Journal of Experimental Microbiology and Immunology (JEMI)

Building an Orthogonal Replication Systems for Performing Directed Evolution in *Escherichia coli*: A Strategic Review and a Summary of the Initial Steps in Cloning Bacteriophage T7 gp4 Primase/Helicase

Wendy Ma, Angela Phan, Riley Walsh, and Kevin Ye
Department of Microbiology and Immunology, University of British Columbia

Techniques for directed evolution are commonly used in industry to induce mutations in a gene of interest (GOI). Rapid mutagenesis of a GOI within a host cell can be used to expedite the process of directed evolution. A system that replicates independently of the host genome is termed orthogonal. An orthogonal replication system can use an engineered error prone DNA polymerase recognizing a specific origin sequence to induce rapid mutagenesis of a GOI. The 2014 University of British Columbia (UBC) iGEM team has designed a strategy using the bacteriophage T7 replisome to introduce an orthogonal replication system into *Escherichia coli*. iGEM UBC have cloned and expressed three of the four genes required for T7 replication (gp1, gp2.5, and gp5). Cloning the fourth gene has been problematic; repeated mutations have been detected in a specific region of the constructs containing gp4 primase/helicase. In this study, we have attempted to correct a Gly^{313}→ Val mutation in one construct but were unsuccessful. Literature review suggests that gp4 primase/helicase expression in *E. coli* is toxic, which may explain the root cause of these mutations as they may alter the structure of the helicase domain. The toxicity may also be due to the presence of two proteins from the same coding region of gp4: 4A and 4B. A review of orthogonal replication systems, the current strategy to engineer an orthogonal replication system in *E. coli*, structural analysis of gp4 and our results suggest that mutations in the gp4 constructs may have been selected due to the toxic effects associated with the expression of the helicase domain in *E. coli*.

Biological systems have complex evolutionary cycles due to natural selection and diversification (1). However, natural selection is limited because it is a slow process and does not necessarily target a specific gene (3, 9, 13). To overcome these limitations and allow study of protein function, the concept of directed evolution arose, which refers to all techniques derived to induce mutagenesis and produce a desired mutant product (22). Techniques for directed evolution are used in industry to increase mutations in a target sequence to study any resulting enhanced or novel protein functions (2, 22). Common steps involved in the process of directed evolution include DNA extraction, random mutagenesis of the target sequence, transformation of the gene into the host, selection of mutants, and screening for the desired function (2, 13). Unfortunately, current methods for this process are quite laborious, specifically due to isolation and transformation of DNA as well as screening large panels of proteins (4, 13).

A system that could induce continuous, rapid mutagenesis inside the host cell rather than isolating DNA for other methods of mutagenesis would greatly reduce the current efforts involved with directed evolution (4). Mutagenesis of a target gene within a host requires engineering a system that replicates separately from the host genome and therefore does not induce mutation in the host genome (2). This could reduce the effort involved with current directed evolution methods by eliminating any defective mutations before screening (2). Such a system can be termed orthogonal and refers to replication of a plasmid carrying the gene of interest (GOI) independent of host genome. Orthogonal replication aims to rapidly mutate the GOI using an engineered error prone DNA polymerase (VanInsberghe, M, unpublished data).

A few studies have been published which describe systems that increase mutagenesis of a target gene within a host. Camps et al. performed a study that induced mutagenesis of a target gene in *Escherichia coli* using a two-plasmid system (2). By introducing mutations in the three domains of DNA polymerase 1 (DNA Pol 1), Camps et al. displayed increased mutagenesis in the Pol 1-dependant plasmid (2). However, this error prone DNA Pol 1 recognizes plasmid sequence relative to chromosome sequence, allowing for possible mutagenesis to the host genome and therefore the system is not orthogonal (2). In another study Esvelt et al. used phage-assisted continuous evolution (PACE) to allow for many rounds of evolution in a single day using *E. coli* (4). A phage reservoir utilizing a modified bacteriophage life cycle is used to transfer evolving genes from host cell to host cell as *E. coli* is passed through the reservoir (4). Despite the ability to observe many cycles of evolution, this method can only be used with genes involved in phage production (4). Furthermore, studies by Ravikumar et al. have been successful using yeast as a host (13). An extranuclear replication system has been created in Saccharomyces cerevisiae, which utilizes an orthogonal DNA plasmid paired with a DNA polymerase (13). The plasmid used in this system was a cytoplasmic plasmid from another yeast species, Kluveromyces lactis, chosen for reasons including high copy number and autonomous replication (13).

To engineer an orthogonal replication system in *E. coli*, the lytic bacteriophage T7 can be employed (5).

Bacteriophage T7 has a simple replication system and the replisome only involves four of its own genes in addition to an origin sequence and a host protein processivity factor called thioredoxin (TrxA) that are needed to recognize and replicate DNA. The four genes in the T7 genome that are involved with replication are gp1: RNA polymerase (RNAP), gp2.5: ssDNA binding protein (ssDNA-BP), gp4: primase/helicase, and gp5 DNA polymerase (DNAP). The resulting proteins form a structure that allows for coupled synthesis of both the leading and lagging strands (5, 6, 8). T7 DNA polymerase recognizes a specific and well defined viral origin sequence that is not present in the host genome, so theoretically, there is promise to develop a mechanism of mutagenesis independent of the host genome (14).

This review aims to organize and present the current knowledge about the use of bacteriophage T7 as an orthogonal replication system in *E. coli*. Hopefully this will provide a starting point for the future development of orthogonal replication systems and ultimately the ability to expedite directed evolution. We include information about a venture that is currently underway engineering an orthogonal replication system in *E. coli*, with a large focus on a point mutation in gp4 primase/helicase and its suspected toxicity. Additionally, previous studies have identified two proteins that can be expressed from gp4 in the same reading frame: 4A and 4B. Furthermore, we aimed to investigate and describe our efforts in reverting the point mutation in the gp4 helicase domain, examining *E. coli* expression of gp4, and knocking out 4B using site-directed mutagenesis (SDM).

