Design Space Exploration of the Violacein Pathway in *Escherichia coli* Based Transcription Translation Cell-Free System (TX-TL)

Phuc H.B. Nguyen[*], Yong Y. Wu [‡], Shaobin Guo[‡], Richard M. Murray[‡]

[*]: California State University, Long Beach, 1250 Bellflower Blvd., Long Beach CA

[‡]: California Institute of Technology, 1200 E. California Blvd., Pasadena CA

Abstract

In this study, an *Escherichia coli* (*E. coli*) based transcription translation cell-free system (TX-TL) was employed to sample various enzyme expression levels of the violacein pathway. TX-TL enables rapid modifications and prototyping of the pathway without complicated cloning cycles. The violacein metabolic pathway has been successfully reconstructed in TX-TL. Analysis of the product via UV-Vis absorption and liquid chromatography-mass spectrometry detected 4.95 mM of violacein. Expression levels of pathway enzymes were modeled using the TX-TL Toolbox. The model revealed the length of an enzyme's coding sequence (CDS) significantly affected its expression level. Finally, pathway exploration suggested an improvement in violacein production at high VioC and VioD DNA concentrations.
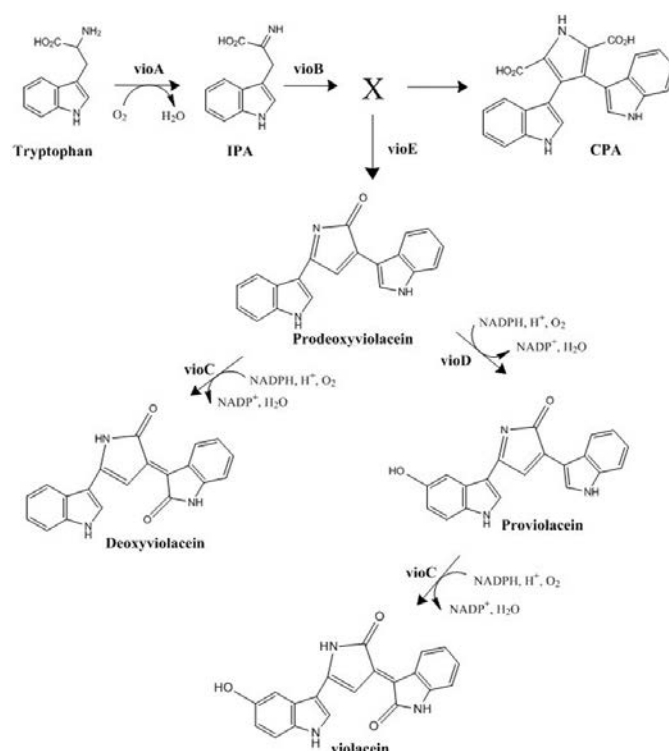
Introduction

Metabolic engineering carries the potential of economical and simplified synthesis of complex biological molecules that would otherwise be too expensive to produce. The field has made important contributions to the industrial synthesis of many biomolecules such as ethanol, glycerol, and lysine [1]. However, many engineered metabolic pathways are poorly optimized, leading to buildup of toxic intermediates and byproducts which limit cell growth. This in turn results in lowered yield and efficiency [2]. To resolve this problem, multiple cycles of design iterations of the synthetic pathway are performed until an optimized version is obtained. Nevertheless, this process is time-consuming and inefficient. Cloning and transforming multiple genes in bacteria can be labor-intensive and take at least 1 week for each iteration [3]. This inevitably lengthens the cycle time required for testing and data collecting.

In this paper, we make use of an *E. coli*-based, cell-free transcription-translation (TX-TL) system that employs both linear and circular DNA. TX-TL enables a fast and simple *in vitro* testing platform for biological genetic networks [4]. First, DNA parts of interest are created using PCR. These pieces are ligated together using Golden Gate cloning method. Complete assembly constructs include promoter, UTR, CDS, and terminator. Finally, these constructs are tested *in vitro*. The system introduces a time-efficient alternative for the conventional *in vivo* expression method. Instead of going through the lengthy transformation and bacterial culturing process, enzyme's DNA sequences can simply be assembled into different linear DNAs and expressed

directly in TX-TL. This significantly reduces the time required for testing each expression variation and allow for a more efficient exploration of metabolic design space.

Violacein is a water-insoluble bacterial pigment with potential applications such as antibacterial, anti-trypanocidal, anti-ulcerogenic, and anticancer drugs [5]. The compound is natively synthesized by *Chromobacterium violaceum* [6]. Nevertheless, its production is too expensive due to the strain's low productivity [5]. The violacein production pathway is shown in Fig. 1. Five core enzymes VioA-E of violacein production pathway have been well-defined [6]. The starting source of violacein synthesis is tryptophan. The amino acid is then processed by a series of enzymes called VioA-E and converted into the following intermediates, respectively, before finally converted into violacein: indole-3-pyruvate, X, prodeoxyviolacein, and proviolacein. Here, X represents an unknown intermediate. Without the presence of VioE, it spontaneously forms chromopyrrolic acid. A common contaminant of the pathway, deoxyviolacein, is created if prodeoxyviolacein is processed by VioC instead of VioD.

