSTRACT

Synthetic biology integrates molecular biology tools and an engineering mindset to address challenges in medicine, agriculture, bioremediation, and biomanufacturing. A persistent problem in synthetic biology has been designing genetic circuits that produce predictable levels of protein. In 2013, Mutalik and colleagues developed bicistronic designs (BCDs) that make protein production more predicable in bacterial cells (*in vivo*). With the growing interest in producing proteins outside of cells (*in vitro*), we wanted to know if BCDs would work as predictably in cell-free protein synthesis (CFPS) as they do in *E. coli* cells. We tested 20 BCDs in CFPS and found they performed very similarly *in vitro* and *in vivo*. As a step toward developing methods for protein production in artificial cells, we also tested 3 BCDs inside nanoliter-scaled microfluidic droplets. The BCDs worked well in the microfluidic droplets, but their relative protein production levels were not as predictable as expected. These results suggest that the conditions under which gene expression happens in droplets result in a different relationship between genetic control elements such as BCDs and protein production than exists in batch CFPS or in cells.

KEYWORDS

Bicistronic Design; Synthetic Biology; Cell-Free Protein Synthesis; Microfluidics

INTRODUCTION

Synthetic biology is a new field that integrates molecular biology, engineering, chemistry, mathematics and computer science.¹ Synthetic biologists have invented many useful devices that produce medicines^{2,3,4}, improve agriculture,⁵ achieve bioremediation⁶ and make biofuels or chemical commodities through biomanufacturing.⁷ Despite these successes, a persistent challenge in synthetic biology has been consistent and reliable protein production. To address this challenge, a team of synthetic biologists invented bicistronic designs (BCDs) that produce predictable levels of protein in *E. coli*.⁸ BCDs are encoded in DNA but function as part of mRNA. A BCD consists of two segments of DNA called cistrons that encode polypeptides. The first cistron begins with a ribosomal binding site (RBS) followed by a start codon that marks the beginning of a coding segment for a non-functional leader polypeptide of 16 amino acids. The stop codon for the coding segment overlaps by one base pair with the start codon for the second cistron (see **Figure 4A**). Upstream of this start codon, and within the 48 bases encoding the leader polypeptide, is an RBS element that directs the production of a protein from the second cistron. This second RBS was the focus of the research by Mutalik and colleagues. They varied the strength of the second RBS and the amount of protein translated varied accordingly. The authors speculated that during translation of the first polypeptide, the large ribosome and its helicase activity eliminated folded mRNA structures formed by base pairing between the second RBS and the beginning of the coding sequence in the second cistron. Thus, the second RBS was allowed to base pair with the 16S rRNA of the ribosome without interference from mRNA folding and a predictable amount of two different reporter proteins was produced with each of the RBSs tested.⁸

Although cells are very good at producing proteins, cells often cannot produce robust amounts of orthogonal proteins and sometimes cannot produce them at all.⁹ For example, *E. coli* will not produce antimicrobial proteins that kill bacteria.¹⁰ Therefore, synthetic biologists sometimes switch from *in vivo* production of proteins to *in vitro* production.^{11,12,13} The purpose of our research was to determine if the BCDs described by Mutalik and colleagues would work as predictably *in vitro* as they do *in vivo*. To conduct this work, we chose to develop an affordable cell-free protein synthesis (CFPS) protocol instead of purchasing expensive premade mixtures. Once we had a functional CFPS protocol, we tested 20 of the BCDs made by Mutalik and colleagues during CFPS. As a step toward developing methods for protein production in artificial cells, we also compared protein production directed by 3 BCDs in CFPS droplets to batch CFPS reactions with the same 3 BCDs.

