APPLICATION NOTE

Arzeda's next-generation protein design workflow powered by the **myTXTL®** cell-free protein expression system enables smarter and faster discovery

Will Skyrud[1], James Leinas[1], Brendan Young[1], Joshua Rhea[1], Maria Rottersman[1], Kenzie Clark[1], Ashwini Devkota[1], Benjamin P. Steil[2], Evelyn Eggenstein[2], Gang Wang[2]

[1] Arzeda, Seattle, WA 98119
[2] Daicel Arbor Biosciences, Ann Arbor, MI 48103

ABSTRACT

Increased demand for biobased solutions for industrial manufacturing has brought synthetic biology to the forefront of modern biotechnology. Success in synthetic biology is driven by the design, build and test cycle. Arzeda's protein design technology is uniquely situated to provide next-generation proteins and enzyme products that are fundamental to biomanufacturing and beyond. The Daicel Arbor Biosciences myTXTL® cell-free protein expression system is tailored to high throughput synthetic biology testing regimes and has the potential to significantly accelerate and improve Arzeda's protein expression platform. To validate this potential, a diverse set of enzymes was expressed and assayed for their desired biocatalytic activity. The results demonstrate that the myTXTL system can produce a wide range of biocatalysts with diverse functions, sizes, and cofactor requirements in a much shorter amount of time than traditional cloning and *in vivo* expression. The success of this system, as well as its availability in a bulk format, all-in-one master mix, make the Daicel Arbor myTXTL cell-free expression kit directly applicable to high throughput testing of protein designs leading to a more efficient design, build, and test cycle for synthetic biology.

INTRODUCTION

Arzeda's computational protein design methodology employs a design-build-test-learn workflow to produce next generation biocatalysts (Fig. 1). High throughput (HT) laboratory strategies are employed to streamline this process and provide the necessary data for machine learning algorithms and future design. As computational capabilities have expanded, the physical production of protein via traditional *E. coli in vivo* methods has become a significant bottleneck. Cell-free protein synthesis (CFPS) systems that can perform *in vitro* transcription-translation (TXTL) have emerged as a low-cost, high-speed alternative to *E. coli-based* or mammalian cell-based expression systems but have suffered from lack of a robust source of off-the-shelf *in vitro* expression material.[1,2] The introduction of the Daicel Arbor Biosciences myTXTL cell-free protein expression system has provided a robust and easy-to-use system that increases the confidence of large-scale utilization of CFPS in a HT setting (Fig. 1). In this study, a diverse set of biocatalysts were expressed and assayed for activity to determine if adoption of the myTXTL kit is viable for companies like Arzeda that work on a wide range of protein scaffolds. The possible material and time savings were also considered in the context of a protein design workflow.

