Expression of *Clostridium acidurici* 9a Glycine Cleavage System in *Escherichia coli* for Formatotrophic Growth via Reductive Glycine Pathway

by

Megha Patel

A thesis submitted in conformity with the requirements for the degree of Masters of Applied Science
Chemical Engineering and Applied Chemistry
University of Toronto

© Copyright by Megha Patel 2016
Expression of *Clostridium acidurici* 9a Glycine Cleavage System in *Escherichia coli* for Formatotrophic Growth via Reductive Glycine Pathway

Megha Patel  
Masters of Applied Science  
Chemical Engineering and Applied Chemistry  
University of Toronto  
2016

Abstract

There has been a marked increase in carbon dioxide levels in the atmosphere over the years, presenting countries with the task of reducing the emissions. This drives the motivation of discovering methods to fix carbon dioxide into valuable commodities. Microorganisms can be engineered to express pathways that do not exist naturally, and more efficient at providing high yields and titre of a particular product. This principle serves as the basis for this research project. A computationally designed carbon fixation pathway for growth on formate, known as the reductive glycine pathway, can theoretically produce pyruvate, a central to *E. coli* metabolites, through the expression of recombinant genes. The goal of the project is to test the expression of a bidirectional glycine cleavage system that has been cloned from *C. acidurici* 9a in *E. coli*, which is predicted to be a major pathway expression challenge for the reductive glycine pathway.
Acknowledgments

The past two years at University of Toronto, St. George campus has been a great journey. My time spent in Biozone of Chemical Engineering, under the supervision of Professor Radhakrishnan Mahadevan and Professor Alexei Savchenko, holds great value in my life. I must thank both my professors for giving me this opportunity to learn and be part of their lab groups. My project would not have been at this stage without their guidance, support, and patience.

I have also met very skilled and kind-hearted people in both labs. I would like to acknowledge Vik who has helped me at every step of the project. Everyone was very nice and made me feel part of the team since the very beginning. Many of the students from Professor Mahadevan’s lab have also helped me, including Kayla, Naveen, and Kevin. From Professor Savchenko’s lab, I have to thank Veronica and Rosa for all their suggestions and expertise in cloning; I have learned a lot from both of them. I would also like to thank Biozone for all the great conservations and food during teatime.

Lastly, I need to thank my family, and close friends for all their moral support during the past two years. They have been a great help to keep my personal life very positive and bright, allowing me to bring the positivity to the lab and the project.
# Table of Contents

Acknowledgments.................................................................................................................... iii

List of Tables................................................................................................................................. vi

List of Figures............................................................................................................................... vii

List of Appendices......................................................................................................................... ix

Glossary........................................................................................................................................... x

Chapter 1 ..................................................................................................................................... 1

1 Background Information ........................................................................................................... 1

  1.1 Designing of Computational Metabolic Pathways................................................................. 1

  1.2 Importance of Enzyme Characteristics.................................................................................. 2

  1.3 Necessity for Green Products / Green Processes................................................................. 3

  1.4 Role of E. coli in the Industry............................................................................................... 4

Chapter 2 ..................................................................................................................................... 6

2 Literature Review ....................................................................................................................... 6

  2.1 Carbon Fixation Pathways .................................................................................................. 6

  2.2 Reductive Glycine Pathway ............................................................................................... 10

  2.3 Glycine Cleavage System .................................................................................................. 11

Chapter 3 ..................................................................................................................................... 15

3 Motivation and Statement of Objectives .................................................................................. 15

Chapter 4 ..................................................................................................................................... 18

4 Methods .................................................................................................................................. 18

  4.1 Ligase Cycling Reaction to Construct Plasmids with Individual C. acidurici 9a GCV proteins. 18

    4.1.1 Amplification of Plasmid and Insert ............................................................................. 18

    4.1.2 Ligating Plasmid with Insert (LCR) .......................................................................... 18

    4.1.3 Transformation of Recombinant Plasmids in DH5α.................................................. 19

  4.2 Small Scale Test Expression of C. acidurici 9a GCV Proteins in E. coli BL21 (DE3) .......... 19

    4.2.1 Transformation of Recombinant Plasmids in BL21-DE3 RIL ...................................... 19

    4.2.2 Inducing Protein Expression ....................................................................................... 20

    4.2.3 Lysing of Cells and Visualizing on SDS-PAGE GEL................................................ 20

  4.3 Lambda red to excise GCV proteins of E. coli.................................................................. 20
List of Tables

Table 1: Overview of the five autotrophic pathways besides photosynthesis also known as carbon fixation pathways (Bar-Even et al., 2010; Fuchs, 2011) 7

Table 2: A Comparison of Glycine Cleavage System proteins between E. coli and C. acidurici 9a (italics). 14

Table 3: Primers used in this study. 44
List of Figures

Figure 1: Flow-chart of recombinant protein expression. Redesigned from Broadview, N. (2012) .................................................................3

Figure 2: 4 of the 5 alternative native non-calvin cycle carbon fixation pathways. Carbon incorporation is through CO₂ (yellow circles) or Bicarbonate ions (green circles). End products of the pathways, pyruvate and acetyl-CoA are represented by blue circles. These pathways cycle around Succinyl-CoA making them interconnected. Permission obtained from (Ducat & Silver, 2012) ...............................................................................................................8

Figure 3: A schematic of the reductive glycine pathway. Redrawn from Bar-Even, 2013......11

Figure 4: The reaction mechanism of the cleavage of glycine by the glycine cleavage system. ........................................................................................................................................................................13

Figure 5: Expression screening of glycine cleavage system proteins from C. acidurici 9a in E. coli BL21-CodonPlus(DE3)-RIL on a small-scale basis with pTRC99a backbone. BB stands for backbone only (pTRC99a) and HPT means construct with all 3 genes ligated. The expected sizes of H, P subunits, and T are 14 kDa, 50 kDa (Pa), 54 kDa (Pb), and 41 kDa, respectively. ........................................................................................................................................................................24

Figure 6: Individual construct design for the C. acidurici 9a glycine cleavage system proteins in the IPTG-inducible pTRC99a backbone. UNS stands for unique nucleotide sequence and RBS for ribosomal binding site. ...............................................................................................................26

Figure 7: The construct layout of plasmid pTRC99a harbouring C. acidurici 9a glycine cleavage system genes. The individual fragments were amplified from plasmids designed in Figure 5. ........................................................................................................................................................................28

Figure 8: 1% agarose DNA gel representing the deletion of E. coli glycine cleavage system at size approximately 5500 base pairs. The band is seen in the wild type (WT) E. coli BW 2511 ΔserA and not in the strain where the chloramphenicol cassette has replaced the glycine cleavage system (ΔserAΔgcv:Chl-). The chloramphenicol cassette is seen in the ΔserAΔgcv:Chl- strain at approximately 2000 base pairs which is not seen in the wild type strain. ........................................................................................................................................................................29

Figure 9: The growth of E. coli BW 25113 strains harbouring pTRC99a backbone (square) and pTRC99a_HPT (triangle) in M9 media with glucose and glycine as carbon sources. Media
recipe described in section 4.5. Figure A, B, and C are E. coli BW 25113 WT, ΔserA, and ΔserAΔgcv:Chl-. Figure D is of same cultures from Figure C, rediluted in 50 mL media to capture the growth curve.  

**Figure 10:** Characterization of E. coli BW 25113 ΔserAΔgcv:Chl- harbouring pTRC99a backbone (square) and pTRC99a_HPT (triangles).

**Figure 11:** Colony screening of E. coli BW 25113 ΔserAΔgcv:chl- carried out at annealing temperature of 55°C. 8 screens out of 12 were positive for the insertion of the chloramphenicol resistance gene as amplified PCR product was seen around the expected size of 869 base pairs. Verification of the location of the chloramphenicol resistance gene in the genome is shown in Figure 8.

**Figure 12:** Exponential phase data points plotted of growth of E. coli BW 15113 ΔserAΔgcv:chl- from Figure 9 (D) to calculate growth rate. The slope of the best fit line, $\mu$, is calculated to be 0.175 h$^{-1}$, also known as the growth rate.
List of Appendices

Appendix A: Lambda Red Colony Screening of Chloramphenicol gene insertion_______42
Appendix B: Growth rate calculation of E. coli BW 25113 ΔserΔgcv:chl___________43
Appendix C: Primers used in the Study_________________________________________43
### Glossary

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>GCV</td>
<td>Glycine Cleavage System</td>
</tr>
<tr>
<td>gDNA</td>
<td>genomic DNA</td>
</tr>
<tr>
<td>H</td>
<td>Glycine Cleavage System H Protein</td>
</tr>
<tr>
<td>L</td>
<td>Glycine Cleavage System L Protein (dihydrolipoyl dehydrogenase)</td>
</tr>
<tr>
<td>LCR</td>
<td>Ligase Cycling Reaction</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny Broth (Luria-Bertani medium)</td>
</tr>
<tr>
<td>P</td>
<td>Glycine Cleavage System P Protein (glycine dehydrogenase)</td>
</tr>
<tr>
<td>pTRC99a</td>
<td>IPTG-inducible plasmid with pBR322 origin of replication</td>
</tr>
<tr>
<td>pTRC99a_HPT</td>
<td><em>Clostridium acidurici</em> 9a Glycine Cleavage System HPT construct with pTRC99a plasmid backbone</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosomal Binding Site</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RPM</td>
<td>Rotations per Minute</td>
</tr>
<tr>
<td>SOC</td>
<td>Super Optimal Broth with Catabolite Repression</td>
</tr>
<tr>
<td>T</td>
<td>Glycine Cleavage System T Protein (aminomethyltransferase)</td>
</tr>
</tbody>
</table>
Chapter 1

1 Background Information

1.1 Designing of Computational Metabolic Pathways

Metabolic engineering refers to the study of integrated metabolic pathways along with their regulatory networks, different from genetic engineering and synthetic engineering which is the study of single genes and enzymes (47). The objective in metabolic engineering is to obtain the expression of genes and biochemical pathways as needed for synthesis of particular products that can be industrially, medically or pharmaceutically relevant. This requires optimization such that the titer, rate, and yield are large enough for chemical plants, while restricting the process to be cost-effective.