MATERIALS AND METHODS

Bacterial Strain and Growth Conditions. *E. coli* DH5α was provided by the MICB 421 culture collection from the Department of Microbiology and Immunology at the University of British Columbia (UBC). The strain lacks endonucleases, and thus is frequently used for cloning applications according to the information on the Life Technologies website (http://www.lifetechnologies.com/ca/en/home/life-science/cloning/competent-cells-for-transformation/chemically-competent/dh5alpha-genotypes.html). Overnight cultures were grown in either Luria-Bertani (LB) broth for the preparation of competent cells or LB supplemented with 25ug/mL chloramphenicol (Sigma®) for plasmid isolation. They were incubated at 37^{O}C with shaking at 150 RPM for 18 to 24 hours.

Plasmids. The isolated pSB1C3-A plasmid is a high-copy plasmid containing a multiple cloning site, chloramphenicol resistance and the gp4 primase/helicase. It was assumed to have gp4 under the control of the pBAD promoter but after sequencing, it was confirmed that the plasmid was an intermediate vector lacking pBAD promoter. The plasmid was provided by Michael VanInsberghe of the 2014 UBC iGEM team (Figure 1). A map of the pSB1C3-B plasmid with the pBAD promoter can be found in Supplementary Figure 1. The pSB1C3-A plasmid is 4.071 kb in size, encodes for the bacteriophage T7 gp4 primase/helicase gene with a single base pair mutation where guanine (G)938 was substituted for thymine (T)938. The isolated pUC19 plasmid encodes a lacZ gene.

Calcium Chloride Treated Competent Cells Preparation. Cells were incubated in LB broth to OD600 of 0.2 to 0.3 and transferred to Oak Ridge tubes (NalgeneTM ¬¬- 3114-0030) to be centrifuged at 14,000 RPM for 5 minutes at 4^{O}C. Cells were resuspended in 100 mM CaCl2 at 1/20th volume, incubated on ice for 30 minutes, and stored in 15% glycerol and 0.1 M CaCl2 at 1/500th volume.

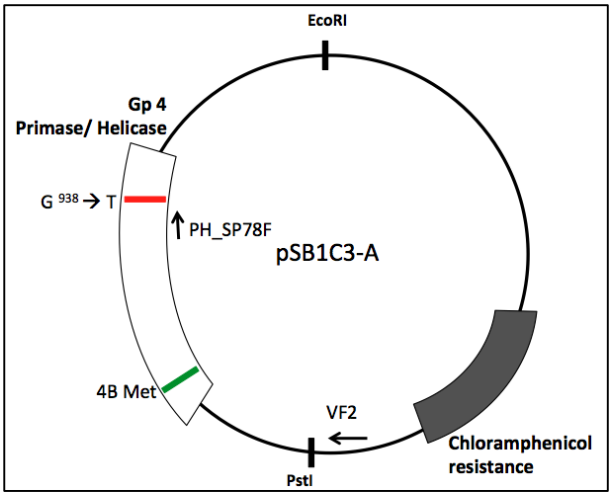

FIG 1 Plasmid pSB1C3-A The map of the plasmid provided by Michael VanInsberge from the 2014 UBC iGEM team. This plasmid vector has chloramphenicol resistance and gp4 on it. This specific construct has a G^{938}→T point mutation, which is the focus of our study and shown in red. The green line marks the start site for gene 4B within the gp4 primase/helicase. The PstI and EcoRI restriction sites are shown because they were used for a restriction digest of the plasmid products. The sequencing primers PH_SP78F and VF2 are primers to confirm the reversion of point mutation on the pSBIC3-A and creation of the 4A' coding sequence respectively.

Synthetic Oligonucleotides used as Primers and Provided Primers. Synthetic oligonucleotides were ordered through UBC-NAPS Portal at Integrated DNA Technologies and they were designed to overlap in the region of the desired base pair changes. The primers' design and melting temperatures calculations (Tm = 81.5 + 0.41(%GC) – 675/ N – % mismatch) were made according to Stratagene QuikChange Site-Directed Mutagenesis Kit - Primer Design Guidelines (http://sevierlab.vet.cornell.edu /resources/Stratagene-QuikchangeManual.pdf), and PrimerX (http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi). The forward primer (1F) and reverse primer (1R) were used to revert the point mutation to wild-type (WT). The sequence of 1F was 5'-CATTATGGTCACTTCCGGTTCCGGTATGGGTAAG-3' and 1R was 5'-CTTACCCATACCGGAACCGGAAGTGACCAT AATG-3'. The forward primer (2F) and reverse primer (2R) were used to knock out 4B expression where the Met64 (TAA) substitution to Leu64 (TAA), which were specifically at the 190 to 192 bases in the pSBIC3-A plasmid. The sequence of 2F was 5'-CCCTCAGGAGGTAAACCATTAACTTACAACGTGTGGAA C-3' and 2R was 5'-GTTCCACACGTTGTAAGTTAATGGTTTA CCTCCTGAGGG-3'. The forward primer (CF) and reverse primer (CR) were used to knockout the lacZ gene expression by introducing a STOP codon in the pUC19 plasmid and were designed in accordance to the TagMaster® Site-Directed Mutagenesis Kit protocol (http://www.gmbiosciences.com/ TagMaster%20Site-Directed%20Mutagenesis%20Kit_Manual. pdf). The sequence of CF was 5'-GTACCCGGGGATC CTCTCAAGTCGACCTGCAGGCAT-3' and CR was 5'-ATGCCTGCAGGTCGACTTGAGAGGATCCCCGGGTAC-3'. Michael VanInsberghe supplied additional primers, such as VF2, PH_SP787, FPH_SP294F, for sequencing or PCR (polymerase chain reaction) (Figure 1).

