Retrobiosynthesis of D-Glucaric Acid in a Metabolically Engineered Strain of *Escherichia coli*

by

TAE SEOK MOON

MS., Chemical Technology
Seoul National University, Seoul, Korea, 2000

BS., Chemical Technology
Seoul National University, Seoul, Korea, 1998

Submitted to the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
November 2009

Signature of Author .....

TAE SEOK MOON
Department of Chemical Engineering
Nov. 30, 2009

Certified by .....

Kristala Jones Prather
Assistant Professor of Chemical Engineering
Thesis Advisor

Accepted by .....

William M. Deen
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Retrobiosynthesis of D-Glucaric Acid in a Metabolically Engineered Strain of *Escherichia coli*

by

TAE SEOK MOON

Submitted to the Department of Chemical Engineering on Nov 30, 2009
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Synthetic biology is an evolving field that emphasizes synthesis more than observation which has been and is the scientific method for traditional biology. With powerful synthetic tools, synthetic biologists seek to reproduce natural behaviors (and eventually to create artificial life) from unnatural molecules or try to construct unnatural systems from interchangeable parts. Accompanied with this recent movement, metabolic engineers started to utilize these interchangeable parts (enzymes in this case) to create novel pathways. In addition, engineering biological parts including enzymes, promoters, and protein-protein interaction domains has led to productivity improvement. However, understanding behaviors of a synthetic pathway in an engineered chassis is still a daunting task, requiring global optimization.

As the first step to understand pathway design rules and behaviors of synthetic pathways, a synthetic pathway for the production of D-glucaric acid has been designed and constructed in *E. coli*. To this end, three disparate enzymes were recruited from three different organisms, and the system perturbed by this introduction of heterologous genes was analyzed. Limiting flux through the pathway is the second recombinant step, catalyzed by *myo*-inositol oxygenase (MIOX), whose activity is strongly influenced by the concentration of the *myo*-inositol substrate. To increase the effective concentration of *myo*-inositol, synthetic scaffold devices were built from protein-protein interaction domains to co-recruit all three pathway enzymes in a designable complex. This colocalization led to enhancement of MIOX activity with concomitant productivity improvement, achieving 2.7 g/L of D-glucaric acid production from 10 g/L of D-glucose input.

Secondly, retrobiosynthetic approach, a product-targeted design of biological pathways, has been proposed as an alternative strategy to exploit the diversity of enzyme-catalyzed reactions. The first step in a glucaric acid pathway designed retrosynthetically involves oxidation of the C-6 hydroxyl group on glucose, but no glucose oxidase in nature has been found to act on this hydroxyl group on glucose. To create glucose 6-

oxidase, a computational design and experimental selection was combined with the help of DNA synthesis technology. To this end, the sequence space of candidate mutations was computationally searched, the selected sequences were combinatorially assembled, and the created library was experimentally screened and characterized. Furthermore, the structure-activity relationship of the newly created glucose oxidases was elucidated, and the kinetic model mechanism for these mutants was proposed and analyzed.

Collectively, parts, devices, and chassis engineering were applied to a synthetic pathway for the production of D-glucaric acid, and this synthetic biology approach was proven to be effective for new pathway design and improvement.

Thesis supervisor: Kristala Jones Prather
Title: Assistant Professor of Chemical Engineering

Dedicated to Hannah and Hea Young

Acknowledgments

Without the support and help of professors, MIT staff members, collaborators, labmates, friends, and family, this thesis would be impossible. Listing their names would not be enough to express my thanks, but I do so as the first step to deliver my deepest gratitude to them. This work was also supported financially by the Office of Naval Research Young Investigator Program (Grant No. N000140510656) and the National Science Foundation Synthetic Biology Engineering Research (Grant No. EEC-0540879).

First of all, I thank Professor Kristala Prather, my advisor, for showing me one of the greatest advisors. She always believes in me, gives me maximum freedom to expand my horizon, guides me into a right direction, and encourages me whenever I am in distress. Her devotion to education led me to seek for my career path as an educator, and she will continue to be my role model for my future career.

I also thank Professor Sarah O'Connor and Professor Narendra Maheshri for their wonderful advice and guidance for my research as well as my future career. Their feedback inspires me to look at problems at various angles, resulting in alternative solutions.

I have also learned tremendously from my collaborators. Without their expertise and their collaboration, my learning process would have been limited. Among these wonderful collaborators, I would like to express my special thanks to Dr. Sang-Hwal

Yoon, Prof. John Dueber, Prof. David Nielsen, and Codon Devices. Dr. Sang-Hwal Yoon taught me how to construct genomic DNA library and provided one of the key enzymes for my pathways. Prof. John Dueber introduced me to the fascinating world of scaffold devices, inspired me to see my pathway engineering problems through a biochemist view, and led me to seek for my next career in a parts/devices engineering field. Prof. David Nielsen taught me how to approach kinetic mechanism problems numerically. Among Codon Devices people, I would like to thank Dr. Shaun Lippow, Dr. Daša Lipovšek, and Dr. Subhayu Basu for introducing me to computational enzyme engineering and DNA synthesis technology. Without their contribution, new engineered enzymes could not have been created.

I am also grateful to other people who helped my projects. I thank Xiazhen Li, Brad Chapman, Keith Robison, Kathy Galle, Paulo Gouveia, Christina Lee, Gina Prophete, Farhana Syed, Ashagari Habtegabriel, Roland Telson, and Ben Cerio for their support to the glucose oxidase project. I thank Dr. Koli Taghizadeh of Center for Environmental Health Sciences at MIT for supporting analysis by LC-MS. I thank Pooya Iranpour for his initial work on uronate dehydrogenase and Kevin Solomon and Eric Shiue for continuing glucaric acid projects that I proposed but did not finish. I thank Sang Taek Jung of the Georgiou lab for valuable advice and discussion on protein characterization. I also acknowledge Prof. Frances Arnold at Caltech, Prof. Seon-Won Kim at Gyeongsang National University, Prof. Frederick Ausubel at the Massachusetts General Hospital, and Weston Whitaker of the Dueber Lab for providing plasmids and strains. I am grateful to my undergraduate researchers Amanda, Joseph, Theresa, Mary-

Jane, Yunhong, and Veniece for working with me and giving me the wonderful chance for teaching and mentoring experience.

Without people who made my life at MIT enjoyable, my successful completion of the PhD program would not be possible. I would like to express my gratitude to all the Prather lab and Maheshri lab members, Gwen, all the members of Korean Graduate Students Association in Chemical Engineering at MIT, and my friends for their wonderful support and encouragement.

Finally, none of my work would have been possible without the love and support of my parents, brother, my parents in law, brother in law, my daughter, and my wife. All my and my wife's family members in Korea always believe in me and encourage me to go forward whenever I have a difficult time. I express my deepest thanks to them. Above all I can't express my thanks enough to Hannah and Hea Young. Hannah grew along with my projects and she is and will be my number one project. She gives all the inspiration and motivation for my work. Hea Young is my meaning of existence. Without her, my quest for a PhD degree would mean nothing. Of course, without her love and devotion, I could not have stepped forward. I am forever indebt for her endless love and support, and I dedicate this thesis to my lovely wife Hea Young.

Table of Contents

List of Figures

Fig. 4-3. *In vitro* activity of recombinant Ino1, MIOX, and Udh expressed in BL21(DE3) harboring the three genes. Cultures were grown at 30°C in LB medium supplemented with 10 g/L glucose and induced with 0.05 mM IPTG. MIOX activity is presented as net activity to account for background. Data are the average and standard deviation of three replicates. ▲ = Ino1; ■ = MIOX; ♦ = Udh.

Fig. 5-1. Calibration curves for D-glucuronate by the enzymatic method using Udh. The reaction mixture (1ml) contained 0.8 mM NAD^+, 100 mM Tris-Cl (pH 8.0), 20 µl sample containing D-glucuronate, and 0.06 U (1U = 1 µmol/min) Udh prepared as described previously (Yoon et al., 2009) and was incubated at room temperature for 30 min. The absorbance increase was measured at 340 nm using a Beckman DU-800 spectrophotometer (Beckman Coulter, Fullerton, CA), and the D-glucuronate concentration was calculated using the extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ for NADH.

Fig. 5-2. Enzymatic assay of MIOX using Udh. MIOX was expressed from pTrc-MIOX (Moon et al., 2009a) in *E. coli* DH10B grown in LB medium with 0.1 mM IPTG at 30°C for 24 hrs. Lysates were prepared by resuspending cell pellets from 50 mL culture in 8 mL sodium phosphate buffer (100 mM, pH 7) with 1 mg/mL lysozyme. EDTA-free protease inhibitor cocktail tablets (Roche Applied Science, Indianapolis, IN) were added to the resuspension buffer according to the manufacturer's instructions. Cell solutions were lysed by sonication and the resulting solutions were centrifuged at 14,000 rpm at 4°C for 15 min to remove insolubles. The reaction buffer consisted of 77 mM sodium phosphate buffer (pH 7), 2 mM L-cysteine, 1 mM $FeSO_4$, and 60 mM *myo*-inositol. Samples were pre-incubated without substrate for 5 minutes at 30°C to activate the MIOX enzyme. Reactions were incubated for 1 hr at 30°C, and the D-glucuronate produced was quantified using Udh and 4 mM NAD^+.

Fig. 6-1. Effect of induction levels (IPTG for pathway enzyme expression; aTc for scaffold expression) on the production of D-glucaric acid using the scaffold $GBD_1SH3_4PDZ_4$ (JT9). For the 0.025 mM IPTG/215 nM aTc condition, no experiment was performed. Data are the averages of three replicates and the standard deviations are not higher than 23% of the averages.

Fig. 6-2. Effect of various scaffold architectures (constructs JT1 to JT9) on the production of D-glucaric acid at 0.05 mM IPTG and three different aTc concentrations as indicated. Δ = 54 nM aTc; ■ = 108 nM aTc; ◊ = 215 nM aTc.

Fig. 6-3. Effect of the number of Ino1-recruiting SH3 domains on D-glucaric acid titer with 0.05 mM IPTG in $GBD_1SH3_cPDZ_2$, where c is 0, 1, 2, 3, 4, 6, and 8. Experiments were performed at four different aTc concentrations as indicated. The scaffolds with degenerate coding sequences were used for c = 2, 3, 4, 6, and 8. Lines are drawn to indicate trends. ● = 27 nM aTc; Δ = 54 nM aTc; ■ = 81 nM aTc; ◊ = 108 nM aTc.

Fig. 6-4. Effect of IPTG concentration on D-glucaric acid titer at 54 nM aTc in $GBD_1SH3_cPDZ_2$, where c is 3, 6, and 8. Experiments were performed at four different

inducer concentrations: from left to right, 0.05, 0.075, 0.1, and 0.2 mM IPTG. Data are the averages and standard deviations of two replicates.

Fig. 6-5. Correlation between D-glucaric acid titer and MIOX activity across various scaffold architectures. Data are from JT1-10 at 0.05 mM IPTG, and 54 or 215 nM aTc; and JTK1-4, JT5, and JT10 at 0.05 mM IPTG and 108, 81, 54, or 27 nM aTc. Data points represent the average values of MIOX activity and D-glucaric acid titer. The standard deviations are not higher than 57% (MIOX activity) and 37% (D-glucaric acid titer) of the averages.

Fig. 7-1. A structural model of wild type galactose 6-oxidase with docked galactose.

Fig. 8-1. A proposed model mechanism (Bates and Frieden, 1973) for GlcOx.

Fig. 8-2. Comparison of the normalized change in product concentration as a function of dimensionless time as obtained experimentally (A.) for Des3-10 (dashed), Des3-2 (dot-dashed), Des3-7 (solid), Des3-6 (dot-dot-dash) and M-RQW (dotted), and as predicted with the presented model (B.) for *k2f* values of 0.005 (dashed), 0.05 (dot-dashed), 0.5 (dotted), and 5 s^{-1} (solid).

Fig. 8-3. The rate of product concentration (*P*) with respect to time predicted as a function of dimensionless time for *k2f* values of 0.005 (dashed), 0.05 (dot-dashed), 0.5 (dotted), and 5 s^{-1} (solid).

Fig. 8-4. Approximation of the lag time, Θ_{lag}, associated with mutant Des3-10. Inset tables summarize the experimental lag times associated with each of the studied mutants, as well as the lag times associated with the model predictions as approximated via the same procedure.

Fig. 8-5. Normalized sensitivity coefficients predicted as a function of dimensionless time for *k1f* (dashed line), *k2f* (dot-dash line), *k2r* (dotted line), *k3f* (open circles), *k3r* (solid circles), *k4f* (open triangles), and *k5f* (open squares).

List of Tables

Chapter 1

Introduction

1.1 Motivation

Synthetic organic chemists approach a synthesis scheme for a compound, based on the structure of the target compound. This retrosynthetic design is a powerful tool, giving organic chemists a lot of flexibility: a wide range of target products and a large set of possible synthetic steps. Despite this flexibility, chemical synthesis is not efficient when the target compounds are complex and have many functional groups. To synthesize these challenging compounds, biochemists and metabolic engineers have recruited enzymes and have shown the potential of these biocatalysts. However, most biosynthetic approaches have been made by utilizing "forward biosynthesis," a substrate-focused design of biological pathway, limiting the range of the target products. My research focuses on combining these two approaches: retrosynthesis and biocatalysis. Using enzyme databases and synthetic biology tools, including parts/devices engineering and chassis engineering, I aim to create and improve novel biosynthetic pathways for highly functionalized compounds, not found in *E.coli*. Specifically, I propose simple and economical pathways for D-glucaric acid, a highly functionalized chiral compound, as a retrosynthetic model (Fig. 1-1).

1.2 Retrobiosynthesis

Metabolic engineering, encompassing application of recombinant DNA technology, has shown the potential to optimize cellular functions for many purposes: recombinant protein production, pathway engineering for productivity enhancement, and new pathway design for new product generation. Defined as a sequence of conversions

that is not found in host species, a novel pathway has been designed by utilizing "forward biosynthesis," a substrate-focused approach for producing the target product, limiting the range of the target products. In this approach, substrates and products as well as enzymes are defined: if the enzyme activity is high and balanced in the foreign host, the conversion is almost guaranteed. Thus, this is an optimization problem rather than a design one. An example of a novel pathway construction by this "forward biosynthesis" used a combination of the existing pathways (Nakamura and Whited, 2003). To develop a recombinant *E. coli* capable of producing 1,3-propanediol with high titers, the authors utilized a sequence of three conversions, for which two genes from *Saccharomyces cerevisiae* and *Klebsiella pneumonia* were transformed. As another example of a novel pathway design, consider the novel polyketide production. Polyketide synthases, multifunctional enzymes catalyzing multiple steps, were introduced and modified in *E. coli* to show that novel polyketide derivatives can be synthesized in a foreign host through modification of the existing multi-subunit enzyme, analogous to that of individual enzymes in a multi-step pathway (Pfeifer et al., 2001). Even though those previous works have expanded the ways for novel pathway construction, few approached a synthesis scheme, based on the structure of the target compound.

Retrobiosynthetic approach, a product-targeted design of biological pathways, has been proposed as an alternative strategy to exploit the diversity of enzyme-catalyzed reactions (Prather and Martin, 2008). This retrobiosynthetic approach has many challenges: 1) only limited sets of pathway design tools are available (Hatzimanikatis et al., 2005; Hou et al., 2003; Prather and Martin, 2008); 2) the availability of enzymes for

purely retrosynthetic pathways is even more limited, often requiring creation of new enzymes. Nevertheless, "Synthetic Biology" (discussed in detail in Chapter 2) will provide with a vast array of biological parts including new enzymes, facilitating this retrobiosynthetic research.

Natural evolution has diversified biological catalysts and chemical reactions in the cells. Despite this diversity of nature, only a small portion of this enzyme library is available for human purposes. Directed evolution and DNA shuffling (Arnold et al., 1999; Arnold and Moore, 1997; Arnold et al., 2001; Stemmer, 1994a; Stemmer, 1994b) have revolutionized the ways to expand this limited enzyme pool, but retrosynthetic pathway design has yet to overcome one challenge: enzyme is specific, and specificity change is tremendously difficult. A few examples of directed evolution approaches are worth mentioning here. First, Sun *et al.* have generated galactose oxidase mutants that also show a low activity on glucose (Sun et al., 2002). It is interesting to note that no glucose 6-oxidase exists in nature and glycolysis pathway has been the preferred choice of glucose oxidation for energy and carbon supply. The fact that the evolved galactose oxidases still have higher activity on galactose than glucose (Sun et al., 2002) would not be surprising when the long history of glucose catabolism by glycolysis is considered. Second, the high C-H bond dissociation energy barrier which should be overcome for ethane-to-ethanol conversion was shown not to be an impossible one when engineered cytochrome P450 is used (Meinhold et al., 2005). Again, the activity level was too low for practical purposes. Last, a novel metabolic route was constructed to alleviate the competition for phosphoenolpyruvate (PEP) between sugar transport system (PTS) and a

key enzyme (DAHP synthase) in the shikimate pathway (Ran and Frost, 2007). Instead of PEP, D-erythrose 4-phosphate and pyruvate were condensed by an evolved enzyme, which originally has a very low activity for those substrates, generating DAHP. Even though this evolution approach enhanced the condensation activity and constructed the new route, the activity was not enough to achieve as high production of shikimate as an optimized DAHP synthase-based pathway.