The initial step is to determine how valuable products can be formed from a particular substrate, followed by an analysis of the various pathways through which these products can be formed. The various pathways can be analyzed in terms of number of steps, flux through the pathway, and the control mechanisms that are in place at each step. Kinetically limiting steps in a given pathway can be recognized using metabolic flux analysis, which involves comparing flux from different conditions to a control (3). This information aids in the mapping of potential pathways, starting with a specific substrate leading to a desired product with predicted flux calculations between each step. Other important information that can be extracted is the theoretical yield of products and the equivalence number of co-factors and intermediate.

The last step is to carry out the computationally sufficient pathway experimentally and scale it up for industrial use. Various organisms can be used to produce valuable substances in a cost-effective manner. One of the more popular industrial organisms for production of commodities through biological means is Escherichia coli due to the fact that E. coli is well understood and has a high cell growth rate (34). For production of amino acids and biofuels, Corynebacterium glutamicum and Saccharomyces cerevisiae are considered ideal in the industry, respectively. A good example where metabolic engineering has been successful is in the production of alkaloids. Alkaloids are found naturally in plants and are highly valued in the industry for their role in pharmaceutical drugs. As such, engineering an organism that efficiently produces alkaloids is a
better option than trying to chemically synthesize alkaloids, which may not be industrially favourable (21).

1.2 Importance of Enzyme Characteristics

Metabolic engineering cannot be understood without understanding the enzymes and the reactions involved in the pathways. It has been long understood that enzymes are highly specific to a substrate and so is the reaction they perform. Recently, the focus has shifted to enzyme evolution since changes at the genomic level, by nature or by intention, can allow enzyme functions to be broadened or enhanced. This enables the enzyme to interact with other substrates or be active in a different environment. Metabolic engineering uses this advantage to design pathways that either incorporate enzymes from various species or broaden substrate specificity of the enzyme with potential of evolved enzyme to be functional.

Designing recombinant protein expression is essential to metabolic engineering. However, this process can be challenging, as numerous factors need to be considered for successful recombinant protein expression (38). Some of these factors include the organism from which a particular gene can be cloned, the type of plasmid that needs to be used to express the gene, determining a host for transforming the plasmid, plasmid compatibility, and the number of plasmids that might be required to express the pathway in its entirety. There are multiple methods via which plasmids can be designed to express inserts, which can contain one gene, two genes, or an entire pathway. The conventional method for cloning is through restriction digestion sites, whereas some of the newer methods include ligase cycling reaction, Gibson’s assembly, ligation independent cloning and using gBlocks for assembly of fragments.

Understanding protein structure and protein function is fundamental to metabolic engineering. While computational biologists can use enzyme kinetic data to predict feasible pathways, the challenge lies in the change that may occur in recombinant protein structure and function. Due to change of host organism, full activity of the protein might not be retained. It is important to know the structure of the protein, the properties of the active site, and the interactions with the substrate. With this information, it can be predicted which proteins might have better activity in a certain host organism for a particular substrate. To understand the structure and kinetic activity
of a protein, it needs to first be cloned from a genome, inserted in an expression construct and screen for expression. Then, the protein can be grown in large amounts before being purified, after which in-vitro studies can be undertaken to elucidate structure and kinetics. Figure 1 below describes the steps of protein expression.

Figure 1: Flow-chart of recombinant protein expression. Redesigned from Broadview, N. (2012).

1.3 Necessity for Green Products / Green Processes

An increase in human population relates to greater industrial activities to meet the requirements of the population but at the cost of greater emissions of greenhouse gases and carbon dioxide. Environment Canada states that in 2013 there was a total of 726 megatonnes (Mt) of carbon dioxide equivalent released into the environment, which is only 3% below the 2005 emissions of 749 Mt (16). In 2009, Canada signed the Copenhagen Accord deal, which commits the country to reducing their greenhouse gas emissions up to 17% lower than the 2005 levels by 2020. For Canada to achieve this goal, focus needs to be directed towards the oil and gas, and transportation sectors which account for 25% and 23% of total greenhouse gas emissions respectively (16).
The production of valuable industrial commodities via renewable energy sources is a current popular research topic. Currently, many industries such as pharmaceuticals, agriculture, and food make use of microbes for production of commodities (29). The focus is to create efficient and cost effective methods for production of industrially valuable commodities. An increase in fossil fuel consumption has been raising health concerns due to higher carbon emissions as well as higher water and air pollution (16). Using biological means for industrial processes where possible can allow for such processes to be greener and help maintain a healthier environment.

1.4 Role of *E. coli* in the Industry

The two most common strains of *E. coli* used in the industry are the K-12 and B strains (33) as they serve various purposes, such as in protein, biofuel, and pharmaceutical production. *E. coli* is the dominant organism in the biotechnology industry due to its fast growth in low-cost rich media and is easy to scale up, with the potential to produce high yields of a given product (6). Modifications to the organism’s genome are most often required in engineering a strain to behave as desired. Deletions in the genome are often done through a lambda-red recombination method (13). Recent method known as “clonetegration” was published which can be used for insertions of genes in the genome (42). *E. coli* was first used as an expression host in 1982 for the production of a biopharmaceutical used in the regulation of human insulin which was made to treat diabetes (6).

Biofuels have gained a lot of interest due to depleting fossil fuel reserves. Insight into the production of biodiesel through microbial means is increasing due to its advantage over plant-based biofuels, such as faster production, reduced land use, and lower CO₂ emissions (6). As biodiesel is a fatty-acid based biofuel and can be produced using microorganisms such as *E. coli*, the need for forest clearance as well as the use of fertilizers and certain fermentation processes is eliminated (6). The concern for production via microbial means lies in the transition to its production in large industrial amounts.

Aside from proteins and biofuels, *E. coli* strains have also been engineered to produce valuable commodities such as phenol (25). Kim and et al. (2014) engineered *E. coli* strains that would reduce phenol toxicity while increasing phenol productivity. They were successful in achieving a
high titre of 3.79 g/L and productivity to 0.18 g/L/h of phenol via microbial fermentation. Another example is the overproduction of Taxol precursor (1 g/L), a potent anticancer drug whose current production method is not very cost-efficient (2).
Chapter 2

2 Literature Review

2.1 Carbon Fixation Pathways

The method of fixing carbon dioxide to organic compounds is called carbon fixation. During this process, plants absorb 10 times more carbon dioxide than is emitted to the atmosphere by human activities (8). It is one of the most important biochemical processes occurring in nature. All plants and most prokaryotes make use of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) which is part of the Calvin-Benson cycle to fix carbon dioxide into biomass (17). This is important because carbon fixation can be a rate-limiting factor for plants and organisms whose agricultural environment is abundant in water, light, and nutrients (8).

There are six naturally occurring carbon fixation pathways known as of 2011 which includes photosynthesis in plants, two different pathways in bacteria (22), two in archaea (9) and one pathway is found in both archaea and bacteria (31). Table 1 below provides details of the five alternative carbon fixation pathways where 3-hydroxypropionate/4-hydroxybutyrate cycle and dicarboxylate/4-hydroxybutyrate cycle are variants of the 3-hydroxypropionate cycle. Figure 2 below shows 4 of these alternative pathways; the reductive acetyl-CoA pathway is excluded due to its anaerobic requirements and use of metallo-chemistry, which makes it difficult to express in non-acetogenic microbes (15). The other 4 pathways all make use of carbon in the form of carbon dioxide or bicarbonate ion along with acetyl-CoA and succinyl-CoA to fix carbon into pyruvate or acetyl-CoA. These alternative carbon fixation pathways provide some advantages over the Calvin cycle such as lower ATP requirements, the ability of certain enzymes to utilize bicarbonate ion instead of carbon dioxide, and that the pathways have similar or higher theoretical efficiencies compared to the Calvin cycle (15). Biotechnological applications of these pathways can be problematic because they are either present in unusual organisms and or the carboxylation reactions can be partially rate-limiting. Thus, the fixation of carbon into a valuable product by these native alternative carbon fixation pathways in an industrial microorganism can be a daunting challenge.
Table 1: Overview of the five autotrophic pathways besides photosynthesis also known as carbon fixation pathways (8, 17).