METHODS AND PROCEDURES

The protocols for CFPS and microfluidics are too extensive to give all the details in this paper, so we have compiled all methods into the appendix. However, we summarize the main steps below. Much of our method has been adapted from multiple protocols.^{14,15,16,17}

Cell-free Protein Synthesis for Batch Reactions and Protein Quantification

E. coli encoding the T7 RNA polymerase (BL21 (DE3) *E. coli* cells; NEB # C2527H) were grown in LB broth supplemented with IPTG to induce T7 RNA polymerase expression. Cells were grown overnight and harvested by centrifugation. Cells were lysed using an AVESTIN EmiulsiFlex B-15 high-pressure homogenizer, also known as a French press. Cytoplasmic lysate was aliquoted and stored at -80 °C for later use. The CFPS reaction conditions were as follows: 55 mM PIPES-KOH (pH 7.0); 14 mM magnesium acetate; 50 mM potassium acetate; 155 mM ammonium acetate; 1.5% PEG-6000; 40 mM 3-phosphoglycerate; amino acid cocktail with all 20 naturally occurring protein amino acids at 2.5 mM each; 1.2 mM ATP; 1.0 mM GTP; 0.8 mM UTP; DNA template at 10 ng/ μ L; and 2.5 μ L *E. coli* cell lysate for a final reaction volume of 25 μ L. We have an Excel template file we use to make the reaction mixtures and this tool is part of the appendix.

Once the ingredients are mixed, we placed the $25 \,\mu$ L aliquots into a 384-well plate and incubated it at the desired temperature for variable lengths of time, as indicated. Plates containing CFPS were scanned at the appropriate wavelength for GFP (excitation 485 nm, emission 515 nm) to quantify the relative fluorescence. The fluorometer was programmed to measure the wells at regular intervals and incubate the plate at the desired temperature between measurements.

Cell-free Protein Synthesis in Microfluidic Droplets

We used the OB1 MK3 Microfluidic Flow Control System from Elveflow to generate droplets for CFPS. CFPS reactions for use in microfluidic droplets were prepared as described above for *in vitro* batch reactions. For each reaction, the aqueous (dispersive phase) reservoir for microfluidic droplet generation was filled with 30 µL of this CFPS mixture, 10 µL of nuclease-free water, and 10 µL containing 500 ng of plasmid DNA for BCD9, BCD19, or BCD24 gene expression constructs (or 10 µL of water for the no DNA negative control). The aqueous reservoir was vortexed for about 5 seconds. For each reaction, the oil (continuous phase) reservoir was filled with 50 µL droplet stabilization oil (Droplet Genomics cat. DG-DSO-20). Both the dispersive and continuous channels were set on the OB1 MK3 to 600 mbar and droplets were collected in a 1.5 ml tube. For each reaction, a volume of 50 µL of droplets was loaded into a 384 well microtiter plate for use in the fluorometer. Positive control batch reactions with 100% CFPS mixture and 75% CFPS mixture were also loaded into the plate. The fluorometer was set to 37 °C and GFP expression was measured after 3 hours using excitation and emission wavelengths of 485 nm and 528 nm respectively.

RESULTS

Producing the *E. coli* cell lysate went smoothly but optimizing CFPS took considerable effort (**Figure 1**). The first challenge was making the amino acid cocktail without precipitating any of the components. The original protocol called for neutralizing the acidic pH of the mixtures, but the only way we could get CFPS to work was when we did not adjust the pH of the amino acid solution (**Figure 1A**).¹⁸ We tested a range of temperatures and found that 30 °C was optimal, but we used 37 °C for potential applications using different proteins. The literature reported the use of a range of PEG-6000 concentrations,¹⁶ but we chose 1.5% as a compromise between maximum protein production (**Figure 1C**) and fastest induction time (**Figure 1D**), defined as the time at which GFP production was twice the background level.



Figure 1. Characterization of CFPS. A. Varying preparations of the 20-amino acid solution for use in the CFPS reaction mixture by adjusting the pH to 7.4, omitting the pH adjustment step, or raising the pH by increasing KOH concentration to 1 M resulted in variable CFPS performance. B. Characterization of CFPS performance across a range of incubation temperatures. C. CFPS reactions with varying concentrations of PEG-6000 (% v/v) affected CFPS output. D. CFPS induction time also varied with PEG-6000 concentrations. CFPS performance verification experiments used the March lysate while PEG-6000 optimization experiments used the July lysate. Error bars represent ± 1 S.E (n=9).