Biological parts can be gathered for retrosynthetic pathway design purposes in different ways. Metagenomics is the study of genetic material from specific environments, such as soil or marine samples and has expanded available enzyme library and pathways (Daniel, 2004; Handelsman, 2004; Rondon et al., 2000). Inspired by metagenomics, synthetic biologists have recently proposed "synthetic metagenomics" which involves extensive sequence database search and DNA synthesis technology (Bayer et al., 2009). This bioinformatics-based approach with help of DNA synthesis technology has proved that methyl halides can be produced from biomass using a synthetic pathway. More computation-rigorous and structure-based enzyme design has shown its potential to redesign enzyme activity even though the catalytically relevant interactions at the active site have yet to be accurately represented for successful enzyme design (Chen et al., 2009). Quantum mechanical transition state calculation and recapitulation of the active sites have led to *de novo* computational design of enzymes (Jiang et al., 2008; Rothlisberger et al., 2008; Zanghellini et al., 2006), promising that *in silico* development of enzyme activities will expand the available enzyme library for retrobiosynthesis. Of course, various computational algorithms (Voigt et al., 2002;

Zanghellini et al., 2006) have been and will be facilitating the search for new enzymes even though computational algorithm development is beyond the scope of this thesis.

1.3 D-Glucaric Acid

D-Glucaric acid is found in fruits, vegetables, and mammals. D-Glucaric acid has been studied for therapeutic purposes including cholesterol reduction (Walaszek et al., 1996) and cancer chemotherapy (Singh and Gupta, 2003; Singh and Gupta, 2007) and it was also identified as a "top value-added chemical from biomass" . It is mainly used as a starting material for hydroxylated nylons (Werpy and Petersen, 2004) and currently produced by chemical oxidation of glucose, a nonselective and expensive process using nitric acid as the oxidant. There is a known route for the production of D-glucaric acid from D-glucose in mammals. However, this is a lengthy pathway, consisting of more than ten conversion steps (Fig. 1-2).

1.4 Thesis Objectives

The ultimate objective of my thesis is to recruit enzymes and to assemble these enzyme-catalyzed reactions in a microbial host just as synthetic organic chemists have employed catalysts and reagents for retrosynthesis of target compounds. This case study for producing glucaric acid will initiate the construction of enzyme libraries for retrobiosynthesis and provide guidelines for novel pathway design.

Among the five pathways (Fig. 1-1), construction of PW1 and PW2 is discussed in this thesis. Using enzyme databases, such as BRENDA (http://www.brenda-enzymes.org/), KEGG (http://www.genome.jp/kegg/kegg2.html), UniProt (http://ca.expasy.org/sprot/), and Metacyc (http://www.metacyc.com/), each pathway was investigated to find suitable enzymes. For the enzyme selection process, the following criteria were used: (1) the class of enzymes, based on EC numbers, (2) availability of gene information, (3) similarity in substrate structure and promiscuity of enzymes, (4) natural host, and (5) other considerations including reversibility of reactions, cofactors, inhibitors, and posttranslational modification. PW1 was selected because all the steps are found in nature. PW2 is the shortest and simplest one among the remaining "unnatural" pathways, and the intermediates have no known metabolic route in *E. coli*, allowing construction of this pathway without any interference. More importantly, PW2 is likely to mimic the chemical oxidation process most closely among the five pathways. PW5 will be least considered for construction because it has the same conversion step as PW4 (glucose → glucono-1,5-lactone) and has more unnatural intermediates (thought to be unstable) than PW4. Even though PW3 is the longest one, one interesting feature is worth mentioning. Noting that EC 1.1.99.28 can initiate PW3 and PW4 simultaneously (D-glucose + D-fructose → D-glucono-1,5-lactone + D-sorbitol) and that there is no inter-pathway route between them, we will be allowed to compare the two pathways in one host if the remaining genes for PW3 and PW4 enzymes are assembled in the host.

1.5 Specific Goals

To purse the above mentioned general goals, the following specific aims were achieved:

1) proposing feasible biosynthetic pathways to produce glucaric acid;

2) obtaining suitable genes for each conversion step from various organisms by using cloning techniques such as PCR amplification and codon-optimized synthesis;

3) developing the screening method of cloned enzyme activities and expanding the enzyme specificities by enzyme engineering;

4) characterizing and analyzing the intermediates and the target product, glucaric acid;

5) maximizing glucaric acid production by analyzing the metabolic flux and by controlling enzyme expression and activity;

1.6 Thesis Organization

This thesis consists of two main themes: parts/devices/chassis engineering and its applications to pathway engineering. Alternatively, two main efforts compose this thesis: construction and improvement of a glucaric acid pathway (PW1 in Fig. 1-1); design and exploration of alternative retrosynthetic pathways (PW2 ~ 5 in Fig. 1-1). According to the first classification, Chapter 3 and 6 ~ 8 describe efforts to recruit, create, and

characterize parts/devices/chassis, while its applications to pathway engineering are demonstrated in Chapter 4 ~ 6. Based on the second categorization, Chapter 3 ~ 6 deal with PW1 construction and its optimization, and Chapter 7 and 8 cover projects toward construction of the remaining retrosynthetic pathways PW2 ~ 5. Even though each chapter has been formatted as an independent research paper, all the chapters are connected toward one main goal: construction and improvement of glucaric acid synthetic pathways. Two extra chapters Chapter 2 and 9 serve as overview chapter of Synthetic Biology and conclusion chapter, respectively.

Chapter 2 overviews Synthetic Biology, the evolving field, which combines classical engineering principles with biology. In this chapter, the potential and limitation of this field is discussed. Chapter 3 deals with cloning and characterization of uronate dehydrogenase, which is a critical part to complete the synthetic glucaric acid pathway (PW1). This chapter describes how this long-waiting debut was made possible and how the identity of gene was confirmed. In Chapter 4, new strategy to make value-added compounds is introduced and a synthetic pathway to produce glucaric acid in *E. coli* makes its successful debut. New assay methods to analyze the synthetic pathway have been developed and their procedures and validity are discussed in Chapter 5. As a great example of synergy between synthetic biology and metabolic engineering, Chapter 6 demonstrates that synthetic parts/devices are so robust and reusable that they can improve biosynthetic pathways. Chapter 7 is dedicated to part engineering: computational design and screening of new glucose 6-oxidases which have never been found in nature, followed by Chapter 8 which proposes and analyzes a kinetic model as the first step to

better characterize the newly created glucose 6-oxidases. Finally, Chapter 9 offers conclusions and future directions.

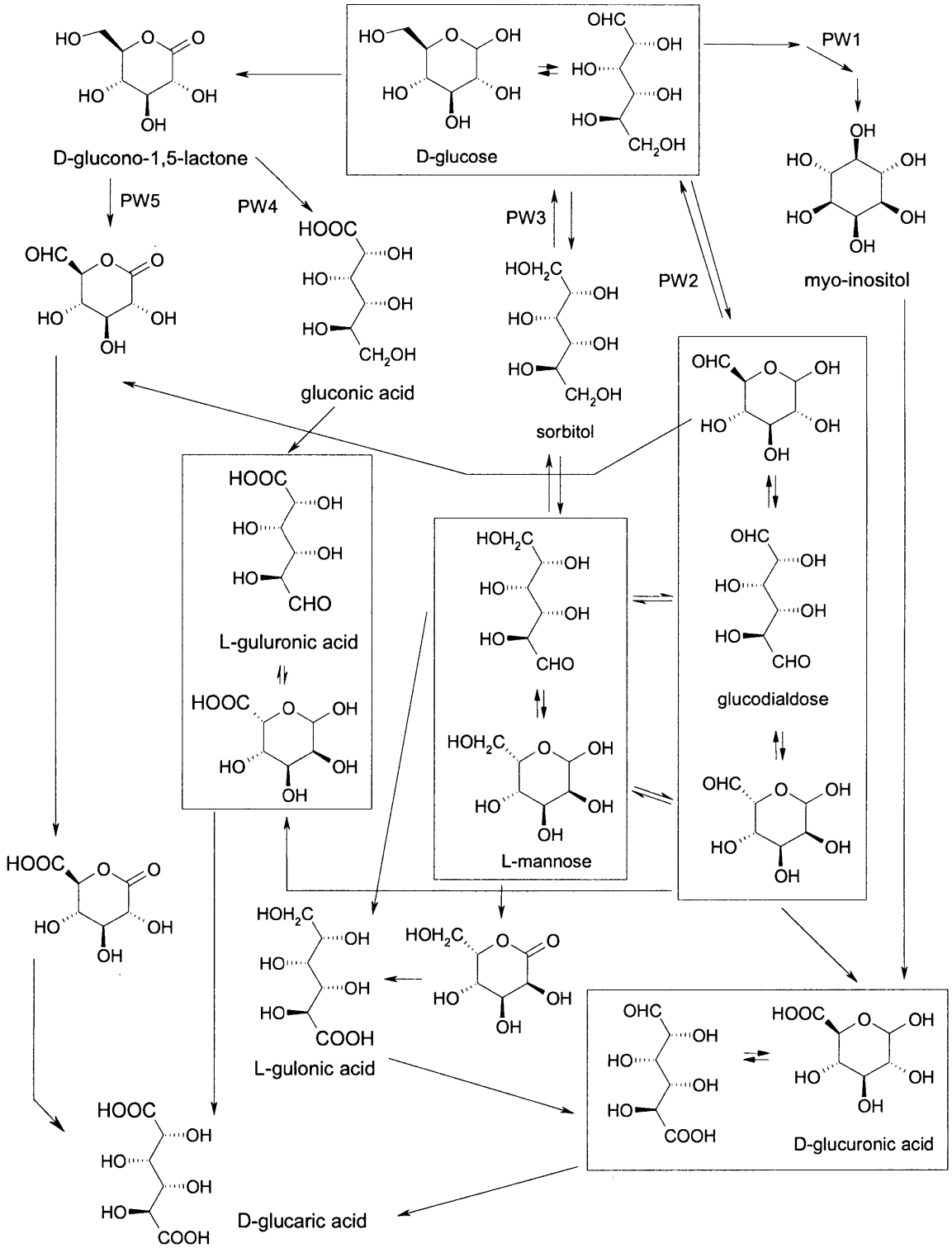


Fig. 1-1. Proposed biosynthetic pathways for producing D-glucaric acid.

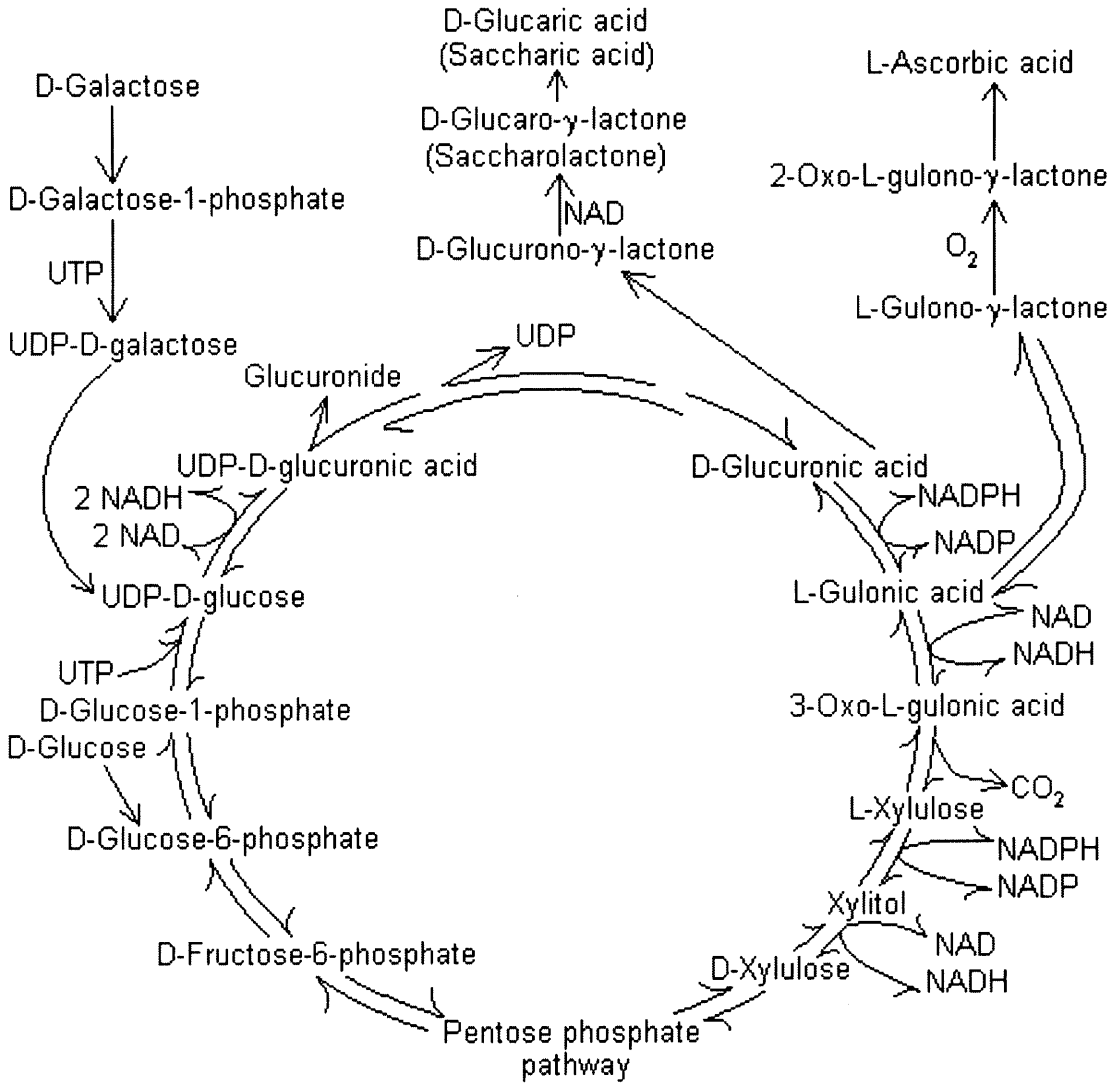


Fig. 1-2. D-glucuronic acid pathway in mammals. Adapted from *Encycl. of Chem. Tech.*

Chapter 2

Synthetic Biology Overview

2.1 Observation and Synthesis for Biology and Bioengineering

Biology has been a scientific field of observations that emphasizes unfiltered, value-neutral experiments. However, the outcome of biological science is not dependent on how strictly scientists meet this largely fictional idea, but rather how reasonably they deal with the "filters" that accompany naturally with human thought (Benner and Sismour, 2005). This seemingly dilemmatic situation could be reconciled (partially or sometimes entirely) by synthesizing the target systems. For example, synthetic organic chemists have expanded our understanding of chemical bonding that was not possible to be achieved by simply observing the structure of natural compounds (Benner and Sismour, 2005). With synthetic power, synthetic biologists are constructing existing or new biological parts that might be assembled to create artificial life or biological systems that mimic natural systems. Several examples of accomplishments by synthetic biologists include the construction of non-natural enzymes by computational enzyme design (Rothlisberger et al., 2008), viral particles from scratch (Cello et al., 2002), oscillating biological system consisting of three repressors (Elowitz and Leibler, 2000), edge-detecting biological program by utilizing sensors (Tabor et al., 2009), and synthetic biological pathway where enzymes were recruited by synthetic metagenomics approach (Bayer et al., 2009).

2.2 Synthetic Biology – Goal and Challenge

Synthetic biology is an evolving field involving the creation of new biological components and systems, such as enzymes, signaling molecules, and metabolic pathways,

using their corresponding building blocks, such as amino acids, protein interaction domains, and enzymes respectively (Benner and Sismour, 2005; Keasling, 2008; Leonard et al., 2008). Synthetic biologists seek to design and characterize interchangeable parts to build synthetic devices and systems that help to understand and to harness biological systems. This field has shown its potential to rewire signaling pathways (Park et al., 2003) and to develop microbes that could synthesize bulk chemicals (Nakamura and Whited, 2003), fuels (Atsumi et al., 2008), and drugs (Martin et al., 2003).