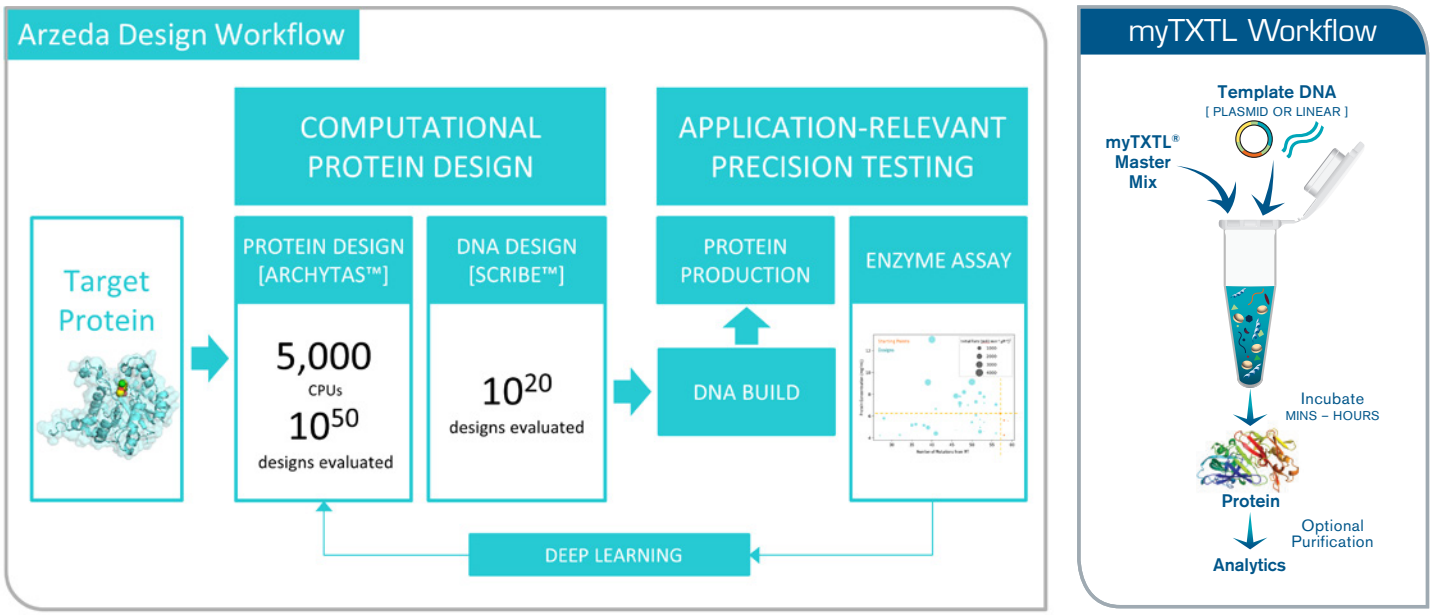


Figure 1. Depictions of the typical workflows for Arzeda protein design and myTXTL protein production.

The myTXTL system was chosen for this work because it has a unique applicability to HT synthetic biology workflows. Most notably, it is provided as an all-in-one master mix in bulk formats. This vastly reduces the handling steps needed and potential for operational errors compared to kits that require mixing small volumes of multiple, sensitive components prior to expression. Its miniaturization capabilities have also been previously validated with successful protein expression in volumes as low as 1 uL in 384-well plates.[3] Comparing this to widely employed *E. coli in vivo* expression workflows, there are significant advantages to the myTXTL system due to labor savings, multiplexed expression, turnaround time, and reduced consumables use/cost, resulting in a lower environmental footprint (Fig. 2).