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Class</th>
<th>Year</th>
<th>Discovered By</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reductive Acetyl CoA Pathway</strong></td>
<td>Anaerobic bacteria (acetogens) and archaea (methanogens)</td>
<td>1965</td>
<td>Ljungdahl and Wood</td>
</tr>
<tr>
<td><strong>Reductive Citric Acid Cycle</strong></td>
<td>Anaerobic and microaerobic bacteria</td>
<td>1966</td>
<td>Buchanan and Arnon</td>
</tr>
<tr>
<td><strong>3-Hydroxypropionate</strong></td>
<td>Green Nonsulfur Bacteria</td>
<td>2002</td>
<td>Herter and Eisenreich</td>
</tr>
<tr>
<td><strong>3-Hydroxypropionate/4-Hydroxybutyrate Cycle</strong></td>
<td>Archaea</td>
<td>2007</td>
<td>Berg and Fuchs</td>
</tr>
<tr>
<td><strong>Dicarboxylate/4-hydroxybutyrate cycle</strong></td>
<td>Anaerobic Archaea</td>
<td>2008</td>
<td>Huber and Fuchs</td>
</tr>
</tbody>
</table>
Photosynthetic organisms are not the ideal option to generate liquid fuel to accommodate a large industry due to low rate of solar energy conversion, and greater land use (20). Thus, the attention
has now been turned towards non-photosynthetic organisms that can fix CO$_2$ to produce liquid fuels also known as electrofuels. Acetogenic microorganisms also have the ability to fix CO$_2$ into acetate, but studies have shown that with some modifications other products can be produced (44). *Clostridium ljungdahlii* has been described as a possible organism for the production of butanol (28) and butyrate, as well as ethanol from syngas metabolism (44).

Microorganisms can use electrical energy to produce fuels and valuable commodities from carbon dioxide; this idea is based on the movement towards electric economy from an oil-based economy (32). Despite this being a fairly new concept some industries have already made considerable advances, such as the invention of electric cars by the automobile industry. There are many advantages of having an electric economy through production of commodities via microbes. This includes no competition for land use with food production, a smaller land area requirement than biomass production, as well as the possibility for the processes to be carried out in remote and extreme environments, and most importantly, it does not require environmental degradation as with biomass production.

Carbon dioxide can be electrochemically reduced to electron donors such as H$_2$, formate, ammonia, sulfide, and Fe(II) (32). H$_2$ and formate are more convincing donors because they have a lower redox potential that can allow microorganisms to gain energy for growth from the production of organic compounds derived from reduction of carbon dioxide. Microorganisms that can use formate for growth are known as formatotrophic microorganisms. Formate is considered a good electron carrier because it is very stable and non-toxic in all forms. It can easily enter the cell through the lipid membrane as it has a high permeability coefficient and can undergo passive movement, with the help of a specific formate channel, FocA (7).

Engineering these types of organisms to express non-native pathways relies on the knowledge of synthetic biologists, metabolic engineers, and microbiologists. Thus, the work on a CO$_2$ fixation pathway expressed in *E. coli* described hereon in the thesis has been supervised by Dr. Radhakrishnan Mahadevan whose expertise is in synthetic biology and metabolic engineering and by Dr. Alexei Savchenko whose expertise lies in microbiology and protein chemistry areas.
2.2 Reductive Glycine Pathway

Carbon fixation pathways are the dominant method of storing energy and producing biomass. Since there are limited natural pathways via which this process can occur, computational biologists use metabolomics and modeling to design pathways that are more efficient and adaptable to the organism’s environment.

Bar-Even et al. described few formatotrophic growth pathways for *E. coli* involving the use of electrosynthesis (7). They show that it is better to engineer a model industrial organism for growth on formate despite the availability of natural organisms that can grow on formate solely, as these organisms are not suitable for bulk cultivation or industrial use. Taking into consideration biomass yield, thermodynamic favouribility, enzyme kinetics, and the number of foreign enzymes required for a pathway, they concluded that the reductive glycine pathway has the most positive feedback with the least potential for expression problems. Some of the formate molecules produced by electrosynthesis can be used for energy via formate dehydrogenase to produce NADPH and CO₂ with metal co-factors. Native formate dehydrogenase of *E. coli* cannot transfer electrons to NADP due to which a foreign protein must be expressed instead. Remaining formate molecules not used in electrosynthesis can be used as metabolites in *E. coli* via the reductive glycine pathway.

With *E. coli* as the host organism, the reductive glycine pathway, in which all of the steps are reversible, makes use of two foreign enzymes, formate dehydrogenase and formate tetrahydrofolate ligase. The pathway also does not contain oxygen sensitive enzymes, theoretically has high biomass yield, and maintains high chemical motive force throughout, making it the most ideal pathway out of the ones analyzed (7). The first two steps of the pathway produce the intermediate 5,10-methylene-THF, which is used by the glycine cleavage system (GCV) to synthesize glycine. It is further used to synthesize serine from glycine after the THF is recycled in the first step to make 10-formyl-THF from formate. Thus, two equivalents of 5,10-methylene-THF are required in the process with the first two steps in the pathway occurring twice.

Despite the reductive glycine pathway being the most feasible, it is not without its challenges. Native glycine cleavage system of *E. coli*, while proven to be reversible, has not been shown to produce glycine in-vivo (7), thus requiring the expression of a foreign glycine cleavage system.
Another concern could be the limitation of the system kinetics due to the low affinities towards carbon dioxide and ammonia. A suggested solution for this is to grow \textit{E. coli} under high carbon dioxide and ammonia concentration until formatotrophic growth is achieved, after which concentrations can be lowered.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{glycine_pathway.png}
\caption{A schematic of the reductive glycine pathway using formate to fix CO\textsubscript{2} to pyruvate. Redrawn from Bar-Even, 2013. ‘F’, ‘D’, ‘GCV’, ‘G’, ‘S’ represents formate tetrahydrofolate ligase (EC 6.3.4.3), bifunctional methenyltetrahydrofolate cyclohydrolase/dehydrogenase (EC 3.5.4.9), glycine cleavage system, serine hydroxymethyltransferase (EC 2.1.2.1) and serine deaminase (EC 4.3.1.17), respectively. Enzyme F is foreign in \textit{E. coli} and enzyme complex GCV natively acts in the opposite direction in \textit{E. coli}, cleaving glycine instead of synthesizing glycine.}
\end{figure}

2.3 Glycine Cleavage System

The glycine cleavage system (GCV) is a complex of 4 proteins that work together and is found widely amongst animals, plants, and bacteria (23). It is also known as the glycine synthase when it functions in the direction of producing glycine instead of the cleavage of glycine. The three
specific GCV proteins are H, P (glycine dehydrogenase – EC 1.4.4.2), and T (aminomethyltransferase – EC 2.1.2.10) which are found alongside each other in the genome, often under the same operon. L, also known as dihydrolipoamide dehydrogenase (EC 1.8.1.4) is not specific to GCV and is located far from the GCV operon in the genome. Sagers and Gansulus first annotated the glycine cleavage system reactions using the anaerobic bacterium, *Diplococcus glycinophilus*, which was shown to ferment glycine to acetate, carbon dioxide, and ammonia (40). They showed that both the carbons of acetate are incorporated from the α-carbon of glycine, whereas the carbon in carbon dioxide comes from the carboxyl carbon of glycine.

The reaction mechanism of the glycine cleavage system is annotated below in Figure 4. H protein is known as the “arm of the complex” as all the reactions occur on its side chain. The first step is the decarboxylation of glycine to carbon dioxide carried out by P protein. Then, T-protein methylenates THF and releases ammonia. L protein is involved in oxidizing the arm of H protein by reducing NAD$^+$ to NADH allowing H protein to be in reduced form to perform cleavage of glycine again. The methylene group bound to THF can be used for one-carbon metabolism and NADH is needed for many purposes such as energy production through electron transport chain (23).
Figure 4: The reaction mechanism of the cleavage of glycine by the glycine cleavage system shown in black arrows represents the native activity of GCV in *E. coli*. Carbon dioxide and ammonia is released during the cleavage of glycine. The arrows in green show the opposite direction of glycine cleavage system, the synthesis of glycine (glycine synthase), which is the native activity of the GCV in *C. acidurici* 9a.

Glycine and C1 units are important metabolites for the cell and are used for the synthesis of many products such as proteins, purine biosynthesis, methionine, thymine, and other methylated products (18). The major synthesizing pathway for glycine and C1 units synthesis is from the degradation of serine to glycine and 5,10-methylenetetrahydrofolate reaction, which is carried out by glyA gene encoding, serine hydroxymethyltransferase. The other pathway is catalyzed by the glycine cleavage system that cleaves glycine to carbon dioxide, ammonia, and 5,10-methylenetetrahydrofolate (40). The glycine cleavage system is highly regulated using Lrp, PurR, and GcvA proteins and is shown to be inducible with exogenous glycine (35). Lrp, known as leucine-responsive regulatory protein, regulates the transcription of many amino acid metabolism genes and is required for normal GCV activity (30). PurR is a repressor protein that
represses purine biosynthetic genes by repressing their transcription when purine concentrations are high (24). GcvA also acts as a repressor to GCV activity when exogenous purine is present but it is an activator for GCV activity with the presence of exogenous glycine (45,46). The *E. coli* glycine transport system encoded by *cyc*, transports glycine and is known to induce GCV by increasing endogenous glycine (18). With so many effectors involved in the regulation of the glycine cleavage system, it can be understood that the system must be of high importance to the cell’s functioning.

In *E. coli*, the GCV functions in the direction of glycine cleavage. For the reductive glycine pathway (section 2.2) to function in the correct direction, GCV of *E. coli* would need to operate in the reverse direction, which is known to exist but not yet proven in-vivo. Thus, it seems a better option to express another bacterium’s glycine cleavage system that naturally has flux towards synthesis of glycine. After some consideration, attention was drawn to *Clostridium acidurici 9a* which has glycine synthase activity, as well as evidence for serine production from glycine and the deamination of serine to pyruvate (39). The glycine reductive pathway is thought to be completely present in *C. acidurici 9a* but as it is not an ideal industrial organism, the interest would be to express the glycine synthase activity of *C. acidurici* in *E. coli*. Table 2 below compares the glycine cleavage system of both organisms.