The ultimate goal of synthetic biology is creation of systems that behave just as electronic circuits do in a programmable and robust way. However, to this end, the complex cellular environment should be understood. Firstly, unlike electronic parts, DNA sequences tend to mutate and recombine. In an evolutionary point of view and for the design purpose, domain recombination and mutation could be advantageous, expanding choice of biological components. However, mutability of biological systems should be minimized for robust and predictable performance of synthetic devices. Secondly, truly interchangeable parts, like electronic parts, are yet to exist. DNA can be manipulated or synthesized to combine various promoters, RBSs, ORFs, and terminators, but the expression level cannot be predicted accurately by current biological knowledge. In amino acid level, this prediction becomes even more difficult: to my knowledge, only few synthetic enzyme-like proteins have been created from purely computational design (Rothlisberger et al., 2008). Lastly, synthetic biologists have yet to be able to predict how well synthetic circuits and pathways would perform in different chassis. Without

understanding the interactions of these synthetic systems with the other components in chassis, reusable and robust systems would be impossible to construct.

2.3 Synthetic Biology Framework

Synthetic biologists focus on modularity of biological components and try to re-construct biology to support the engineering of biology. To this end, biological systems are divided into a hierarchy of abstractions: parts, devices, and systems (Endy, 2005). Parts are defined as basic biological functions (e.g. promoter, ribosome binding site, open reading frame, and terminator). Devices are combination of parts to achieve a human-defined function (e.g. combination of promoter, ribosome binding site, open reading frame, and terminator for expression of a gene). Systems consist of devices to execute a task, such as a biosynthetic pathway that can produce compounds. These systems require a chassis, a host cell, which provides resources for proper operation of these systems.

Modularity in biology enables synthetic biologists to assemble parts into devices and systems. For example, synthetic biologists have recombined biological components such as enzymes and protein interaction domains in order to create novel pathways (Atsumi et al., 2008; Martin et al., 2003; Moon et al., 2009a; Nakamura and Whited, 2003; Niu et al., 2003; Prather and Martin, 2008) and scaffold-mediated systems (Dueber et al., 2009; Park et al., 2003) respectively. Transcriptional repressors were combined to make an oscillator and a toggle switch, taking advantage of modular characteristics of repressors and promoters (Elowitz and Leibler, 2000; Gardner et al., 2000). Modular

structure of proteins has been exploited to construct a new chimaeric sensor by fusing the photoreceptor from Cph1 and the histidine kinase and response-regulator from EnvZ-OmpR (Levskaya et al., 2005). Despite these successful examples, synthetic biology needs to resolve two issues: the robustness of the interchangeable parts and their composability. Without robust parts and composable systems, the synthetic biology framework would be a failed engineering trial.

2.4 Synthetic Biology for Microbial Chemical Factories

Synthetic biology intersects with metabolic engineering in areas such as the construction of novel pathways for product generation, in which enzymes are considered as interchangeable parts, and the improvement of those pathways for increased productivity. Until now, metabolic engineering has focused more on global optimization of metabolic flux to improve existing natural pathways in heterologous hosts than new pathway design. This optimization problem should be also addressed once new pathways are designed and constructed because perturbations such as the introduction of new pathways into a host often lead to problems, limiting the productivity of the synthetic pathways. The problems include competition for endogenous metabolites, toxicity of newly generated intermediates, and side-reactions by introduced promiscuous enzymes, and these problems are common for both natural and synthetic pathways that are introduced into heterologous hosts. Thus, traditional metabolic engineering approaches could solve these problems.

In contrast, new pathway design, rather than reconstruction of natural pathways in heterologous hosts, is a new engineering problem. Currently, not many examples of new pathway construction exist because limited enzyme availability prevents pathway designers from expanding their imagination. However, recent advance in synthetic biology and enzyme engineering have opened a new era toward *de novo* pathway design. The first example of new pathway design is combination of natural activities of enzymes. In this design approach, individual step is found in nature, but the combined pathway is new. One great example of this strategy is biosynthesis of 1,3-propanediol (Nakamura and Whited, 2003). In this pathway, genes from *Saccharomyces cerevisiae* and *Klebsiella pneumonia* are introduced into *E. coli*, where a glycolysis intermediate (DHAP) is converted into 3-hydroxypropionaldehyde by the two step conversions, followed by endogenous enzyme-catalyzed conversion to 1,3-propanediol. Similar examples include synthesis of artemisinin precursor amorphadiene (Martin et al., 2003; Ro et al., 2006), polyketide derivatives (Pfeifer et al., 2001), and methyl halides (Bayer et al., 2009). The last example is worth noting in that the authors used total gene synthesis and activity screening to find natural versions of a key enzyme from putative methyl halide transferase sequences which are identified from NCBI sequence database (Bayer et al., 2009). The second way of new pathway design is harnessing promiscuous enzymes to generate new routes. This approach is an excellent expansion of versatile abilities of promiscuous enzymes and may be a starting point for generalized biocatalysts analogous to chemical catalysts. A few success examples are production of 1,2,4-butanetriol (Niu et al., 2003) and branched-chain higher alcohols (Atsumi et al., 2008). The third approach is using engineered enzymes to create new pathways. This strategy will be facilitated by

advance in *de novo* enzyme design (Rothlisberger et al., 2008), directed evolution (Arnold and Moore, 1997), and DNA shuffling technology (Stemmer, 1994b). A couple of examples are biosynthesis of triacetic acid lactone from glucose (Zha et al., 2004) and shikimic acid from pyruvate and D-erythrose 4-phosphate (Ran and Frost, 2007). Of course, these three categories are not exclusive, but combination of the three design approaches will broaden the horizon for pathway designers. In addition, biosynthetic pathway design will be facilitated by using computational tools for pathway design, such as the University of Minnesota Biocatalysis and Biodegradation Database (UM-BBD) (Ellis et al., 2006), Biochemical Network Integrated Computational Explorer (BNICE) algorithm (Hatzimanikatis et al., 2005), and Retro-Biosynthesis Tool (ReBiT, http://www.retro-biosynthesis.com) (Prather and Martin, 2008).

2.5 Perspectives

Scientists saw the letters of DNA for the first time, and DNA have been described in a reductionist sense as a duplex assembled from four nucleotide building blocks and their base-pairing. Scientific innovation then allowed biological researchers to read the letters, words, and sentences embedded in DNA strands. Now, synthetic biologists try to write their own masterpieces, and DNA synthesis technology accelerates publishing of the masterpieces. Synthetic biology is a transformative view of biology from "observation approach" to "synthesis approach." Until now, however, the focus of synthetic biology has been the creation and characterization of genetic parts/devices and small modules which consist of these devices. These efforts have led to increase in the

availability of these biological components and better understanding of biological systems and their design principles. However, to design and construct biological systems in a more predictable and robust way, synthetic biologists need to create new design and computational tools that consider uncertainty in biology such as mutability and composability. In addition, synthetic biology should move into higher levels of engineering (Purnick and Weiss, 2009) that allows researchers to assemble devices and modules into larger but application-oriented systems rather than "toy" systems. Innovative synthetic approaches will pave a wide range of applications to transform our views on production of sustainable energy and renewable chemicals, environmental problems, and human disease treatments.

Chapter 3

Part Recruitment – Cloning and Characterization of Uronate Dehydrogenase

S-H Yoon, TS Moon, P Iranpour, A Lanza and KJ Prather. **Cloning and characterization of uronate dehydrogenases from two Pseudomonads and *Agrobacterium tumefaciens* str. C58**, *J. Bacteriol.* 191, 1565-1573 (2009)

Abstract

Uronate dehydrogenase has been cloned from *Pseudomonas syringae* pv. tomato DC3000, *Pseudomonas putida* KT2440, and *Agrobacterium tumefaciens* str. C58. The genes were identified by using a novel complementation assay employing an *Escherichia coli* mutant incapable of consuming glucuronate as the sole carbon source but capable of growth on glucarate. A shotgun library of *P. syringae* was screened in the mutant *E. coli* by growing transformed cells on minimal medium containing glucuronic acid. Colonies that survived were evaluated for uronate dehydrogenase, which is capable of converting glucuronic acid to glucaric acid. In this manner, a 0.8 Kb open reading frame was identified and subsequently verified as *udh*. Homologous enzymes were identified in *P. putida* and *A. tumefaciens* based on a similarity search of the sequenced genomes. Recombinant proteins from each of the three organisms expressed in *E. coli* were purified and characterized. For all three enzymes, the turnover number, k_{cat}, was higher for glucuronate as a substrate than for galacturonate; however, the Michaelis constant, K_m, was lower for galacturonate. The *A. tumefaciens* enzyme was found to have the highest rate constant ($k_{cat} = 1.9 \times 10^2\ s^{-1}$ on glucuronate), which was more than 2-fold higher than both of the Pseudomonad enzymes.

Keywords: uronate dehydrogenase, glucuronate, glucarate, *Pseudomonas syringae, Pseudomonas putida, Agrobacterium tumefaciens*

3.1 Introduction

Pathway design and construction is often hampered by limited availability of enzymes which perform the desirable sequence of reactions. The first step to recruit these enzymes is searching enzyme database, such as BRENDA (http://www.brenda-enzymes.org/), KEGG (http://www.genome.jp/kegg/kegg2.html), UniProt (http://ca.expasy.org/sprot/), and Metacyc (http://www.metacyc.com/), and publications demonstrating the suitable catalysis. If the target genes have already been cloned and identified, the recruitment only involves obtaining the source strains or DNA to subclone into the expression vector of interest. However, the gene information is often unavailable, and the only information is circumstantial evidence that the target reaction may exist in some strains. Uronate dehydrogenase, the last step enzyme for PW1 and PW2 (Fig. 1-1), is known to be the key enzyme of one of the aldohexuronate catabolism pathways, but the gene information has not been available.

Aldohexuronate catabolism in bacteria is reported to involve two different pathways, one initiating with an isomerization step and the other with an oxidation step. In the isomerization pathway, aldohexuronate (glucuronate and galacturonate) is isomerized to ketohexuronate by uronate isomerase and ultimately degraded to pyruvate and 3-phosphoglyceraldehyde. The isomerization pathway has been previously reported to occur in bacteria including *Escherichia coli* (Cynkin and Ashwell, 1960), *Erwinia carotovora* (Kilgore and Starr, 1959a) and *Erwinia hrysanthemi* (Hugouvieux-Cotte-Pattat and Robert-Baudouy, 1987), *Areobacter aerogenes* (Farmer and Eagon, 1969; Mc and Novelli, 1958), and *Serratia marcescens* (Payne and Mc, 1958). In the oxidation

pathway, aldohexuronate is oxidized to aldohexarate by uronate dehydrogenase and further catabolized to pyruvate (Ashwell et al., 1958; Chang and Feingold, 1970; Cynkin and Ashwell, 1960; Farmer and Eagon, 1969; Kilgore and Starr, 1959a; Kilgore and Starr, 1959b; Mc et al., 1959). Uronate dehydrogenase (Udh), the key enzyme of this pathway, has been investigated in two plant pathogen bacteria, *Pseudomonas syringae* and *Agrobacterium tumefaciens*. To date, only limited studies pertaining to the properties of Udh have been reported in the literature (Bateman et al., 1970; Chang and Feingold, 1969; Wagner and Hollmann, 1976; Zajic, 1959), and no sequence has yet been identified. Udh is classified as an NAD-linked oxidoreductase (EC 1.1.1.203), with a total molecular weight of about 60,000. It is a homo-dimer composed of two subunits of about 30,000 molecular weight each (Wagner and Hollmann, 1976). Udh is a thermally unstable, reversible enzyme, with an optimum pH of about 8.0 (Bateman et al., 1970; Chang and Feingold, 1969; Wagner and Hollmann, 1976).

In *E. coli* MG1655 with the isomerization pathway for aldohexuronate catabolism, glucuronate is transported by an aldohexuronate transporter encoded by *exuT* and converted to fructuronate by uronate isomerase, encoded by *uxaC* (Mata-Gilsinger and Ritzenthaler, 1983; Portalier et al., 1974). Fructuronate is transferred to the Entner-Doudoroff pathway to be utilized as an energy source via 2-keto-3-deoxy-6-phospho-gluconate (Cynkin and Ashwell, 1960; Neidhardt and Curtiss, 1996; Robert-Baudouy et al., 1982; Robert-Baudouy and Stoeber, 1973). Therefore, *E. coli* MG1655 with an *uxaC* deletion can not use glucuronate as a carbon source. In this same strain, glucarate is converted to 5-keto-4-deoxy-D-glucarate by D-glucarate dehydratase, encoded by *gudD*,

and then transferred to glycolysis via pyruvate or 2-phosphoglycerate (Neidhardt and Curtiss, 1996; Roberton et al., 1980). Recently, a number of bacterial genome sequences have been published, including those of the Udh containing *P. syringae* pv. tomato DC3000 and *A. tumefaciens* str. C58 (Buell et al., 2003; Goodner et al., 2001). A shotgun library of *P. syringae* was constructed to identify the gene encoding Udh. Screening for Udh was conducted in *E. coli* MG1655 *ΔuxaC*. Since uronate dehydrogenase converts glucuronate to glucarate, *E. coli ΔuxaC* strains harboring the shotgun library of *P. syringae* that can grow in a minimal medium containing glucuronate as a sole carbon source may carry the gene encoding Udh (Fig. 3-1). Once an initial Udh is identified from *P. syringae*, a BLAST homology search may lead to the identification of Udhs from other bacteria.

Various xenobiotic substances, such as pollutants and drugs, can induce the liver microsomal enzyme system (Jung et al., 1981; Mocarelli et al., 1988). Urinary excretion of D-glucaric acid has proven to be a suitable index of microsomal enzyme activity of the liver (Addyman et al., 1996; Jung et al., 1981; Mocarelli et al., 1988). Clinical interests in urinary D-glucaric acid have led to the development of its quantification methods. Enzymatic assays have been the preferred methods for the determination of the urinary level of D-glucaric acid in clinical labs (Jung et al., 1981; Mocarelli et al., 1988). Two different principles have been applied to these method development efforts (Mocarelli et al., 1988). The first principle is based on the following observations: 1) D-glucaric acid is in equilibrium with its lactone forms and can be transformed into 1,4-glucarolactone when boiled at an acidic pH; 2) 1,4-glucarolactone competitively inhibits β-

glucuronidase; thus, D-glucaric acid concentration can be measured indirectly by determining the extent of inhibition of β-glucuronidase on substrates, such as phenolphthalein-β-D-glucuronide and 4-nitrophenyl-β-D-glucuronide (Colombi et al., 1983; MacNeil et al., 1986; Marsh, 1963; Simmons et al., 1974; Steinberg et al., 1985). The second method utilizes a bacterial extract containing glucarate dehydratase and ketodeoxyglucarate aldolase. These two enzymes convert quantitatively glucarate into pyruvate, which can be measured by using lactate dehydrogenase (Marsh, 1985). These assay methods could be applied to the quantification of D-glucaric acid from microbial production, but these methods have disadvantages for this purpose: 1) enzymes need to be prepared or acquired; 2) there could be other potential inhibitors in bacterial culture broth that complicate the first assay method; 3) for the second method, pyruvate generated by the production host should be differentiated from one converted from glucarate.

Determination of D-glucaric acid concentration by chromatography has also been reported (Ishidate et al., 1965; Laakso et al., 1983; Poon et al., 1993). This approach is usually time consuming and requires the appropriate instrumentation, so it has been less adopted in clinical labs. In contrast, chromatographic methods have generally been the choice for metabolite assays because different metabolites in a sample can be separated and quantified at the same time. However, there are two main issues that should be addressed before choosing the chromatographic method for the determination of D-glucaric acid produced from microbial fermentation. Firstly, D-glucaric acid is known to spontaneously form lactones (Horton and Walaszek, 1982) especially when the solution

pH is low. This spontaneous lactonization may complicate the quantification of D-glucaric acid. Because lactones are hydrolyzed at high pH, this complication can be resolved by base treatment. Secondly, D-glucuronic acid, the common intermediate in PW1 ~ 3 (Fig. 1-1), was reported to have almost the same retention time as that of D-glucaric acid (Poon et al., 1993). However, these two compounds can be separated by using boronic acid affinity gel since boric acid binds specifically with coplanar adjacent cis hydroxyl groups (Fig. 3-2) which glucaric acid contains (but glucuronic acid does not). For my thesis projects, chromatography-based assay methods are selected and developed.

3.2 Materials and Methods

3.2.1 Bacterial strains, plasmids, and growth conditions

Strains, plasmids, and primer sequences used in this study are indicated in Table 3-1. Media and chemical reagents were purchased from Sigma (St. Louis, MO, USA) or BD Biosciences (San Jose, CA, USA). *P. syringae* pv. tomato str. DC3000 was used as the source of the genomic library and was donated by Dr. Frederick Ausubel of Massachusetts General Hospital. *P. syringae* was grown in LB (Luria-Bertani) medium with 50 μg/mL rifampicin at 30°C. *Pseudomonas putida* KT2440 (ATCC 47054) was purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA) and grown in LB medium at 30°C. *E. coli* strains were grown in 2YT medium (16 g tryptone, 10 g yeast extract, and 10 g sodium chloride per liter) at 37°C. As required, ampicillin and kanamycin were added to the medium at 100 and 25 μg/mL, respectively.

