BaNCE: A Simplified Model for *in vivo* Bacterial Non-Continuous Evolution

Authors:

James Andon, Biochemistry

Katherine Vasconcelos, Biochemistry and Professional Writing

Submitted to:

Dr. Destin Heilman, Advisor, Department of Chemistry and Biochemistry

Abstract

Directed evolution is a powerful tool that has been honed to create new and improved proteins. One form of directed evolution, Phage-Assisted Continuous Evolution (PACE), developed by Dr. David Liu, elegantly links favorable mutations in an arbitrary gene of interest to a bacteriophage's ability to transfect a host. Inspired by PACE, this paper explores the creation and viability of a Bacterial Non-Continuous Evolution model (BaNCE) that aims to conserve the mutagenic and selective properties of PACE while eliminating some of its complexities. BaNCE employs the use of three plasmids within E. coli. The first contains an error-prone polymerase that will make mutations to a second plasmid that encodes for a gene of interest. The final plasmid provides BaNCE's selective pressure via an auxotrophic selection model where prototrophy is restored by generation of favorable mutations in the gene of interest. Through preliminary research, we assembled the components of BaNCE and began to test this model's viability. In designing BaNCE, we hope to introduce a simplified version of PACE to increase directed evolution accessibility to academic labs.

Acknowledgements

We would like to thank the following people:

- Dr. Destin Heilman, not only for his guidance, advice, feedback, and encouragement throughout our MQP, but also for his support throughout our academic careers at WPI. Our future careers in the biochemistry field would not be possible without him.
- Dr. David Liu and Chris Podracky, from Harvard University, for their supply of pMdh2 that was used in this MQP, as well as Mr. Podracky's help in pMdh2 sequencing efforts.
- Dr. Peter Shultz and Lubica Supekova, from Scripps College, for their supply of the pBAD metA WT plasmid that was used in this MQP to engineer pMetA.
- Ruben Bonilla-Santiago, Paula Moravek, and Silvana Reid for their support in helping keep our MQP run smoothly through supplying our lab.

Table of Contents

Introduction

While it has not always been an in-lab technique, people have participated in directed evolution for centuries. About 11,000 years ago, Middle Eastern settlers began to farm their own crops to avoid relying on gathering. Their farming techniques present one of the first instances of directed evolution. For example, as they noticed that certain crops of wheat were more favorable than others, they began to save the seeds of these crops to sew better wheat, thus undertaking their own "artificial selection". Over time, farmers began to modify wheat traits to make more favorable crops, in ways such as selecting seeds that did not fall off the crop when they were ripe and ones that were more easily separable from the plant. Similar evolutionary processes have been used to transform other crops (The National Academies of Sciences, Engineering, and Medicine, 2008) and the idea of directed evolution has persisted since.

Laboratory directed evolution of proteins did not appear until 1993, when Chen and Arnold evolved the protein subtilisin E *in vitro*. To do this, they used error prone PCR to introduce random mutations into the subtilisin E gene. After mutation, proteins were expressed and screened for increased catalytic activity. Genes exhibiting favorable mutations were then subjected to further mutagenesis and screening. After three sequential rounds of mutagenesis and screening, they observed a 256 fold increase in the protein's catalytic efficiency. This experiment was the first to demonstrate the advantages of using a sequential mutagenesis-selection model to evolve proteins (Cobb *et al.,* 2013; Chen & Arnold, 1993).

One year after Chen and Arnold demonstrated the ability of error-prone mutagenesis to evolve proteins, Willem Stemmer demonstrated the evolutionary power of recombination. Stemmers' method involved creation of a library of analogous genes, which was then digested into smaller fragments. These fragments were then ligated together to recombine them into new genes. Favorable mutations were selected for by cloning the recombined genes into plasmids and expressing them to assess favorability. Stemmer used this method to evolve β-lactamase to confer a 320,000 fold increase in resistance to the antibiotic cefotaxime. The work done by Chen and Arnold and by Stemmer laid the framework for random mutation driven directed evolution. Despite the promise exhibited by this type of directed evolution, the large amount of time required to generate successive cycles of mutation and screening have impeded the usefulness of these methods. This issue would not be addressed for decades until researchers began devising ways of coupling mutagenesis and selection using *in vivo* systems (Cobb *et al.,* 2013; Stemmer, 1994).

In 2011, Dr. David Liu and colleagues conducted research on continuous evolution through their own method called PACE (Phage-Assisted Continuous Evolution), which involves "evolving genes transferred from host cell to host cell through a modified bacteriophage life cycle in a manner that is dependent on the activity of interest" (Carlson et. al., 2011). The PACE model involves the use of two DNA plasmids and a phage genome. The first plasmid, called the mutagenesis plasmid (MP), encodes for an error-prone polymerase that will generate random mutations in the system. The phage genome, called the selection phage (SP), contains the system's gene of interest (GOI). To create the SP, the pIII gene was excised from the phage genome and the GOI was inserted in its place (Carlson et. al., 2011)

Without pIII, phage infection rate drops by nine orders of magnitude. Consequently, pIII function must be restored for the phage to propagate and infect a greater number of host cells. Thus, to restore function, favorable mutations of the gene of interest must be linked to restoration of pIII expression, which is done by the system's second DNA plasmid. This plasmid, called the accessory plasmid (AP), encodes for a constitutively repressed pIII that provides the system's selective pressure. Expression of pIII is inhibited until favorable mutations in the GOI produce some change in the system that restores pIII expression to the AP. This has been done in various ways and largely depends on the type of GOI. In the first example of PACE, the GOI was a T7 polymerase which was evolved to recognize a novel promoter. This promoter was placed upstream of pIII in the AP, allowing pIII expression only to be restored if T7 could evolve to recognize the novel promoter (Carlson et. al., 2011).

In order to combine the MP, SP, and AP into one system, Liu and colleagues devised a "lagoon" (a fixed volume vessel) system. *E. coli* host cells are continuously pumped through the "lagoon" that contains replicating phages. The added *E. coli* cells contain the MP and AP. In order to regain infectiousness, pIII production must be restored to the phage faster than the lagoon can wash away the phages. To accomplish this, the phage will infect the *E. coli* host cells, which will then carry the MP, AP, and SP. Thus, the goal is for the error-prone polymerase from the MP to make random mutations to the GOI in the SP. Through selective pressure provided by the AP, a GOI with a gain of function mutation will be selected to continue to replicate, as the system can only survive with a mutation that increases the GOI's function enough so that it can restore pIII production in the AP (Carlson et. al., 2011).

The beauty of PACE is its ability to evolve any gene that can be linked to pIII production in *E. coli* (Carlson et. al., 2011) and it has shown promising results across a large number of publications. In an even more recent paper (Roth et. al., 2019), Dr. David Liu and co-workers demonstrated PACE's ability to evolve the *Bacillus methanolicus* methanol dehydrogenase protein, Mdh2, to improve its catalytic rate. Liu *et al.* demonstrated that their modified Mdh2 was faster than any other modified Mdh2 previously described. Liu's goal in improving Mdh2 catalytic activity had implications for developing methylotrophic *E. coli* to convert methane into metabolites for metabolism integration. Thus, his research could have implications for greenhouse gas reduction.

Due to greenhouse gases, such as methane, the Earth's temperature is rising as the greenhouse effect is worsening (Climate Action Reserve, 2019). Average surface temperature of the Earth has risen 0.9 degrees Celsius since the late 19th century (NASA, 2019). It is predicted that, without interference, global temperatures could increase by 10 degrees Fahrenheit by 2100, causing irreversible changes to Earth's climate (Climate Action Reserve, 2019). Methane has a particularly detrimental effect on the atmosphere as its greenhouse effects are 34 times that of carbon dioxide. Fossil fuels and livestock are two major sources of atmospheric methane (UNFCCC, 2019), which, in 2017, accounted for 10% of the United States' greenhouse gas emissions (EPA, 2019). Currently, there is no easy way to rid the atmosphere of methane other than prevention or combustion. However, as shown by Dr. Liu, directed evolution is a possible solution to reduce methane emissions.