Site-Directed Mutagenesis and DpnI Digestion. SDM involved amplification of DNA fragments from the plasmid template, which was achieved with 25 uL reactions consisting of 0.02 U/uL Phusion High-Fidelity PCR DNA polymerase (New England Biolabs Inc.), 350 nM of each mutagenic forward and reverse primers (1F, and 1R or CF and CR), 0.2 mM dNTP, 3% dimethyl sulfoxide (DMSO), 2 ng/uL of template DNA, and water. The reactions containing pSBIC3-A plasmid were referred to as SDM products. The negative controls included reactions without the DNA polymerase and reactions without template DNA. The positive control included reactions with the pUC19 plasmid. The reactions were subjected to an initialization step (30 seconds (s), 98^{O}C), 30 cycles of denaturation (15 s, 98^{O}C), annealing (60 s, 55^{O}C), and extension (2 minutes, 70^{O}C), and a final elongation step (10 minutes, 72^{O}C). With the exception of the negative controls without template DNA, the SDM products were subsequently digested with 1 uL of DpnI in order to remove parent plasmid (New England Biolabs Inc.) at 20 U/uL and incubated for 1 hour at 37^{O}C.

Preparation of Plates. Luria agar plates were prepared containing 25 µg/mL chloramphenicol (Sigma®) and supplemented with either 0.2% D-glucose or L-arabinose for the *E. coli* transformants containing the SDM gp4 products. Luria Agar plates were prepared containing 100 µg/mL ampicillin (Sigma-Aldrich®) and with 40 uL of 0.1M Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 uL of 20mg/mL 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal), which were spread on the plates 30 minutes prior plating the cells, for the *E. coli* transformants containing the SDM lacZ gene products.

Transformation of Site-Directed Mutagenesis Products. 20 uL of competent cells was transformed with 2 uL of the SDM products, negative controls or the positive controls, which were now the original pSB1C3-A plasmid in addition to the pUC19 plasmid containing the lacZ gene with the STOP codon. The cells were then incubated on ice for 20 to 40 minutes, heat shocked in 42°C water bath for 45 to 60 s. 180 uL of LB broth was added to the mixture and incubated at 37°C room for 1 hour according to the protocol by the Hancock Lab (http://cmdr.ubc.ca/bobh/methods/CaCl2TRANSFORMATIONOFECOLI.htm). The cells containing the SDM products, negative controls or the original pSB1C3-A plasmid were subsequently incubated on the Luria Agar plates containing 25 µg/mL chloramphenicol with either 0.2% D-glucose or L-arabinose. The cells containing the lacZ gene positive control was plated on the Luria Agar plates containing 100 µg/mL ampicillin, IPTG and X-gal, and would be examined for white colonies. The plates were incubated over two nights and restreaked on the same types of plates for purity.

Plasmid Isolation. A colony was obtained from each Luria Agar plates containing 25 µg/mL chloramphenicol with either 0.2% D-glucose or L-arabinose and incubated in LB broth containing 25 µg/mL of chloramphenicol overnight. The plasmids were isolated as per the instructions using the Invitrogen PureLink Quick Plasmid Miniprep Kit. The concentrations of the plasmids were determined using Thermo Scientific Nanodrop 2000C Spectrophotometer.

Proposed Sequencing. The primer PH_SP787 was used to confirm the reversion of point mutation on the pSBIC3-A plasmid while the primer V2F was used to confirm the creation of the 4A' coding sequence. Sequencing of the plasmids were to be performed using Sanger DNA Sequencing services of GeneWiz.

Plasmid Digestion and Agarose Gel Electrophoresis. 20 ng/uL of SDM products or 20 ng/uL plasmids, were digested with 0.2 U/uL of EcoRI, and 0.2 U/uL of PstI, 1X NEBuffer 3.1, and water. The reactions was incubated at 37^{O}C for 1.5 hours and subsequently resolved on 1% agarose gel using 1X TBE buffer at 100 volts for an hour.

RESULTS

Cloning Strategy to Create an Orthogonal Replication system using the bacteriophage T7 replisome and *E. coli*. The 2014 UBC iGEM team has utilized the components of the bacteriophage T7 replication system towards engineering an orthogonal replication system in *E. coli*. Upon completion, this system was designed to operate using two plasmids: the first containing the T7 DNA replication machinery, and the second containing both the T7 replication origin and a gene of interest (http://2014.igem.org/Team:British_Columbia/ProjectChassis).

To construct the plasmid containing the replication machinery, the T7 DNAP, RNAP, ssDNA-BP and primase/helicase were first independently cloned into the pSB1C3-A and then the pSB1C3-B vector with the pBAD promoter (Figure 2B). Once all of the genes can be expressed independently they will then collectively be assembled on a single pSB1C3-B plasmid, each under the pBAD promoter and a terminator. Once this recombinant plasmid is transformed into *E. coli*, these cells will be able to express the components necessary to recapitulate the T7 replisome. The GOI will be transformed into these cells on a separate plasmid vector called JH-pSB2K3, which contains the T7 origin sequence (Figure 2B). This viral origin sequence is recognized by the T7 DNAP, which will be engineered to be error prone in the final orthogonal replication system as described on the 2014 iGEM UBC website (http://2014.igem.org/Team:British_Columbia/ProjectChassis).

Cloning and expression of each gene was done in the *E. coli* strain DH5α. The DNAP and the ssDNA-BP expressed at high levels, however the RNAP expression was not detectable on an SDS-PAGE gel and the primase/helicase seemed to be defective according to the 2014 iGEM UBC website (http://2014.igem.org/Team:British_Columbia/ProjectChassis). The problems encountered with the primase/helicase appear to be due to mutations occurring in the gene itself after it is cloned into the DH5α cells. Sequence analysis of four independent clones of the construct revealed a 200 bp deletion, a point mutation, and two clones had a 3-base deletion in the same area of gp4 (http://2014.igem.org/Team:British_Columbia/ProjectChassis). The point mutation is shown in both Figure 1 and Figure 3. Regardless of the setback with gp4, iGEM will continue to work on this project in order to address the following questions; Do the individual components express in *E. coli*; Does the system function orthogonally and recognize the specific T7 origin sequence; Does the system affect non-target genes in the host genome; and finally, does the T7 DNAP increase mutagenesis in the target gene (VanInsberghe, M, unpublished data)?