**Table 2: A Comparison of Glycine Cleavage System proteins between *E. coli* and *C. acidurici 9a* (italics).**

<table>
<thead>
<tr>
<th>GCV Protein</th>
<th>Length (basepair)</th>
<th># of Amino Acid</th>
<th>Sequence Identity (Amino Acid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>390/381</td>
<td>129/126</td>
<td>52.3%</td>
</tr>
<tr>
<td>L</td>
<td>1425/1383</td>
<td>474/460</td>
<td>34.0%</td>
</tr>
<tr>
<td>P</td>
<td>2874/2812</td>
<td>957/935</td>
<td>30.6%</td>
</tr>
<tr>
<td>T</td>
<td>1095/1113</td>
<td>365/370</td>
<td>12.7%</td>
</tr>
</tbody>
</table>
Chapter 3

3 Motivation and Statement of Objectives

Biochemistry and metabolic engineering are the basis of understanding production of chemicals, biofuels, pharmaceuticals, and other industrially valuable commodities through biological means. It is predicted that by 2052, there would be no oil deposits left if current depletion rate of 4 billion tonnes per year continues and the shortage can occur even earlier if population growth rate is considered (12). Thus, alternative energy sources and production methods need to be discovered that are economically feasible, environmentally-friendly, and have the potential to be produced on a large scale. Using microorganisms as production factories for an end product or an important intermediate has revolutionized many production lines in various industries such as food, pharmaceuticals, and polymers as described in section 1.4.

The reductive glycine pathway is a carbon fixation pathway that will allow for production of pyruvate, which is a central metabolite in E. coli for many pathways. If one can improve the productivity of pyruvate in E. coli, it will be useful to combine this pathway with another one currently established in the industry to produce a valuable commodity.

To express the reductive glycine pathway in E. coli, the foreign glycine cleavage system and enzyme tetrahydrofolate ligase from Clostridium must be active in E. coli. Also, the native GCV of E. coli must be excised or inactivated so that it does not compete with GCV of C. acidurici 9a as the flux for both is in opposite directions. Lastly, a strain that is already a serine and glycine auxotroph must be used to prove the expression of the glycine cleavage system. There are two known pathways of L-serine production in E. coli, the major route being from 3-phospho-D-glycerate via serA, serC, and serB enzymes (37). The other pathway uses the glycine cleavage system to synthesize serine from glycine. However, for this to occur, native glycine cleavage system of E. coli would have to flux in the less favourable direction. If both of these pathways were not available for serine synthesis, the strain would not be able to survive. Using this to our advantage, if the glycine cleavage system of Clostridium is expressed in a strain where both the serine biosynthesis pathways have been removed, the strain can survive due to the serine production from the expressed foreign glycine cleavage system. The expression and optimization of the overall reductive glycine pathway in E. coli can seem daunting if the details are not
covered first. The focus of this project will be on expression of *C. acidurici 9a* glycine cleavage system in *E. coli*.

**Statement of Objectives and Hypothesis:**

1. *Clone C. acidurici 9a GCV proteins in separate plasmids and preform small-scale expression analysis in E. coli.*

   It is important to check that the foreign proteins are expressed in *E. coli* when expressed individually before trying to express all three glycine cleavage system proteins at once in the same construct. The expressed *C. acidurici 9a* GCV proteins are hypothesized to be soluble at 14 kDa for gcvH, 50 and 54 kDa for gcvP subunits, and 41 kDa for gcvT.

2. *Carry out deletion of native GCV proteins of E. coli using lambda red recombinase on E. coli BW 25113 ΔserA strain.*

   Native glycine cleavage system of *E. coli* must be removed to avoid competition and counter balancing of flux from glycine cleavage system of *C. acidurici 9a*. Excising native GCV system of *E. coli* should hypothetically create a glycine and serine auxotroph strain, replacing a 5500 base pair fragment with 2000 base pair fragment of chloramphenicol antibiotic.

3. *Design construct with C. acidurici 9a GCV proteins and transform in E. coli BW 25113 ΔserAΔgcv.*

   A construct containing three GCV proteins, each with a separate but same promoter and RBS site will allow expression of the glycine cleavage complex with just one plasmid. This reduces the stress applied on the strain compared to multiple plasmids as one antibiotic resistance only constricts the strain. Also, expressing using the same promoter and RBS site for each gene provides a base expression result for future modifications on strength of promoter to increase efficiency of the complex. Therefore, it can be hypothesized that equal amounts of each gene should be transcribed and can be qualitatively analyzed on the SDS-PAGE gel.

4. *Characterize the growth of the strain E. coli BW 25113 ΔserAΔgcv expressed with C. acidurici 9a GCV proteins.*

   Growth characterization of the strain and metabolite analysis through HPLC will
determine if the *Clostridium* glycine cleavage system is capable of rescuing the glycine and serine *E. coli* auxotrophic strain and what metabolites, if any, are accumulated. The metabolites of interest are pyruvate, glycine, and serine. It can be hypothesized that the wild type strain harboring the foreign GCV proteins would grow slower under more stress due to plasmid expression and flux competition between native GCV and foreign GCV complexes. The native *E. coli* GCV- strain should not grow and expressing the foreign GCV complex should rescue the strain.

5. *Clone and express Clostridium ljungdahlii* formate tetrahydrofolate ligase in parallel with GCV proteins in order to test expression of the reductive glycine pathway.*

Formate tetrahydrofolate ligase will need to be cloned in a plasmid compatible with the plasmid used to clone the glycine cleavage system. Baseline expression results can be obtained by having the same promoter for both the plasmids once expression of the pathway on formate has been confirmed. This can be modified to obtain optimal expression results. Expressing foreign formate tetrahydrofolate ligase and GCV complex should hypothetically induce expression of the reductive glycine pathway on formate.
Chapter 4

4 Methods

4.1 Ligase Cycling Reaction to Construct Plasmids with Individual \textit{C. acidurici} 9a GCV proteins

4.1.1 Amplification of Plasmid and Insert

Plasmid pTRC99a was amplified excluding sites 265-321 using primers designed for Ligation Cycling Reaction. The reverse primer contained a UNS sequence as an overhang obtained from a recent publication on DNA parts for synthetic biology (43). Each plasmid for each insert was amplified using the following PCR program with the Pfx system: 2:30 min at 95°C, 15 sec at 95°C, 30 sec at 63°C, 4 min at 68°C, go to step 2 (repeat 29x), 10 min at 72°C, and hold at 4°C forever. Primers are listed in Appendix C. PCR clean-up was performed and the following concentrations were measured on nanodrop for amplified plasmid H, plasmid P, and plasmid T respectively; 345.4 ng/uL, 184.3 ng/uL, 170.1 ng/uL. The inserts were amplified from genomic DNA of \textit{Clostridium acidurici} 9a obtained from DSMZ-Deutsche Sammlung von. The forward primer included the RBS sequence with a translational initiation rate of 20000 and the reverse primer was designed to consist another UNS sequence. The inserts were amplified using the following PCR program with the Pfx system: 2:30 min at 95°C, 15 sec at 95°C, 30 sec at 56°C, 4 min at 68°C, go to step 2 (repeat 29x), 10 min at 72°C, and hold at 4°C forever. For insert P and T, less amount of gDNA was used (0.2 uL instead of 0.4 uL), 1 uL of Pfx was used (instead of 0.5 uL), the annealing temperature was set to 53°C and the elongation time to 3 min. PCR clean-up was performed on the amplified inserts and the following concentrations were measured on nanodrop for insert H, P, and T respectively; 50.6 ng/uL, 48.8 ng/uL, 88.4 ng/uL.

4.1.2 Ligating Plasmid with Insert (LCR)

3 reactions were set up to phosphorylate the 5’ ends of the DNA parts involved in the LCR mix. For pTrc99a_H and pTrc99a_T, a 20ul mix was prepared containing 100 fmol of the insert and the plasmid, 1 uL of 100mM ATP, 1 uL of 10U T4 polynucleotide kinase, 2 uL of 10x Ampligase thermostable DNA ligase reaction buffer, in water. For pTrc99a_P, 1:4 ratio of plasmid:insert was needed for successful ligation. The mix was incubated at 37°C for an hour, followed by incubation at 65°C for 20 minutes to inactivate the enzyme.
The 25 uL LCR mix was prepared using 15 uL of the phosphorylation mix, 0.5 uL of each 1.5uM bridging oligos (2 bridges in total), 1 uL of 10x Ampligase thermostable DNA ligase reaction buffer, 1.5 uL of 5U/uL Ampligase, 2.25 uL of 5M betaine, 2 uL of 100% DMSO in water. The PCR program used for the ligation reaction is as such: 2 min at 94°C, 10 sec at 94°C, 30 sec at 55°C, 60 sec at 66°C, go to step 2 (repeat 49x), and hold at 4°C forever. The bridging DNA fragments are listed in the Appendix C.

4.1.3 Transformation of Recombinant Plasmids in DH5α

5 uL of LCR mix was added to 40 uL of DH5α chemically competent cells and it was allowed to chill on ice for 15 minutes. The cells were heat shocked at 42°C for 1 minute, followed by 2 minutes on ice after which the cells were recovered in 850 uL of SOC at 37°C with shaking at 220 rpm for 1.5 hours. The cells were centrifuged for 1 minute at 13.3 rpm; supernatant was discarded leaving about 100 uL of media in the eppendorf. The pellet was resuspended in the remaining 100 uL of media, 55 uL was plated and spread using glass beads on LB and ampicillin plates.