While PACE is a powerful method, the complexity of the system may deter other labs with fewer resources from investigating PACE on their own. Specifically, we found that the use of phages and a complex pump system would be too time consuming to establish this model in our own lab. To address this, we designed our own method of continuous directed evolution with simplicity and ease of use in mind, called the BaNCE (Bacterial Non-Continuous Evolution) method. In order to assess the viability of BaNCE, results from the Mdh2 PACE experiments were attempted to be replicated. To assess the viability of this proposed system relative to PACE, we aimed to compare the catalytic efficiency of a BaNCE evolved Mdh2 to PACE's Mdh2.

The main aspect of PACE we wanted to eliminate was the use of viruses. PACE uses the M13 filamentous bacteriophage to link favorable mutations to the propagation of progeny. The crux of PACE hinges on connecting these favorable mutations of the gene of interest to expression of *pIII*. In Liu's Mdh2 PACE experiments, this was done by using a formaldehyde sensitive transcription factor (FrmR) to repress *pIII* expression. Since Mdh2 converts methanol to formaldehyde, increased enzymatic activity correlates to an increase in formaldehyde concentration. In Liu's model, Mdh2 improvement caused an increase in formaldehyde concentration, causing decreased FrmR repression of *pIII* and thus phage reproduction (Carlson *et al.,* 2011; Rother *et al.,* 2019).

Overall, the hypothesized BaNCE system is composed of three plasmids and is performed within *E. coli*. The first plasmid, called pEP, encodes for the Error Prone (EP) DNA polymerase involved in making mutations. It will mutate a second plasmid, called pMdh2, which encodes for methanol dehydrogenase (Mdh2), BaNCE's gene of interest. The goal is for the EP DNA polymerase to introduce a gain of function mutation into Mdh2 that will improve the rate at which Mdh2 can convert methanol into formaldehyde. Sensing an increase in formaldehyde formation will be done through BaNCE's third plasmid, pMetA. This plasmid contains *MetA*, which is a gene essential for methionine biosynthesis. The system will be performed in *E. coli* cells with *MetA* knocked out, inhibiting the methionine biosynthetic pathway. Regaining of *MetA* is essential for the *E. coli* cells to survive in a methionine deficient environment. In pMetA, *MetA* is regulated under the same formaldehyde (Frm) promoter/operator as used in PACE, which is bound by the formaldehyde repressor (FrmR). Only a large enough increase in formaldehyde will result in binding of formaldehyde to FrmR, thus releasing it from the operator and allowing transcription of *MetA* and subsequent *E. coli* growth restoration. As this system eliminates the use of phages, the hope is that it will make a form of directed evolution more accessible to other research labs, thus combating the complexity of a phage-centric system.

Materials and Methods

MetA- Competent Cell Preparation

100 mL of LB broth with a kanamycin (Kan) concentration of 25 μg/mL was inoculated with 5 μL of liquid MetA- cell culture. Cells, along with 500 mL of LB broth, were placed in a 37°C incubator overnight. The following day, 400μL of 50μg/mL Kan was added to the LB broth. Target OD600 of the cell culture is between 0.400 and 0.600, so the OD600 of the current culture was measured and a portion of the culture was added to the 500mL LB+Kan broth to achieve an OD600 of 0.100. Cells were then incubated and measured every 20 minutes until the OD600 measured close to the target. Cells were then incubated on ice for 10 minutes and 2, 50 mL aliquots were placed into tubes and spun down at 3500x g at 4°C for 15 minutes. The supernatant was decanted and cells were resuspended with 2.5 mL of 50 mM MgCl2 and vortexed. The total volume of the tube was then raised to 25 mL with 50 mM MgCl2. Cells were centrifuged for 15 more minutes at 3500x g at 4°C. The supernatant was again decanted and cells were resuspended again with 2.5 mL of 50 mM MgCl2 and vortexed to mix. The total volume was then raised to 12.5 mL with MgCl2. Cells were incubated on ice for 25 minutes and then pelleted at 3500x g for 10 minutes at 4°C. 80% of the supernatant was decanted and cells were resuspended in the remaining liquid. The culture was then transferred to clean tubes and pelleted at 3500x g for 15 minutes at 4°C. The supernatant was aspirated and the cells were resuspended in 2.5 mL of 50 mM CaCl2. Glycerol was added to make a 15% stock solution and tubes were snap-freezed in liquid nitrogen and stored at -80°C.

Dropout Media

Since *E. coli* cells lacking *MetA* were used, a dropout media lacking methionine was needed. To make the dropout media, the following components were mixed: 200 uL 5x M9 Salts (to make M9 salts, in 500 mL of DI water, combine and mix 32g $Na_2HPO_4•(H_2O)_7$, 7.5g KH_2PO_4, 2.5g NH_4Cl, and 1.25g NaCl), 200 uL 1M $MgSO_4$, and 10 uL 1M $CaCl_2$, 0.192g dropout mix (amino acids minus methionine). DI water was added to bring the total volume up to 98 mL. 2 mL of 20% glucose was then added to the solution to bring the final volume to 100 mL. When needed, for control experiments, 8.2 mL of 78.5 mM methionine was added to the media, prior to the addition of DI water.

Transformations

Approximately 100 μL of appropriate cells for transformations were thawed on ice. 15 ng of target DNA was added to the cells, gently mixed, and incubated for 30 minutes on ice. Cells were then heat shocked at 42°C for 20 seconds and placed back on ice for 5 minutes. 450 μL of LB broth was added to the cells, which were then incubated for 1 hour at 220 rpm and 37°C. Following incubation, between 75 and 150 μL of cells were spread onto an appropriate plate and incubated overnight at 37°C.

Table of Transformations

Plasmid	Cell Line(s)	Antibiotic(s)
pMdh2	JM109, BL21	Streptomycin
pEP	JM109, BL21 MetA-	Chloramphenicol
pBAD MetA WT	MetA-	Kanamycin
pStart-T2	MetA-	Tetracycline
pMetA	MetA-	Tetracycline
None	MetA-	Kanamycin (natural resistance)

DNA Mini Preps

To extract and purify plasmids throughout this project, Promega miniprep kits were used. 3mL of c3mL of cells with target DNA were grown overnight. The next day, 600 µL of culture were added to a 1.5 mL tube. To the tube, 100 µL of cell lysis buffer was added and inverted six times to mix. After mixing, 350 µL of cold neutralization buffer was added and mixed until cloudy and yellow. The mixture was centrifuged at 20,000 rcf for 3 minutes. Following centrifugation, the supernatant was transferred to a minicolumn with a collection tube and was again centrifuged at 20,000 rcf for 20 seconds. Flowthrough was discarded and 200 µL of Endotoxin removal wash was added to the tube, which was then centrifuged again at 20,000 rcf for 20 seconds. 400 µL of Column wash was then added and centrifuged at 20,000 rcf for a 30 second interval. The column was transferred to a clean tube and 30 µL of Elution buffer was added and allowed to sit at room temperature for 1 minute. Following incubation, the tube was centrifuged for 20 seconds at 20,000 rcf to elute the DNA.

Mutagenesis Test

An experiment was performed to test for EP DNA Pol I's mutagenesis. 3mL of LB broth, 2.2 µL of Cm (Chloramphenicol, 34 g/L), and 30 µL of Strep (Streptomycin, 10 g/L) were added to a culture tube. Colonies pre-picked from an overnight culture of double transformant cells with pMdh2 and pEP DNA pol I were inoculated into the tubes. Tubes were grown overnight at 220 rpm and 37°C. In addition, single transformants of pMdh2 and pEP were inoculated into LB+Strep and LB+Amp culture tubes respectively as a control and also grown overnight. Following overnight growth, cells were passed and again grown overnight. The next day, cells were passed, induced with IPTG and arabinose for induction, and scaled up into 40 mL of liquid culture to prepare for protein extraction and DNA purification.