gp4 encodes two gene products: 4A and 4B which have primase and helicase activity, and helicase activity respectively. To proceed with this strategy proposed by iGEM, it is important to develop an understanding of the T7

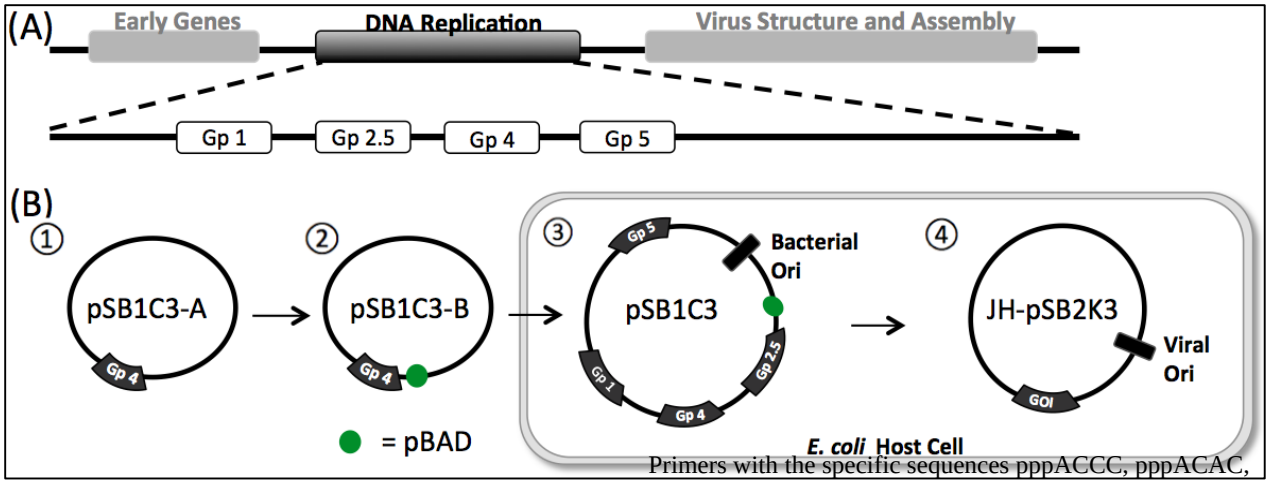

FIG 2 Cloning strategy of an orthogonal replication system from the bacteriophage T7 genome into *E. coli*. (A) The bacteriophage T7 genome is composed of early genes, replication genes as well as virus structure and assembly genes. Only four genes are involved in DNA replication, Gp1 RNAP, Gp2.5 ssDNA-BP, Gp4 Primase/Helicase and Gp5 DNAP. Other requirements for this replication system are the specific viral origin sequence (Ori) and thioredoxin, which is a host protein that serves as a processivity factor. Together, the T7 replisome is formed with the T7 genes. **(B)** The cloning strategy put forth by iGEM to introduce the T7 replisome into *E. coli* as an orthogonal replication system. This strategy starts by (1) introducing each T7 genes involved in DNA replication into the pSB1C3-A plasmid, where this is done individually with each gene. Then (2), each gene is introduced into the pSB1C3-B plasmid with the pBAD promoter. Once each gene has been cloned and expressed using the pSB1C3-B plasmid, (3) all of the genes can be assembled onto a single recombinant pSB1C3 plasmid with the bacterial origin of replication, pBAD promoter and terminator. This recombinant plasmid can then be transformed into *E. coli* and the replisome can be expressed. (4) The gene of interest (GOI) will be inserted into the JH-pSB2K3 plasmid that contains the viral origin of replication. The error prone DNAP present in the replisome will recognize this viral origin sequence and induce mutagenesis in the target gene.

primase/helicase structure and function. There are two proteins that can be expressed from gp4 in the same reading frame: 4A and 4B. 4B is transcribed and translated from the methionine at the 64th amino acid residue within the coding sequence of 4A as shown in Figure 3 and Figure 4B (15, 19). Gene 4A is 63 kDa in size and exhibits both primase and helicase activities, whereas 4B is 56 kDa in size and exhibits only helicase activity (12, 15, 19). At the N terminus of the 4A protein is comprised of the primase domain, including the zinc binding domain (ZBD) and RNA polymerase domain (RPD), connected together by a flexible linker (7, 12, 19). Following the primase domain is another linker of 26 residues and a helicase domain at the C terminus (7, 12). Due to the truncation of the first 63 amino acids in 4B, it contains only the RBD and the helicase domain (7, 12, 15, 19). During T7 infection of *E. coli*, equal amounts of 4A and 4B are expressed (19, 23).

The helicase domain is responsible for the unwinding of double stranded DNA and translocation of single stranded DNA in the 5' to 3' direction during DNA synthesis (16). The helicase domains assemble into hexamers before binding to the lagging strand (Figure 3) and hydrolyze nucleotides for energy (5, 16). dTTP is the preferred substrate for the six NTP binding sites present and enables the formation of gp4 hexamers (5, 16). The DNA polymerase – TrxA pair then binds to the hexamer for synthesis of DNA, while approximately 60 to 80 bases of the ssDNA are protected by the hexamers (10).

DNA polymerase requires primers provided by the primase activity of 4A and 4B to produce Okazaki fragments on the lagging strand during DNA synthesis (7). Primers with the specific sequences pppACCC, pppACAC, and pppACCA are made upon the recognition of template sequences such as 5'-(G/T)(G/T)GTC-3' by the ZBD (5, 7, 11, 24). The flexible linker brings the two subdomains together with the DNA template and movement of the template DNA occurs upon synthesis of diribonucleotides, all in correlation with a conformational change of the ZBD (12). The RPD is involved with the formation of phosphodiester bonds between nucleotides thereby extending the oligoribonucleotides (5, 12). Accordingly, without the ZBD present in 4B, the protein cannot synthesize primers and only contains the helicase activity (7). Heterohexamers, which contain alternating 4A and 4B proteins shown in the inset crystallographic structure Figure 4A, form more readily in contrast to hexamers of 4A proteins, due to lack of steric interference by adjacent ZBDs (19, 23). The assembly of six helicases allow for the interactions of different ZBDs with different RPDs (12, 23).