Figure 2. Comparison of cell lysate preparations. A, B. July and September lysates were used in CFPS reactions with six buffers that varied from pH 6.5 to 8.0: PIPES-KOH (pH = 6.5); PIPES-KOH (pH = 7.0); HEPES-KOH (pH = 7.0); PIPES-KOH (pH = 7.5); HEPES-KOH (pH = 7.5); and HEPES-KOH (pH = 8.0). C. CFPS induction time compared across the six buffers and two lysate treatments. Error bars represent ± 1 S.E (n=9).

An unexpected complication was optimizing the pH of the CFPS reaction mixture. We found that each lysate preparation had its own optimal buffer and optimal pH (**Figure 2**). For example, the lysate produced in July worked best in HEPES-KOH pH 7.0 whereas the September lysate worked best in PIPES-KOH pH 7.0. We do not have an explanation for these observations. Induction times for GFP production also varied depending on the buffer and pH employed. For most experiments, we chose PIPES-KOH pH 7.0 as our standard reaction condition since this buffer was incorporated in functional CFPS across multiple lysate preparations.

We were curious about why all the time-course fluorescence readings in **Figures 1 and 2** began with a decline in fluorescence. We hypothesized that CFPS did not commence immediately and that some chemical reaction that altered the pH had to take place first. We suspected the pH might need to be closer to 7.4, so we withheld the DNA for variable amounts of time from reactions incubated at 37 ° C and measured the pH during the entire experiment (**Figure 3**). If the CFPS reaction is allowed to pre-incubate for 15 minutes, then the final protein output is substantially greater. It was striking to us that whether the reaction contained DNA or not, the pH gradually fell from about 7.9 to near 7.0. The decrease in fluorescence between pH measurements was shown to be temperature-dependent because it diminished as we reduced the incubation temperature within the fluorometer. As described in the literature, when we added maltose to a final concentration of 12 mM after 2 hours of reaction, we doubled the final protein output.¹⁸



time since incubation initiation (hh:mm:ss)

Figure 3. Analysis of CFPS GFP production and pH changes. CFPS reaction mixture pre-DNA incubation times are indicated in the legend. Dotted sections of the graph represent the duration of measurement periods to account for the time the CFPS reactions spent out of the plate reader during pH sampling. CFPS employed the July lysate. Error bars represent ± 1 S.E (n=9).

Once we had optimized CFPS, we tested 20 of the BCDs from Mutalik and colleagues (**Figure 4**).⁸ We ranked the GFP output from highest (BCD19) to lowest (BCD22). When we compared our GFP fluorescence levels from **Figure 4** to those from the Mutalik paper with Spearman's Rank Order Correlation, we found a high degree of correlation (r = 0.88). From the data in Figure 4B and the high degree of concordance with Mutalik's data, we conclude that the BCDs first developed for *in vivo* use work similarly *in vitro*. This is the first publication showing BCDs function *in vitro*.



Figure 4. CFPS using bicistronic designs A. Diagram of BCD components as part of a gene expression cassette that also includes a T7 promoter and a GFP reporter gene (not drawn to scale). B. Fluorescence output after three hours of CFPS reactions with 20 different BCDs, a strong RBS positive control, and a no DNA negative control. Error bars represent \pm 1 S.E (n=9). The 3 boxed BCDs were used for droplet CFPS in **Figure 5**.

After establishing the functionality of BCDs in CFPS batch reactions, we explored their potential for use in nanoliter-scale microfluidic droplets (Figure 5A). To realize our goal of using microfluidic droplets for metabolic engineering will require further exploration of the function of gene regulatory elements such as BCDs in nanoscale droplets. Through experimentation, we discovered the normal CFPS reaction mixture was too viscous to produce reliable droplets, so we diluted the droplet reaction mixture with water to 75% of the concentration of CFPS batch reactions (Figure 5B). We tested three of the BCDs (19, 9 and 24) from Figure 4 that represent a range of strengths. We also compared 75% reaction mixtures in batch reactions and droplet CFPS. The 100% CFPS batch reactions performed as we expected, though the 75% pooled CFPS varied a bit more than we had predicted. However, we were surprised to see that 75% concentration CFPS within the droplets outperformed the same concentration in pooled CFPS. We observed that the measured collection of droplets was about 75% oil by volume and 25% aqueous CFPS by volume. Therefore, the overall observed output of GFP from CFPS in droplets is actually 4-fold higher than is represented in the batch results of Figure 4. We offer two possible explanations for this result. One is that the reflective nature of the oil and the geometry of the droplets enhances GFP detection, and the other is that CFPS happened more efficiently in droplet CFPS than in CFPS batch reactions. We also found that in droplet CFPS, the relative fluorescence outputs from BCD19 and 9 are reversed compared to the results in vivo or in CFPS batch reactions in vitro. These results suggest that the reaction conditions under which gene expression happens in droplets result in a different relationship between BCDs and protein production than exists in batch CFPS or in cells. Due to the complexity of microfluidic droplet formation, there also could be other reasons for these unexpected results.