Agarose Gel Electrophoresis

50 mL of a 0.9% agarose solution in 1x TAE was prepared and microwaved for 1 minute. Once the solution was cool enough to touch, 1µL of Ethidium Bromide was added and swirled to mix. The agarose solution was poured into a gel mold and allowed to harden. 10 µL of DNA ladder was added to the first lane with DNA samples added to subsequent ones. It was then run at 90V for 1 hour and products were viewed on a UV transilluminator.

Restriction Cloning

Restriction cloning was used to insert *MetA* into the pStart-T2 plasmid. First, a ~200 ng load of DNA (6.3 uL of 30.1 ng/uL pStart-T2 or 38.3 ng/uL *MetA* insert) was digested by incubating it with 2 uL 10x CutSmart buffer, 2 uL ApaI (NEB), 2 uL XhoI (Promega), and 7.7 uL ddH_2O at 37°C for 1 hour. Each reaction mixture was then heat inactivated by incubating at 65°C for 20 minutes. Next, the restricted DNA was recovered using a Promega Wizard PCR Cleanup Kit using the supplied protocol. To ligate *MetA* into pStart-T2, 6.6 uL of restricted pStart-T2 (15.7 ng/uL), 1.3 uL of restricted insert (12.8 ng/uL), 1 uL Ligase 10x buffer (Promega), and 1 uL T4 Ligase (Promega) were incubated overnight at 4°C. To test ligation efficiency, 2 uL of ligation mixture and 2 uL of pStart-T2 (30 ng/uL) were each transformed into competent MetA- *E. coli* and growth levels compared.

Protein Quantification Assay

A protein quantification assay was used to normalize protein concentration for SDS-PAGE. First, a standard curve was prepared by performing a 5x dilution of 10 mg/mL bovine serum albumin (BSA), and then five 2x serial dilutions. Likewise, each protein extract sample was diluted using three 2x serial dilutions. Then, 50 uL of each sample was loaded into a 96-well plate and mixed with 250 uL Pierce™ 660nm Protein Assay Reagent (Thermo Scientific). A blank was prepared by mixing 50 uL water with 250 uL reagent. The samples were incubated at room temperature for 20 minutes and optical density was observed at 595 nm wavelength.

Protein SDS-PAGE

12% Polyacrylamide gels were cast using the following protocol. The resolving gel was cast by mixing 4 mL of 30% acrylamide/1% bis-acrylamide, 2.5 mL resolving buffer (1.5 M Tris-HCl, pH 8.8), 3.4 mL ddH_2O, 0.1 mL 10% SDS, 550 uL 20% ammonium persulfate, and 7 uL TEMED. For the stacking gel, 0.65 mL 30% acrylamide/1% bis-acrylamide were mixed with 1.25 mL stacking buffer (0.5 M Tris-HCl, pH 6.8), 3 mL ddH_2O, 50 uL 10% SDS, 25 uL 20% ammonium persulfate, and 5 uL TEMED. In addition to in-house gels, 16% BioRad Mini-PROTEAN precast gels were also used. Samples were diluted five-fold in the loading buffer (0.1% bromophenol blue, 0.1 M EDTA pH 8.0, 30% glycerol) to a final volume of 40 uL and heated at 95 °C for 5 minutes. Handcast gels were run at 25 mA until the dye ran off the gel and precast gels were run 30 mA. Following, gels were stained with a blue staining solution (0.1% Coomassie Brilliant Blue R-250, 50% methanol, 10% acetic acid). Gels were gently mixed in the solution for approximately 30 minutes. Following staining, gels were destained overnight in destaining solution (50% methanol, 10% acetic acid). Gels were then imaged on a BioRad gel imager.

Results

Plasmid Design

From research done in Dr. David Liu's lab, phage-assisted continuous evolution (PACE) has emerged as a powerful tool to evolve proteins. This system hinges upon the rapid replication rate of the M13 bacteriophage and its inability to replicate without the *pIII* gene. In PACE, *pIII* is replaced with a gene of interest that can restore *pIII* function if it evolves favorably. In one application of PACE, methanol dehydrogenase II (Mdh2) was evolved for better catalytic efficiency by linking formaldehyde production to regaining of *pIII* expression via a formaldehyde sensitive transcription factor. To ensure a constant supply of host bacteria for the phage to replicate in, PACE employs a complex pump system called a lagoon. Due to the complex nature due to the use of phages and a lagoon pump system, we wanted to simplify this model through designing a bacterial non-continuous evolution method which we called BaNCE.

Instead of the phage-bacteria combination used by PACE, we decided to only use *E. coli*. Despite the two-fold reduction in reproduction time, its ease of use and extensive library of research makes *E. coli* a prime candidate for the BaNCE model. To create a directed evolution scheme for *E. coli*, a way to link favorable mutations in Mdh2 to cell survival was devised. To do this, an amino acid dropout model was designed. In order to survive, *E. coli* requires access to all 20 amino acids. In this model, a non-essential amino acid was pursued so, through the BaNCE model, its biosynthesis could be impeded and subsequently restored within the cell. By knocking out a protein in the biosynthetic pathway for an amino acid, *E. coli* cells will only have access to 19 of the 20 required amino acids, thus causing cell death. To regain the function of the amino acid dropout, favorable mutations in Mdh2 will need to be linked to production of a specific amino acid to ensure cell survival.

To identify the best knockout candidate, a literature review was conducted to research past experiments that identified auxotrophic *E. coli* available from the Coli Genetic Stock Center (CGSC). Eleven auxotrophic *E. coli* were examined to identify single gene knockouts that produced bacteria incapable of growing on media lacking a specific amino acid. Nine of the eleven mutants tested were incapable of survival without their amino acid of interest supplemented in growth media. Of those nine, MetA (from the methionine biosynthetic pathway) and PheA (from the phenylalanine biosynthetic pathway) had the highest range of growth linearity with respect to amino-acid-of-interest concentration. Of those two candidates, the *MetA* knockout was chosen for its smaller gene size compared to PheA (Bertels *et al.*, 2012).

The next piece of the design process was to find a method of *in vivo* mutagenesis. PACE uses a specialized, low-fidelity DNA polymerase to make mutations in phage DNA. This polymerase is nonspecific and will reproduce any and all DNA it can access. This is not an issue in PACE since the phages replicate faster than the bacteria, meaning that any mutations to bacterial genomic DNA will not matter because the phage will reproduce before the bacteria, and any mutated bacteria will be washed out and replaced by the lagoon. However, since the BaNCE method requires that the bacteria can reproduce continuously, mutations to genomic DNA are not tolerable. To fit this restriction, a DNA polymerase that is origin specific was required. A highly error prone DNA polymerase I (EP-DNAP I) that exhibits preference for ColE1 origins of

replication was chosen as the best candidate. This polymerase produces mutations at a rate of 8.1 x 10^{-4} mutations per base pair, an 80,000 fold increase relative to wild type DNA polymerase I (Camps *et al.*, 2003).

Figure 1 shows the hypothesized BaNCE system. The first plasmid involved is called pEP (the error-prone polymerase plasmid). This plasmid encodes the Error Prone DNA Polymerase I protein, which will introduce random mutations into the DNA of the second plasmid, called pMdh2 , which encodes methanol dehydrogenase. Once EP DNA Pol I is transcribed from pEP, shown by the purple protein, it will recognize the ColE1 origin of the pMdh2 plasmid and bind to it. The EP DNA Pol I will then initiate pMdh2 replication and introduce random mutations into the growing pMdh2, shown by the wavy red circle. The introduced mutations within *Mdh2* are designated by a red star. Overall, the goal of introducing mutations is to create a gain of function mutation that will improve the rate at which Mdh2 can convert methanol to formaldehyde. Selection of a favorable mutation is based upon sensing an increase in formaldehyde production. This is done via this system's third and final plasmid, pMetA, which contains *MetA* and has been knocked out of the MetA- *E. coli* (Strain JW3973-1 from the Coli Genetic Stock Center) used in this system. Regaining of *MetA* is essential for these auxotrophic cells to survive. *MetA* is regulated under the formaldehyde (Frm) promoter/operator which is bound by the formaldehyde sensitive transcription factor FrmR. A favorable mutation in Mdh2 resulting in a significant increase in formaldehyde concentration causes FrmR to be irreversibly bound to formaldehyde, thus releasing it from the operator and allowing transcription. Once *MetA* is transcribed, methionine biosynthesis is restored to the bacterium.