Escherichia coli DH10B (F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*) 7697 *gal*U *gal*K λ⁻ *rps*L *nup*G) was used as the host strain for the genomic library as well as for subcloning of screened genes (Invitrogen Corp, Carlsbad, CA, USA). *E. coli* MG1655 *ΔuxaC* was provided from Dr. F. R. Blattner of the *E. coli* Genome Project at University of Wisconsin-Madison. For M9 minimal agar, 22 mM glucose, glucuronate, or glucarate were used as carbon sources. Plasmid vectors pTrc99A and pTrc99SE were used for construction of the genomic library and as an expression vector for candidate genes, respectively (Table 3-1). The plasmid pTrc99SE was donated by Prof. Seon-Won Kim at Gyeongsang National University, Korea. pBluescript (Invitrogen, Carlsbad, CA, USA) was used as a general cloning vector.

3.2.2 Genomic DNA preparation and construction and screening of *P. syringae* genomic library

Genomic DNA preparation and general cloning procedures were carried out as described in Sambrook *et al.* (Sambrook and Russell, 2001). The genomic DNA of *A. tumefaciens* str. C58 was purchased from the ATCC (ATCC Number 33970D). Restriction enzymes and T4 ligase were purchased from New England Biolabs (Beverly, MA, USA). *P. syringae* genomic DNA was partially digested with *Bfu*CI, and then loaded onto a 0.8 % agarose gel. Fragments of 2-6 Kb were purified from the gel, and then ligated to pTrc99A with dephosphorylated *Bam*HI ends. After ligation for 2 days at 4°C, the reaction mixtures were used to transform *E. coli* DH10B. Successful transformant clones were collected and pooled from agar plates, followed by storage at -

80°C. Plasmid pools isolated from the colony pools were used to transform *E. coli* MG1655 *ΔuxaC* to screen for Udh activity. Transformed strains were cultured on M9 minimal agar plates with 22 mM glucuronate for 4 days at 30°C. Surviving clones from plates were screened by purifying and sequencing their plasmids. The sequencing results were compared with the genome sequence of *P. syringae* pv. tomato str. DC3000, as reported in GenBank, Accession Number NC_004578 (http://www.ncbi.nlm.nih.gov/).

3.2.3 Construction of expression plasmid vectors containing *udh* genes

PCR amplification was carried out using *Pfu* Turbo AD as described by the manufacturer (Stratagene, La Jolla, CA, USA). The three candidate genes of *iolE*, *iolB*, and PSPTO_1053 were each amplified from the genomic DNA using primers as listed in Table 3-1. PCR products were blunt-end ligated to *Eco*RV-digested pBluescriptII, resulting in pBiolE, pBiolB, pBiolEB and pB1053, which were each sequenced to confirm their identities. *iolE*, *iolB*, and *iolEB* were each cleaved by digestion with *Eco*RI and *Sal*I, and then ligated to pTrc99A digested by same enzymes to construct pTiolE, pTiolB, and pTiolEB, respectively. PSPTO_1053 from pB1053 was cleaved by digestion with *Nco*I and *Sac*I, and then ligated to pTrc99A digested by the same enzymes, resulting in pT1053.

Putative *udh* genes from genomic DNA of *A. tumefaciens*, *P. putida*, and *P. syringae* were amplified using the primer pairs ATudh2-F/ATudh-R, PPudh-F/PPudh-R and PSudh-F/1053-R, respectively (Table 3-1). PCR products were blunt-end ligated to pBluescriptII digested with *Eco*RV, resulting in plasmids pBATudh2, pBPPudh and

pBPSudh. To construct plasmids pTATudh2, pTPPudh, and pTPSudh, the corresponding genes were excised with *Eco*RI and *Sac*I from pBATudh2, pBPPudh, and pBPSudh, respectively, and were inserted into the same sites of pTrc99SE.

3.2.4 Protein purification and determination of kinetic parameters

The *udh* genes from genomic DNA of *A. tumefaciens*, *P. putida*, and *P. syringae* were amplified using primers ATuEQ-F/R, PPuEQ-F/R, and PSuEQ-F/R as listed in Table 3-1. The PCR products were digested with *Sac*I and *Hin*dIII and inserted into the same sites of pET21b containing a 6X His-Tag to construct pETATu, pETPPu, and pETPSu, respectively (Table 3-1). These plasmids were used to transform *E. coli* BL21 (DE3) to use for protein expression. The recombinant *E. coli* BL21 strains were cultivated at 30°C, 250 rpm for 6 hours after IPTG induction. Protein purification was carried out using the ProBond^{TM} Purification System as described by the manufacturer (Invitrogen Corp, Carlsbad, CA, USA). SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was performed as described in Sambrook *et al.* (Sambrook and Russell, 2001). Enzyme activities on substrates of purified proteins were measured by monitoring initial NADH generation at 340 nm and room temperature. Kinetic analysis on glucuronate and galacturonate was carried out using 0 to 10 mM glucuronate or galacturonate and 1.2 mM NAD^+ in 100 mM Tris-HCl, pH 8.0. Kinetic analysis on NAD^+ was performed using 0 to 2 mM NAD^+ and 10 mM glucuronate in 100 mM Tris-HCl, pH 8.0. A series of enzymatic assays were conducted to estimate the initial activity as a function of starting substrate concentration. These data were used to fit the parameters of the Michaelis-Menten kinetic model, k_{cat} and K_m, by nonlinear least-

squares regression. Nonlinear least-squares regression analyses were performed via the Gauss-Newton method as implemented using the intrinsic Matlab® function *nlinfit*.

3.2.5 LC-MS and CD analysis for determination of glucarate produced from glucuronate by Udh

The reaction mixture for producing glucarate from glucuronate by Udh consisted of 20 mM glucuronate, 21.6 mM NAD^+, 40 mM sodium phosphate buffer, pH 8.0, and bacterial lysate prepared as described above. The enzyme reaction was performed by addition of either crude lysate or purified proteins to the reaction mixture and incubation at room temperature for 60 minutes, and then stopped by addition of 1M sodium hydroxide. Glucarate was separated from the reaction mixture using a column packed with boronic acid affinity gel (Affi-gel boronate gel, Bio-Rad Laboratories, Hercules, CA, USA) which is able to bind to the coplanar adjacent *cis*-hydroxyl groups of glucarate (Poon et al., 1993). Glucuronate can not bind to the gel due to its *trans*-diol groups. After loading the Affi-gel column with reaction mixture, the column was washed with 80 mM potassium phosphate-20mM boric acid buffer (pH 7.0), and then glucarate was eluted by the addition of 0.1 M HCl. The eluent was neutralized by the addition of 5 M NaOH and then analyzed by LC-MS using an Agilent 1100 series LC/MSD (Agilent Technologies, US) equipped with an Aminex HPX-87H column (300×7.8 mm, Bio-Rad Laboratories, Hercules, CA USA) and an electron spray ionization detector. Mass spectra were obtained in both the positive and negative ion detection modes. 0.1% (v/v) Trifluoroacetic acid, pH 2.0, was used as the mobile phase at a flow rate of 0.5 mL/min, at room temperature.

The stereochemistry of glucarate formed from glucuronate was confirmed by comparing its circular dichroism (CD) spectrum with that of an authentic glucarate standard. CD was performed on an Aviv Model 202 CD Spectrometer (Aviv Biomedical, Lakewood, NJ). Reaction mixtures contained 20 mM glucuronic acid, 7 mM NAD^+, 100 mM potassium phosphate buffer (pH 8.0), and the purified enzymes prepared as described above. Glucarate was separated from glucuronate using boronic acid affinity gel as described above.

3.2.6 Computational analysis including sequence identification and alignment analysis

Biocyc™ (http://biocyc.org/) was used to identify relevant metabolic pathways and metabolites. DNA sequences for *P. syringae*, *P. putida* and *A. tumefaciens*, were obtained from NCBI (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/), with Accession Numbers NC_004578, NC_002947 and NC_003063, respectively. Sequence management and alignment were carried out using Vector NTI software (Invitrogen, Carlsbad, CA, USA).

3.2.7 GenBank Accession Numbers for *udh* sequences

The *udh* gene sequence from *P. syringae* has been deposited with GenBank, Accession Number EU377538. The corresponding genes from *A. tumefaciens* and *P. putida* were deposited with Accession Numbers BK006462 and BK006380, respectively.

3.3 Results

3.3.1 Cloning of *udh* gene from *Pseudomonas syringae*

The screen established to identify the gene corresponding to Udh activity in *P. syringae* utilized a mutant strain of *E. coli* MG1655. A deletion of *uxa*C prevents growth on glucuronate while retaining the ability to grow on glucarate as a sole carbon source. Since Udh catalyzes the conversion of glucuronate to glucarate (Bateman et al., 1970; Wagner and Hollmann, 1976), *E. coli* MG1655*ΔuxaC* clones harboring *udh* genes from a *P. syringae* genomic library should grow on glucuronate as the sole carbon source. *E. coli* DH10B and pTrc99A were used as the host strain and plasmid vector, respectively, for initial construction of the *P. syringae* genomic library. A library plasmid pool was prepared from the *E. coli* DH10B clone pool, and then used to transform the *ΔuxaC* strain. Transformed *ΔuxaC* clones were incubated on M9 minimal agar containing glucuronate for 4 days at 30°C.

From ten agar plates, 28 clones were selected for further screening, each of which contained an inserted fragment of 2 - 5 kb. From these, 8 clones with different sized inserts were sequenced for comparison with the *P. syringae* genome sequence (GenBank Accession Number NC_004578). Six of these clones included *iolE*, *iolB*, or both of them, while one clone contained the unassigned PSPTO_1053 open reading frame. The final clone included a chimera of the *iolEB* and PSPTO_1053 regions. The open reading frames from the library fragments were PCR-amplified and inserted into expression vector pTrc99A, yielding plasmids pTiolE, pTiolB, pTiolEB and pT1053. Clones

containing these vectors were used to determine which gene corresponded to uronate dehydrogenase activity. *E. coli* MG1655, the *ΔuxaC* derivative, and four *ΔuxaC* clones transformed with the candidate genes were incubated on M9 minimal agar containing glucuronate as the sole carbon source. Wild type, *ΔuxaC* (pTiolB), *ΔuxaC* (pTiolEB), and *ΔuxaC* (pT1053) strains grew on M9-glucuronate agar, while the *ΔuxaC* (pTrc99A) and *ΔuxaC* (pTiolE) strains did not. Therefore, *iolB* and PSPTO_1053 were responsible for growth on glucuronate as the sole carbon source, identifying them as candidate *udh* genes.

To further discriminate between the two candidate genes, plasmids pTiolB and pT1053 were used to transform *E. coli* DH10B to express the recombinant genes. The resulting clones were grown in LB medium with 0.1 mM IPTG. Analysis of Udh activity in crude lysates from these two clones suggested that the strain harboring pT1053 exhibits Udh activity, but not pTiolB. The assay employed glucuronate as a substrate and monitored production of NADH at 340 nm. Consequently, the 828 bp PSPTO_1053 gene was deduced to encode uronate dehydrogenase. The gene is hereafter referred to as *udh* and was registered to Genbank (http://www.ncbi.nlm.nih.gov/Genbank/) as Accession Number EU377538.

3.3.2 Cloning and identification of *udh* genes from *P. putida* and *A. tumefaciens*

The translated protein sequence of *udh* from *P. syringae* was analyzed using BLASTP from NCBI (http://www.ncbi.nlm.nih.gov/blast/) to identify putative homologues. The Udh activity of *A. tumefaciens* has been studied previously (Chang and

Feingold, 1969; Zajic, 1959). The translation of open reading frame Atu3143 of *A. tumefaciens* had the highest sequence identity of 47.8% and was considered a candidate for a homologous Udh. Another candidate open reading frame, PP1171 of *Pseudomonas putida* KT2440, was also found to have high similarity to *P. syringae* Udh, with a sequence identity of 75.6%. Atu3143 and PP1171 were PCR-amplified from their respective genomes and, along with *udh* from *P. syringae*, were integrated into plasmid vector pTrc99SE to create plasmids pTATudh2, pTPPudh, and pTPSudh, respectively, for comparison of relative activities of the expressed recombinant proteins. Transformed DH10B clones were cultivated in LB with or without 0.1 mM IPTG before preparing crude lysates to carry out enzymatic analysis. These assays confirmed a NAD^+-consuming activity in the presence of glucuronate as a substrate for the recombinant proteins of *A. tumefaciens* and *P. putida*, similar to that previously obtained with *P. syringae*. The two *udh* genes from *A. tumefaciens* and *P. putida* were also deposited to Genbank as Accession Numbers BK006462 and BK006380, respectively.

3.3.3 Purification and characterization of recombinant Udh, and analysis of the reaction product

Enzyme reactions using crude *E. coli* lysates containing the *P. syringae udh* gene confirmed the presence of an activity that utilized glucuronate as a substrate, with the reaction rate proportional to glucuronate concentration for low substrate loads (data not shown). The activity also utilized NAD^+ but not $NADP^+$ as a co-factor (data not shown). These results indicated that the substrate was oxidized. An examination of the structure of glucuronate suggests two possible points of oxidation: the conversion of an alcohol to

a ketone, or the conversion of the aldehyde to carboxylic acid, the latter reaction producing glucarate. The difference in these two products should be evident from a mass spectrum, as the former would result in a mass difference of -2 relative to the substrate, while the latter would produce a mass difference of +16. To confirm the product of the enzyme reaction as glucarate, a sample was analyzed by LC-MS. The spectra of the eluent separated from the enzyme reaction and a glucarate standard are in agreement, suggesting glucarate as the product of the Udh reaction (Fig. 3-3).

Each of the three *udh* genes were expressed in *E. coli* with 6X-His tags and purified to determine the kinetic parameters of the corresponding enzymes. Purified enzymes were analyzed by SDS-PAGE to confirm molecular weight of the monomer and estimate purity (Fig. 3-4). The Udh of *P. syringae* and *P. putida* were both approximately 30 KDa molecular weight, which is consistent both with the translation of the cloned gene and previous reports (Wagner and Hollmann, 1976). The *A. tumefaciens* Udh is slightly larger, at 32 KDa.

The purified preparations were used to determine the kinetic parameters, k_{cat} and K_m, for each of the enzymes. Both glucuronate and galacturonate were used as substrates, and NAD^+ co-factor concentration was also varied to determine the corresponding K_m (Table 3-2). Measurements of k_{cat} obtained by varying co-factor concentration were within 20% of the values obtained using glucuronate as the substrate (data not shown). In all cases, k_{cat} was higher for glucuronate than for galacturonate. The highest rate constant was found for the *A. tumefaciens* enzyme utilizing glucuronate as substrate (k_{cat} = 1.9 x

10^2 s^{-1}), which was more than 2-fold higher than the rate for the *Pseudomonas* enzymes. However, the Michaelis (affinity) constant was lower for galacturonate in all cases, with the lowest K_m, 0.04 mM, found for the *P. syringae* enzyme utilizing galacturonate as substrate. The first order rate constants, k_{cat}/K_m, are highest for galacturonate as substrate, with the largest difference between glucuronate and galacturonate observed for *P. syringae*.

The responses of the enzyme activities to changes in pH and temperature were also investigated (Fig. 3-5). A pH optimum of 8.0 was observed for both the *A. tumefaciens* and *P. syringae* enzymes, although the activity was relatively unchanged between pH~7 and pH~8 for *P. syringae* Udh (Fig. 3-5a). This pH behavior is consistent with previous reports for *P. syringae* Udh (Bateman et al., 1970). The *P. putida* enzyme exhibited highest activity at pH~7.0. In general, enzyme activities varied approximately 10% between pH~5 and pH~8, with significant drops in activity observed for pH values greater than 8 for all three enzymes.

The impact of temperature was evaluated in two ways. First, the thermal stability was examined by exposing enzyme preparations to various temperatures for 30 minutes, then performing the enzyme assay under standard conditions. The *A. tumefaciens* Udh was found to exhibit a significantly higher thermal stability than either of the *Pseudomonas* enzymes (Fig. 3-5b). The activity remained near 80% of maximum after exposure of the *A. tumefaciens* preparation to 37 °C, while the corresponding activities for both of the other enzymes was below 20% of maximum. The stability profiles for

both *Pseudomonas* enzymes were similar to one another. Finally, enzyme activity was evaluated for assays conducted under increasing temperatures. These activities followed a general trend of increasing activity with increasing temperature between 4 and 42 °C, which is consistent with an Arrhenius-type dependence of the catalytic rate constant on temperature (Fig. 3-5c).