Possible toxic effects of gp4 primase/helicase due to internal promoter in 4A coding sequence for 4B. It has been observed that there is regulation of the primase and helicase in order to prevent toxic effects during replication and growth in *E. coli* (15). A study by Studier involved cloning the 4B gene under the Φ10 promoter for T7 RNAP into *E. coli*, which led to cell death and lysis, suggesting it could be toxic (18). Possible explanations include accumulations of T7 proteins being toxic or activation of the DE3 prophage present in the strain (18). Unsuccessful cloning of 4A and 4B may have been due to an internal promoter for *E. coli* RNA polymerase, found to contain a similar consensus sequence, found near the N terminal of 4A coding sequence, thus allowing for expression of 4B (15). Due to difficulties in obtaining pure 4A, 4A' was

created by eliminating the expression of 4B through a substitution of the methionine at amino acid 64 with a leucine, which was predicted to be a tolerable mutation (15). The 4A' was found to contain the same activities as the original 4A and elimination of 4B and its helicase activity (10, 15). However, efficiency of replication is increased upon supplementation of 4A' or 4B, suggesting that the limiting factor is the helicase activity (15). Also, the 4A' primase activity was enhanced by three fold once 4B was added and the two possible explanations were a conformational change between 4A' and 4B or a doubling of activity through formation of heterodimers (11).

Crystallographic analysis of bacteriophage T7 primase/helicase revealing that the Gly313 → Val mutation may lead to reduction of dTTP hydrolysis. The bacteriophage T7 primase/helicase consists of a ring formed by helicase domains, with primase domains attached and projecting away from the ring, resulting in a crown-like shape overall as seen in Figure 4C. Forming a right-handed permutation, the primase domains are swapped onto the neighbouring helicase domains. With few interactions, each of the primase domain active sites point towards its neighbouring primase domain and all are oriented tangential from the circumference of the ring (19). The flexibility obtained through the few interactions between the primase and helicase allows for the helicase domains to push the DNA through the central channel of the ring during replication. Before the replication loop is formed by the helicase and lagging strand polymerase, the primase domains are able move away because of the flexible connections to prevent interference (17, 19).

As a hexameric ring, the T7 DNA helicase translocates along a single stranded DNA using energy from dTTP hydrolysis. The dTTPs bind between pairs of adjacent subunits as the DNA helicase separates the two strands. Near the center of the hexameric ring structure, the binding of dTTPs stabilizes the DNA-binding motif. The C-terminal face of the helicase domain is negatively charged and interacts with the T7 DNA Polymerase, whereas the N-terminus interacts with the primase domain (17).

The Gly313 → Val mutation seen in one of the clones produced by iGEM is proposed to be present as a way of avoiding toxicity in *E. coli*, as the mutation may have compromised helicase the activity. According to Trevino et al., glycine residues lack a side chain, which influences conformational stability for particular β-turn positions (20). The observed Gly313 → Val mutation is located on a turn between an alpha helix and a β-sheet as observed on the Protein Data Base website, therefore a replacement of the glycine residue with a valine residue may be detrimental to overall protein structure and function (PDB ID 1E0K, [http://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=1E0K&bionumber=1]). In accordance with this observation, previous studies have shown that a replacement of T7 helicase Gly317 → Val led to loss of the helicase activity through the reduction of dTTP hydrolysis (21). Interestingly, these residues are located within the same β-turn as the observed Gly313 → Val mutation.

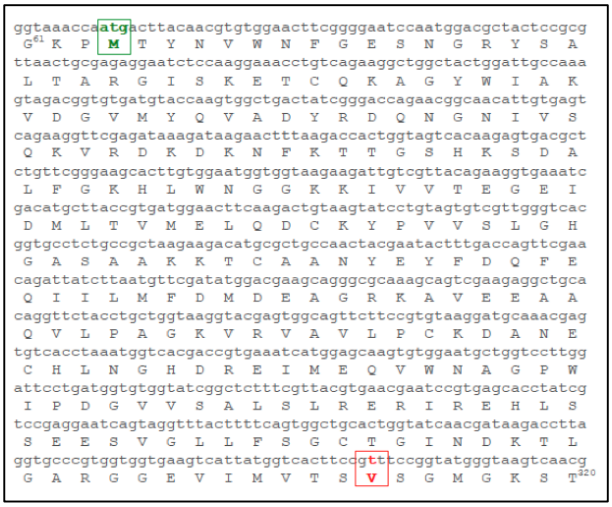

FIG 3 Bacteriophage T7 gp4 sequence with the gene 4B start codon and the Gly^{313} → Val point mutation. Highlighted in green is the methionine start codon located at amino acid 64 within the 4A gene, and which is used for the expression of 4B gene. Additionally, a section of the mutated Bacteriophage T7 gp 4 primase/helicase gene is shown from amino acids 61 to 320. The mutated gene 4 sequence, showing where the Gly^{313} → Val point mutation is highlighted in red The remainder of the gene 4 sequence had complete homology to the native sequence and the point mutation located in the 3D structure can be seen in Figure 4B.

Reduction of dTTP hydrolysis activity is attributed to steric hindrance of dTTP binding by Met316, which was driven to a new position due to the Gly317→ Val mutation (17).

Our attempts to revert Gly313 → Val point mutation in the pSB1C3-A plasmid using SDM. The construct that we received contained the gp4 with a point mutation causing Gly313 → Val and the aim of this experiment was to revert it to wild-type (WT) using the primers designed to revert the T938 back to the original G (Figure 3). The gp4 did not have a promoter at the N terminus so we expected no expression of the primase/helicase. The controls for the SDM were the negative PCR control, which contained no template and therefore should result in no amplification, and the negative SDM control, which contained no polymerase. When DpnI is used, the DNA in negative SDM control should be digested, as it should only have parental methylated plasmid. Results will be seen in the transformants and all data from our efforts can be found in Supplementary Table 1.