4.2 Small Scale Test Expression of C. acidurici 9a GCV Proteins in E. coli BL21 (DE3)

4.2.1 Transformation of Recombinant Plasmids in BL21-DE3 RIL

The recombinant plasmids obtained from LCR reaction in section 4.1.3 were used for test expression. 5 uL of the LCR mix was added 40 uL of BL21-DE3 RIL cells and it was allowed to chill on ice for 15 minutes. The cells were heat shocked at 42°C for 1 minute, and then placed on ice for 2 minutes. 850 uL of SOC was used to recover the cells at 37°C with shaking at 220 rpm for 1.5 hours. The cells were centrifuged for 1 minute at 13.3 rpm, supernatant was discarded leaving about 100 uL of media in the eppendorf. The pellet was resuspended in the remaining 100 uL of media, 55 uL was plated and spread using glass beads on LB, ampicillin and chloramphenicol plates.
4.2.2 Inducing Protein Expression

3 colonies from each transformation were picked using a tip and placed in 750 uL of ZYP-5052 rich Studier medium without lactose containing ampicillin and chloramphenicol antibiotics. The culture was allowed to incubate at 37°C for 5 hours with shaking at 220 rpm, before switching to 18°C without shaking for 20 minutes. The tips were removed and 2.25x10^{-4} mmol of IPTG was added to each culture which was then allowed to incubate overnight at 18°C with shaking at 230 rpm.

4.2.3 Lysing of Cells and Visualizing on SDS-PAGE GEL

The culture was centrifuged at 3000 rpm for 15 minutes, and the pellet was chilled at -80°C for 15 minutes after discarding the supernatant. The pellet was thawed at room temperature before suspending in 500 uL of Lysis buffer. The cells were incubated at 5°C with shaking at 230 rpm for 1.5 hours. 40 uL of lysate was mixed with 10 uL of 5X sample buffer for SDS-PAGE, this is the total expression sample. The remaining lysate was centrifuged at 13.3 rpm for 10 minutes. 40 uL of the supernatant was mixed with 10 uL of 5x sample buffer for SDS-PAGE, this is the soluble fraction. The total expression samples and the soluble fractions were incubated at 95°C for 10 minutes before loading 10 uL of sample on SDS-PAGE gel.

The electrophoresis was run at 120 V for 90 minutes. The gel was washed three times in water, each time microwaving for 1 minute, shaking for 1 minute at room temperature, and discarding the water. After the last wash, the gel was stained with SimplyBlue SafeStain dye by microwaving for a minute and shaking at room temperature for 1.5 hours before destaining the gel with water. The gel can be seen in Figure 4.

4.3 Lambda red to excise GCV proteins of E. coli

Chloramphenicol cassette in pKd3 was amplified with overhangs homologous to regions outside of the GCV operon in E. coli genome. Primers are listed in Appendix C. 1 mL of an overnight culture of E. coli BW 25113 ΔserA in LB was put in fresh LB and allowed to grow until OD_{600} = 0.36. 1 mL of the culture was centrifuged at 13.3 rpm for 1 minute and supernatant was discarded before being washed with 1 mL of 10% glycerol. This was repeated 3 times. After last
wash, pellet was suspended in liquid left over after supernatant was discarded. 27 uL of the cells was added to 3 uL of pKd46 and allowed to chill on ice for 15 minutes before electroporating. 1 mL of SOC was added to recover the cells at 30°C for 1.5 hours. 55 uL of the cells was plated on LB plate with ampicillin and incubated at 30°C overnight. Many colonies had grown overnight.

An overnight culture of 1 colony of *E. coli* BW 25113 ΔserA with pKd46 was started in 3 mL LB and 1000 ug/mL ampicillin at 30°C and allowed to grow in fresh 100 mL of LB with ampicillin the next morning. 0.001 mol of L-arabinose was added to the culture to induce lambda red recombinase expression when OD$_{600}$ = 0.08. At OD$_{600}$ = 0.4, 1 mL of the cells were centrifuged, washed, and electroporated with 3 uL of pKd3 fragment using same protocol as mentioned for pKd46. The cells were allowed to recover with 1 mL of SOC at 37°C for an hour. 55 uL of the cells were plated on LB plus ampicillin and chloramphenicol plates and allowed to incubate at 37°C overnight. There were no colonies on the plate the next day. 100 uL of the remaining recovered cells were plated onto another LB plate with ampicillin and chloramphenicol and colonies were seen the following day after incubation at 37°C. Colony screening was preformed using Taq polymerase with the following PCR program: 2 min at 94°C, 30 sec at 94°C, 30 sec at 50°C, 2 min at 68°C, go to step 2 (repeat 29x), 10 min at 72°C, and hold at 4°C forever. This protocol was used for checking of chloramphenicol gene insertion shown in Figure 12. Primers spanning outside of the chloramphenicol cassette were used to check the location of the insertion to verify it has replaced the GCV operon. The same protocol was used but the elongation time was set to 6 min at 68°C as shown in Figure 7.

### 4.4 Designing construct to ligate GCV proteins together

#### 4.4.1 Amplification of H, P, T gene inserts

H, P and T inserts were amplified from their respective plasmids designed in section 4.1.3. Gene H was amplified from the UNS regions surrounding the gene. Forward primers for Gene P and Gene T consisted of UNS sequence as overhang. The primers are listed in the appendix. The Pfx system was used to amplify using the following PCR program: 2:30 min at 95°C, 15 sec at 95°C, 30 sec at 58°C (P was at 62°C), 3.5 min at 68°C, go to step 2 (repeat 29x), 10 min at 72°C, and hold at 4°C forever. DpnI treatment and PCR clean-up was preformed and the following
concentrations were measured on nanodrop for insert H, P, and T respectively; 98.9 ng/uL, 60.1 ng/uL, and 47.8 ng/uL.

4.4.2 Ligating pTRC99a backbone with H, P, and T gene inserts (LCR)

A phosphorylation reaction was set up to phosphorylate the 5’ ends of the DNA parts involved in the LCR mix. A 20ul mix was prepared containing 100 fmol of each insert and the open plasmid, 1 uL of 100mM ATP, 1 uL of 10U T4 polynucleotide kinase, and 2 uL of 10x Ampligase thermostable DNA ligase reaction buffer, in water. The mix was incubated at 37°C for 1 hour and inactivated at 65°C for 20 minutes. The LCR reaction was done as described in section 4.1.2. The ligated plasmid was transformed in DH5alpha cells using same protocol as in section 4.1.3. Miniprep concentration was measured on nanodrop to be 278.0 ng/uL.

4.5 Growth Characterization of E. coli strains

E. coli BW 25113 was obtained from Keio collection at Yale and serA deletion was preformed by A. V. Pandit in Dr. Mahadevan’s lab. Transformation of pTRC99a backbone and pTRC99a_HPT was completed via electroporation method in E. coli BW 25113, ΔserA, and ΔserAΔgcv:chl- strains. In each case, the strain was inoculated overnight in LB with appropriate antibiotics and rediluted in fresh 50 mL of LB to achieve OD$_{600}$ around 0.5. 1 mL of the cells were centrifuged at 3000 rpm for 4 minutes, supernatant was discarded and the pellet was resuspended in ice-cold 10% glycerol. The glycerol wash was preformed 3 times and after the last wash, the pellet was resuspended in the remaining solution in eppendorf. 40 uL of the cells were chilled on ice with ~300 ng of the plasmid for 15 minutes before electroporating the cells and recovering in 1 mL of SOC for 1 hour at 37°C. For the double mutant strain, 2 mL of the cells were centrifuged before the glycerol washes to obtain more cells for electroporation. Also, it was recovered for longer time, 1.5 hours at 37°C.

All strains characterized were started overnight in LB with appropriate antibiotics from a colony and transferred to M9 media the next morning. 1 mL of the overnight culture was centrifuged at 3000 rpm for 4 minutes and the supernatant was discarded. The pellet was washed three times with 10% glycerol, each time spinning at 3000 rpm for 4 minutes and discarding the supernatant. Then, the pellet was resuspended in 50 mL of M9 media containing 35 mL of autoclaved water,
10 mL of M9 salts, 0.1 mmol of MgSO₄, 0.56 mmol of glucose, 0.5 mol of glycine, 0.005 mmol of CaCl₂, 50 µL of trace metals mix and 1.125x10⁻⁴ mmol of IPTG was added were applicable along with appropriate antibiotics (50 µL of 100 mg/ml ampicillin, 50 µL of 12.5 mg/ml chloramphenicol). A stock solution of M9 salts was prepared by making a 1000 mL solution using 800 mL H₂O, 64.0 g Na₂HPO₄·7H₂O, 15.0 g KH₂PO₄, 2.5 g NaCl, and 5.0 g NH₄Cl, which was then stirred and adjusted to 1000 ml by addition of water. HPLC was used to measure the metabolites (Aminex HPX-87H, 5mM sulfuric acid, 0.4 mL/min, 36°C).
5 Results and Discussion

5.1 Expression of *C. acidurici* 9a GCV Proteins in *E. coli*

Figure 5: Expression screening of glycine cleavage system proteins from *C. Acidurici* 9a in *E. coli* BL21-CodonPlus(DE3)-RIL on a small-scale basis with pTRC99a backbone. BB stands for backbone only (pTRC99a) and HPT means construct with all 3 genes ligated. The expected sizes of H, P subunits, and T are 14 kDa, 50 kDa (Pa), 54 kDa (Pb), and 41 kDa, respectively. The first column for each protein shows all the proteins expressed in that sample, whereas the second column shows all the soluble proteins.