When fully assembled, the BaNCE model consists of one *E. coli* strain with three transformed plasmids described in Figure 2. The pEP plasmid, responsible for generating mutations in the gene of interest, contains EP DNA Pol I under the IPTG inducible lac promoter. The lac promoter was chosen for its low expression as high levels of mutagenic polymerase could extensively mutate the *E. coli* genome and could risk cell survival. The pMdh2 plasmid, which contains the gene of interest, encodes Mdh2 under a ColE1 origin of replication and is the only plasmid in the system regulated by ColE1. Lastly, pMetA, responsible for linking evolutionary pressure to favorable mutagenesis, contains *MetA* with a FrmR promoter.

In principle, EP-DNA Pol I will mutate *Mdh2*, which will produce varying levels of formaldehyde depending on whether the mutation is favorable or not. If the mutations result in an increase in formaldehyde concentration, *MetA* will be transcribed via release of the FrmR repressor. Increased *MetA* expression will result in increased cell survival, creating a cycle where only favorable mutations are propagated.

Replicon Selection

Plasmid incompatibility poses a potential problem to a three plasmid system. Competition for the same replication machinery occurs between plasmids within the same incompatibility group, meaning that two plasmids from the same group cannot coexist. Furthermore, small RNA produced by a plasmid to regulate copy number can interfere and prevent the coexpression of two plasmids with the same origin in a single bacterium. To avoid this, we identified three unique replicons from different incompatibility groups: ColE1, a high copy number origin to be

used for pMdh2, pSC101, a low copy number origin to be used for pEP, and p15A, a medium copy number origin to be used for pMetA (Rosano & Ceccarelli, 2014).

The agarose gel shown in Figure 3 pictures replicon compatibility of pEP, which contains a pSC101 origin, and pMdh2, which contains a ColE1 origin. Plasmids that are too closely related will be incompatible, so this gel was run to ensure that the two plasmids were able to coexist in the same cell line without replicon competition. Both plasmids were co-transformed into MetA- *E. coli* cells. The top band shown in lane 2 appears at around 6kb, while the bottom band is seen at about 5kb. These are the expected lengths of both pEP (6.2 kb) and pMdh2 (4.7 kb). Thus, as both plasmids were apparent on the gel, it proves that both the pEP plasmid and pMdh2 plasmid can be co-expressed in the MetA- cell line.

Methionine Auxotrophy *E. coli* Selection

To ensure that the auxotrophic *E. coli* could not grow in the absence of methionine, MetA- cells were grown on custom dropout media lacking methionine or dropout media with methionine supplemented back in. No MetA- cells grew on the dropout media, however, growth was robust on the dropout media with methionine. Interestingly, MetA- cells that grew on dropout media were cloudy and less defined than MetA- cells grown on LB agar plates.

pMetA Construction

A pBAD MetA WT plasmid was received from Scripps College, care of the Dr. Peter Schultz lab, to excise the MetA gene for construction of the pMetA plasmid. The MetA gene needed to be cloned out of the pBAD vector and into a p15A replicon plasmid. pStart-T2 was chosen because it contains the p15A origin and an antibiotic selection (tetracycline) which was compatible with pMdh2 and pEP (streptomycin and chloramphenicol, respectively). The insert needed to include restriction sites for insertions, the Frm promoter and operator, and the MetA gene. In order to successfully clone this from the pBAD vector, an overhang PCR was conducted. A set of forward and reverse primers, shown in Figure 3, were designed for the PCR reaction and for successful cloning. Primers included the Frm promoter region, restriction sites, and overhangs for enzymes to sit atop.

Following PCR, an agarose gel, shown in Figure 3, was run to ensure primer amplification. In lane 2, a band was observed near 100 bp, which is the expected size of the MetA insert, showing that PCR and primer amplification were successful.

A DpnI restriction digest was run on the remaining PCR product to ensure that any remaining genomic DNA was removed. Following the digest, the MetA insert underwent restriction cloning into the pStart-T2 plasmid to create pMetA and to excise the ccdB gene, which encodes for the CcdB protein that causes cell death, from the pStart-T2 plasmid. After ligation, MetA- cells were transformed with the ligation reaction mixture and grown overnight in parallel with MetA- cells transformed with pStart-T2 as a negative control. Three colonies grew on the pMetA plate, while no growth was seen on the pStart-T2 plate. Since cells grew, this tentatively shows that the restriction/ligation was a success.

Mutagenesis Experiments

To test the efficiency of EP DNA Pol I, mutagenesis experiments were run to check if it could not only mutate pMdh2, but that it was possible for *E. coli* to express both plasmids. Chemically competent MetA- cells were transformed with varying plasmids plated on LB+Agar plates with associated antibiotics. Sample 1 contained pEP with chloramphenicol resistance, sample 2 with pMdh2 and streptomycin resistance, and Sample 3 with both pEP and pMdh2 with chloramphenicol and streptomycin.

Following transformations, transformed pEP cultures were plated on LB+Cm plates, pMdh2 on LB+Strep plates, and pEP+pMdh2 on LB+Cm+Strep plates to prepare for mutagenesis experimentation. Colonies were grown overnight and growth, with consistent phenotype, was seen on all three plates. On day zero, a single colony from each plate was picked and grown overnight in 3mL of LB broth with associated antibiotics. The following day, day one, cells were passed and grown again overnight to give pEP the chance to further mutate pMdh2. On the second day of growth, cells were passed, induced with IPTG and arabinose for induction, and scaled up into 40 mL of liquid culture and again grown overnight to prepare for a protein extraction and DNA purification.

Following mutagenesis and prior to protein extraction, an aliquot of each sample was taken for DNA purification. Following purification, the selected colonies were sent to be sequenced to determine if the pEP and pMdh2 plasmids were present in *E. coli* in samples 1 and 2 and if pEP was able to successfully mutate pMdh2 in sample 3. The plasmids were sequenced using primers flanking the *Mdh2* gene, which is 1,158 bp. All sequencing results showed an exact match between the sequenced plasmid and the *Mdh2* gene, meaning that we did not observe any mutations in the *Mdh2* coding region resulting from cotransformation with pEP.

A PAGE gel was then run using the 3 samples to test whether proteins were expressed and if co-expression was possible. Figure 7 shows the results of this PAGE gel. Lane 1 contains a protein ladder with weights labeled. Lane 2 contains Sample 2 with EP DNA Pol I, where a band would be expected at about 109 kDa. A dark band can be seen at around 15 kDa. Lane 3 contains Sample 3 with Mdh2, with a band seen at the expected 40.7 kDa. Lane 4 contains both EP DNA Pol I and Mdh2, with a band seen at the appropriate weight for Mdh2, but no band seen for EP DNA Pol I.

This gel shows that it is possible for Mdh2 to be expressed in MetA- *E. coli* cells, but since EP DNA Pol I is not seen at the expected weight, a new set of protein extracts, along with a control, were prepared. In addition, prior to a PAGE gel, a Pierce Assay was run to measure protein concentration in the case that the extracts were either too concentrated or not concentrated enough for EP DNA Pol I to appear clearly on a PAGE gel.