For final characterization of the products of these reactions, the boronic acid affinity gel was used to isolate the putative glucarate produced from all three enzymes in *in vitro* reactions using purified proteins. Samples of the three products were then subjected to circular dichroism (CD) analysis to examine the stereochemistry of the compounds. All three spectra were in agreement with a glucarate standard, confirming the identity of the product as glucaric acid and the identity of the three genes as those encoding uronate dehydrogenases (data not shown).

3.4 Discussion

Uronate dehydrogenase (Udh) catalyzes the first step of an oxidation pathway for aldohexuronate catabolism in bacteria. In bacteria, only limited studies of Udh in *P. syringae* and *A. tumefaciens* have been reported. Moreover, Udh has been even more rarely studied in eukaryotes. A Udh sequence was reported in the wine grape *Vitis vinifera*, where it was identified as galacturonate reductase (EC 1.1.1.203; BRENDA Accession Number A1Y2Z0, GenBank Accession Number DQ843600). We synthesized this gene with codon usage optimized for expression in *E. coli* (DNA 2.0, Menlo Park,

CA), and expressed the recombinant protein. However, no activity related to Udh was observed when using either NAD^+ or $NADP^+$ as a cofactor (data not shown). An alignment of this sequence with the *P. syringae* Udh identified in the current work reveals only 10% identity between them. I can not rule out the possibility that the *V. vinifera* enzyme could not be functionally expressed in *E. coli*; however, based on the alignment, I conclude that the reported sequence from *V. vinifera* is either not uronate dehydrogenase, or it is a highly divergent version of the enzyme.

A shotgun library of *P. syringae* was introduced into *E. coli ΔuxaC* to screen for the *udh* gene encoding uronate dehydrogenase, and PSPTO_1053 and *iolB* gene were identified and screened as possible Udh candidates. By enzymatic analysis, PSPTO_1053 was ultimately identified to be the *udh* gene encoding uronate dehydrogenase. In a *uxa*C deletion mutant of *E. coli*, where glucuronate catabolism is abolished, glucuronate was converted to glucarate by uronate dehydrogenase, and then degraded to pyruvate or 2- phosphoglycerate from which it can be used as an energy source (Neidhardt and Curtiss, 1996; Roberton et al., 1980). In *E. coli ΔuxaC*, introduction of the *iolB* gene allowed for growth on M9 agar containing glucuronate as a sole carbon source as well, but this gene did not possess Udh activity. IolB has previously been reported as a protein related to myo-inositol catabolism in *Bacillus subtilis* and *Lactobacillus casei* (Yebra et al., 2007; Yoshida et al., 2008). IolB belongs to the *iol* operon used for myo-inositol degradation in *Bacillus subtilis* and converts 5-deoxy-glucuronate to 2-deoxy-5-keto-D-gluconate (Yoshida et al., 2008). IolB of *P. syringae* has about 48% homology with that of *B. subtilis*. The precise mechanism of

glucuronate consumption in cells harboring IolB in our screen is unclear. Presumably, this protein is able to convert glucuronate to an analogous compound that is compatible with *E. coli* metabolism.

We have screened and sequenced three uronate dehydrogenases from *A. tumefaciens*, *P. putida*, and *P. syringae*, which can effectively convert glucuronate to glucarate. While this enzyme is important for the catabolism of uronic acids in several types of bacteria, it may also be useful in the development of biosynthetic pathways for the production of aldaric acids, such as glucaric acid. Glucarate is the end-product of nucleotide sugar metabolism and is found naturally in mammals and plant (Marsh, 1986; Walaszek, 1990). Glucarate and its derivatives such as glucaro-1,4-lactone have been studied previously as detoxifying and natural anti-carcinogenic compounds (Duff, 2002; Marsh, 1986; Singh and Gupta, 2003; Walaszek, 1990), as well as a building block for polymer synthesis (Ibert et al., 2002). It has also been designated as a potential "top value-added" chemical to be produced from biomass (Werpy and Petersen, 2004). Presently, glucarate is synthesized from glucose by chemical oxidation using a strong oxidant such as nitric acid or nitric oxide (Merbouh et al., 2001). We have used the *udh* of *P. syringae* identified in this study to successfully produce glucaric acid from a synthetic pathway in *E. coli* (Moon et al., 2009a). In Chapter 4, this synthetic pathway construction is discussed.

Table 3-1. Strains, plasmids, and primers used in this study.

Plasmids and Primers	Description	Reference or source
Strains		
Pseudomonas syringae pv. tomato strain DC3000	Wild type	
Pseudomonas putida KT2440	Wild type	(ATCC 47504)
Escherichia coli DH10B	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*) 7697 *gal*U *gal*K λ⁻ *rps*L *nup*G	(Invitrogen Corp., Carlsbad, CA, USA)
Escherichia coli MG1655 *Δ uxaC*	Wild type with deletion of *uxaC* gene encoding D-glucuronate isomerase	(Kang et al., 2004)
Escherichia coli BL21 (DE3)	F- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* Ã«(DE3) T1R	(Invitrogen Corp., Carlsbad, CA, USA)
Plasmids		
pBluescriptII	*lac* promoter, ColE1 origin, Ampicillin resistance, *lacZ*	(Stratagene, La Jolla, CA, USA)
pTrc99A	*trc* promoter, pBR322 origin, Ampicillin resistance, *lacI*^{q}	(Amann et al., 1988)
pET21b	T7 promoter, ColE1 origin, Ampicillin resistance, *lacI*	(Novagen, Darmstadt, Germany)
pTrc99SE	pTrc99A containing RBS sequence of AGGAGGTAATAAAT	(Seon-Won, Kim)
pTiolE	pTrc99A with *iolE* of *P. syringae*	This study
pTiolB	pTrc99A with *iolB* of *P. syringae*	This study
pTiolEB	pTrc99A with *iolE* and *iolB* of *P. syringae*	This study
pT1053	pTrc99A with PSPTO_1053 of *P. syringae*	This study
pTATudh2	pTrc99SE with *udh* of *A. tumefaciens*	This study

pTPPudh	pTrc99SE with *udh* of *P. putida*	This study
pTPSudh	pTrc99SE with *udh* of *P. syringae*	This study
pETATu	pET21b with *udh* of *A. tumefaciens*	This study
pETPPu	pET21b with *udh* of *P. putida*	This study
pETPSu	pET21b with *udh* of *P. syringae*	This study

Primers [a]

iolE-F	5'- CGAATTCAGG**AGGTACAACCATGCCTGTTTCAG** -3'
iolE-R	5'- CGTCGACT**TATCGCGCATCGGCCAGCAGTTG** -3'
iolB-F	5'- CGAATTC**AGGAGGATTGAATCATGAGTC** -3'
iolB-R	5'- CGTCGAC**TTAAAGATCCAGCAGCCAGC** -3'
1053-F	5'- G**CCATGGCATCGGCTCATACCAC** -3'
1053-R	5'- CGAGCTC**TTATTTATCGCCGAACGGTCC** -3'
ATudh2-F	5'-CTAGAATTC**ATGAAACGGCTTCTTGTTACC**-3'
ATudh-R	5'-CTAGAGCTC**TTAGCTCTGTTTGAAGATCGGGTTG**-3'
PPudh-F	5'-GTCGAATTC**ATGACCACTACCCCCTTCAATC**-3'
PPudh-R	5'-CTAGAGCTCCGTGGGG**TTAGTTGAACGGGC**-3'
PSudh-F	5'-CTAGAATTC**ATGGCATCGGCTCATACCACTC**-3'
ATuEQ-F	5'-TCAGAGCTCG**AAACGGCTTCTTGTTACCGGTGC**-3'
ATuEQ-R	5'-CTGAAGCTT**GCTCTGTTTGAAGATCGGGTTGTCG**-3'
PPuEQ-F	5'-TCAGAGCTCG**ACCACTACCCCCTTCAATCGCC**-3'

PPuEQ-R	5'-CTGAAGCTT**GTTGAACGGGCCGGCCACGGCG**-3'
PSuEQ-F	5'-TCAGAGCTC**GGCATCGGCTCATACCACTCAAACTCC**-3'
PSuEQ-R	5'-CTGAAGCTT**TTTATCGCCGAACGGTCCGGACGC**-3'

[a]Primer binding sites, restriction sites, start or stop codons were indicated as bold letters, double and single underlines, respectively.

Table 3-2. Turnover numbers (k_{cat}) and Michaelis constants (K_m) of uronate dehydrogenases from *A. tumefaciens*, *P. putida*, and *P. syringae*. Glucu = Glucuronate; Galactu = Galacturonate.

	A. tumefaciens			*P. putida*			*P. syringae*		
	Glucu	Galactu	NAD^+	Glucu	Galactu	NAD^+	Glucu	Galactu	NAD^+
k_{cat} (10^2 1/s)	1.9 ± 0.1	0.92 ± 0.14	-	0.55 ± 0.03	0.30 ± 0.03	-	0.74 ± 0.03	0.24 ± 0.01	-
K_m (mM)	0.37 ± 0.12	0.16 ± 0.12	0.18 ± 0.03	0.25 ± 0.07	0.10 ± 0.06	0.21 ± 0.02	0.28 ± 0.07	0.04 ± 0.01	0.17 ± 0.07
k_{cat}/K_m (10^2 1/(s-mM))	5.2	5.7	11	2.2	3.0	2.6	2.6	6.0	4.3

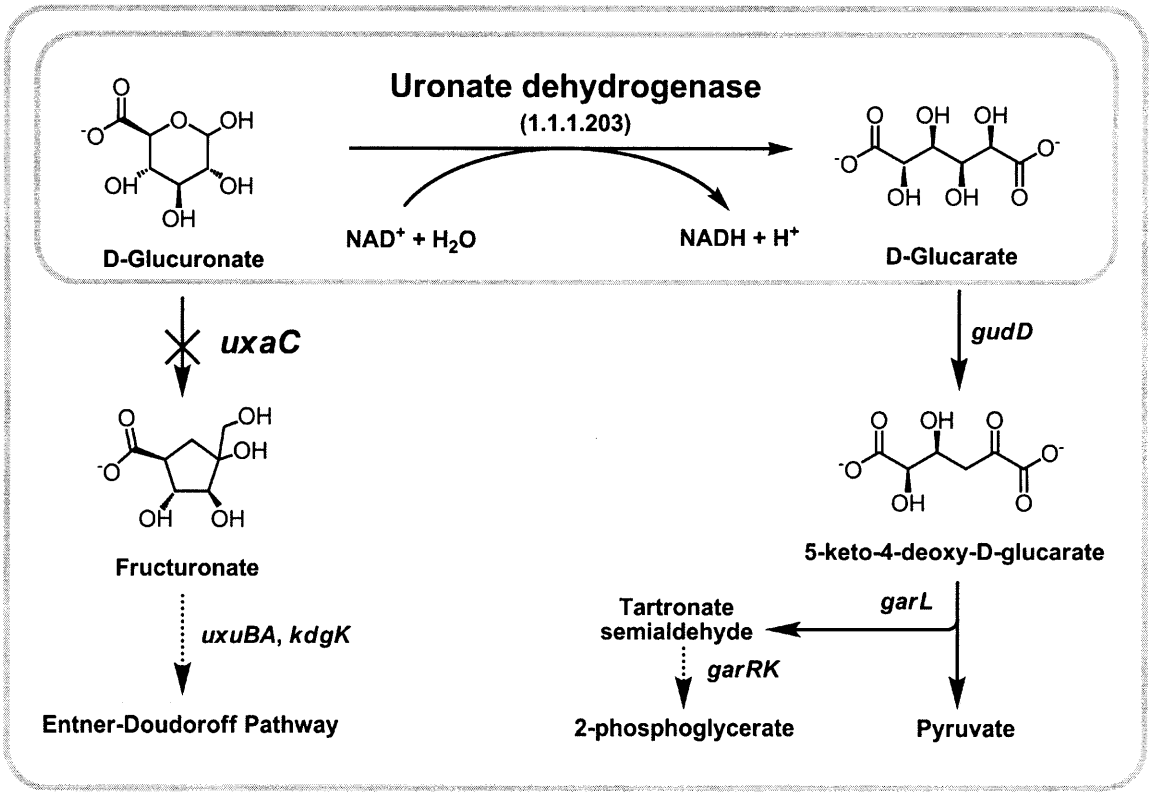


Fig. 3-1. Catabolism of glucuronic and glucaric acids in bacteria. Glucuronic acid consumption is prevented by knock-out of the *uxa*C gene. Presence of uronate dehydrogenase in an *uxaC* knock-out enables growth of *E. coli* on glucuronic acid.

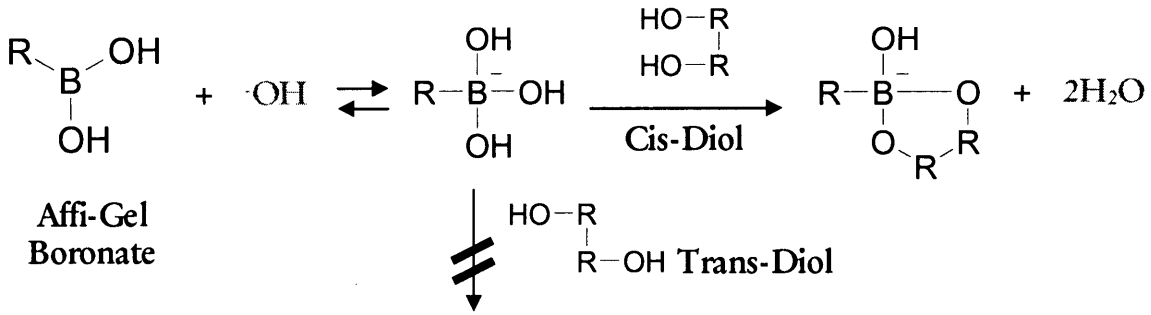
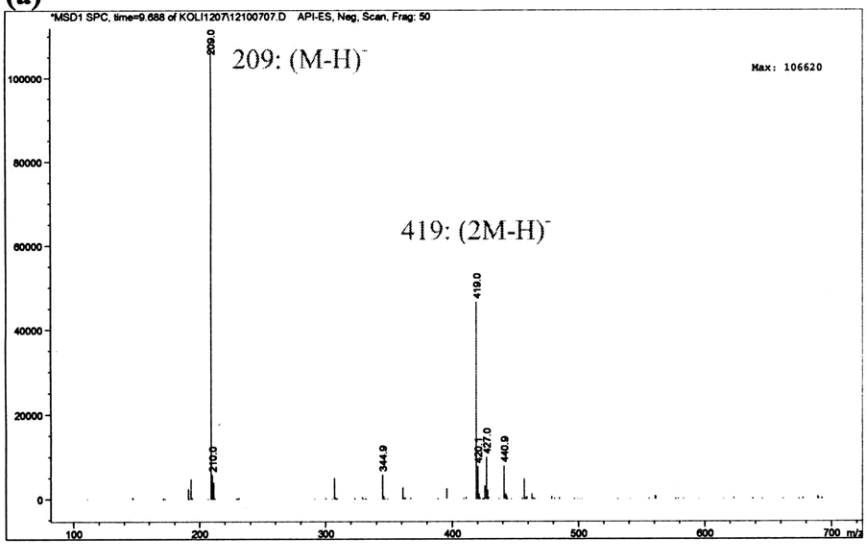


Fig. 3-2. Mechanism of binding between diol and boronic acid affinity gel (http://www.bio-rad.com).

(a)



(b)

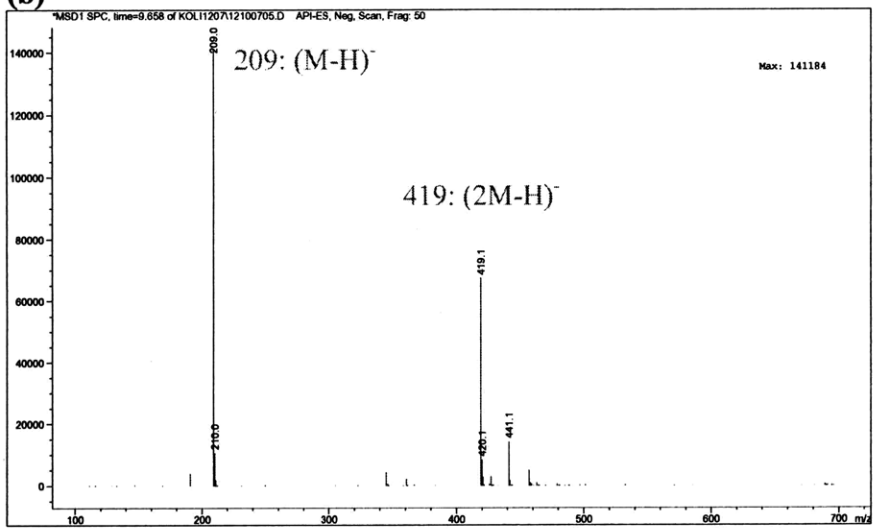


Fig. 3-3. ES mass spectra of glucarate, **(a)** as separated from the enzymatic reaction mixture and **(b)** standard. Glucarate was characterized by its masses (m/z = 209 and 419) and peaks of samples corresponded to masses of the glucarate standard.