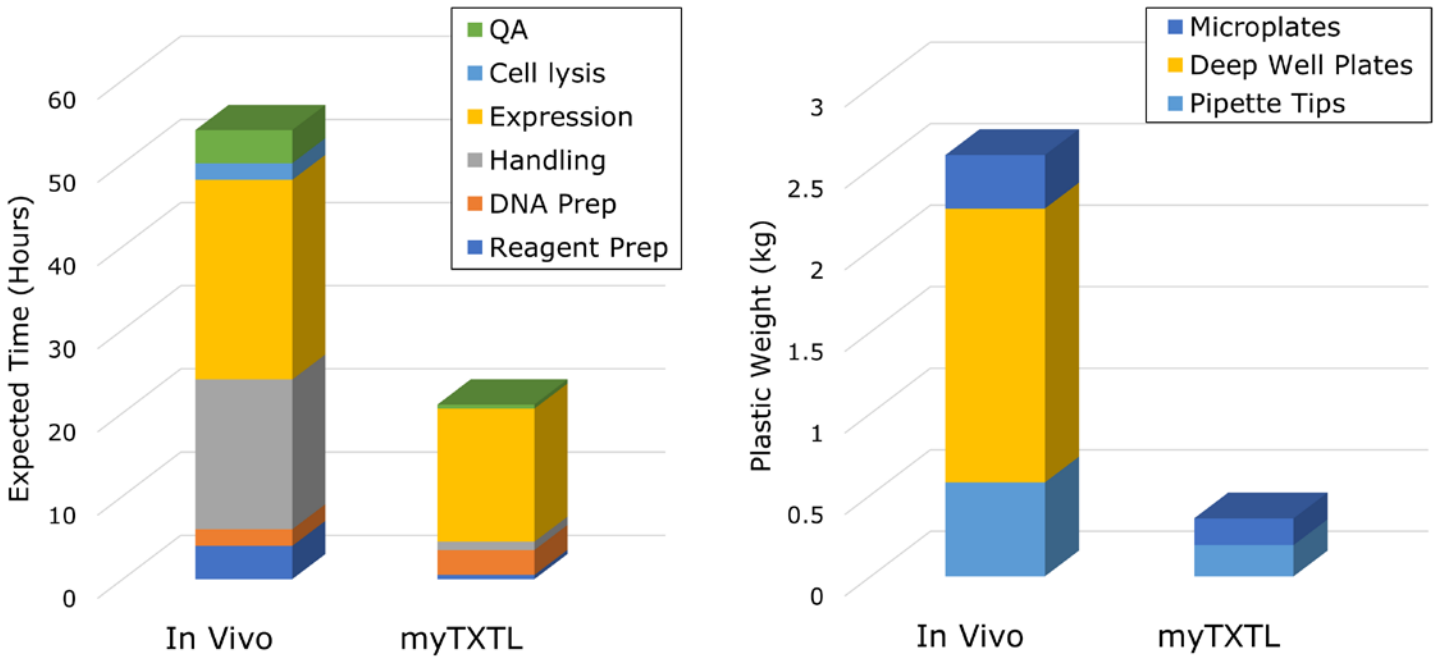


Figure 2. Comparison of protein expression time and consumable usage between the traditional *E. coli in vivo* expression and the myTXTL *in vitro* system. Representative time and material for expression of 4 x 96 well plates.

Establishment of a faster, more miniaturized test cycle via the myTXTL system increases the total number of samples tested while simultaneously decreasing the experimental time. This has direct effects on the success of projects that rely on computational protein design. First, it reduces the total time required in the test phase of the design-build-test-learn workflow by providing data on a tighter timeline. Second, the throughput capability significantly increases the number of designs that can be tested and learned from within the same timeframe (Fig. 3). This will allow synergistic strategies to be employed in parallel, dramatically decreasing the time required to deliver a new protein or enzyme.

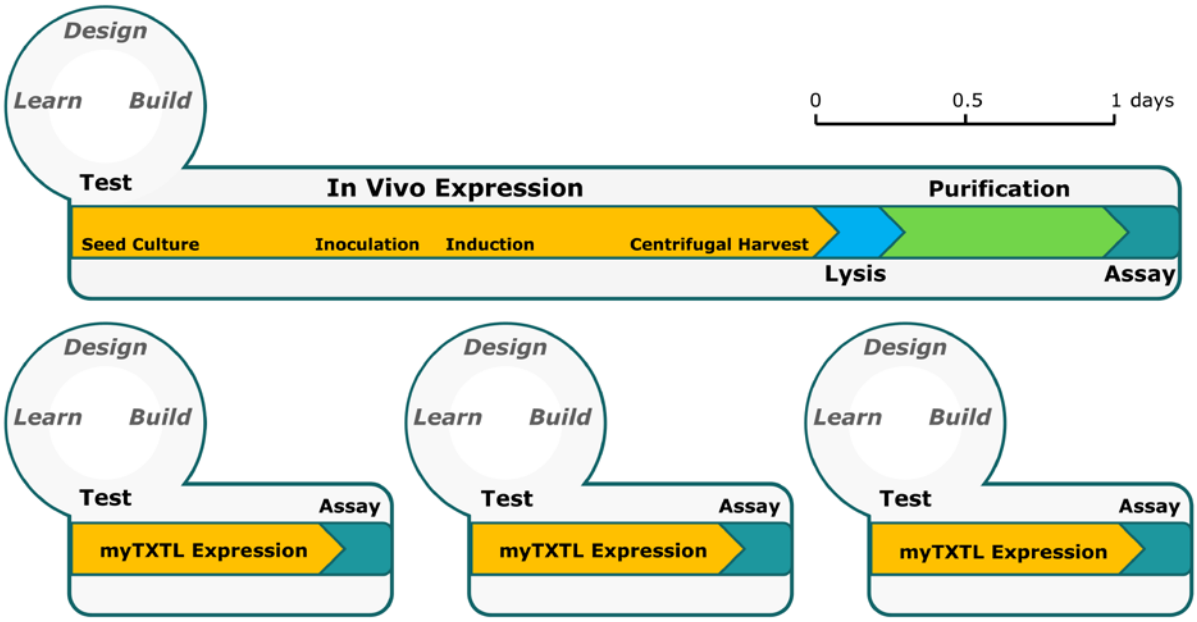


Figure 3. Total testing (expression/isolation/assay) time comparison of a generic *E. coli in vivo* expression workflow vs. the myTXTL cell-free expression system demonstrating the potential for 3x improvements in turnaround time using myTXTL.