Figure 8 shows the results of this Pierce Assay. For this assay, the standard was first diluted fivefold, and followed by five 2x serial dilutions. The samples were tested at their original undiluted concentration as well as after they underwent three 2x serial dilutions. Sample 1, which is a negative control of protein extract from MetA- cells with no transformed plasmids, had a concentration of 1.8 mg/mL. Sample 2, cells transformed with pEP, had a concentration of 3.0

mg/mL. Sample 3, cells with pMdh2, had a concentration of 8.9 mg/mL. Sample 4, cells with both pEP and pMdh2, had a concentration of 14.0 mg/mL. Table 3 shows OD values corrected for the blank measurement. From this data, to create a uniform protein concentration, Sample 1 was left undiluted, Sample 2 was diluted to a 2-fold, Sample 3 was diluted 4.5-fold, and Sample 4 was diluted seven-fold.

The resulting PAGE gel from the protein extract dilutions can be found in Figure 9. Lane 1 of the gel contains a protein ladder. Lane 2 contains Sample 1, the MetA- cell extract negative control. Lane 3 contains Sample 2, lane 4 contains Sample 3, and lane 5 contains Sample 4. In both lanes 3 and 4, the pMdh2 plasmid can be seen with a band at the expected 40.7 kDa. While no band can be seen for pEP at the expected 109 kDa, a band can be in boths lanes 3 and 5 at around 20 kDa. In lanes 2 and 3 of Figure 7, this band can also be seen at around 20 kDa. Since this band does not appear in lane 2 of Figure 9, the control extract, it is likely that this band correlates to the presence of pEP. Further experimental evidence needs to be collected to make this conclusion.

Discussion

Directed evolution has proven itself as a powerful tool for creating new and improved proteins. In this report, we outline a novel system for non-continuous directed evolution which aims to conserve the elegance of the PACE model while simplifying some of its more complex features. In addition to the design of a Bacterial Non-Continuous Evolution (BaNCE) system, we report preliminary findings which test the feasibility of this model. We demonstrated that the JW3973-1 strain of *E. coli* (referred to as MetA- in this paper) is capable of being transformed with two plasmids using a ColE1 and pSC101 replicon. Additionally, we have shown that these cells are incapable of growing on our custom dropout media which lacks methionine, but that growth can be restored by supplementing methionine back into the media. We also present data that suggests the successful construction of BaNCE's third and final plasmid, pMetA, which includes a p15A replicon and the *MetA* gene regulated under the Frm promoter/operator. Lastly, we present data which assesses the ability of MetA- cells to coexpress multiple proteins from separate plasmids and the ability of EP DNA Pol I to generate mutations *in vivo* under low expression conditions. These data inform us on the potential of the new BaNCE model, possible modifications to the system, and future experiments to further appraise the feasibility of BaNCE.

Unfortunately, due to the COVID-19 pandemic, we were unable to complete the research we had hoped to finish since WPI's campus was put on lockdown and students were asked to return home and take classes remotely for the remainder of the academic year. For this reason, the results of this project ended more abruptly than expected, leaving unfinished work due to the sudden closure of the lab. As a result, the discussion section will place a heavy focus on speculations and possible conclusions of the completed research, next steps that had been planned to be completed, as well as possible future directions for this project.

Through methionine auxotrophy selection experiments, shown in Figure 4, we were able to prove that MetA- cells cannot grow without a methionine supplement in pre-made growth media. This shows that, if the pMetA plasmid were to express *MetA*, thus restoring methionine biosynthesis to the MetA- cells, they should grow on dropout media. Regardless, further experimentation of MetA- cells should be run as a phenotype change in the cells can be seen in Figure 4 when they are grown on dropout media with a methionine supplement (4B) versus LB broth (4C) to ensure that this change does not affect BaNCE experimental results. Once all three BaNCE plasmids are transformed into the cell line, a growth test on dropout media will be needed to show that the system works as predicted.

Prior to halting of experimentation, the final plasmid in the BaNCE system, pMetA, was constructed. The data shown in Figure 6, whose gel proves that primer design and cloning were successful. In addition, growth testing following ligation showed that MetA- cells transformed with pMetA were able to grow, while those in the unrestricted pStart-T2 plasmid did not, proving that the ccdB gene was excised. While construction of pMetA should be confirmed through sequencing, this promising data suggests that all three BaNCE plasmids have been constructed, and the system can be tested to completion.

Since all three plasmids are available for BaNCE system experimentation, a triple transformation into MetA- cells needs to be performed to confirm replicon compatibility. Figure 3 shows an agarose gel that proves that not only can both pEP and pMdh2 be transformed into MetA- cells, but the two distinct bands in lane two prove replicon compatibility. Through a triple transformation of the three BaNCE plasmids, three distinct bands on a gel would prove that all three replicons are compatible. Replicon compatibility is needed for overall success of the BaNCE model as, if the plasmids were competing for the same replicon, transformation and plasmid perpetuity would not be possible.

Despite attempting several mutagenesis experiments, we were unable to observe the generation of any mutations in ColE1 plasmids. In these experiments, MetA- *E. coli* were cotransformed with pEP and a ColE1 target plasmid, for which we used pMdh2. Although sequencing showed no mutations in the *Mdh2* coding region, this does not disprove the introduction of mutations elsewhere in the plasmid. Furthermore, because the mutations generated by EP DNA Pol I are random, the average sequence of a collection of plasmids may appear unchanged, even though the plasmids may individually have each accumulated a small number of mutations. For sequencing of a population of plasmids to be a useful tool to determine the mutagenicity of EP DNA Pol I, it may be necessary to concentrate plasmids with the same mutations so that these may be observed using sequencing. The use of a ColE1 plasmid encoding GFP could help identify when mutations are generated by a change in the bacteria's phenotype. If colonies transformed with both pEP and a GFP/ColE1 plasmid change from glowing to dark, this could be indicative of a mutation in the coding region of GFP. However, it is important to note that a change in phenotype could also result from mutations in the promoter/operator of GFP, which could hinder transcription, or from mutations in the replicon, which would hinder plasmid replication. The latter would also require the cells to gain resistance to associated antibiotic selection. Regardless of the method, proving EP DNA Pol I's ability to generate targeted mutations *in vivo* is essential to establishing BaNCE's potential as a system for directed evolution.

Likewise, the ability for MetA- cells to express the key proteins encoded in each of BaNCE's plasmids is critical to the model's success. Figure 7 shows the results of a SDS-PAGE gel using protein extracts of MetA- cells transformed with either pEP, pMdh2, or both plasmids. In lane 3, labeled Mdh2, there is a large band located around 40 kDa. Given the size of the band and its location, this was determined to be Mdh2, which has a molecular weight of 40.7 kDa. A fainter band can be seen at the same location in lane 4, labeled EP DNA Pol I and Mdh2, suggesting that Mdh2 is successfully expressed in both pMdh2 and pEP/pMdh2 transformed cells. However, because the bands in lanes 3 and 4 differ greatly in intensity, it was concluded that protein concentrations between samples differed greatly. In contrast to Mdh2, EP DNA Pol I, which has a weight of 109 kDa, is not apparent in lane 2, labeled EP DNA Pol I, nor in lane 4. There are heavy bands around 15 kDa and 5 kDa that do not appear in the Mdh2 only lane, however, the nonuniform concentrations make it difficult to determine whether these are a product of pEP's presence or a result of different concentrations of host cell proteins. Although EP DNA Pol I's absence on this gel is alarming, it was expected to have lower expression levels than Mdh2 given the differences in copy number and induction between pEP and pMdh2. pEP employs the low

copy number pSC101 origin and regulates EP DNA Pol I expression under the lac system, which has a history of temperamental induction. pMdh2 uses the high copy number ColE1 replicon and arabinose induction. These distinctions help explain the differences between Mdh2 and EP DNA Pol I concentrations on the gel, but they do not explain the complete absence of EP DNA Pol I. To further investigate this, the experiment was reattempted with the addition of a Pierce 660 assay to normalize protein concentrations between samples as well as a negative control to help visualize differences in protein concentrations.