We included a pSB1C3-A as another positive control during the transformation of SDM products into competent DH5α. Few and small colonies resulted from the transformation of DH5α competent cells with the SDM products. Numerous colonies resulted from the positive pSB1C3-A plasmid control. The positive SDM lacZ gene control and negative controls displayed the expected results, which can be seen in Supplementary Table 1. With the few colonies from the transformation of SDM products, the plates were incubated another night for more growth, and restreaked to ensure purity and presence of the chloramphenicol resistance gene.

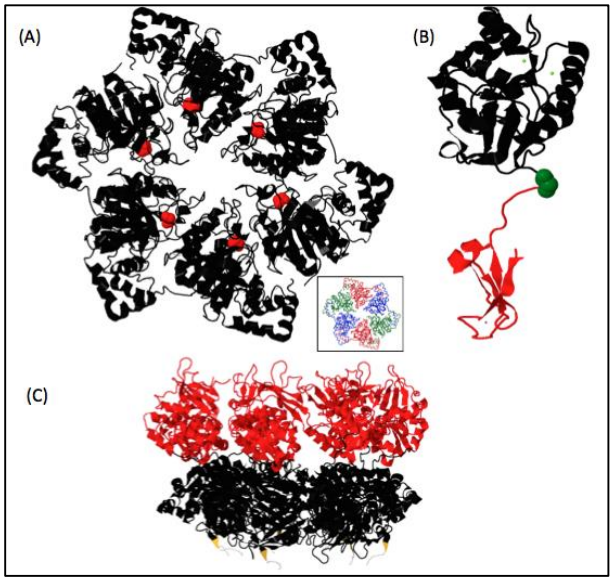

FIG. 4. Crystallographic picture of the T7 primase/helicase (A) The three-dimensional structure of the T7 helicase, consisting of six identical subunits, each highlighted in a different colour in the inset photo. The $Gly^{313}\rightarrow$ Val mutation is highlighted in red. As a hexameric ring, the T7 DNA helicase translocates along a single stranded DNA using energy from dTTP hydrolysis (17). **(B)** The three-dimensional structure of the zinc binding domain (ZBD), shown in red and RNA polymerase domain (RPD), shown in black, both which are part of the T7 primase**.** There are two proteins expressed from gp4 in the same reading frame are 4A and 4B. As shown in green, 4B is transcribed and translated from the methionine at the 64th amino acid residue within the coding sequence of 4A, and therefore does not contain the ZBD. This residue also separates the RPD and ZBD. Gene 4A is 63 kDa in size and exhibits both primase and helicase activities, whereas 4B is 56 kDa in size and exhibits only helicase activity (12, 15, 19). **(C)** The crown-like structure formed between the primase domains in red and the helicase domains in black. Forming a right-handed permutation, the primase domains are swapped onto the neighbouring helicase domains. The flexibility obtained through the few interactions between the primase and helicase allows for the helicase domains to push the DNA through the central channel of the ring during replication (17, 19). The RCSB Protein Data Base ID for primase/helicase, helicase only and primase only are 1Q57, 1E0K and 1NUI, respectively.

Attempts to isolate the corrected pSB1C3-A plasmid. The plasmids from transformants containing either the SDM products or positive control were isolated after overnight incubation in the presence of chloramphenicol or ampicillin, respectively. Plasmid concentration consistently resulted in concentrations below 50 ng/uL, which is not enough to obtain sequencing results. However, the positive pSB1C3-A plasmid control had a similar concentration as the original pSB1C3-A plasmid. We further assessed why there were low plasmid concentrations from our samples.

PCR analysis of pSB1C3-A shows isolated plasmid concentrations from transformants were too low to sequence. Troubleshooting efforts began by restriction digest of the pSB1C3-A plasmids from the transformants containing SDM products, positive pSB1C3-A control, and the original pSB1C3-A plasmid. There were no bands on the subsequent gel for the pSB1C3-A plasmids from the transformants containing either SDM products, which corresponds to the low concentrations determined previously. The lanes containing the positive pSB1C3-A plasmid control, and the original pSB1C3-A displayed the expected bands at 2 kb. We concluded that the isolated plasmid demonstrated consistent low plasmid concentrations.

A PCR involving the forward primer PH_SP294F and the primer 1R was done to confirm that pSB1C3-A plasmid was suitable for amplification. A band of approximate size of 500 bp was expected, and confirmed on a gel (Supplementary Table 1). The band confirmed that the pSB1C3-A was suitable for amplification by PCR. However, the negative control, which should have no plasmid initially, was also present.

DISCUSSION

Currently, there is an orthogonal replication system in yeast involving a DNA polymerase that targets a plasmid independent of the host genome (13). iGEM endeavors to create a similar orthogonal replication system in *E. coli* using the Bacteriophage T7 replication machinery. The goal is to obtain a plasmid construct with four specific genes required for orthogonal replication of the plasmid under an inducible promoter. The iGEM team currently has cloned and expressed three of the genes in *E. coli*, but are experiencing mutations arising in a common area of the primase/helicase gene according the 2014 UBC iGEM website (http://2014.igem.org/Team:British_Columbia/ProjectChassis). Due to these mutations, it is thought that the primase/helicase protein product is toxic to the cell, resulting in selective pressure for mutations that disrupt the function of primase/helicase.