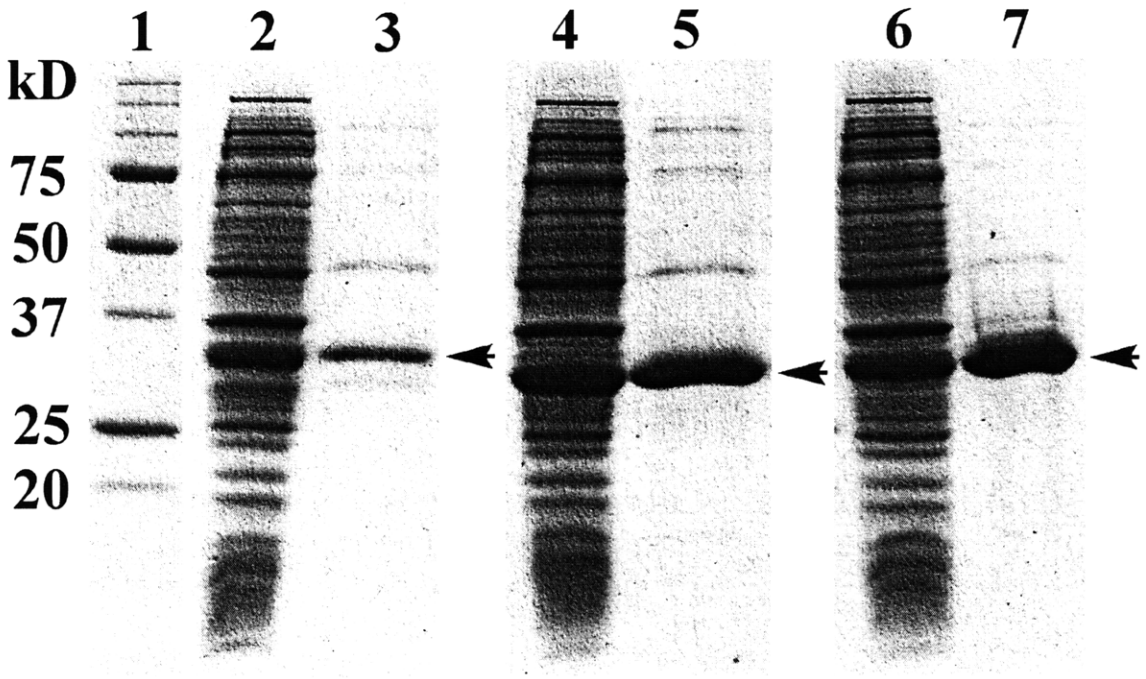
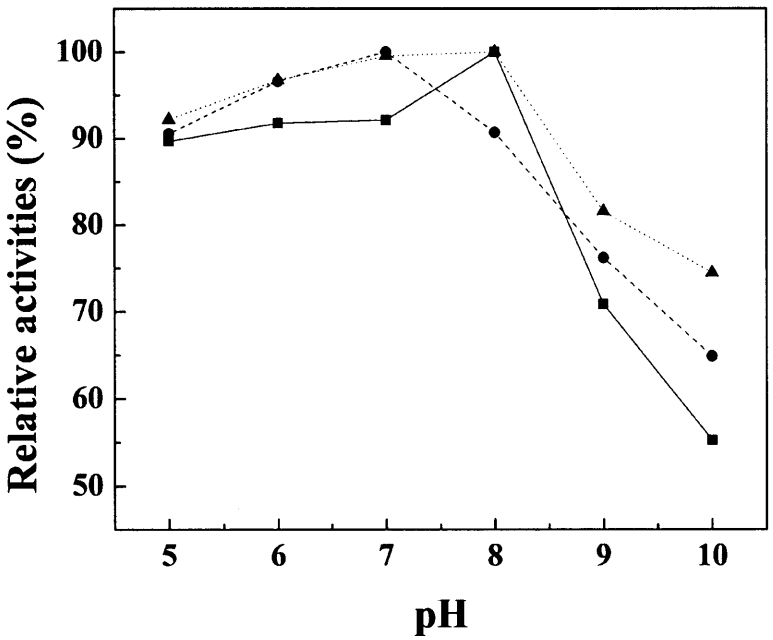
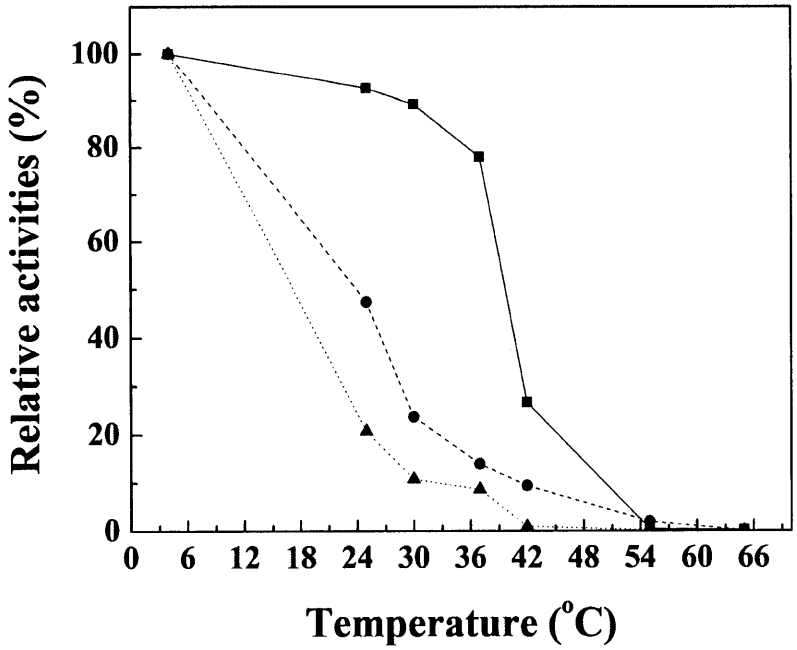


Fig. 3-4. SDS-PAGE analysis of purified Udhs. The purified Udhs were subjected to electrophoresis in a 12% sodium dodecylsulfate polyacrylamide gel under denaturing conditions. Lane 1, molecular weight markers; lanes 2 and 3, crude extract and purified *A. tumefaciens* Udh of *E. coli* BL21(DE3) expressing pETATu; lanes 4 and 5, crude extract and purified *P. putida* Udh of *E. coli* BL21(DE3) expressing pETPPu; lanes 6 and 7, crude extract and purified *P. syringae* Udh of *E. coli* BL21(DE3) expressing pETPSu. The purified Udhs are indicated by the arrow symbols.

(a)


Relative activities (%)
100
90
80
70
60
50
5
6
7
8
9
10
pH

(b)


Relative activities (%)
100
80
60
40
20
0
0
6
12
18
24
30
36
42
48
54
60
66
Temperature (°C)

(c)

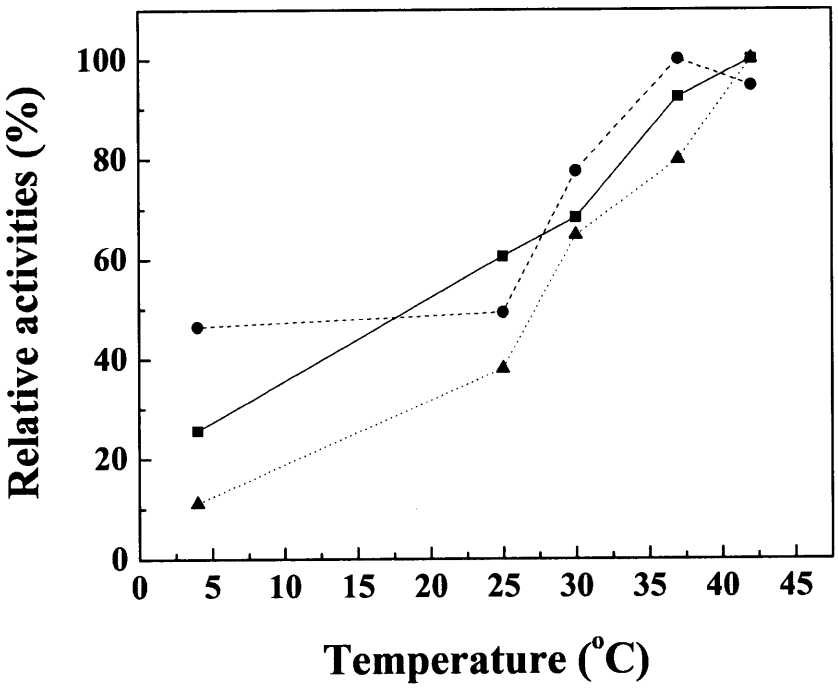


Fig. 3-5. Effect of pH and temperature on activities of Udhs from *A. tumefaciens*, *P. putida*, and *P. syringae udh*. **(a)** Relative activities as a function of pH. **(b)** Relative activities after incubation for 30 minutes at indicated temperatures. **(c)** Relative activities as a function of assay temperature. Square with plain line, *A. tumefaciens* Udh. Circle with dashed line, *P. putida* Udh. Triangle with dotted line, *P. syringae* Udh.

Chapter 4

Part Assembly – Construction of Synthetic Glucaric Acid Pathway

TS Moon, S-H Yoon, A Lanza, J Roy-Mayhew and KJ Prather. **Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli***, *Appl. Environ. Microbiol.* 75, 589-595 (2009)

Abstract

A synthetic pathway has been constructed for the production of glucuronic and glucaric acids from glucose in *Escherichia coli*. Co-expression of the genes encoding *myo*-inositol-1-phosphate synthase (Ino1) from *Saccharomyces cerevisiae* and *myo*-inositol oxygenase (MIOX) from mouse led to production of glucuronic acid through the intermediate *myo*-inositol. Glucuronic acid concentrations up to 0.3 g/L were measured in the culture broth. The activity of MIOX was rate-limiting, resulting in the accumulation of both *myo*-inositol and glucuronic acid as final products, in approximately equal concentrations. Inclusion of a third enzyme, uronate dehydrogenase (Udh) from *Pseudomonas syringae*, facilitated the conversion of glucuronic acid to glucaric acid. The activity of this recombinant enzyme was more than two orders of magnitude higher than that of Ino1 and MIOX and increased overall flux through the pathway such that glucaric acid concentrations in excess of 1 g/L were observed. This represents a novel microbial system for the biological production of glucaric acid, a "top-value added chemical" from biomass.

Keywords: glucaric acid; glucuronic acid; uronate dehydrogenase; inositol-1-phosphate synthase; *myo*-inositol oxygenase; biosynthetic pathway.

4.1 Introduction

The discipline of metabolic engineering was defined more than fifteen years ago as "the improvement of cellular activities by manipulations of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology" (Bailey, 1991). It has shown its potential to optimize cellular functions for many purposes including recombinant protein production and pathway engineering for productivity enhancement. More recently, the field has produced several important accomplishments in the area of pathway design for new product generation. I consider a designed pathway as one in which a sequence of conversions is assembled from the individual enzymes of multiple hosts, in a sequence that is not native to any single organism. As examples, I highlight the production of 1,3-propanediol (Nakamura and Whited, 2003), amorphadiene (Martin et al., 2003), and 1,2,4-butanetriol (Niu et al., 2003) in *E. coli*. In these approaches, the conversion steps were designed based on enzyme availability, the recruited enzyme activities from various organisms were identified, and the designed pathways were constructed in *E. coli* by assembling these enzymatic steps. The general concept incorporated in these metabolic engineering examples is to consider biological components, including enzymes, as interchangeable "parts," and the term "synthetic biology" has been used to describe this concept (Benner, 2003; Yeh and Lim, 2007). The ability to conceive of and assemble new biosynthetic pathways by combining protein parts from multiple sources leads to increased possibilities for the construction of microbial chemical factories to produce compounds that do not have a microbial origin.

One such compound of interest is D-glucaric acid. It is available as a dietary supplement in the form of calcium D-glucarate and has been studied for therapeutic purposes including cholesterol reduction (Walaszek et al., 1996) and cancer chemotherapy (Singh and Gupta, 2003; Singh and Gupta, 2007). D-glucaric acid was also identified as a "top value-added chemical from biomass" in a report of the Pacific Northwest National Laboratory (PNNL) and the National Renewable Energy Laboratory (NREL) (Werpy and Petersen, 2004), which also described its potential use as a building block for a number of polymers, including new nylons and hyperbranched polyesters. Indeed, D-glucaric acid produced from D-glucose has been successfully utilized to produce a hydroxylated nylon, resulting in a presumably biodegradable fiber (Kiely et al., 1994). D-glucaric acid is a highly functionalized compound with four chiral carbons and is currently produced by chemical oxidation of glucose, a nonselective and expensive process using nitric acid as the oxidant (Werpy and Petersen, 2004). New catalytic processes using biological systems may lead to higher yield and selectivity.

To construct a biological system for the synthesis of D-glucaric acid, it is reasonable to start with an examination of the naturally existing mammalian pathway, which has been elucidated (Kuellmer, 2001). D-glucaric acid and L-ascorbic acid are both end products of the D-glucuronic acid pathway. This pathway is a cycle that is initiated with either D-galactose or D-glucose and interacts with the pentose phosphate pathway. D-glucuronic acid is converted to D-glucaric acid in three successive steps, through the intermediates D-glucurono-γ-lactone and D-glucaro-γ-lactone. Therefore,

there is a known route for the production of D-glucaric acid from D-glucose. However, this is a lengthy pathway, consisting of more than ten conversion steps.

Here I describe the construction of a synthetic pathway for the production of D-glucaric acid in *E. coli* by combining "biological parts" from disparate organisms (Fig. 4-1). *myo*-Inositol 1-phosphate synthase (Ino1, also known as "MIPS"), encoded by the *INO1* gene of *Saccharomyces cerevisiae*, is used to produce *myo*-inositol-1-phosphate from glucose-6-phosphate (Dean-Johnson and Henry, 1989). Ino1 requires NAD^+ for activity, but the co-factor is regenerated in the isomerization reaction. The Ino1 substrate is present in *E. coli* as the result of glucose transport by the PTS system, which consumes one molecule of phosphoenolpyruvate per molecule of glucose transported (Postma et al., 1993). *myo*-Inositol-1-phosphate, the product of the Ino1 reaction, is hydrolytically de-phosphorylated to produce *myo*-inositol by an endogenous phosphatase (Matsuhisa et al., 1995). In yeast, *myo*-inositol is a constituent of membrane phospolipids, and its derivatives are important for cell signaling. A second recombinant enzyme, *myo*-inositol oxygenase (MIOX) can convert *myo*-inositol to D-glucuronic acid using molecular oxygen. This enzyme is present primarily in mammalian sources and represents the first step of *myo*-inositol catabolism (Charalampous and Lyras, 1957). Co-expression of these two enzymes in *E. coli* enables the production of D-glucuronic acid from D-glucose. As an alternative to the 3-step mammalian route for producing D-glucaric acid from D-glucuronic acid, uronate dehydrogenase from *Pseudomonas syringae* can perform this conversion utilizing NAD^+ as a co-factor and a water molecule (Bateman et al., 1970; Wagner and Hollmann, 1976). The *udh* gene encoding this activity was recently cloned

and characterized (Chapter 3) (Yoon et al., 2009). Expression of this third gene with *INO1* and MIOX enables the production of D-glucaric acid from D-glucose.

4.2 Materials and Methods

4.2.1 Strains, growth media, and plasmids

E. coli strain DH10B [F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*Δ139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L (Str^R) *nup*G] was used for all molecular biology manipulations. DH10B and BL21 Star™ (DE3) [F^- *omp*T *hsd*S_B (r_B^- m_B^-) *gal dcm rne1*31 (DE3)] were used as hosts for production of organic acids. Competent cells of both strains were purchased from Invitrogen Corporation (Carlsbad, CA). Cultures were propagated in either LB or M9 media. LB (Miller) medium was prepared from dehydrated powder according to manufacturer's instructions (BD Biosciences, San Jose, CA). M9 was prepared as described (Sambrook and Russell, 2001), and consisted of 1X M9 salts (12.8 g/L $Na_2HPO_4 \cdot 7H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 10 g/L (1%) glucose. Leucine was added to a final concentration of 105 μg/mL for DH10B. Kanamycin was added to a final concentration of 20 μg/mL and ampicillin to a final concentration of 100 μg/mL where desired to provide selective pressure for plasmid maintenance.