Figure 9 shows the results after protein concentration normalization. Once again, Mdh2 has a clear presence in both the lanes containing pMdh2, however, now band intensities are comparable between the two samples. Nevertheless, there is once again an absence of any significant bands around 110 kDa where EP DNA Pol I is expected to be found. Interestingly, there is a significant band present at 20 kDa, which is noticeably darker in both samples containing pEP (lanes 3 and 5), than it is in the control. This band seems to correlate to the presence of pEP, but we cannot conclude that it signifies the expression of EP DNA Pol I. Although these results do not confirm expression of EP DNA Pol I, they do not disprove it either. The previously described conditions causing low EP DNA Pol I expression may make this experiment non-ideal for identifying expression of EP DNA Pol I since it may not be present in concentrations high enough to distinguish itself from other proteins in the extract. Introducing a protein tag to EP DNA Pol I and subsequent purification may be required to determine conclusively whether or not the current plasmid system allows for expression of the polymerase.

Through the use of the BaNCE system, we hoped to create an easier, more user-friendly model for directed evolution that could be performed in academic labs without the complexity of phages. Seeing as the PACE method poses implications for methane pathways, we hope that the simplified BaNCE method could have potential to create pathways that could convert methane into energy sources for bacteria.

Figures and Tables

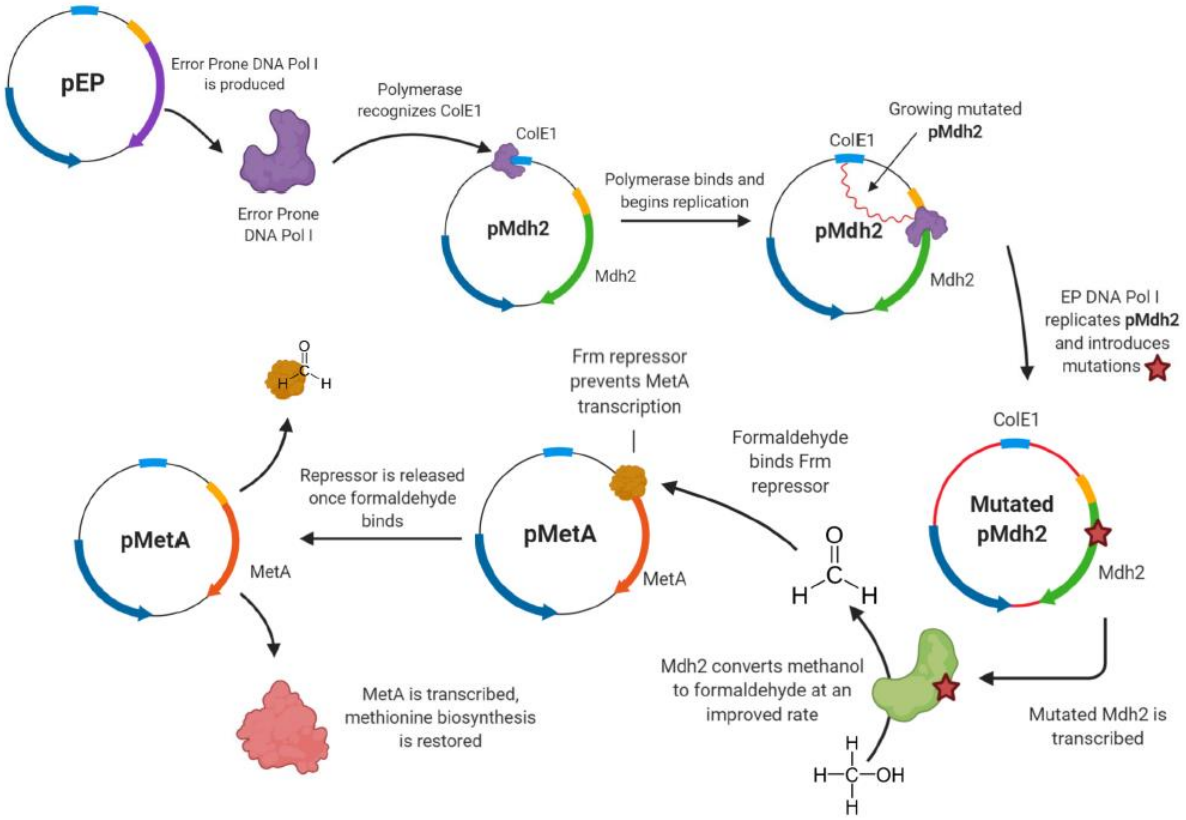

Figure 1. BaNCE experimental design. Beginning with the pEP plasmid, the error prone (EP) DNA Polymerase I will be transcribed. This polymerase will then recognize the ColE1 ORI of the pMdh2 plasmid and bind to it. The EP DNA Pol I will then replicate the pMdh2 plasmid and introduce random mutations, thus creating a mutated pMdh2, delineated by the wavy red circle. The introduced mutations within the Mdh2 gene are shown by a red star. The goal is to create a gain of function mutation that will allow Mdh2 to convert methanol to formaldehyde at an improved rate. As a result, the mutated gain of function Mdh2 will be selected for continued replication and transcribed. Thus, the formaldehyde from Mdh2 converted methanol will bind to the formaldehyde (Frm) repressor on pMetA, shown by the orange molecule. Once formaldehyde binds to the Frm repressor, it will be released from pMetA, thus allowing for the transcription of the MetA gene. As a result, methionine biosynthesis will be restored and the MetA- *E. coli* cells will be able to grow and propagate.

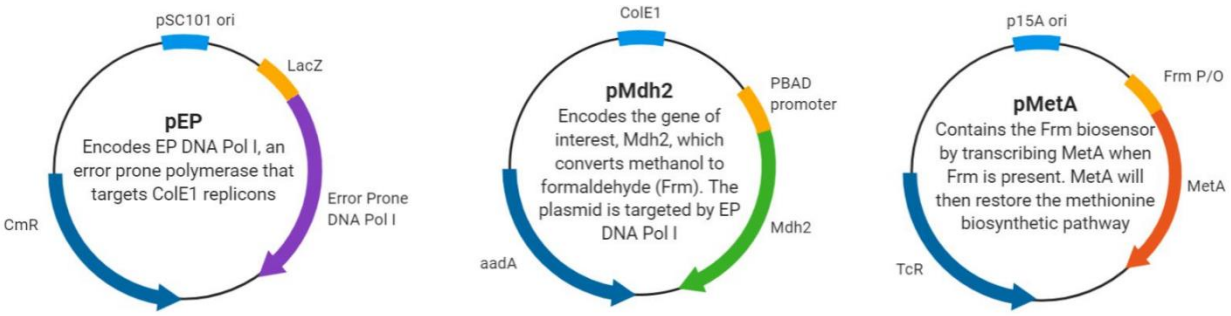

Figure 2. Simplified Plasmid Maps of plasmids used. The pEP plasmid that will make random mutations to the pMdh2 plasmid through the use of Error Prone Polymerase I. The pMdh2 plasmid contains the gene of interest, Mdh2 (methanol dehydrogenase), that will convert methanol to formaldehyde. pMdh2 mutations from the pEP plasmid will produce varying levels of formaldehyde depending on a favorable mutation. The pMetA (Methionine) plasmid links evolutionary pressure to the favorable mutagenesis of the pMdh2 plasmid.