The current construct we are working with has a single base pair mutation that results in an amino acid change from $Gly^{317}\rightarrow$ Val. This particular amino acid is located on a turn between a β-strand and an α-helix (http://www.rcsb.org/pdb/explore/remediatedSequence.do?structureId=1E0K&bionumber=1). The difference between glycine and valine is in the flexibility of the bond; glycine allows rotation whereas valine is more sterically hindering due to the side chain (25, Oliver, D, personal communication). It is possible that this change affects the overall 3D structure, inhibiting the formation of the β-turn. Since primase/helicase normally forms hexamers *in vivo*, this point mutation could interfere with the formation of hexamers and prevent proper function of the primase/helicase protein. Without proper protein function, the potential toxic effects of gp4 could be bypassed when functioning *in vivo* (23). Alternatively, a study by Washington *et al.* showed that the replacement of glycine with valine at residue 318 resulted in a sharp drop in the helicase activity caused by steric hindrance of the dTTP hydrolysis site (21). Since the mutation is similar and not far from codon 313, it is possible that the

point mutation in pSB1C3-A had an effect similar to mutation of residue 318 which inhibited the function of the helicase.

iGEM UBC's inability to clone primase/helicase suggests possible toxicity of the gene product. Since the pSB1C3-A plasmid does not have pBAD promoter for the expression of gp4 we expected either no expression or expression at a low level from a promoter further upstream. While T7 primase/helicase is known to be toxic from previous studies, the exact mechanism of toxicity is not clear (15, 18). It has been proposed that the presence of certain T7 genes may induce a prophage present in the genome and thereby cause cell death, suggesting that gp4 may be one of these genes (18). However, since our strain of *E. coli* DH5α does not contain a prophage within its genome, this is unlikely due to the supporting information found on the OpenWetWare website (http://openwetware.org/wiki/E._coli_genotypes).

SDM was performed to potentially revert the $Gly^{313} \rightarrow$ Val mutation in the pSBC13-A plasmid to WT. The SDM was properly executed due to the success of the positive control, producing a mutation of the *lacZ* gene on the pUC19 plasmid as evidenced by a change from blue to white colonies on selective media containing X-gal and IPTG. Transformation of the SDM products containing the plasmid with the suspected WT primase/helicase (WTPH) gene into *E. coli* DH5α cells resulted in very few transformants. However, these are believed to contain the aforementioned plasmid, as these transformants were able to survive on agar containing chloramphenicol. The negative positive controls for transformation displayed the expected results, suggesting that DpnI digestion and transformation was successful. A PCR reaction was run to ensure the pSB1C3-A plasmid was suitable for amplification. The expected bands were present while the negative control also showed a band, but our results should still be valid. One reason could be a potential contamination while preparing the PCR samples.

Plasmid concentrations from the WTPH plasmid isolations were determined to be too low for sequencing. Additionally, the plasmid isolation process was assumed successful because of high plasmid concentrations from the positive pSB1C3-A plasmid control transformation and the original pSB1C3-A plasmid. After restriction digest, the gel confirmed the results of the controls but bands were absent in the WTPH samples. The Nanodrop readings and lack of bands from the restriction digest suggests that the concentrations of the WTPH plasmids are very low.

There are two possible reasons for the low concentration of plasmids. Firstly, the WTPH plasmids may have been toxic to the cells, and rather than expressing the protein the *E. coli* cells maintained the plasmids at low levels to prevent extensive toxicity of the T7 primase/helicase products (26). Accordingly, there still may have been minimal expression of the WTPH plasmids from an upstream promoter. Secondly, the chloramphenicol resistance gene on the WTPH plasmids may have been integrated into the genome of the *E. coli* cells, therefore allowing them to grow in the presence of chloramphenicol. It is unlikely that cells obtained chloramphenicol resistance without the transformation being successful, so it is assumed that all cells that grew in the presence of chloramphenicol have successfully been transformed with the WTPH plasmid. This suggests that between the transformation and plasmid isolation steps the plasmid was altered and the cells retained their chloramphenicol resistance. Moreover, the pSB1C3-A plasmids that did not undergo SDM remained present in the cells, when the only difference was the $Gly^{313} \rightarrow$ Val mutation, suggesting that the reversion to WT had a negative impact on the plasmid, presumably in the form of toxicity to the cell.

In conclusion, the results demonstrate that the pSB1C3-A plasmids containing the suspected WT gp4 corrected through SDM, were low in concentration. The gp4 primase/helicase may have been expressed at a low level from an upstream promoter within the plasmid. This suggests that the expression of WT T7 primase/helicase may be toxic to *E. coli* DH5α cells, and thus, a low concentration of the pSB1C3-A plasmid could be favorable to them. Alternatively, the chloramphenicol resistance may have recombined into the genome of the *E. coli* cells. By reviewing current literature, we speculated that the point mutation in gp4 of the original pSB1C3-A plasmid lead to a disruption of the helicase activity, by either preventing hexamer formation or dTTP hydrolysis. The resulting reduction in helicase activity may have allowed *E. coli* to tolerate the toxic effects of gp4.

FUTURE DIRECTIONS

With regards to what has been found, regulating gp4 expression under the pBAD promoter would allow for better determination of whether the reversion to WT would be toxic. The main focus would be to correct the gp4 $Gly^{317} \rightarrow$ Val mutation in the pSB1C3-B plasmid containing pBAD promoter by SDM with the provided primers, 1F and 1R. Ultimately, the aim would be to obtain the WT sequence. The pSB1C3-B plasmid has the promoter upstream of gp4 for tight regulation of the particular T7 gene, where glucose will be used for repression and arabinose will be used for induction. Because of the tight regulation, it is possible that the subsequent plasmid isolation will result in higher copy numbers of the pSB1C3-B plasmids, and therefore allow for sequencing results. The plating should be done consecutively after the correct results have been determined on either media containing glucose or arabinose. Additionally, analysis of 4A' expression and function on the WT gp4 pSB1C3-B plasmid could further provide evidence whether gene 4B is needed for T7 replication. 4A' can be

created using the potential WT gp4 pSB1C3-B plasmid and primers 2F and 2R.

Determining whether gp4 is toxic after reversion to WT on the pSB1C3-B plasmid, and if 4A' proves to be sufficient and less toxic will allow the UBC iGEM team to continue engineering an orthogonal replication system in *E. coli*.

Proposed second experiment to create 4A' using SDM. A proposed second experiment involved creating the 4A' protein through SDM with our designed primers 2F and 2R. We had aimed to knockout 4B expression by changing the Met^{64} to Leu^{64} on the WT pSB1C2-A plasmid. Changing the start codon on the pSB1C3-A plasmid that was reverted back to WT was expected to remove function of the gene 4B product, and hopefully any associated toxicity. Unfortunately, we were unable to confirm the corrected gp4 on the construct before proceeding with the second experiment.