All molecular biology manipulations were performed according to standard practices (Sambrook and Russell, 2001). The *INO1* gene encoding *myo*-inositol 1-phosphate synthase (Ino1, also known as MIPS) was PCR-amplified from a genomic

DNA preparation of *Saccharomyces cerevisiae* using the following primers: forward – 5'-GAATTCATGACAGAAGATAATATTGCTC-3'; reverse – 5'-AAGCTTCTACAACAATCTCTCTTCG-3'. *Eco*RI and *Hind*III restriction sites included in the 5' ends of the primers are underlined. The mouse MIOX gene encoding *myo*-inositol oxygenase was synthesized with codon optimization for expression in *E. coli* by DNA 2.0 (Menlo Park, CA) based on GenBank Accession Number AF197127. Optimization of the 858 nucleotide (286 codon) sequence was performed by the vendor, with the results summarized as follows: 19.2% of the nucleotides were altered, affecting 153 of the 286 codons (53.5%). Among the optimized codons, 144 (94.1%) were only altered at the third nucleotide position. All three nucleotides were changed in 3 of the codons. The synthetic gene was received as plasmid pJ2-MIOX. *Eco*RI and *Hind*III restriction sites were included in the 5' and 3' ends of the gene, respectively. Both genes were sub-cloned into the IPTG-inducible plasmids pMMB206 (Morales et al., 1991) and pTrc99A (Amann et al., 1988) to confirm activity of the expressed enzymes. The resulting plasmids were designated pMMB-INO1, pTrc-INO1, pMMB-MIOX, and pTrc-MIOX. For co-expression of both genes, the pRSFDuet-1 vector from Novagen containing two T7 promoters was used (Gibbstown, NJ). The *INO1* gene was sub-cloned into the first position using the *Eco*RI and *Hind*III sites, producing plasmid pRSFD-IN. To introduce the MIOX gene into the second position, the *Hind*III site of pJ2-MIOX was end-filled using Klenow enzyme prior to digestion with *Eco*RI. pRSFD-IN was first digested with *Xho*I and end-filled and then digested with the *Eco*RI-compatible *Mfe*I prior to ligation with the MIOX gene fragment. The resulting plasmid was designated as pRSFD-IN-MI. We previously isolated the *udh* gene encoding uronate dehydrogenase

from *Pseudomonas syringae* (GenBank Accession Number EU377538) and sub-cloned the gene into pTrc99A to produce pT1053 (hereafter referred to as pTrc-udh) as described (Yoon et al., 2009).

4.2.2 Enzyme assays for Ino1, MIOX, and Udh activity

Functional expression of the *INO1*, MIOX, and *udh* genes was confirmed through *in vitro* assays of enzyme activity. Crude lysates were prepared by first re-suspending cell pellets from 1-2 mL culture in 100-200 μL 10 mM Tris-Cl (pH 8.0) with 1 mg/mL lysozyme. Cell solutions were lysed by alternating freezing in liquid nitrogen with thawing in 30-40°C water for 5 cycles. The resulting solutions were centrifuged at 14,000 rpm at 4°C for 15 minutes to remove insolubles. The total protein concentration of lysates was determined using the Bradford method (Bradford, 1976).

Assays for *myo*-inositol 1-phosphate synthase activity were performed as described previously (Adhikari et al., 1987; Barnett et al., 1970). Briefly, glucose-6-phosphate substrate was converted to *myo*-inositol-1-phosphate in a reaction buffer consisting of 50 mM Tris-acetate (pH 7.5), 0.8 mM NAD^+, 14 mM NH_4Cl, 5 mM mercaptoethanol, and 5 mM glucose-6-phosphate. Reactions were initiated with the addition of lysate and incubated for 1 hr at 37°C. Reactions were terminated with the addition of 0.4 volume 20% trichloroacetic acid. To quantitate product, inorganic phosphate was removed from the *myo*-inositol-1-phospate by oxidation with equal volume 0.2 M $NaIO_4$. Excess periodate was destroyed with the addition of equal volume

1 M Na_2SO_3. Control reactions were established without glucose-6-phosphate and without addition of periodate.

Assays for *myo*-inositol oxygenase activity were performed as described previously (Arner et al., 2001; Reddy et al., 1981a; Reddy et al., 1981b). The reaction buffer consisted of 50 mM Tris-Cl (pH 8.0), 2 mM L-cysteine, 1 mM $Fe(NH_4)_2(SO_4)_2$, and 60 mM *myo*-inositol. Samples were pre-incubated without substrate for 10 minutes at 30°C to activate the MIOX enzyme. Reactions were incubated for 1 hr at 30°C, then terminated with the addition of 1/10 volume 30% trichloroacetic acid. The glucuronic acid produced was quantified using an orcinol reagent (Charalampous and Lyras, 1957). The reagent consisted of 40 mg orcinol in 10 mL concentrated HCl containing 5.4 mg $FeCl_3$. One volume sample was mixed with two volumes orcinol reagent and incubated for 30 minutes in boiling water. After cooling to room temperature, absorbance at 670 nm was measured to determine glucuronic acid concentration. Control reactions were established without *myo*-inositol to account for background.

Assays for uronate dehydrogenase activity were performed by monitoring NADH co-factor generation at 340 nm as described previously (Wagner and Hollmann, 1976; Yoon et al., 2009). The reaction mixture contained 100 mM sodium phosphate buffer (pH 8.0), 2.5 mM glucuronic acid, 0.9 mM NAD^+, and bacterial lysate prepared as described above.

4.2.3 Growth conditions for acid production

Cultures were grown in LB medium supplemented with 10 g/L glucose and induced with IPTG as indicated in Results. An inoculum was prepared in LB medium, and 1 or 2% (v/v) was used to inoculate 250-mL baffled flasks containing 50 or 100 mL of medium. The cultures were incubated at 30°C and 250 rpm, with periodic sampling to determine cell density and product concentration in the culture medium.

4.2.4 Detection and quantification of organic acids

Metabolites including glucuronic acid and glucaric acid were quantified by high-performance liquid chromatography (HPLC). For glucaric acid assays, samples were pre-treated as previously described (Poon et al., 1993; Yoon et al., 2009) to separate glucaric acid from other metabolites including glucuronic acid. Briefly, boronic acid affinity gel (Affi-gel boronate gel, Bio-Rad Laboratories, Hercules, CA), which has an affinity for the coplanar adjacent *cis*-hydroxyl groups present in glucaric acid (Poon et al., 1993), was mixed with samples and washed with 80 mM potassium phosphate - 20 mM boric acid buffer (pH 7.0). Glucaric acid was eluted with 0.1 M hydrochloric acid. The eluate was neutralized by adding 10 M NaOH and then analyzed by HPLC. HPLC analyses were performed on an Agilent 1100 series instrument equipped with an Aminex HPX-87H column (300 mm x 7.8 mm, Bio-Rad Laboratories, Hercules, CA) and refractive index and diode array detectors under the following conditions: mobile phase, 5 mM sulfuric acid in water; flow rate, 0.5 mL/min; injection volume, 50 μL; temperature, 55°C; UV wavelength, 210 nm.

4.3 Results

4.3.1 Verification of recombinant Ino1 and MIOX activities

The use of *myo*-inositol 1-phosphate synthase (Ino1) from *Saccharomyces cerevisiae* to produce high concentrations of *myo*-inositol through *E. coli* fermentation has been previously reported (Hansen et al., 1999). Product titers up to 21 g/L were obtained under high cell density, fed-batch fermentations operated for 54 hrs. To confirm Ino1 performance in shake flasks, the corresponding gene was amplified, inserted into a compatible vector, then sub-cloned into both high- and medium-copy plasmids for expression in the common laboratory strain DH10B. Plasmid pTrc-INO1 contains the modified ColE1 replicon that results in copy numbers of several hundred, while pMMB-INO1 is based on the RSF1010 replicon with a copy number of the order of 10. Two plasmids were evaluated to explore the potential for co-expression of the *INO1* and MIOX genes in a single strain using compatible vectors. *In vitro* activity of 344 nmol/hr/mg and 128 nmol/hr/mg was measurable for cultures harboring pTrc-INO1 and pMMB-INO1, respectively, and incubated at 30°C, indicating successful expression of the enzyme. However, only expression from the high-copy plasmid resulted in accumulation of measurable quantities of *myo*-inositol in the culture medium, 0.37 g/L. Activity was also a strong function of temperature, with none detectable for cultures grown at 37°C. *myo*-Inositol production was also tested in M9 minimal medium. I postulated that growth in minimal medium with glucose as the only carbon source might increase glucose flux and accordingly increase *myo*-inositol production. However, only half the amount of *myo*-inositol was produced, suggesting that while glucose flux may indeed be higher, the Ino1 enzyme expressed under these conditions does not compete as

effectively against glycolysis for substrate. Subsequent experiments were conducted in LB medium supplemented with glucose.

MIOX is a protein of primarily eukaryotic origin, and the homologues from human, mouse, rat, and pig have been best characterized (Arner et al., 2004; Arner et al., 2001; Reddy et al., 1981a; Reddy et al., 1981b). *myo*-Inositol oxygenase (MIOX) has been functionally expressed in *E. coli* and purified for characterization of the enzyme's properties; however, to our knowledge, mammalian MIOX has not been used in a whole cell, recombinant system to produce glucuronic acid. The mouse version of the enzyme had been found to have the most favorable properties upon expression in *E. coli* (Arner et al., 2004) and was chosen for investigation. A synthetic version of the gene was purchased from DNA 2.0, with codon optimization for *E. coli*. This gene was also sub-cloned into both the high-copy and low-copy vectors used to evaluate Ino1 activity in DH10B. MIOX activity was initially evaluated at 37°C since the enzyme is of mammalian origin.

The MIOX enzyme is known to require Fe^{2+} and cysteine for activation *in vitro* (Arner et al., 2001). The addition of these compounds to the culture medium did not improve the expression of the enzyme from pTrc-MIOX as measured in the *in vitro* assay but rather resulted in a decrease in activity (Table 4-1). Glucuronic acid was still measured in the culture medium, though at a lower concentration. The observed decrease in enzyme activity coincided with a significant decrease in cell density, indicating toxicity of these compounds to the host. As reported previously (Reddy et al., 1981a;

Reddy et al., 1981b), MIOX activity is inhibited by Fe^{2+} and cysteine at high concentrations. While the extracellular concentrations were set at a level that activates the enzyme in the *in vitro* assay, the corresponding intracellular concentrations are unknown. It was also reported previously that inclusion of *myo*-inositol in the culture medium improved soluble expression of MIOX in *E. coli* (Arner et al., 2004). We also observed this behavior, noting a sharp decrease in activity of the enzyme when expressed in the absence of *myo*-inositol supplementation (Table 4-1). One striking feature of recombinant MIOX is its apparent instability (Arner et al., 2004). High activity was observed in samples taken during exponential phase (6 hrs after inoculation) but dropped substantially in stationary phase (24 hrs after inoculation) (Table 4-1). The background activity of the assay, as measured in control samples containing empty pTrc99A plasmid, generally increases with time. Note that the high background of the assay results from the non-specificity of the orcinol reagent, which is known to react with other biological compounds, though to a smaller extent. As a result, the assay may not be reliable for precise quantification of enzyme activity. However, the differences observed between samples with and without *myo*-inositol, and between samples with *myo*-inositol at early and late time points are sufficiently large that the trends can be considered significant.

Neither *in vitro* enzyme activity nor *in vivo* production of glucuronic acid was observed in cultures containing the lower copy pMMB-MIOX construct, suggesting that high expression levels are required to achieve measurable MIOX activity. Because *INO1* is only actively expressed at 30°C, *in vivo* MIOX performance was also evaluated at this

temperature from the high copy plasmid. A comparable amount of glucuronic acid, 0.40 g/L, was produced after 24 hr in culture, with titers doubling to 0.78 g/L after 48 hr.

4.3.2 Production of glucuronic acid from glucose

Production of glucuronic acid from glucose requires the co-expression of both *INO1* and MIOX in the same strain. The compatible plasmids pTrc99A and pMMB206 were both investigated, with the expectation that a doubly transformed strain containing either pTrc-INO1 and pMMB-MIOX or pMMB-INO1 and pTrc-MIOX could be used for production. However, our results indicated that reasonable *in vivo* activities, as determined by accumulation of each desired product in the culture medium, were only achievable with expression of both genes from high-copy plasmids. To address this issue, we introduced both enzymes into the high-copy pRSFDuet vector, which contains a pair of multi-cloning sites, each behind a T7 promoter. Enzyme activities were confirmed as described previously and expression was verified by SDS-PAGE (data not shown). In this manner, an IPTG concentration of 0.1 mM was determined to be preferred. The host strain was also changed from DH10B to BL21(DE3), to enable expression from the T7 promoter. We had previously observed that DH10B was incapable of consuming glucuronic acid for growth (data not shown). BL21(DE3) can metabolize glucuronic acid; however, its consumption appeared to be subject to catabolite repression (data not shown). Therefore, cultivation of the strain in excess glucose prevents consumption of the desired product.

The BL21(DE3) strain carrying pRSFD-IN-MI is capable of producing glucuronic acid from glucose, though to levels of only ~270 mg/L (Fig. 4-2). The culture profile shows that glucuronic acid is present after 24 hrs with no intermediates detectable, and the concentration increases by 50% in 4 days. However, after 48 hr, significant quantities of *myo*-inositol appear in the culture medium. *myo*-Inositol continues to accumulate in the medium and is present in concentrations slightly higher than the desired end product, glucuronic acid, by the end of the experiment. The final glucuronic acid concentration, 0.27 g/L, is lower than that observed with direct conversion of *myo*-inositol in the DH10B(pTrc-MIOX) system above (0.78 g/L). The accumulation of *myo*-inositol suggests that MIOX activity is the limiting factor in production of high concentrations of glucuronic acid. *In vitro* assays confirmed that Ino1 activity is significantly higher than the vector-only control throughout the course of the experiment, with only marginal background activity appearing after 3 days (data not shown). In contrast, MIOX activity was only slightly higher than background after 1 day and was subsequently indistinguishable from background. This is consistent with the results summarized previously (Table 4-1) that indicate that MIOX activity drops sharply after 24 hrs. Additionally, it is likely that the activity of MIOX in this system is limited by the concentration of *myo*-inositol produced by Ino1. While an extracellular supplementation of 60 mM (10.8 g/L) *myo*-inositol does not mean the intracellular concentration is also this high, it is reasonable to suspect that the intracellular concentrations of *myo*-inositol that result from Ino1 activity are likely to fall short of the equivalent concentration.

4.3.3 Production of glucaric acid from glucose

We recently cloned and characterized the gene encoding uronate dehydrogenase activity from *Pseudomonas syringae* pv. tomato DC3000 (Yoon et al., 2009). The *udh* gene was found to be very well-expressed in *E. coli,* resulting in high enzyme activities. For the production of glucaric acid, we utilized a previously constructed vector harboring the *udh* gene in pTrc99A, which is compatible with pRSFD-IN-MI. Both vectors were introduced into BL21(DE3) to construct an *E. coli* strain carrying *INO1*, MIOX, and *udh*. Productivity of this strain was measured under several different induction conditions (Table 4-2). To our surprise, up to 1 g/L of glucaric acid was produced although only 0.27 g/L of glucuronic acid was previously observed in the system harboring the first two genes. Under induction conditions identical to those previously used for glucuronic acid (Table 4-2, Condition A), 0.72 g/L of glucaric acid was produced. To further characterize the system, enzyme activities in crude lysates were measured after each day of culture (Fig. 4-3). Udh activity was highest, more than two orders of magnitude higher than Ino1 activity, and three orders of magnitude higher than MIOX activity. The high activity of Udh thus appears to pull glucose flux through the glucaric acid pathway, leading to a relatively higher titer of glucaric acid. In these samples, MIOX activity does not appear to decrease over time as observed previously; however, the magnitude of the activity remains quite low. Additionally, the first data point here is after one day, when MIOX activity was previously shown to have decreased significantly from that observed during exponential growth (Table 4-1). No glucuronic acid was detected after three days culture time while *myo*-inositol accumulated, confirming that the MIOX-catalyzed step is limiting.

The three induction conditions tested resulted in glucaric acid concentrations that ranged from 0.72 to 1.13 g/L. In general, higher induction levels, i.e., higher IPTG concentration, resulted in a higher yield of glucaric acid on glucose but lower product concentration (compare, for example, Conditions A and B in Table 4-2). Higher induction levels also led to less glucose consumption and a lower cell density, indicating a metabolic burden associated with higher expression of the three enzymes. However, in the case of lower glucose consumption rate, a higher fraction of glucose flux was directed towards glucaric acid production versus biomass. I also observed that poorer aeration, resulting from doubling the total culture volume from 50 to 100 mL in 250-mL baffled flasks, led to a decrease in the glucaric acid titer by half, while growth was not affected (data not shown). This reduced titer is likely attributed to the fact that MIOX, the enzyme for the limiting step, uses molecular oxygen as a co-substrate (Charalampous, 1959; Xing et al., 2006c). Finally, production of glucaric acid was tested in M9 minimal medium; however, a negligible amount of glucaric acid was produced.