We seek to employ TX-TL to rapidly prototype the violacein pathway for improved product yield. By testing different expression levels of violacein pathway enzymes, valuable information such as rate-limiting step as well as the pathway's dynamic can be obtained. Here, we successfully reconstructed the violacein pathway in TX-TL. Violacein production was confirmed using LC-MS. The expressions of the Vio enzymes were modeled *in silico* using the TX-TL Toolbox [7]. We demonstrated that the CDS length of each enzyme significantly affects its expression level. Finally, prototyping results suggest high VioC and VioD DNA concentrations result in an improvement in violacein production while minimizing intermediate buildup. These results can be used for the future engineering of more efficient violacein producing *E.coli* strains.

Fig. 1: Overview of violacein production pathway. Tryptophan is sequentially converted into indole-3-pyruvate, X, and then prodeoxyviolacein by VioA, VioB, and VioE respectively. X represents an unknown intermediate. Both prodeoxyviolacein and proviolacein are substrates for VioC. The former case resulted in synthesis of deoxyviolacein. The latter case resulted in violacein production.

Results

TX-TL reactions of violacein pathway and its variants yielded various colored products.

Coding sequences of VioA-E enzymes [2] were cloned under the same promoter and RBS to minimize variation in expression levels at the same DNA concentration. Here, the violacein pathway and its modifications were reconstructed by simply leaving out specific enzyme DNA in TX-TL. The reactions were ran using equimolar amount of each enzyme's linear DNA. The products were found to be insoluble and extracted using 70% ethanol. The results indicated production of different metabolites as evident from the extracts' various colors (Fig. 2). As expected, the extract from the complete, violacein producing pathway had a deep violet color. The pathway lacking VioD yielded a product with a similar color to the above pathway. According to Fig. 1, this product is anticipated to be deoxyviolacein. Extracts from those lacking VioE,C,D, VioC,D, and VioC were colorless, pink, and orange, respectively. We further investigated the violacein and deoxyviolacein pathways in TX-TL. The appropriate linear DNAs

3

were expressed in TX-TL at an equimolar amount. The products were extracted with 70% ethanol, and a UV-Vis spectrum of each was obtained (Fig. 3). The spectra demonstrated similar shapes and peak positions to those found in the literatures [5], [6].
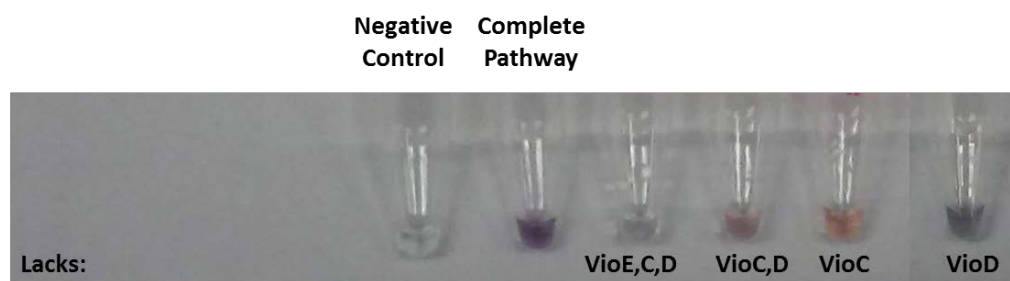