To facilitate the adoption of the myTXTL system as an integral part of Arzeda's HT testing regime it is necessary to demonstrate that the myTXTL system is successful at producing a wide range of enzymes that have the desired catalytic activity. Additionally, with the wish to assay directly in the myTXTL background, it is important to verify its compatibility with fluorometric, colorimetric and LC-MS detection methods. To accomplish this, 25 different enzymes were tested for expression and catalytic activity as well as two green fluorescent protein (GFP) variants and two red fluorescent protein (mCherry) variants (Fig. 4,5) for a total of 29 proteins tested. The results suggest that the myTXTL system is rapid, high performing, and yields enough active protein to facilitate HT testing of designed protein variants over a wide range of enzyme commission (EC) numbers. Additionally, its easy-to-use format proves the proposed improvements (Fig. 2,3) over *E. coli-based in vivo* expression are achievable as outlined.

METHODS

I. DNA Preparation

The plasmids of interest were cloned into *E. coli* DH5α and grown up in liquid culture. Plasmids were purified in a 96-well miniprep kit format and then were cleaned and concentrated using a 96-well PCR clean up kit. Final DNA concentrations ranged from 30-80 ng/uL and were not normalized prior to expression.

II. myTXTL Cell-Free Expression

Protein expression was carried out according to the Daicel Arbor Biosciences myTXTL protocol using the Sigma 70 Master Mix kit. Since coding sequences were under control of a standard inducible T7 system, the Daicel Arbor T7RNAP Helper Plasmid (HP) was added as well as isopropyl β-D-1-thiolgalactopyranoside (IPTG). The final components of each expression well were: 9 uL myTXTL Sigma 70 master mix, 2.5 uL plasmid DNA template, 0.012 uL 1M IPTG (1 mM final), 0.5 uL T7 RNAP HP (0.1 nM). An ExpressMix was built with all the components except the DNA. The DNA was aliquoted to a v-bottom 96-well plate and the ExpressMix was added to give a final volume of 12 uL. Plates were covered and incubated at 29°C for 16 hours. No specific cofactors were added during expression.

III. Fluorescent Quality Assurance (QA) and Enzyme assays

Upon completion of expression GFP and mCherry reactions were diluted and their relative fluorescence units were measured. Expressed enzymes were diluted as necessary and assayed according to Arzeda in-house validated protocols. All necessary cofactors were added at the time of assay.

RESULTS AND DISCUSSION

The goal of this study was to independently determine if the myTXTL system can successfully produce active protein regardless of scaffold identity, and if the kit background would limit the type of detection method that could be employed. Test enzymes were chosen that rely on either colorimetric, fluorometric, or LC-MS detection. These enzymes require a suite of cofactors for activity and range in molecular weight. In total, 25 different enzymes were tested (Fig. 5). These comprised 8 separate types of enzymes with 2-4 representatives per type spanning EC1-4. Wild type as well as Arzeda designed enzymes were tested. The percentage of identity within each enzyme type ranged from 0-91% with an average of 47% identity within types. Additionally, GFP and mCherry were included to be validated for future QA strategies. In all cases plasmid DNA was used under a typical IPTG inducible T7 promoter. Since this is the same DNA type used in *E. coli in vivo* expression, it demonstrates the potential for the myTXTL system to be directly incorporated into any current workflow without altering the DNA format.