Expression screening should be the first test performed when expressing a foreign protein. Many strains are used as expression strains for such screening, one of them being BL21-CodonPlus(DE3)-RIL derived from the BL21-Gold competent cell line. This strain allows for a high level of heterologous protein expression in *E. coli* by containing extra copies of the argU,
ileY, and leuW tRNA genes that recognize the arginine codons (AGA and AGG), isoleucine codon (AUA), and leucine codon (CUA) (27). This protocol shows whether the protein is expressed when the correct band size is seen in the total lysate (column labelled). The column beside the total lysate column is the soluble column that shows the proteins that are found to be soluble. Soluble proteins can be further studied for structural analysis (10).

In this result, both subunits of P were seen as expected at 50 kDa and 54 kDa, along with protein T at 41 kDa in the total lysate fraction as well as soluble fraction. Protein H was seen at 17-18 kDa, which is higher than expected at 14 kDa. A darker band at the approximate expected size compared to the rest of the sample is a good indication the expression of the plasmid was induced and the protein is present. The plasmids were constructed using pTRC99a backbone and the plasmid backbone was used as a control plasmid (BB). None of the proteins that were expressed using individual constructs were seen when the control plasmid was expressed. Inducing the plasmid with the glycine cleavage protein construct (HPT) showed soluble expression of protein P subunits and protein T as seen with the respective individual plasmids but expression of protein H was not seen.

Two unexpected observations relating to protein H were noticed through the small-scale expression analysis of C. Acidurici 9a glycine cleavage system proteins in E. coli BL21-CodonPlus(DE3)-RIL. When protein H was expressed individually, the protein was seen at a higher size than expected. Initially, the hypothesis was that protein H is bound to a ligand or the protein sequence is being read past the stop codon. To answer these questions, soluble protein H band was excised using the protocol described by Shevchenko and colleagues and analyzed through mass spectroscopy (41). The sequence obtained from mass spectroscopy analysis was blasted against Clostridium proteins, E. coli proteins, and the pTRC99a plasmid to see if there was any over-reading of the gene. The sequence only resulted in high confidence level with C. Acidurici 9a glycine cleavage system H protein. Other studies have also shown similar results; Motokawa and colleagues from the University of Tokushima also obtained a band size of 18.5 kDa when detected against antibody for E. coli H protein expressed in E. coli (36). Their reasoning was that the over expression of E. coli H protein in E. coli causes most of the expressed protein to remain as inactive apoform and leading to a slightly higher band size than expected. This can also explain why protein H band in this result is slightly higher and also at the size they observed (18.5 kDa).
The second unexpected observation was the absence of protein H all together in the expression of construct HPT. An explanation for this could be the reduced activity of all three proteins when compared to their individual counterparts. H protein, when expressed individually, showed a lighter band compared to P and T protein so when the construct was expressed, the expression of H protein was reduced enough not to be detectable on the SDS-PAGE gel. The reduced activity of the construct proteins could be due to lower cell growth as a larger plasmid has more stress on the growth of the cells, leading to less number of plasmids being induced leading to less expression.

**Significance of Finding:** *Clostridium acidurici 9a* glycine cleavage system proteins can be expressed in *E. coli* and are seen to be soluble, indicating the construct can be used as part of the reductive glycine pathway. Also, the proteins can be purified if in-vitro studies are to be done.

5.2 Construct Design

![Figure 6: Individual construct design for the *C. acidurici 9a* glycine cleavage system proteins in the IPTG-inducible pTRC99a backbone. UNS stands for unique nucleotide sequence and RBS for ribosomal binding site.](image)

Each individual glycine cleavage system gene from *C. acidurici 9a* was cloned into the IPTG-inducible plasmid, pTRC99a, using a relatively recent cloning method called ligase cycling reaction (LCR) (14). The plasmid was amplified at the promoter site with a UNS sequence as an overhang and the gene was amplified from *C. acidurici 9a* genomic DNA with an RBS sequence as overhang on the forward primer and UNS sequence as overhang on the reverse primer. Thus, two oligio-bridges were designed that ligated the UNS sequence from the 5’ end of the plasmid
to the RBS site and the UNS sequence from the gene to the 3’ end of the plasmid. The purpose of placing UNS sequence before and after the gene is if rearrangement of genes is required for a future multi-gene construct, it can be done using the appropriate bridges.

Plasmid H and plasmid T were constructed using standard LCR protocol. The protocol to ligate backbone pTRC99a and gene P needed to be optimized to a ratio of 1:4 molar plasmid to insert, maybe because the insert was much longer than gene H or gene T.

From these individual constructs, each gene was amplified starting at the promoter region with a UNS sequence as an overhang on the forward primer and reverse primer at the UNS sequence after the gene. Gene H was amplified starting at UNS 6 as the promoter was already part of the final plasmid. Figure 6 shows the individual fragments amplified to design the final plasmid. Only the pre and post UNS sequence parts of the gene fragment are shown but each fragment contains all the parts in between the UNS sequences as shown in Figure 5, including the promoter, middle UNS sequence and the RBS site. 4 oligo bridges were required to ligate 3 inserts plus the backbone. Each insert has its own promoter and RBS site so the expression level of each gene is under the same strength, allowing the order of the proteins to be less significant. The preference was to have protein H first in the construct because it is the “arm” of the complex, one of the most critical proteins. It is often seen that expression of the latter genes is weakened when multiple genes are expressed under one promoter (26).

Other methods were also considered such as Gibson Assembly, co-expression of duet plasmids, and LCR by ligating one gene to backbone per cycle. Gibson assembly was fairly new as well but de Kok et al. described Gibson as less efficient than LCR due to which it was disregarded (14). Separating the proteins onto duet plasmids did not seem convincing as this doubles the stress on the cell to express two plasmids as well as overcome the stress of two deletions in the genome. Thus, this method was kept as a back-up plan. Multiple rounds of LCR can also be done where only one gene is ligated to the backbone each time but this is not a preferred method as each round would involve the amplification of the plasmid, increasing possibility of introducing mutations (4).

Significance of Finding: Ligase cycling reaction along with UNS sequences was helpful in constructing plasmids containing individual GCV proteins and the construct with all three
proteins. Although, it was noticed modifications to the procedure had to be done to ligate larger gene fragments, contradicting the literature.

Figure 7: The construct layout of plasmid pTRC99a harbouring *C. acidurici* 9a glycine cleavage system genes. The individual fragments were amplified from plasmids designed in Figure 5.
5.3 Deletion of *E. coli* GCV using Lambda Red Recombinase

Figure 8: 1% agarose DNA gel representing the deletion of *E. coli* glycine cleavage system at size approximately 5500 base pairs. The band is seen in the wild type (WT) *E. coli* BW 2511 ΔserA and not in the strain where the chloramphenicol cassette has replaced the glycine cleavage system (ΔserAΔgcv:Chl-). The chloramphenicol cassette is seen in the ΔserAΔgcv:Chl- strain at approximately 2000 base pairs which is not seen in the wild type strain.

The native glycine cleavage system of *E. coli* BW 25113 ΔserA has been removed from the genome using phage lambda Red recombinase. Primers were designed to amplify the glycine cleavage system with the forward and reverse primers 500 base pairs upstream and downstream of the operon, making the amplified fragment size 5500 base pairs. This fragment is not seen in the colony screened of the strain predicted to be serA- and GCV- phenotype, confirming it does have both the deletions. A noticeable band is seen around 2000 base pairs in the serA- and GCV-strain, which is the amplified fragment for the chloramphenicol cassette that is not seen in the
serA- strain, confirming the chloramphenicol cassette has replaced the glycine cleavage system. The other bands seen in the gel are due to mispriming of the primers. To verify this, the serA- and GCV- strain did not grow in minimal M9 media with glucose and glycine as carbon source as discussed in the next section.

**Significance of Finding:** Lambda red recombinase can be used to excise large genomic fragments of size more than 5000 base pairs. An E. coli strain that is GCV- and serA-, produces a glycine and serine auxotroph strain.

### 5.4 Growth Characterization of E. coli strain harbouring C. acidurici 9a GCV proteins

It was predicted that when *C. acidurici* 9a glycine cleavage system is expressed in the presence of native *E. coli* glycine cleavage system, it will have a toll on the cell’s growth and this is exactly what was observed as shown in Figure 9 (A and B) below. When *C. acidurici* GCV is expressed, not only does the cell have the burden of expressing those proteins but it also functions in the opposite direction, creating a loop hindering the growth of the cell. For both the wild type *E. coli* BW 25113 strain (Figure 9A), and the strain with the serA deletion (Figure 9B), the growth of the strain expressing the GCV proteins is seen to be plateauing when the strain expressing control plasmid is still in exponential phase, indicating the strength of stress applied by the expression of GCV proteins. The strain with the serA deletion has greater stress because it has to overcome the burden of gene deletion and is going from two pathways of serine production to only one pathway.

The opposite is seen in the strain where the native GCV has been deleted and as predicted. *E. coli* BW 25113 ΔserAΔgcv:Chl- acts as a serine and glycine auxotroph. As seen in Figure 9C, the strain expressing the control plasmid did not show any growth whereas the strain with *C. acidurici* 9a GCV proteins showed rescuing effects as OD$_{600}$ reached approximately 1.2 within 47 hours. To characterize the growth rate, the triplicates of both the strains were rediluted in fresh 50 ml M9 media as described in section 4.5 after the first transfer from LB to M9 media was allowed to grow until lag phase. It is understood that the transfer of cells to fresh media will decrease the lag phase as cells will adapt to media faster and some growth is seen in the lag
phase consistently. If more transfers are done, it is predicted that the cells will adapt better and decrease the lag phase. The readings between 28-31 hours are considered as exponential phase and they were plotted using the following relationship to determine growth rate, as shown in Figure 13 in appendix: \( \ln(\text{OD}_t/\text{OD}_0) = \mu(t-t_0) \) (19). The growth rate, \( \mu \), is calculated to be 0.175 h\(^{-1}\), is considered to be slow.