Fig.2: Metabolite Extraction from TX-TL Reactions. The metabolites were pelleted by centrifugation, washed with water, and re-dissolved in 70% ethanol.
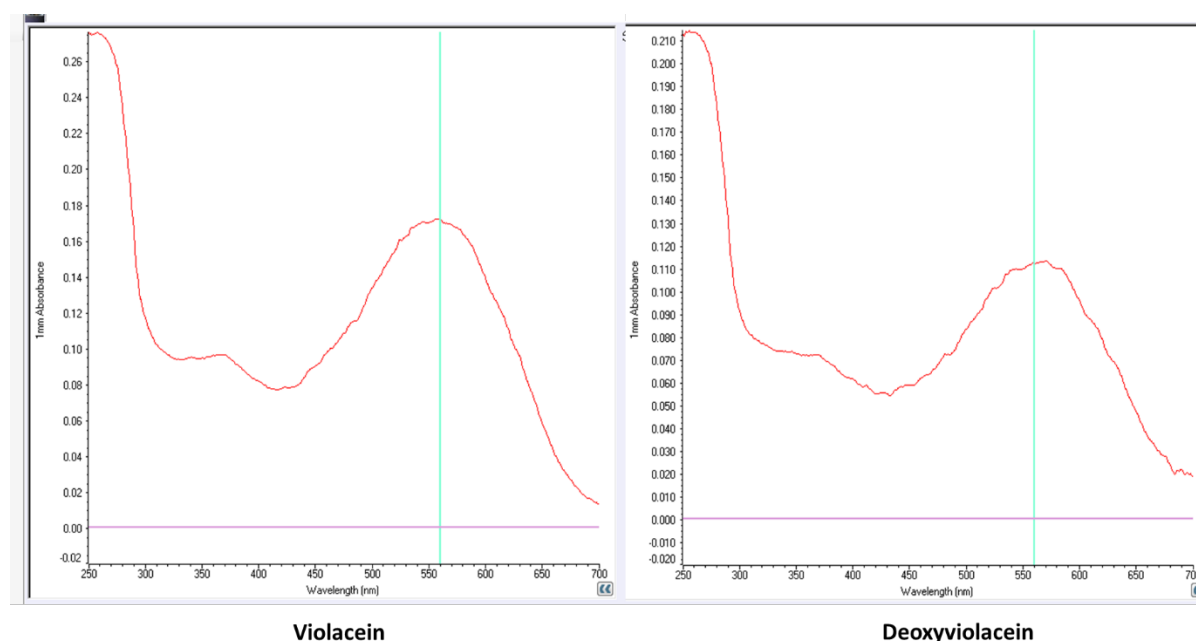
Fig. 3: UV-Vis Spectra of TX-TL Produced Violacein and Deoxyviolacein. Metabolites from TX-TL reactions were pelleted by centrifugation, washed with water, and dissolved in 70% ethanol. The blue lines indicate the literature reported peaks [5].
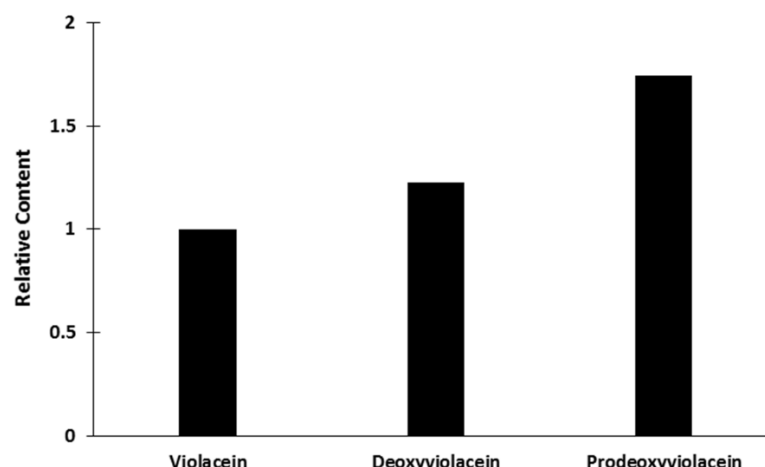
LC-MS analysis of violacein pathway reactions suggested successful synthesis of violacein in TX-TL