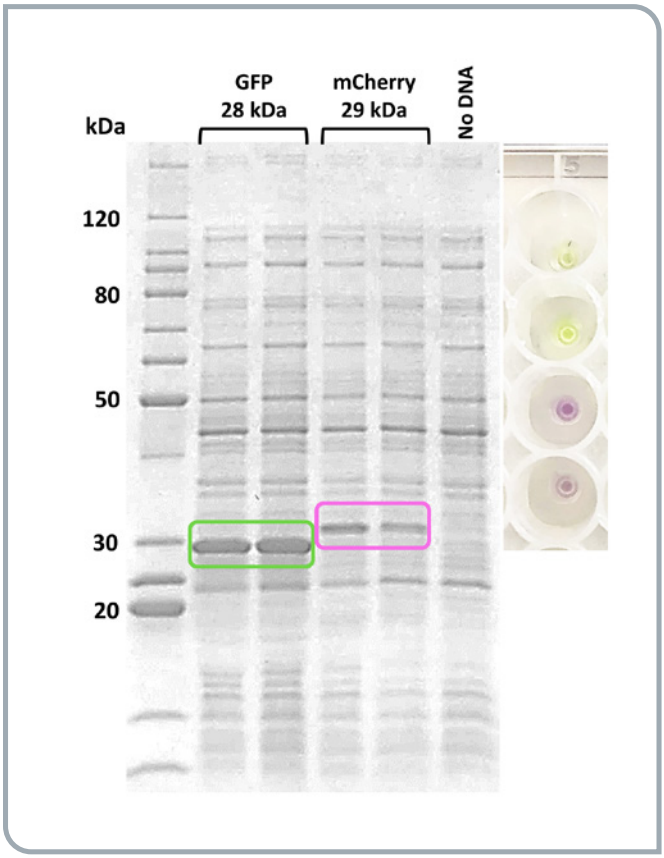


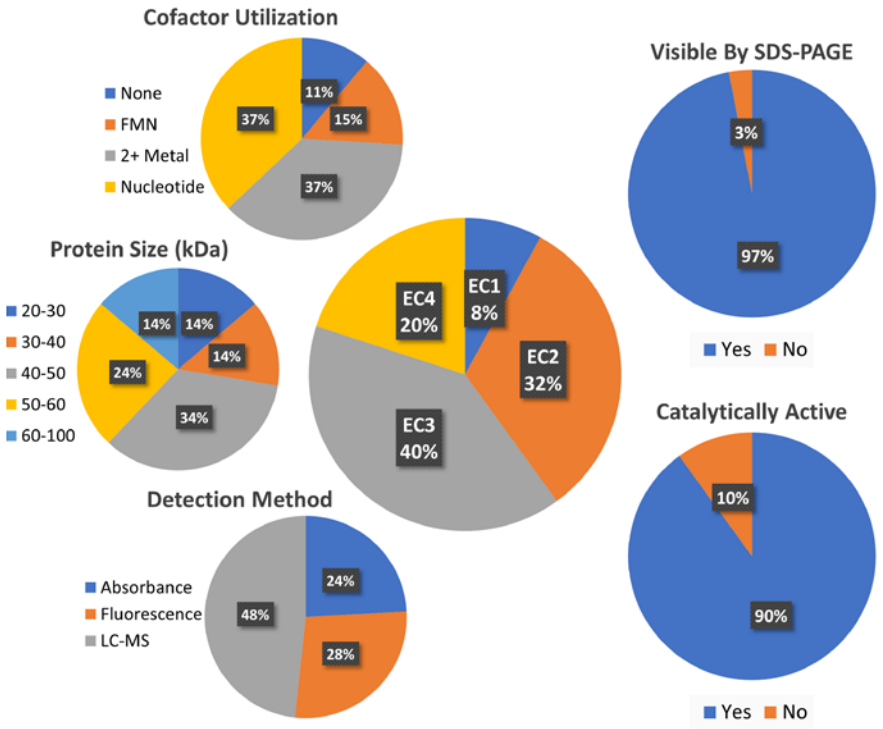
Figure 4. Example SDS-PAGE of expressed fluorescent protein compared to the myTXTL kit background without any template DNA provided. An image of the expressed GFP and mCherry that produced the accompanying gel is included for comparison.

Successful expression of a protein of interest was determined by the presence of a band at the expected molecular weight that was not present in the myTXTL kit background (Fig. 4). The background was determined by running an expression sample that had not been provided with DNA template. Using this method, it was determined that 28 of the 29 proteins tested, or 97%, successfully expressed (Fig. 5). Successful scaffolds included GFP and mCherry (also visible by eye, Fig. 4) as well as at least one representative of each tested enzyme type. Enzyme sizes ranged from 25-100 kDa and there was no apparent preference for expression of a specific size as judged by band intensity on SDS-PAGE (data not shown).