Figure 5. CFPS using 3 BCDs in microfluidic droplets. A. The left micrograph shows a perpendicular junction PDMS droplet-generation microchip with labels that indicate the flow of the aqueous CFPS dispersive phase, the oil continuous phase, and droplets immersed in oil. The right micrograph was taken on a fluorescence microscope and shows various levels of GFP produced by CFPS. B. Fluorescence output after three hours with BCD19, BCD9, and BCD24 using 100% CFPS solutions in pooled reactions, 75% CFPS solutions in pooled reactions, 75% CFPS solutions in populations of microfluidic droplets, and no DNA (DNA-) negative controls. Error bars represent \pm 1 S.E (n=9).

CONCLUSIONS

In this paper, we describe a less expensive approach to CFPS using cell lysate produced locally, as spelled out in the detailed protocols of the appendix (**Figures 1 – 3**). The goal of this research was to determine if BCDs performed similarly *in vitro* and *in vivo*. As shown by the results presented in **Figure 4** and **Figure 5**, our Spearman's Rank Order Correlation calculations show the BCDs function equivalently in pooled CFPS and droplet CFPS. Furthermore, we found that GFP detection is amplified in droplets, either because the reflection of light enhanced the GFP signal detected in a droplet or oil mixture or CFPS in droplets was more efficient. Although the BCDs worked well in the microfluidic droplets, their relative expression levels were not as consistent as expected. These complex results are not surprising given that Figure 5 is the first published example of BCD gene expression in microfluidic droplets. Additional research will be required to determine if CFPS with BCDs is more efficient or if

GFP light reflection produces enhanced light detection in the absence of more protein production. Furthermore, more research will be required to determine the source of droplet variability for GFP production.

ACKNOWLEDGMENTS

Funding for this multi-year research project was provided by NSF RUI grants MCB-1613203 to Davidson College and MCB-1329350 to Missouri Western State University, and the James G. Martin Genomics Program.