4

To further analyze the TX-TL reaction of violacein pathway, produced metabolites from the reaction were analyzed using LC-MS method. Commercially available violacein was used as a standard. Equimolar of 4 nM for each enzyme's DNA was used in TX-TL. Products were extracted with methanol and applied onto a reverse phase column. The metabolites were eluted with water. Mass windows of 344.103 Da, 328.108 Da, and 308.128 Da were extracted for the analysis of violacein, deoxyviolacein, and prodeoxyviolacein, respectively.
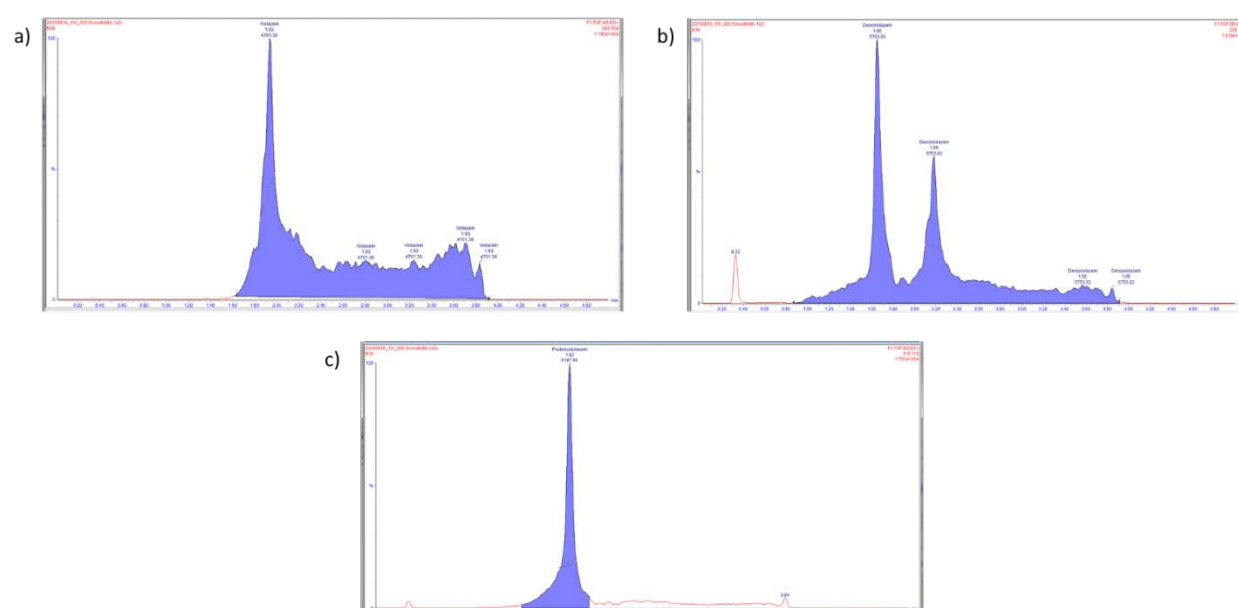
Fig.4: Detection of Violacein (a), Deoxyviolacein (b), and Prodeoxyviolacein (c) in TX-TL Reaction Containing All Pathway Enzymes. The elution time and peak areas were labeled above the on top of the peak. DNA concentrations used were 4 nM for all enzymes. Metabolites were extracted with Methanol. Samples were applied onto Acquity UPLCr BEH C18 1.7 µm. The running buffer was 10 mM Ammonium Formate in 1% ACN. The metabolite was eluted with Acetonitrile. Mass windows of 344.103 Da, 328.108 Da, and 312.113 Da were extracted for the analysis of violacein, deoxyviolacein, and prodeoxyviolacein respectively.
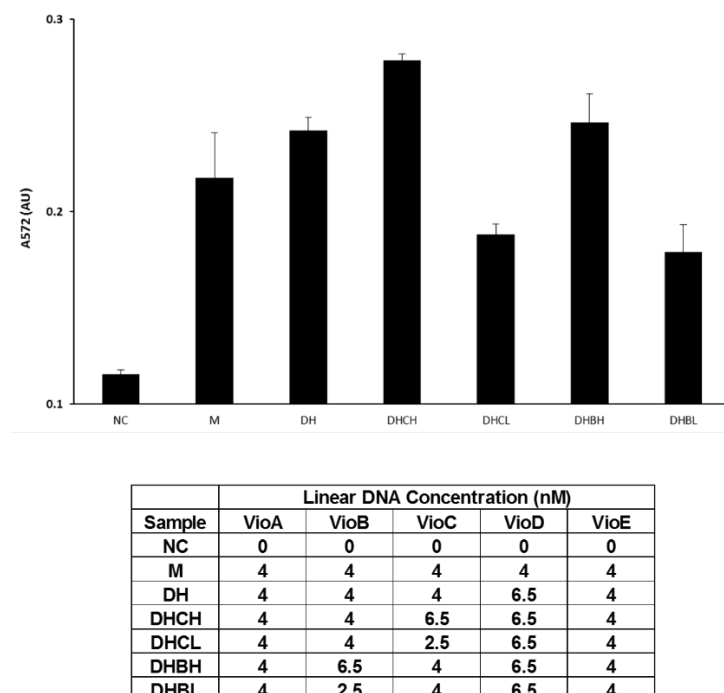
As demonstrated in Fig. 4, violacein was detected in the reaction, confirming the functionality of the pathway in TX-TL. Based on peak area analysis, the sample contained considerable deoxyviolacein byproduct. Prodeoxyviolacein, an intermediate in the pathway was also present in significant quantity (Fig. 5). Using a commercial standard, violacein concentration produced in TX-TL was determined to be approximately 1.7 ug per µl reaction, or 4.95mM.

5

Fig.5: LC-MS Analysis of Violacein, Deoxyviolacein, Prodeoxyviolacein Content in 4 nM Equimolar DNA TX-TL Reaction. Peak area of each metabolite was normalized to that of violacein. LC-MS condition was the same as described in Fig. 4.