After verification of expression, samples were directly assayed without any further purification or clean up steps. For colorimetric assays (24% of test enzymes, Fig. 5) substrate or cofactor depletion was read directly, this included NADPH consumption. Fluorescence (28% of test enzymes, Fig. 5) was also used to detect product formation and for direct measurement of GFP and mCherry. The remainder of the reactions were analyzed via LC-MS (48% of test enzymes, Fig. 5) with reaction products ranging from small, polar to large, hydrophobic compounds. In all cases, reactions without plasmid DNA were used as negative controls. No significant source buildup or metabolite carry over was observed during runs. 90% of the proteins tested were active (Fig. 5). Activity was seen in all enzyme types tested and at least two members of each exhibited the desired activity. There were three enzymes that did not show any catalytic activity. One of these was the enzyme that was not successfully expressed, so it was expected that no activity would be observed. The remaining two were visible by SDS-PAGE, but both require bound cofactors, and one had been previously demonstrated to require a low *in vivo* expression temperature to show enzymatic activity. Other enzymes, including some of the same type as those that failed, also require bound cofactors, and exhibited adequate activity so

it is not a limitation of the myTXTL kit. Therefore, it is presumed these inactive outliers were expressed in either their apo form or an unfolded state and thus lacked catalytic activity when assayed. These specific cases are worth following up on, but the overall result of these experiments is a success. Expression of the majority of the tested enzymes in a catalytically active state show that the myTXTL system is a sufficient expression platform with broad applicability to biocatalysis testing regimes.

Figure 5. Representation of the proteins used in these experiments. Charts represent the percentage of the total in each category.



CONCLUSIONS

This study demonstrates that the Daicel Arbor Biosciences myTXTL cell-free protein expression system can quickly produce a wide range of enzymes covering a diverse sequence space. It is therefore a viable expression system for Arzeda, where the core technology has applications across the proteome. Harnessing the power of cell-free protein synthesis will have significant time and cost savings for current and future design projects at Arzeda. Another aspect of time saving is that the plasmid DNA used for these experiments is under control of a traditional inducible T7 *in vivo* system. This means that any level of scale up will be immediately testable and that in *vitro/in vivo* workflows can operate in parallel when necessary. Also of note, the work presented here was carried out using the manufacturer's protocol without any further optimization and utilized separate myTXTL batch preparations. This ease of use and guaranteed delivery of a robust product make the myTXTL system a promising candidate for even larger scale adoption by Arzeda as a key component of the testing pipeline. High throughput testing has also been successfully demonstrated with miniaturization and automation of the myTXTL system proving that it is straightforward to incorporate into an automated workflow.[3] Moving forward Daicel Arbor Biosciences and Arzeda will continually improve this process and greatly decrease the overall time and costs to deliver next generation biobased products.

REFERENCES

1. Shin J, Noireaux V. **Efficient cell-free expression with the endogenous** *E. Coli* **RNA polymerase and sigma factor 70.** J Biol Eng. 2010 Jun 24;4:8. https://jbioleng.biomedcentral.com/articles/10.1186/1754-1611-4-8.

2. Garamella J, Marshall R, Rustad M, Noireaux V. **The All** *E. coli* **TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology.** ACS Synth Biol. 2016 Apr 15;5(4):344-55.https://pubs.acs.org/doi/10.1021/acssynbio.5b00296.

3. Bailey, J.; Eggenstein, E.; Lesnick, J. (2020) **Miniaturization and Rapid Processing of TXTL Reactions Using Acoustic Liquid Handling.** https://arborbiosci.com/wp-content/uploads/SEED2019-myTXTL-Labcyte_Echo.pdf

ACKNOWLEDGEMENTS

The authors would like to acknowledge the many people who contributed to the material and protocols used in this work, as well as those who provided guidance over the course of the project. Specifically, Cara Tracewell, Sean Sleight, Cristina En, Amandeep Sangha, Niklas Kristiansen, Jim Havranek, Rolando Rodriguez, Jacob Bale, Karen Eaton, Kyle Roberts, and Aaron Korkegian.

CONTACT INFORMATION

Web: www.arborbiosci.com | E-mail: info@arbor.daicel.com
Web: www.arzeda.com | E-mail: info@arzeda.com

DAICEL arbor biosciences | Arzeda. THE PROTEIN DESIGN COMPANY™

myTXTL®
Cell-Free Protein Expression Kits
www.arborbiosci.com/myTXTL
10/2022