ACKNOWLEDGEMENTS

The finances for this study were supplied by the Department of Microbiology and Immunology at the University of British Columbia. We would like to thank Dr. David Oliver for his guidance in the experimental approach and troubleshooting, Jia Wang and Celine Michiels for their assistance in the lab, and Michael Vaninsberghe for his feedback on the iGEM cloning strategy, and for supplying pSB1C3-A plasmid containing the point mutation, and sequencing primers.

REFERENCES

1. **Arnold, FH, Georgiou, G, eds.** 2003. Directed Enzyme Evolution : Screening and Selection Methods. Humana Press, Totowa, NJ, USA. ProQuest ebrary. Web. 27 March 2015.
2. **Camps, M, Naukkarinen, J, Johnson, BP, Loeb, LA.** 2003. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. Proc. Natl. Acad. Sci. U. S. A. **100:**9727-9732. doi: 10.1073/pnas.1333928100.
3. **Drake, JW.** 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U. S. A. **88:**7160-7164.
4. **Esvelt, KM, Carlson, JC, Liu, DR.** 2011. A system for the continuous directed evolution of biomolecules. Nature. **472:**499-503.
5. **Hamdan, SM, Richardson, CC.** 2009. Motors, switches, and contacts in the replisome. Annu. Rev. Biochem. **78:**205-243.
6. **Hamdan, SM, Marintcheva, B, Cook, T, Lee, SJ, Tabor, S, Richardson, CC.** 2005. A unique loop in T7 DNA polymerase mediates the binding of helicase-primase, DNA binding protein, and processivity factor. Proc. Natl. Acad. Sci. U. S. A. **102:**5096-5101. doi: 10.1073/pnas.0501637102.
7. **Kato, M, Ito, T, Wagner, G, Richardson, CC, Ellenberger, T.** 2003. Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. Mol. Cell. **11:**1349-1360.
8. **Lee, S, Richardson, CC.** 2011. Choreography of bacteriophage T7 DNA replication. Curr. Opin. Chem. Biol. **15:**580-586.
9. **Nowak, MA.** 1992. What is a quasispecies? Trends in Ecology & Evolution. **7:**118-121.
10. **Patel, SS, Hingorani, MM.** 1993. Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins. J. Biol. Chem. **268:**10668-10675.
11. **Patel, SS, Rosenberg, AH, Studier, FW, Johnson, KA.** 1992. Large scale purification and biochemical characterization of T7 primase/helicase proteins. Evidence for homodimer and heterodimer formation. J. Biol. Chem. **267:**15013-15021.
12. **Qimron, U, Lee, S, Hamdan, SM, Richardson, CC.** 2006. Primer initiation and extension by T7 DNA primase. EMBO J. **25:**2199-2208.
13. **Ravikumar, A, Arrieta, A, Liu, CC.** 2014. An orthogonal DNA replication system in yeast. Nature Chemical Biology. **10:**175-177.
14. **Romano, LJ, Tamanoi, F, Richardson, CC.** 1981. Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins: requirement for T7 RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. **78:**4107-4111.
15. **Rosenberg, AH, Patel, SS, Johnson, KA, Studier, FW.** 1992. Cloning and expression of gene 4 of bacteriophage T7 and creation and analysis of T7 mutants lacking the 4A primase/helicase or the 4B helicase. J. Biol. Chem. **267:**15005-15012.
16. **Satapathy, AK, Kulczyk, AW, Ghosh, S, van Oijen, AM, Richardson, CC.** 2011. Coupling dTTP hydrolysis with DNA unwinding by the DNA helicase of bacteriophage T7. J. Biol. Chem. **286:**34468-34478. doi: 10.1074/jbc.M111.283796; 10.1074/jbc.M111.283796.
17. **Sawaya, MR, Guo, S, Tabor, S, Richardson, CC, Ellenberger, T.** 1999. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. Cell. **99:**167-177.
18. **Studier, FW.** 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J. Mol. Biol. **219:**37-44.
19. **Toth, EA, Li, Y, Sawaya, MR, Cheng, Y, Ellenberger, T.** 2003. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. Mol. Cell. **12:**1113-1123.
20. **Trevino, SR, Schaefer, S, Scholtz, JM, Pace, CN.** 2007. Increasing protein conformational stability by optimizing β-turn sequence. J. Mol. Biol. **373:**211-218.
21. **Washington, MT, Rosenberg, AH, Griffin, K, Studier, FW, Patel, SS.** 1996. Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. J. Biol. Chem. **271:**26825-26834.
22. **Yuan, L, Kurek, I, English, J, Keenan, R.** 2005. Laboratory-directed protein evolution. Microbiol. Mol. Biol. Rev. **69:**373-392. doi: 10.1128/MMBR.69.3.373-392.2005.
23. **Zhang, H, Lee, SJ, Kulczyk, AW, Zhu, B, Richardson, CC.** 2012. Heterohexamer of 56- and 63-kDa Gene 4 Helicase-Primase of Bacteriophage T7 in DNA Replication. J. Biol. Chem. **287:**34273-34287. doi: 10.1074/jbc.M112.401158; 10.1074/jbc.M112.401158.
24. **Frick, DN, Richardson, CC.** (1999). Interaction of bacteriophage T7 gene 4 primase with its template recognition site. J. Biol. Chem. **274:** 35889–35898. doi: 10.1074/jbc.274.50.35889.
25. **Cormanich, RA, Ducati, LC, Rittner, R.** (2011). Are hydrogen bonds responsible for glycine conformation preferences? Chemical Physics. **387:** 85-91. doi: 10.1016/j.chemphys.2011.07.006
26. **Saida, F, Uzan, M, Odaert, B, Bontems, F.** (2006). Expression of Highly Toxic Genes in *E. coli*: Special Strategies and Genetic Tools. Current Protein and Peptide Science. **7:** 47-56. doi: 10.2174/138920306775474095