Design space exploration of the violacein pathway and *in silico* modeling revealed CDS length effect on enzyme expression levels in TX-TL

To create differences in expression levels in TX-TL, various DNA concentrations of each Vio enzyme were used. For each enzyme, concentrations of 2.5 nM, 4 nM, and 6.5 nM were defined as "low", "medium", and "high" respectively. Multiple combinations of these concentrations were added in TX-TL. The products were extracted with 70% ethanol. Crude violacein yields were determined by measuring A_{572} (Fig. 6). Compared to the equilmolar DNA reaction of 4 nM, an increase of VioD DNA concentration to 6.5 nM resulted in an increase in the crude yield. An additional increase of VioC DNA concentration led to a further improvement in yield. However, an increase in VioB DNA concentration instead of VioC did not improve the yield. Lowering of either VioC or VioB DNA significantly affected crude violacein production. The expression of the pathway enzymes was modeled using TX-TL Toolbox. All DNA concentration parameters were set to 4 nM. Modeling result indicated VioE expression was approximately 5 times higher than those of VioA, VioC, and VioD. Meanwhile, VioB expression was about 19 times lower than those of the above three enzymes (Fig. 7).

6

	Linear DNA Concentration (nM)				
Sample	VioA	VioB	VioC	VioD	VioE
NC	0	0	0	0	0
M	4	4	4	4	4
DH	4	4	4	6.5	4
DHCH	4	4	6.5	6.5	4
DHCL	4	4	2.5	6.5	4
DHBH	4	6.5	4	6.5	4
DHBL	4	2.5	4	6.5	4

Fig.6: Design Space Exploration of Violacein. Metabolites from TX-TL reactions were pelleted by centrifugation, washed with water, and dissolved in 70% ethanol. Ten microliters of each sample was measured for absorbance at 572 nm for crude violacein yield estimations.
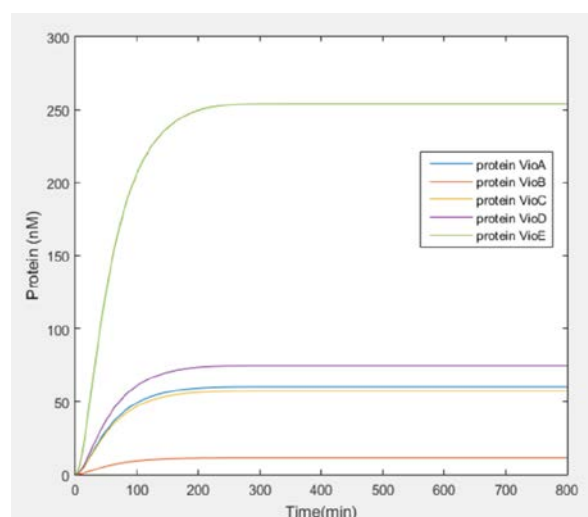
Fig.7: *In Silico* Modeling of Vioenzymes Expressions in TX-TL Using TX-TL Toolbox. DNA concentration parameters for all enzymes were set to 4 nM. The same promoter and 5' UTR parameters were also used to for all enzymes.

7

Western blotting of Vio enzymes in TX-TL was performed to verify the model's prediction. VioA, VioB, and VioE were cloned with a strep tag at the C terminals. VioD and VioC were cloned with a his tag and a 3x FLAG tag, respectively. The tagged enzymes were expressed in TX-TL at an equimolar concentration of 4 nM (Fig. 8). Based on Western blotting result using anti-strep antibody, VioE was expressed significantly more than VioA. VioB was not observed in the equimolar 4 nM reaction. However, a TX-TL reaction with 10 nM VioB DNA by itself showed a clear band. The bands' intensities were quantified using ImageJ. As shown in Fig. 8, the expression levels of Vio enzymes in TX-TL demonstrated the same trend as predicted by the model.
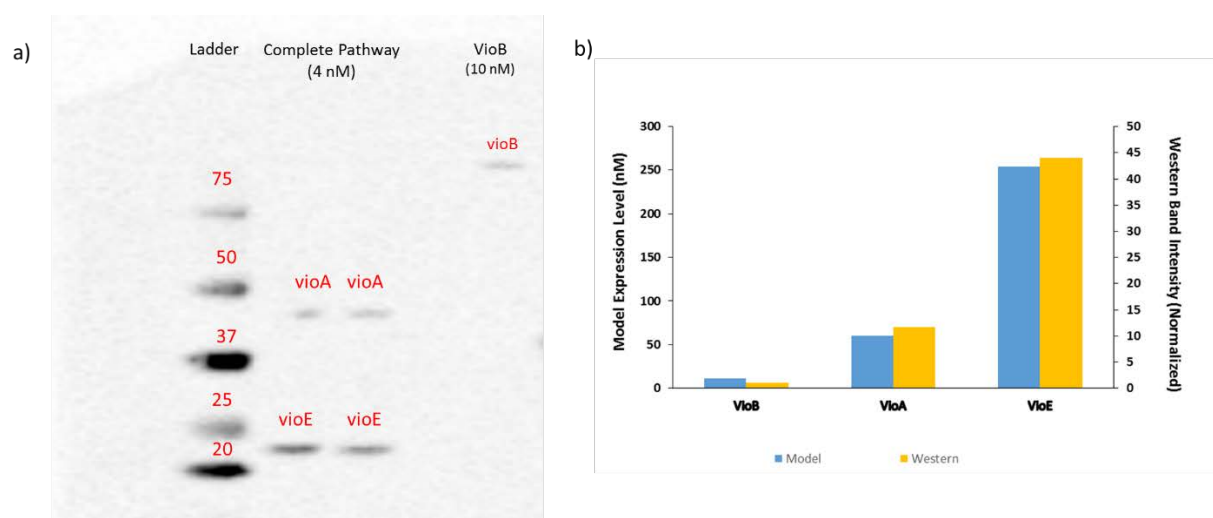