REFERENCES

- 1. Campbell, A. M. (2005) Meeting Report: Synthetic Biology Jamboree for Undergraduates, *Cell Biology Education* 4(1), 19–23. *doi:10.1187/cbe.04-11-0047*
- Denby, C. M., Li, R. A., Vu, V. T., Costello, A., Lin, W., Chan, L. J. G., Williams, J., Donaldson, B., Bamforth, C. W., Petzold, C. J., Scheller, H. V., Martin, H. G., Keasling, J. D. (2018) Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer, *Nat Commun* 9, 965. *https://doi.org/10.1038/s41467-018-03293-x*
- Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast, *Nature* 440(7086), 940–943. *https://doi:10.1038/nature04640*
- Bourdeau, R. W., Lee-Gosselin, A., Lakshmanan, A., Farhadi, A., Kumar, S. R., Nety, S. P., and Shapiro, M. G. (2018) Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts, *Nature* 553(7686), 86–90. https://doi:10.1038/nature25021
- 5. Liu, W., and Stewart, C. N. (2015) Plant synthetic biology, *Trends in Plant Science* 20(5), 309–317. https://doi:10.1016/j.tplants.2015.02.004
- 6. Karig, D. K. (2017) Cell-free synthetic biology for environmental sensing and remediation, *Current Opinion in Biotechnology* 45, 69–75. *https://doi:10.1016/j.copbio.2017.01.010*
- 7. Clomburg, J. M., and Gonzalez, R. (2010) Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology, *Applied Microbiology and Biotechnology* 86(2), 419–434. https://doi:10.1007/s00253-010-2446-1
- Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements, *Nature Methods* 10(4), 354–360. *https://doi:10.1038/nmeth.2404*
- 9. Ruiz, R. C. H., Kiatwuthinon, P., Kahn, J. S., Roh, Y. H., and Luo, D. (2012) Cell-Free Protein Expression from DNA-Based Hydrogel (P-Gel) Droplets for Scale-Up Production, *Industrial Biotechnology* 8(6), 372–377. *https://doi:10.1089/ind.2012.0024*
- Yue, S., Jie, J., Xie, L., Li, Y., Zhang, J., Lai, X., Xie, J., Guo, X., and Zhai, Y. (2020) Antimicrobial peptide CAMA-syn expressed in pulmonary epithelium by recombination adenovirus inhibited the growth of intracellular bacteria, *The Journal of Gene Medicine* 22(3). *https://doi:10.1002/jgm.3149*
- Lagassé, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., and Kimchi-Sarfaty, C. (2017) Recent advances in (therapeutic protein) drug development. F1000Research, 6, 113. https://doi:10.12688/f1000research.9970.1
- 12. Lim, H. J., Park, Y. J., Jang, Y. J., Choi, J. E., Oh, J. Y., Park, J. H., Song, J. K., and Kim, D.-M. (2016) Cell-free synthesis of functional phospholipase A1 from Serratia sp., *Biotechnology for Biofuels* 9(1). *https://doi:10.1186/s13068-016-0563-5*
- Merlin, M., Gecchele, E., Capaldi, S., Pezzotti, M., and Avesani, L. (2014) Comparative Evaluation of Recombinant Protein Production in Different Biofactories: The Green Perspective, *BioMed Research International* 2014, 1–14. https://doi:10.1155/2014/136419
- 14. Krinsky, N., Kaduri, M., Shainsky-Roitman, J., Goldfeder, M., Ivanir, E., Benhar, I., Shoham, Y., and Schroeder, A. (2016) A simple and rapid method for preparing a cell-free bacterial lysate for protein synthesis, *PLOS ONE* 11(10), e0165137. *https://doi.10.1371/journal.pone.0165137*
- Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., and Yokoyama, S. (2004) Preparation of Escherichia coli cell extract for highly productive cell-free protein expression, Journal of Structural and Functional Genomics 5(1/2), 63–68. https://doi:10.1023/bijsfg.0000029204.57846.7d
- Hino, M., Kataoka, M., Kajimoto, K., Yamamoto, T., Kido, J.-I., Shinohara, Y., and Baba, Y. (2008) Efficiency of cell-free protein synthesis based on a crude cell extract from *Escherichia coli*, wheat germ, and rabbit reticulocytes, *Journal of Biotechnology* 133(2), 183–189. *https://doi:10.1016/j.jbiotec.2007.08.008*
- 17. Kim, T.-W., Kim, D.-M., and Choi, C.-Y. (2006) Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system, *Journal of Biotechnology* 124(2), 373–380. *https://doi:10.1016/j.jbiotec.2005.12.030*
- Caschera, F., and Noireaux, V. (2015) Preparation of amino acid mixtures for cell-free expression systems, *BioTechniques* 58(1). *https://doi:10.2144/000114249*

ABOUT STUDENT AUTHORS

The team of student authors were all undergraduates at the time they conducted the research. Some worked at Missouri Western State University (MWSU) whereas others worked at Davidson College (DC). They were funded by a pair of collaborative grants

that allowed them to visit each other's campuses during the summer of their research. Owen Koucky was a genomics major whose senior thesis focused on optimizing CFPS. Students at MWSU specialized in droplet production and droplet CFPS. Students at DC specialized in pooled CFPS. Students on both campuses conducted the DNA cloning.

PRESS SUMMARY

This synthetic biology study examines the reliability of protein production performed *in vitro*. Specifically, we compared the function of 20 bicistronic designs (BCDs) developed for use inside *E. coli* cells to their function in cell-free protein synthesis. We found that BCDs functioned similarly *in vitro* and *in vivo*. When tested within nanoliter scaled droplets produced by microfluidics, the BCDs also performed well, though GFP detection varied. In summary, BCDs can be used interchangeably for *in vivo* and *in vitro* protein production.