HPLC analysis was preformed on the double mutant strain, \( E. coli \) BW 25113 \( \Delta \text{serA} \Delta \text{gcv:Chl} \)-harbouring glycine cleavage system proteins of \( C. acidurici \) 9a. As seen in Figure 10, three main metabolites were detected: glucose, acetate and ethanol. Complete glucose consumption by the double mutant strain expressing the GCV proteins from \( C. acidurici \) 9a, whereas the double mutant strain harbouring only the plasmid backbone did not show significant glucose consumption. This is in accordance with the growth of the strain as increase in growth relates to glucose consumption. As glucose concentration was decreasing, acetate concentration was increasing. Acetate production is produced from pyruvate via two routes, either through pyruvate dehydrogenase complex and pyruvate formate-lyase (PFL) or through direct conversion by pyruvate oxidase (1). It is understood that accumulation of pyruvate is difficult as it is a central metabolite for many pathways. A significant change in ethanol concentration was not seen probably because the only detected concentration is from the chloramphenicol antibiotic solution used in the media.

From this characterization study, it is determined that glycine cleavage system proteins from \( C. acidurici \) 9a rescued the serine and glycine auxotroph strain. Thus, a portion of pyruvate must be produced via the reductive glycine pathway in parallel with production of pyruvate via glycolysis from glucose. It is also evident that glycine cleavage system of \( Clostridium \) is cleaving glycine supplemented in the media to produce 5,10-methylene THF that is further used as a co-factor to produce serine from glycine, rescuing the strain. Thus, it can be assumed GCV of \( C. acidurici \) 9a is working in the opposite direction in \( E. coli \) compared to its native activity in \( C. acidurici \) 9a. C-13 isotope studies can be preformed to verify the amount of acetate produced by pyruvate from the reductive glycine pathway by using labeled glycine.

**Significance of Finding:** Expression of \( Clostridium \) GCV proteins in the presence of \( E. coli \) native GCV complex reduces the growth of the strain due to flux and resource competition whereas expression in the glycine and serine auxotroph \( E. coli \) strain rescues the strain.
Figure 9: The growth of *E. coli* BW 25113 strains harbouring pTRC99a backbone (square) and pTRC99a_HPT (triangle) in M9 media with glucose as carbon sources. Media recipe described in section 4.5. Figure A, B, and C are *E. coli* BW 25113 WT, ΔserA, and ΔserAΔgcv:Chl-, respectively.
Figure 10: Characterization of *E. coli* BW 25113 ΔserAΔgcv:Chl- harbouring pTRC99a backbone (square) and pTRC99a_HPT (triangle).

5.5 Unsuccessful Expression of Reductive Glycine Pathway

After successful evidence of expression and activity of *C. acidurici* 9a glycine cleavage system in *E. coli* leading to rescue of the serine and glycine auxotroph strain, the next main hurdle is to express the full reductive glycine pathway. For this, the first step of the pathway requires a foreign enzyme, formate tetrahydrofolate ligase that was cloned from *Clostridium ljungdahlii* into pACYC184 vector with p15A origin of replication for compatibility with pBR322 origin of replication of pTRC99a vector. This allows for co-transformation of formate tetrahydrofolate ligase along with glycine cleavage system proteins. In this study, co-transformation of pTRC99a_HPT and pACYC184_THF was unsuccessful when both plasmids were electroporated
together, with equal amounts. Other methods do exist that can potentially lead to successful co-transformation; this work is now passed onto a colleague in the laboratory. Transformation of pACYC184_THF in the strain that already carries pTRC99a_HPT is one potential method that could be successful. Other methods include transformation of pACYC194_THF in chemically competent cells that carry pTRC99a_HPT. In the process of electroporating, many cells are killed whereas heat-shock method of transformation does not have a strong shock effect to kill cells. Another approach could be to excise the chloramphenicol cassette in the genome, reducing the antibiotic stress the strain has to overcome. Currently, the strain would have to grow on three antibiotics, expressing two plasmids with two different antibiotics, producing a total of 4 foreign proteins. This must be a large burden on the cell to overcome, thus, not being able to transform both plasmids easily. Adapting the strain few times before transforming the second plasmid might also be another potential method to try. The availability of many transformation methods and tricks seem to provide an optimistic outlook for the co-transformation of the plasmids required in the expression of the reductive glycine pathway.

**Significance of Finding:** Deletion of the full native E. coli GCV complex creates a growth defect strain that is not as stable and useful for downstream applications. Perhaps deleting one gene, such as gcvT from the complex may produce a more stable strain that is more competent for transformations.
6 Conclusions and Future Work

In this study, we saw that the deletion of serA gene and GCV operon in *E. coli* does create a serine and glycine auxotroph as predicted. Glycine cleavage system proteins of *Clostridium acidurici 9a* were seen to be expressed and soluble in *E. coli*, and retained their function to rescue the glycine and serine auxotroph strain, *E. coli* BW 25113 ΔserΔgcv. When the *Clostridium* GCV proteins were expressed in the presence of *E. coli* GCV proteins, the strain faced greater stress and hindered growth due to flux competition between the native and expressed GCV complexes. HPLC results indicate the production of pyruvate but not high enough to be accumulated.

Now that it is certain *C. acidurici 9a* GCV can be expressed in *E. coli* and does rescue the glycine and serine auxotroph strain, the construct can be modified to find the optimal combination of promoters and ribosomal binding site in front of each gene. As seen in the expression test, protein H was not seen when the construct was expressed indicating very low expression of the first gene in the construct. A stronger promoter, or RBS, can be used to increase expression, which may increase GCV activity as protein H is the arm of the complex and is the central protein out of the three-GCV proteins.

In terms of the pathway expression, the next steps would be to co-express GCV proteins and enzyme *F*, tetrahydrofolate ligase, which ligates THF to formate producing 10-formyl-THF. Successful expression of both enzyme *F* and GCV construct should enable the growth of *E. coli* on formate. Once the glycine and serine auxotrophic strain is shown to express both plasmids, focus can be geared towards optimal productivity and titre. As a final major step in this project, a pathway starting from pyruvate leading to a valuable commodity can be joined together. Pyruvate is hard to accumulate and over produce in the cell because there are many pathways that use pyruvate as an intermediate, so the cell consumes it as it is produced. If successful, the following step would be to express formate dehydrogenase with supplements of bicarbonate ion or carbon dioxide. This does come with challenges as expression of more foreign enzymes will increase burden on the cell but newer methods of foreign protein expression may allow for this to occur. A new method known as “clonetegration,” seems promising as it allows for cloning of foreign genes in the genome of the host overcoming plasmid expression stress (42).
References


Appendices

Appendix A: Lambda Red Colony Screening of Chloramphenicol gene insertion

Figure 11: Colony screening of *E. coli* BW 25113 ΔserΔgcv:chl- carried out at annealing temperature of 55°C. 8 screens out of 12 were positive for the insertion of the chloramphenicol resistance gene as amplified PCR product was seen around the expected size of 869 base pairs. Verification of the location of the chloramphenicol resistance gene in the genome is shown in Figure 8.
Appendix B: Growth rate calculation of E. coli BW 25113 ΔserAΔgcv:chl-

Figure 12: Exponential phase data points plotted of growth of E. coli BW 15113 ΔserAΔgcv:chl- from Figure 9 (D) to calculate growth rate. The slope of the best fit line, μ, is calculated to be 0.175 h⁻¹, also known as the growth rate.