Fig.8: Western Blotting of Strep Tagged Vioenzymes TX-TL Reactions (a) And Comparison Between Predicted With Modeling Results (b). Similar to modeling results, Western blotting indicates expression level increases from VioB to VioA and then VioE. VioB's band intensity in 4 nM reaction was obtained by measuring the background's intensity at its location on the gel.

High DNA concentration of both VioC and VioD suggested an increase in violacein yield and limited intermediate buildups in TX-TL

Various expression combinations of the vioenzymes were investigated for an increase in violacein yield. Since a buildup of prodeoxyviolacein was observed in the 4 nM equimolar DNA reaction, VioC and VioD DNA was increased to 6.5 nM to rescue the bottleneck. Additionally, TX-TL reactions with a lowered VioE DNA concentration of 2 nM were run to minimize the buildup. The products were extracted with methanol and analyzed for violacein, deoxyviolacein, and prodeoxyviolacein contents using LC-MS (Fig. 9). The reaction with 6.5 nM DNA

8

concentration of VioD and VioC and 4 nM DNA concentration of the rest showed a high violacein content compared to deoxyviolacein and prodeoxyviolacein. Lowering VioE DNA to 2 nM while keeping VioB, VioC, VioD concentrations high at 6.5 nM resulted in a further improvement in violacein yield. However, the deoxyviolacein and prodeoxyviolacein contents also significantly increased. Other tested reactions did not show noticeable improvement in violacein yield.
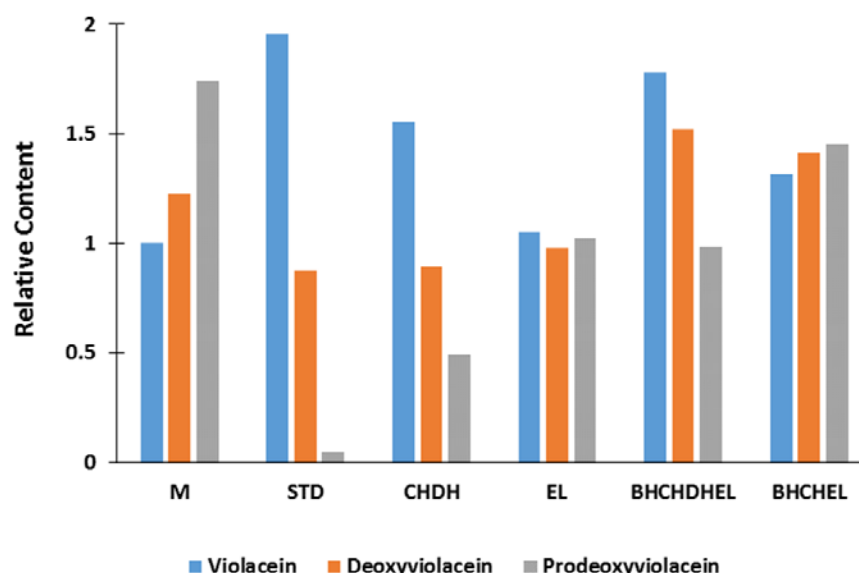
	Linear DNA Concentration (nM)				
Sample	VioA	VioB	VioC	VioD	VioE
STD	Comercially avaiable standard				
M	4	4	4	4	4
CHDH	4	4	6.5	6.5	4
EL	4	4	4	4	2
BHCHDHEL	4	6.5	6.5	6.5	2
BHCHEL	4	6.5	6.5	4	2

Fig.9: LC-MS Analysis of Violacein, Deoxyviolacein, Prodeoxyviolacein Contents in Space Exploration TX-TL Reactions. Peak area of each metabolite was normalized to that of violacein at 4 nM equimolar DNA concentration. LC-MS condition was the same as described in Fig. 4.

Discussion

In this study, the violacein metabolic pathway has been successfully reconstructed in TX-TL. All five pathway enzymes were cloned under a medium-strong promoter with a strong RBS. The complete pathway reaction yielded a product with purple color. Missing each of VioC, VioE, VioD, or a combination of these enzymes triggered the production of different colored metabolites. The result demonstrated the functionality of these enzymes in TX-TL. This also hinted at the possible use of TX-TL to rapidly generate specific intermediates in metabolic pathways. Many intermediates are unstable, however. Therefore, preservation methods need to be investigated beforehand.