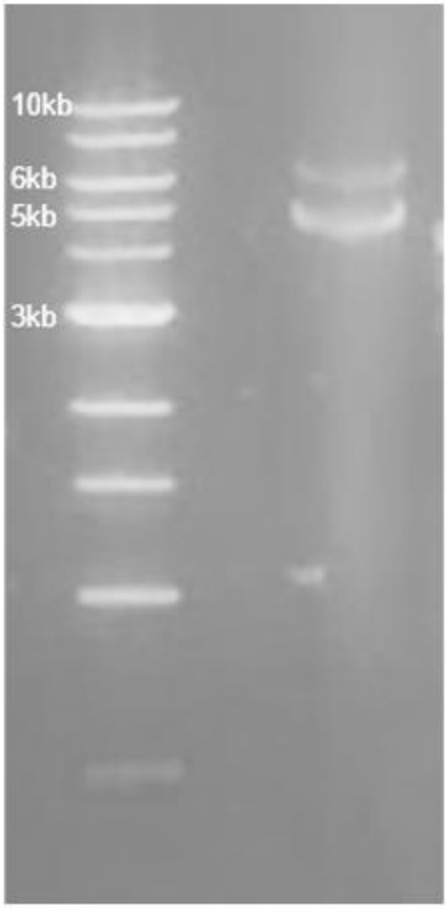

Figure 3. Agarose gel of replicon compatibility. Lane 1 pictures a 2-log DNA ladder with associated lengths labeled. Lane 2 shows a co-transformation of pEP and pMdh2 done in MetA-*E. coli* cells. Bands can be seen at 6kb, which is near the expected length of pEP (expected 6.2 kb), and just shy of 5kb, which is near the expected length of pMdh2 (expected 4.7 kb). These two distinct bands prove replicon compatibility of these two plasmids.

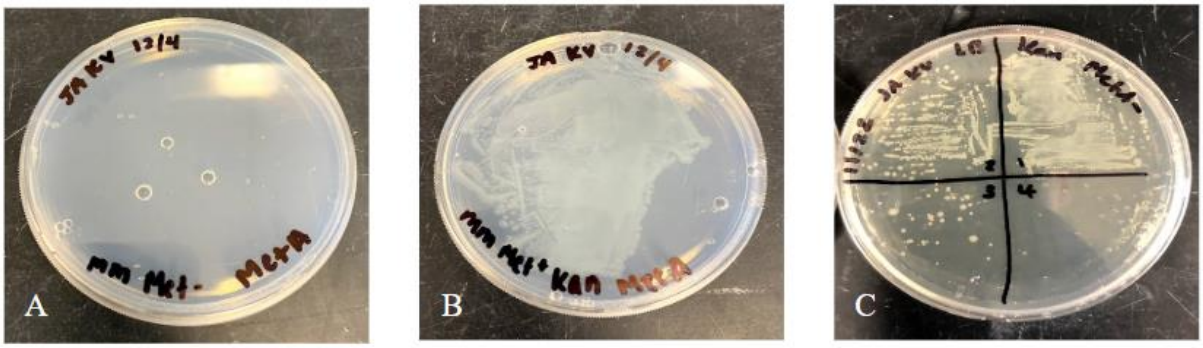

Figure 4. MetA- knockout cell test. To ensure that the auxotrophic *E. coli* did not grow without methionine, the MetA- cells were grown on plates with dropout media that either contained 19 amino acids (excluding methionine) supplemented back into the media or all 20. Figures 3A-3C show the agar plates from this growth test. Figure 3A shows cells grown on dropout media without methionine where, as expected, there was no cell growth. The small circles seen on the plates are air bubbles within the agar. Figure 3B shows cells grown on dropout media with all 20 amino acids, where, as expected, there was cell growth. 3C shows the MetA- cells grown on an LB agar plate for colony phenotype comparison between medias.

Forward Primer:

5'-
GAGGTGGGCCCTTGACATATAGAATACCCCCCTATAGTATATTGCATGCAGATGATG
AGGTGCGAAATGCCGATTCGTGTGCCGGAC -3'

Apa1 Restriction Site

Reverse Primer:

5'-ATTCGCTCGAGTTAATCCAGCGTTGGATTCAT-3'

Xho1 Restriction Site

Figure 5. MetA gene primer design. Both a forward and reverse primer were designed to clone the MetA gene out of the pBAD MetA WT host plasmid. On the forward primer, text highlighted in blue is a part of the frm promoter region. Underlined text in the forward primer is where frmR (the formaldehyde repressor) will bind. In both the forward and reverse primers, text highlighted in red are each primer's corresponding restriction site (Apa1 and Xho1, respectively). Black text on both primers indicate overhangs for the restriction enzymes to sit on top of.

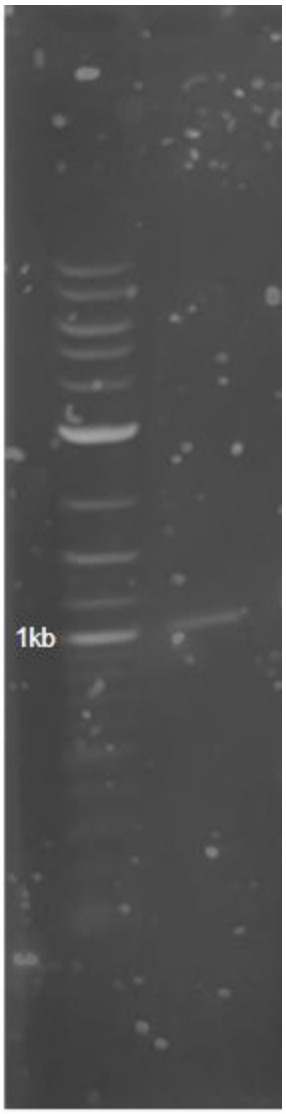

Figure 6. Agarose gel of MetA PCR insert. Through using the designed primers the *MetA* gene of interest from the pBAD MetA WT host plasmid was amplified through PCR. Lane 1 contains a DNA ladder while Lane 2 contains the *MetA* insert. The insert band appears at around 1000 bp, as expected, showing the PCR was successful and *MetA* was successfully cloned.

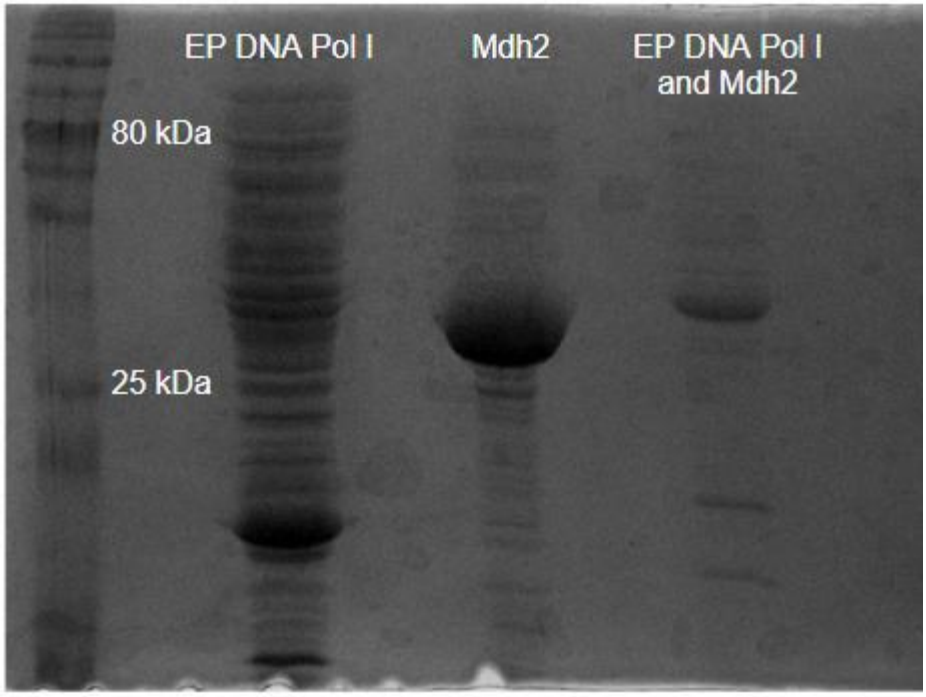

Figure 7. PAGE gel for co-expression tests. Lane 1 contains a protein buffer. Lane 2 contains a sample of protein extract with cells that have been transformed with pEP. A band can be seen in this lane at 15 kDa. Lane 3 contains a sample of protein extract from cells transformed with pMdh2. A band in this lane is seen at 40 kDa. Lane 4 contains a sample with proteins purified from MetA- *E. coli* cells transformed with pMdh2 and pEP. Bands in this lane can be seen at 40 kDa, 15 kDa, and 5 kDa.