Appendix C: Primers used in this study

Table 3: Primers used in this study.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence 5’ → 3’</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_pTRC99a_F</td>
<td>GGTATGGCTGTGAGGTCAGTAAATCAC</td>
<td>Forward primer to sequence insert in pTRC99a plasmid</td>
</tr>
<tr>
<td>Seq_pTRC99a_R</td>
<td>CTACGGCGTTCACGTCTGAGTTCGCC</td>
<td>Reverse primer to sequence insert in pTRC99a plasmid</td>
</tr>
<tr>
<td>Primer_PH_F</td>
<td>CATTACTCGCATCCATTTCAGGCTGTCTGC TCAGGTGAGTGCAGGAAGAGA</td>
<td>Forward primer for amplification of H open plasmid for individual construct</td>
</tr>
<tr>
<td>Primer_PH_R</td>
<td>GTATGTGACCGTAGATTCTTAGGTGGCAG</td>
<td>Reverse primer for</td>
</tr>
<tr>
<td>Primer/Gene Name</td>
<td>Sequence/Description</td>
<td>Function</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Primer1_GeneH_F</td>
<td>ACCTCGACCTTTAAGGATACCTAACAAACATGAGC AAAATAGTAGAAAG</td>
<td>Forward primer for amplification of H gene for individual construct</td>
</tr>
<tr>
<td>Primer1_GeneH_R</td>
<td>TTATTTCTGCACAGAAATACTCTGCTATTTTGACGCATCCATTAATG</td>
<td>Reverse primer for amplification of H gene for individual construct</td>
</tr>
<tr>
<td>BridgeH_5'</td>
<td>GCTGCCACCTAAGAATACTCTACGTCACATAACCTGCACCTTTAAGGATACCTAAACACATGAGC</td>
<td>Bridge used for LCR to ligate gene H and 5’ of H open plasmid</td>
</tr>
<tr>
<td>BridgeH_3'</td>
<td>CTGCAAAATACGAGATAATTCTGCTACAAGAATTACAAAACATGAGC</td>
<td>Bridge used for LCR to ligate gene H and 3’ of H open plasmid</td>
</tr>
<tr>
<td>Seq_H_F_576</td>
<td>GATGACGAACCAGAGACTAGTAAACACATGAGC</td>
<td>Forward primer to sequence insert H</td>
</tr>
<tr>
<td>Primer_PP_F</td>
<td>CTGACCTCCTGCCAGAATAGTAAGAACAACACGCAAAGTCGCTGTGTTTTGCGGATGAGAG</td>
<td>Forward primer for amplification of P open plasmid for individual construct</td>
</tr>
<tr>
<td>Primer_PR</td>
<td>GCTTGAGTTCTGCTGTGTTTTCCGTCTAGAACATG</td>
<td>Reverse primer to amplify P and T open plasmids for individual construct</td>
</tr>
<tr>
<td>Primer1_GeneP_F</td>
<td>ACCTCGACCTTTAAGGATACCTAACAAACATGAGC CATAGATATAATACCAAACATGAGC</td>
<td>Forward primer for amplification of P gene for individual construct</td>
</tr>
<tr>
<td>Primer1_GeneP_R</td>
<td>TTAATTTTGTATATTTTAGATATAGCTCTTTTGCCGCTTACGCTCTGCA</td>
<td>Reverse primer for amplification of P gene for individual construct</td>
</tr>
<tr>
<td>BridgeP_5’</td>
<td>GTAGACCGAAACAAACACGACAGATCCAAAGCAGCCTGACCTTTAAGGATACCTAAACACATGAGC</td>
<td>Bridge used for LCR to ligate gene P and 5’ of P open plasmid</td>
</tr>
<tr>
<td>BridgeP_3’</td>
<td>GAAGCTAATACGGCGAATAAGATCTATACATAAAATAATACAAAGTAACCTGCCAGCATG</td>
<td>Bridge used for LCR to ligate gene P and 3’ of P open plasmid</td>
</tr>
<tr>
<td>Seq_P_F_704</td>
<td>GGACGTAAACATAATACGCTGTCTATATGATG</td>
<td>Forward primer to sequence insert P</td>
</tr>
<tr>
<td>Seq_P_F_1293</td>
<td>GCTACATCAAATATCTGTTTCAAACCAAGGATTAAATG</td>
<td>Forward primer to sequence insert P</td>
</tr>
<tr>
<td>Seq_P_F_1667</td>
<td>CGACAAAGCTAGTATCTGCTATGGAGG</td>
<td>Forward primer to sequence insert P</td>
</tr>
<tr>
<td>Seq_P_F_1967</td>
<td>CAAAATACAACTCCATATCAATCAAGAAGATAGGTACAAAG</td>
<td>Forward primer to sequence insert P</td>
</tr>
<tr>
<td>Primer_PT_F</td>
<td>GAGCCAACCTCCCTTTACACCTCCTCAATGCTGTTAGAGGCGCTGTCTTTTGGCGGATG</td>
<td>Forward primer for amplification of T open plasmid for individual construct</td>
</tr>
<tr>
<td>Primer1_GeneT_F</td>
<td>ACCTCGACCTTTAAGGATACCTAACAAACATGAGCCTAAAGAACAAC</td>
<td>Forward primer for amplification of T gene for individual construct</td>
</tr>
<tr>
<td>Primer1_GeneT_R</td>
<td>TTATTTTGATTTATTGTGCTTTTCTAAGAATCTTCTTTCGCT</td>
<td>Reverse primer for amplification of T gene for individual construct</td>
</tr>
<tr>
<td>BridgeT_5’</td>
<td>GTAGACGGAAACAAACACGACAGATCCAAAGCAGCCTCAGCTTTAAGGATACCTAAACACATGAGC</td>
<td>Bridge used for LCR to ligate gene T and 5’ of T open plasmid</td>
</tr>
<tr>
<td>Name</td>
<td>Sequences</td>
<td>Notes</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>BridgeT_3’</td>
<td>GTTATAAGCAGAAAGTTCTTAGAAAAGACACAA TAAAATCAAAATAAAGAGCCACCTCCCTTACCA CCTCACTCAAG</td>
<td>Bridge used for LCR to ligate gene T and 3’ of T open plasmid</td>
</tr>
<tr>
<td>Seq_T_R_444</td>
<td>CGAAGTCTACTGATAGTTCCATCCTGC</td>
<td>Reverse primer to sequence insert T</td>
</tr>
<tr>
<td>Primer_Amplify_H_F</td>
<td>CTCGTTCCTGCACCTAAGAATACTCTACGGGT CACATACACCTC</td>
<td>Forward primer for amplification of H insert for final construct</td>
</tr>
<tr>
<td>Primer_Amplify_H_R</td>
<td>GAGACGAGAGACAGACGCTGAGAATGGATG GAGTAATG</td>
<td>Reverse primer for amplification of H insert for final construct</td>
</tr>
<tr>
<td>Primer_Amplify_P_F</td>
<td>CCTCGTCTCAACAAAGCAATCAACCCCATTT GACAATTAATCAGCCGC</td>
<td>Forward primer for amplification of P insert for final construct</td>
</tr>
<tr>
<td>Primer_Amplify_P_R</td>
<td>GACTTTGCCTTGTGCTTATACTATTGCTGGAG GAGTCAGTTATTTTTGTATATTATAG</td>
<td>Reverse primer for amplification of P insert for final construct</td>
</tr>
<tr>
<td>Primer_Amplify_T_F</td>
<td>GTTCTTATCATCTGCGAATCGGACCACCTTGAACAATTAATCATCGCCGC</td>
<td>Forward primer for amplification of T insert for final construct</td>
</tr>
<tr>
<td>Primer_Amplify_T_R</td>
<td>CTCTAACCAGGACTTGAGTGTTGAAAGGGA GTTGCTCTTATATTATTAGGTATTTGTC</td>
<td>Reverse primer for amplification of T insert for final construct</td>
</tr>
<tr>
<td>Primer_P2_R</td>
<td>GTATGTGACCGTGAGATATTCTTAGGTTGCA CGAAGCAGTAGGATGAGAAGATTTTCAG</td>
<td>Reverse primer for amplification of open final plasmid</td>
</tr>
<tr>
<td>Primer_P2_F</td>
<td>GGCTGTATTGTCGCGATGAGAGAGAAGATTTTCAG CCGTACAGATAATCAG</td>
<td>Forward primer for amplification of open final plasmid</td>
</tr>
<tr>
<td>BridgeH_5’P2</td>
<td>GCTGCCACCTAAGAATACTCTACGGTGACATA CACCTGACCTTTAAGGTACTTACAAACAT GAGC</td>
<td>Bridge used for LCR to ligate insert H and 5’ of final open plasmid</td>
</tr>
<tr>
<td>BridgeT_3’P2</td>
<td>CTTTTACACCTACTCAAGTCCGTAGAGG CTGTTGTTGCGGATGAGAAGATTTTTC</td>
<td>Bridge used for LCR to ligate insert T and 3’ of final open plasmid</td>
</tr>
<tr>
<td>BridgeHP_P2</td>
<td>CCAATTCTCAGGCTCTGCTCAGTGCTCAGTCTCCCTGACCTAACAAGGGAATCATGACACCC</td>
<td>Bridge used for LCR to ligate insert H and P</td>
</tr>
<tr>
<td>BridgePT_P2</td>
<td>GCCAGGAAATAGTAAAGACACACAGCAAGGTAAGTCCGCTGCTTATCATCTGCGGAATCGGCC</td>
<td>Bridge used for LCR to ligate insert P and insert T</td>
</tr>
<tr>
<td>Pkd3_ecoli_mg1655_F</td>
<td>TAATTTCAACATGAAAGTTTGTAGACGCCCCGC TTATTTCAATGAGGCAAAGTTTAGGTAGCTGAGAC TGCCTC</td>
<td>Forward primer for amplification of chloramphenicol cassette in pkd3 plasmid with overhangs homologous to genome</td>
</tr>
<tr>
<td>Pkd3_ecoli_mg1655_R</td>
<td>CTGACTAAAAAGGGCCGAGGCGGCTTTTAGA AAATGTCGAATCGAATGGAATGGAATTAGCCA TGGTCC</td>
<td>Reverse primer for amplification of chloramphenicol cassette in pkd3 plasmid with overhangs homologous to genome</td>
</tr>
<tr>
<td>Seq_pkd3_ecoli_F</td>
<td>TAATTTCAACATGAAAGTTTGTAGACGCCCCG</td>
<td>Forward primer to check the insertion of chloramphenicol gene</td>
</tr>
<tr>
<td>Seq_pkd3_ecoli_R</td>
<td>CTGACTAAAAAGGGCCGAGGCGGCTTTTAGA AAATGTCGAATCGAATGGAATGGAATTAGCCA TGGTCC</td>
<td>Reverse primer to check</td>
</tr>
<tr>
<td></td>
<td>Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Seq_gcv_F</strong></td>
<td>GAAACTGCGGACACCTTC</td>
<td>Forward primer to check location of chloramphenicol gene insertion</td>
</tr>
<tr>
<td><strong>Seq_gcv_R</strong></td>
<td>GGGTAAACAGATGCGCGTTAT</td>
<td>Reverse primer to check location of chloramphenicol gene insertion</td>
</tr>
</tbody>
</table>