LC-MS analysis of 4 nM equimolar DNA TX-TL reaction suggested the presence of violacein at an approximate concentration of 4.95 mM. Crude violacein productions corresponding to different Vio enzyme's DNA concentrations were investigated. An increase in VioC and VioD DNA both improved the yield. However, increasing VioB DNA had no significant effect. Since VioB is upstream and both VioC and VioD are downstream of VioE, the pathway exploration results supported the LC-MS data of a prodeoxyviolacein build up in this pathway.

Modeling of protein expression at this 4 nM equimolar DNA concentration predicted a high VioE and low VioB expression compared to the rest. The prediction was verified with the Western blotting result. Tagged VioA, VioB, and VioE enzymes showed up at the appropriated positions on SDS-PAGE gel, indicating that they were expressed. The model's prediction of enzyme expression trend likely originated from its built-in transcriptional and translational rate functions [7]. Shorter DNAs will be transcribed and translated faster than the longer ones. VioB's CDS was approximately 2.5 times longer than that of VioA, VioC, and VioD and 5 times longer than VioE's. Additionally, the model did not take into account the effect of multiple RNA Polymerases (RNAPs) and ribosomes binding to one DNA strand and mRNA transcript [7]. Taken together, the information explained the model's prediction of more small enzymes being expressed compared to the larger ones. Despite excluding the multiple RNAPs/ribosomes binding effect, the model's result still fit experimental data's trend. This might be because in TX-TL reaction of the violacein pathway, RNAP concentration is not present in far excess of the total DNA concentration. A study by Karzburn et al. approximated the TX-TL RNAP concentration to be 29 nM [8] while total DNA concentration used was 20 nM. Under this condition, the multiple RNAPs/ribosome binding effect might become less significant. Overall, the result demonstrated the significant effect of CDS length on protein expression in TX-TL. At the same DNA concentration, proteins with longer CDS will be expressed less and vice versa. This needs to be taken into account for future pathway exploration experiments in TX-TL. The model and Western blot results also provided a possible explanation for the buildup of prodeoxyviolacein in the equimolar DNA TX-TL reaction. Since VioE was expressed significantly more than the rest, its metabolic product was more likely to accumulate in the pathway.

Different DNA concentration combinations were sampled for an increase in violacein production in TX-TL. Reactions with VioC and VioD DNA concentrations adjusted to 6.5 nM while

10

keeping the rest at 4 nM resulted in higher violacein yield. At the same time, both deoxyviolacein and prodeoxyviolacein contents were lowered. This is expected since an increase in VioC and VioD expressions led to a higher consumption of prodeoxyviolacein intermediate. Interestingly, modeling at this concentration combination still predicted VioE expression to be higher than the rest, although the gap was closer (not shown). This might suggest VioD and VioC are catalytically more efficient than VioE. The concentration combination with high VioB, VioC, VioD DNA at 6.5 nM and low VioE DNA at 2 nM led to a further increase in violacein yield. However, deoxyviolacein and prodeoxyviolacein productions also significantly increased. The result was unexpected since VioE DNA was two times lower than the previous combination. This might be due the increase in VioB expression which channeled the flux down the pathway's branch point.

In conclusion, we demonstrated the feasibility of using TX-TL as a testing platform for metabolic engineering by successfully reconstructing the violacein metabolic pathway in TX-TL. Through analysis of the pathway using LC-MS and *in silico* modeling, we discovered the proportional relationship between an enzyme's CDS length and its expression level in TX-TL. This effect has important implications to future pathway prototyping experiments in TX-TL. Finally, we established that high VioC and VioD DNA concentrations combined with lower, equimolar concentrations of the rest would lead to a high violacein production while minimizing unwanted intermediates and byproducts. This study contributed to the overall effort of engineering a more efficient violacein production bacterial strain. Future work includes the investigation of CDS length effect on enzymes' expression *in vivo.* If the effect is not observed, linear DNA concentrations and regulatory sequences need to be carefully chosen to ensure TX-TL results best reflect the *in vivo* situation. Finally, expression level combinations with highest violacein yields in TX-TL will be tested *in vivo.*

Methods

Cloning of VioA, VioB, VioC, VioD, and VioE and *in vitro* Linear DNA Assembly.

Yeast plasmid harboring violacein pathway was obtained from Lee et al [2]. CDS of each enzyme was PCR amplified using New England Biolab's Phusion® Hot Start Flex 2X Master Mix. Annealing temperature was set to 60°C for 30 sec. Elongation time was 2 minute at 72°C. The PCR was set to 35 cycles. Each CDS was ligated to its promoter, 5' UTR, and terminator as described in Table 1 using standard Golden Gate cloning protocol described by Sarrion-Perdigones et al [9]. Complete linear DNA fragments were checked for integrity using agarose gel and sequencing.