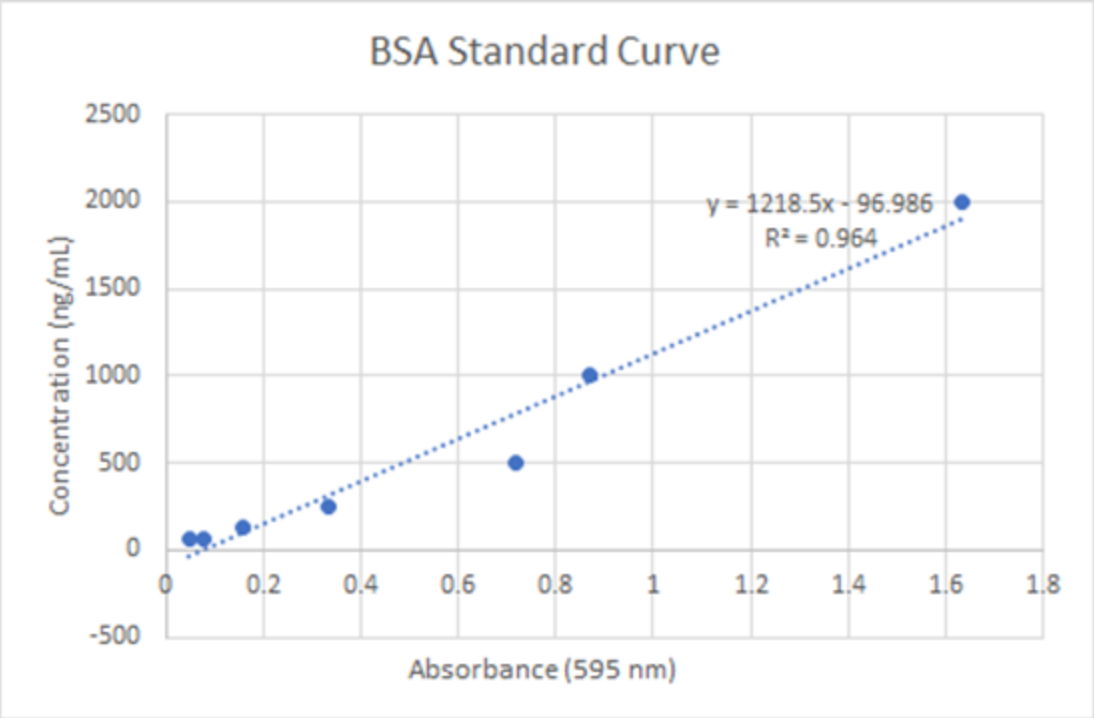

Figure 8. Pierce 660-Assay standard curve.

	1	2	3	4	5	6	7
Standard	1.633	0.871	0.719	0.334	0.158	0.078	0.047
Control	1.329	0.721	0.454	0.156	0.183		
pEP	1.612	1.161	0.705	0.376	0.368		
pMdh2	2.381	1.903	1.512	0.994	1.036		
pEP & pMdh2	2.092	1.897	1.806	1.521	1.591		

Table 1. Pierce 660-Assay OD values corrected for blank measurement.

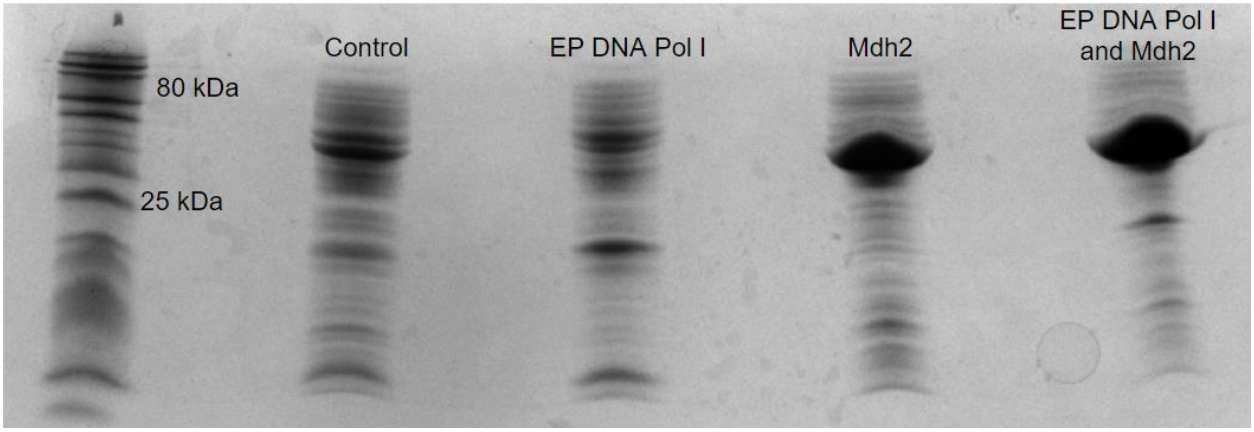

Figure 9. PAGE gels for co-expression tests with normalized protein concentrations. Lane 1 contains a protein ladder. Lane 2 shows a control of only MetA- cell protein extract with no transformed plasmids. Lane 3 contains protein extract of MetA- cells transformed with pEP. ALane 4 contains protein extract of MetA- cells transformed with pMdh2. Lane 5 contains protein extract of MetA- cells transformed pEP and pMdh2. Expected size of Mdh2 and EP DNA Pol on are 40.7 kDa and 109 kDa, respectively. Mdh2 appears at the expected length in both lanes 4 and 5, while EP DNA Pol I appears near 20 kDa.

References

Bertels, F., Merker, H., & Kost, C. (2012). Design and Characterization of Auxotrophy-Based Amino Acid Biosensors. PLoS ONE, 7(7). doi: 10.1371/journal.pone.0041349

Camps, M., Naukkarinen, J., Johnson, B. P., & Loeb, L. A. (2003). Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. Proceedings of the National Academy of Sciences of the United States of America, 100(17), 9727–9732. doi:10.1073/pnas.1333928100

Carlson, J., Esvelt, K., and Liu, D. (2011). A System for the Continuous Directed Evolution of Biomolecules. *Nature*, 472, pp. 499-503). doi:10.1038/nature09929.

Climate Action Reserve. (2019, September 29). *Climate Change Facts*. Retrieved from https://www.climateactionreserve.org/resources/climate-change-facts/.

Chen, K., & Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5618–5622. doi:10.1073/pnas.90.12.5618

Cobb, R. E., Chao, R., & Zhao, H. (2013). Directed Evolution: Past, Present and Future. *AIChE journal. American Institute of Chemical Engineers*, *59*(5), 1432–1440. doi:10.1002/aic.13995

EPA. (2019, September 30). *Overview of greenhouse gases.* Retrieved from https://www.epa.gov/ghgemissions/overview-greenhouse-gases

NASA. (2019, July 9). *Climate Change Evidence: How Do We Know?*. Retrieved from https://climate.nasa.gov/evidence/.

The National Academies of Sciences, Engineering, and Medicine. (2008). *Evolution in agriculture: the domestication of wheat.* Retrieved from http://nationalacademies.org/evolution/EvolutioninAgriculture.html.

Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172. Published 2014 Apr 17. doi:10.3389/fmicb.2014.00172

Roth, T. B., Woolston, B. M., Stephanopoulos, G., & Liu, D. R. (2019). Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. ACS Synthetic Biology, 8(4), 796–806. doi: 10.1021/acssynbio.8b00481

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature, 370(6488), 389–391. doi: 10.1038/370389a0

UNFCCC. (2019, September 30). *Why Methane Matters*. Retrieved from https://unfccc.int/news/new-methane-signs-underline-urgency-to-reverse-emissions.