Preparation of crude cell ("E31") extract

Standard *BL21-Rosetta* strain was used for extraction. Cells were grown in 1 L cultures and pelleted according to the protocol stated by Sun et al. [10] with the following modification: The 2xYT media was supplemented with 6.2% (v/v) phosphates solution (11:20 dibasic : monobasic potassium phosphate). Collected cells were homogenized and extracted according to the method

described by Kwon, et al. [11] with the post-homogenization incubation period extended to 80 min instead of 60 min.

TX-TL reactions of Violacein pathway and metabolite extraction

TX-TL buffer and mastermix were prepared as described in Sun et al [10]. Extract "E31" was used in the entirety of this study. TX-TL reaction was setup by adding Mastermix to DNA mixture to the appropriate DNA concentrations. TX-TL reaction product was pelleted. The supernatant was discarded, and the pellets were washed with water once. The precipitated products from each sample were re-dissolved in 70% ethanol and measured for absorbance at 572 nm using BioTeK Synergy H1 microplate reader.

LC-MS analysis of TX-TL produced violacein.

TX-TL reaction products were pelleted and washed as described above. The metabolites were re-dissolved in methanol. The sample was injected onto Acquity UPLCr BEH C18 1.7 µm reverse phase column. Buffer A was 10 nM Ammonium Formate in 1% ACN. Buffer B was Acetonitrile. The flowrate was 0.5 ml/min. Applied gradient was from 0 to 90% B in 3 min and hold for 0.2 min. Mass windows of 344.103 Da, 328.108 Da, and 312.113 Da and elution times of 1.93, 1.65, 1.83 min were extracted for the analysis of violacein, deoxyviolacein, and prodeoxyviolacein respectively using MassLynx 4.1 software.

Western blotting of strep-tagged vioenzymes and i*n silico* modeling of vioenzymes expression in TX-TL

Overnight TX-TL reaction mixture was separated using Invitrogen's 4-12% Bis-Tris SDS PAGE gel. The proteins was blotted onto a PVDF membrane soaked in semi-dry transfer buffer (36.8 mM Tris-HCl, 39 mM glycine, 3.5 mM SDS, and 20% methanol). The blotted membraned was blocked in 5% Milk in TBST buffer (136 mM NaCl, 2.7 mM KCl, 19 mM Tris-HCl, 0.1 % Tween-20) for 1 hr. The membrane was washed 3 times with TBST. Protein bands were visualized using Bio-Rad ChemiDoc gel Imager. Vioenzymes expression modeling was done using the TX-TL Simulation Toolbox described by Tuza, et al [7]. Simulation's parameters were chosen as described by Karzbrun, et al [8]. Each enzyme's initial DNA concentration was set to 4 nM.

Faculty Program and by the Gordon and Betty Moore Foundation through Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative.

Reference

1. Stephanopoulos, G.; Aristidou, A. A.; Nielsen, J. *Metabolic Engineering: Principles and Methodologies*; Academic Press, 1998.

2. Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **2013**, *41*, 10668–10678.

3. Green, M. R.; Sambrook, J. *Molecular cloning : a laboratory manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y, 2012.

4. Sun, Z. Z.; Yeung, E.; Hayes, C. A.; Noireaux, V.; Murray, R. M. Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. *ACS Synth. Biol.* **2013**, *3*, 387–397.

5. Jiang, P.; Wang, H.; Zhang, C.; Lou, K.; Xing, X.-H. Reconstruction of the violacein biosynthetic pathway from Duganella sp. B2 in different heterologous hosts. *Appl. Microbiol. Biotechnol.* **2009**, *86*, 1077–1088.

6. Balibar, C. J.; Walsh, C. T. In Vitro Biosynthesis of Violacein from l-Tryptophan by the Enzymes VioA−E from Chromobacterium violaceum†. *Biochemistry (Mosc.)* **2006**, *45*, 15444–15457.

7. Tuza, Z. A.; Singhal, V.; Kim, J.; Murray, R. M. An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular breadboard; system. In *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*; 2013; pp. 1404–1410.

8. Karzbrun, E.; Shin, J.; Bar-Ziv, R. H.; Noireaux, V. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Phys. Rev. Lett.* **2011**, *106*, 048104.

9. Sarrion-Perdigones, A.; Vazquez-Vilar, M.; Palací, J.; Castelijns, B.; Forment, J.; Ziarsolo, P.; Blanca, J.; Granell, A.; Orzaez, D. GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiol.* **2013**, *162*, 1618–1631.

10. Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., Noireaux, V. Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. J. Vis. Exp. (79), e50762, doi:10.3791/50762 (2013).

11. Kwon, Y.-C.; Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* **2015**, *5*.