4.4 Discussion

We have assembled a biosynthetic pathway for the production of glucaric acid using enzymes from three disparate sources: Ino1 from *S. cerevisiae*, MIOX from mouse, and Udh from *P. syringae*. An endogenous phosphatase also participates in the pathway. The *suh*B gene product of *E. coli* has been shown to possess inositol monophosphatase activity *in vitro* and is therefore a reasonable candidate for this endogenous activity (Matsuhisa et al., 1995). This pathway is attractive from a

thermodynamics perspective, since the standard free energy changes (ΔG) for all three steps, as estimated by group contribution theory (Li et al., 2004; Mavrovouniotis, 1991) and considering molecular oxygen as the ultimate oxidant, are all negative: -14.3 Kcal/mol for the glucose to *myo*-inositol step; -86.8 Kcal/mol for the *myo*-inositol to glucuronic acid step; -55.9 Kcal/mol for the glucuronic to glucaric acid step. However, as Khosla and Keasling have indicated (Khosla and Keasling, 2003), metabolic engineering is more than simply recruiting various enzymes. It also involves global optimization of metabolic flux when perturbations such as the introduction of new pathways into a host organism are made. Issues of metabolic burden associated with the maintenance of plasmids and expression of plasmid-encoded genes are of particular interest in this case (Bentley et al., 1990; Birnbaum and Bailey, 1991; Jones et al., 2000). In our system, a detectable amount of glucuronic acid was produced *in vivo* only by high-copy number plasmids. Glucose-6-phosphate, the first substrate, should not be limiting for central metabolism because LB medium supplemented with excess glucose was used for growth. Therefore, it appears that high expression levels of the recombinant genes are needed in order to compete with the fast and robust glycolysis pathway and to divert glucose-6-phosphate towards glucuronic acid. The result that only small amounts of *myo*-inositol and no detectable amount of the organic acids was produced in M9 medium implies that when glucose is the sole carbon and energy source, almost all of the substrate enters endogenous cellular metabolism. This competition may also explain why the yield of glucaric acid on glucose during the first two days of the process, when glucose concentration is higher in the medium, is generally higher than that of the later days when the concentration is lower (data not shown). The requirement for *myo*-inositol to achieve

high MIOX activity suggests that low productivity from the Ino1 enzyme may ultimately be the limitation towards formation of the organic acids in M9 medium. Alternatively, MIOX may be poorly expressed in minimal medium. It should be noted that previous studies with Ino1 have resulted in high levels of *myo*-inositol production in an alternative chemically-defined medium and also employing a high-copy number plasmid for gene expression; however, these experiments were conducted in larger-scale, fed-batch fermentations for several days (Hansen et al., 1999). During the initial batch period prior to the onset of glucose feeding (approximately 10 hours), the *myo*-inositol concentration was less than 1 g/L. Thus, it is worth exploring the extent to which cultivation under fed-batch conditions could improve the productivity of our system.

Plasmid copy number is not the only factor related to expression level that affects the performance of our synthetic system. As shown in Table 4-2, increasing the inducer concentration to increase expression resulted in lower product concentration. IPTG concentrations below 0.05 mM did not improve glucaric acid production even though glucose consumption rate and growth rate were enhanced due to the reduced metabolic burden (data not shown). *E. coli* growth is better at 37°C than at 30°C and the activity of the rate-limiting enzyme MIOX should be higher at 37°C. However, fermentation was performed at 30°C because Ino1 was only functionally expressed at this lower temperature. Considering the reported unusual thermal instability of Udh (Bateman et al., 1970; Wagner and Hollmann, 1976), a temperature lower than 30°C may be better for its activity; however, I observed that the Udh activity at 30°C was much higher than that

of either Ino1 or MIOX (Fig. 4-3) and selected 30°C as the culture temperature to maximize the functional expression of Ino1.

In considering overall limitations on productivity of this system, potential inhibition by intermediates in the pathway should be examined. MIOX from hog kidney was reported to be inhibited *in vitro* by D-glucaric acid but not by D-glucuronate and D-glucuronolactone (Reddy et al., 1981a; Reddy et al., 1981b). Given that MIOX activity dropped sharply at the stationary phase even in the absence of D-glucaric acid (Table 4-1), low MIOX activity is more likely due to its intrinsic instability than inhibition by intermediates (Arner et al., 2004). It should also be noted that I did not overexpress the *suh*B gene or a homologous phosphatase. However, no *myo*-inositol-1-phosphate was detected among the culture products, while *myo*-inositol did accumulate. Therefore, I conclude that the phosphatase activity is not limiting flux through the pathway. *E. coli* also contains the D-glucarate catabolic pathway (Hubbard et al., 1998). Indeed, the ability of *E. coli* to consume D-glucarate as the sole carbon source for growth was used to develop a screen to identify uronate dehydrogenase activity (Yoon et al., 2009). BL21(DE3) can also metabolize D-glucuronic acid. However, the consumption of both organic acids appears to be subject to catabolite repression, preventing the undesirable loss of products in the presence of glucose (data not shown). The theoretical limit of D-glucaric acid titer therefore seems to be determined by the toxicity of the acids and the kinetics of each step. We have observed that *E. coli* growth and glucose consumption are not affected by the addition of potassium glucarate and sodium glucuronate at concentrations as high as 10 g/L (data not shown); thus, there is room for improvement of

titers by focusing on improving the kinetics of the rate-limiting steps. We have been investigating further optimization for enhancing glucose flux to our synthetic pathway by recruiting better enzymes from different sources and down-regulating the competing pathways. These approaches have failed to improve the productivity. Considering that the MIOX step is rate-limiting and that the activity assay method for MIOX by orcinol has specificity issues, I have developed a specific and simple MIOX assay, which is the topic of Chapter 5.

Table 4-1. Activity of recombinant MIOX expressed from high-copy pTrc-MIOX in *E. coli* under various culture conditions. Cultures were grown at 37°C in LB medium and induced with 1.0 mM IPTG. Glucuronic acid was measured at 24 hr. Supplements: MI = *myo*-inositol (60 mM, 10.8 g/L), Fe = $Fe(NH_4)_2(SO_4)_2$ (1 mM), Cys = L-cysteine (2 mM). N/D = not detectable. N/A = not measured.

Culture Conditions	Activity at 6 hr (nmol/min/mg)	Activity at 24 hr (nmol/min/mg)	Glucuronic Acid (g/L)
pTrc99A control	N/D	82	N/D
+ MI	430	76	0.44
+ MI, + Fe, + Cys	180	42	0.33
- MI	28	15	N/A

Table 4-2. Production of glucaric acid in BL21(DE3)(pRSFD-IN-MI, pTrc-udh) after 3 days culture. Cultures were grown at 30°C in LB medium supplemented with 10 g/L glucose and induced with IPTG. Data are the average and standard deviation of three independent experiments. OD_{600} = optical density at 600 nm, Yield (%) = 100 x glucaric acid produced / glucose consumed (mol/mol). Condition A = 0.1 mM IPTG at 0 hr; Condition B = 0.05 mM IPTG at 0 hr; Condition C = 0.05 mM IPTG at 0 hr and 0.1 mM IPTG at 17.5 hr. N/D = not detectable.

Condition	OD_{600}	Glucose (g/L)	*myo*-Inositol (g/L)	Glucuronic Acid (g/L)	Glucaric Acid (g/L)	Yield (%)
A	5.10±0.27	5.69±0.85	0.10±0.02	N/D	0.72±0.09	17.4±5.1
B	6.13±0.31	1.43±0.81	0.18±0.05	N/D	1.13±0.17	13.1±1.0
C	5.80±0.39	2.47±1.00	0.23±0.07	N/D	0.82±0.06	11.0±2.4

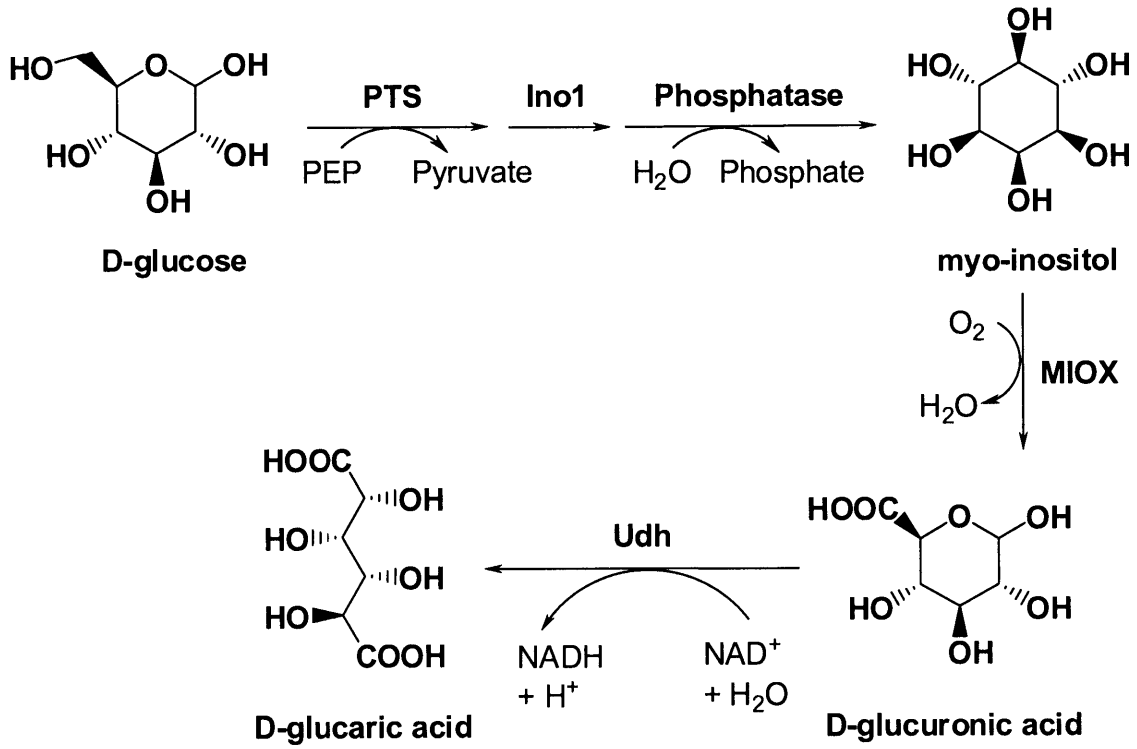


Fig. 4-1. Designed pathway for the production of glucaric acid in *E. coli*. PTS = phosphoenolpyruvate-dependent phosphotransferase system; Ino1 (MIPS) = *myo*-inositol 1-phosphate synthase from *Saccharomyces cerevisiae*; Phosphatase = SuhB, endogenous *E. coli* enzyme (Matsuhisa et al., 1995); MIOX = mouse version of *myo*-inositol oxygenase with codon optimization; Udh = uronate dehydrogenase from *Pseudomonas syringae*; PEP = phosphoenolpyruvate.

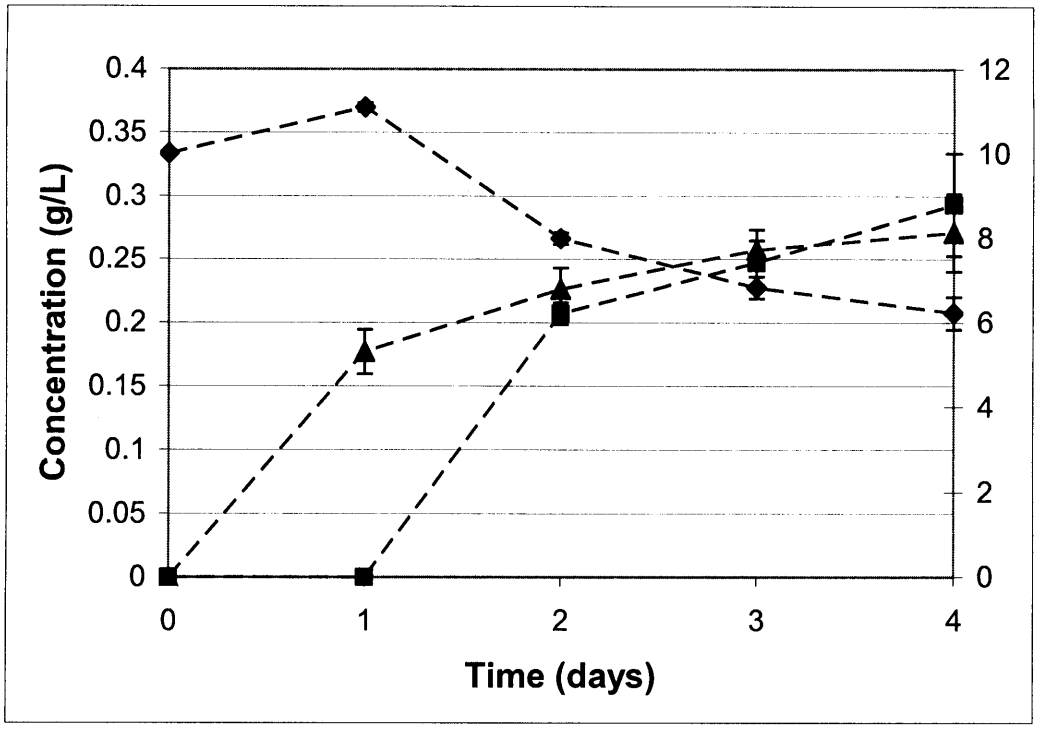


Fig. 4-2. Production of glucuronic acid in BL21(DE3)(pRSFD-IN-MI). Cultures were grown in triplicate at 30°C in LB medium supplemented with 10 g/L glucose and 0.1 mM IPTG. Data points are the average and standard deviation of the three biological replicates. ▲ = Glucuronic acid (left axis); ■ = *myo*-inositol (left axis); ♦ = Glucose (right axis).

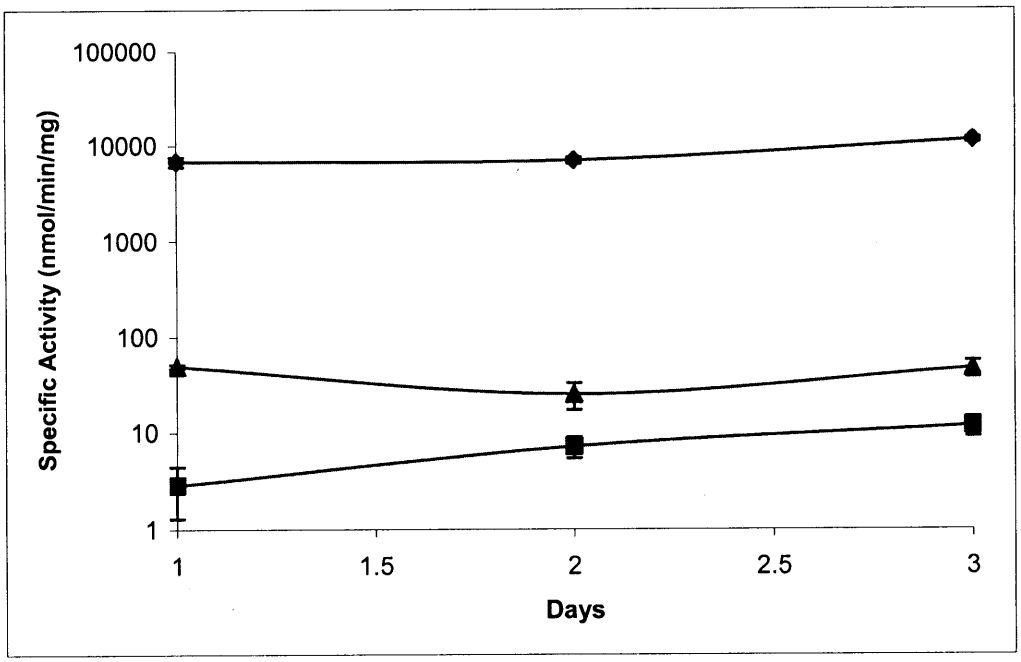


Fig. 4-3. *In vitro* activity of recombinant Ino1, MIOX, and Udh expressed in BL21(DE3) harboring the three genes. Cultures were grown at 30°C in LB medium supplemented with 10 g/L glucose and induced with 0.05 mM IPTG. MIOX activity is presented as net activity to account for background. Data are the average and standard deviation of three replicates. ▲ = Ino1; ■ = MIOX; ♦ = Udh.

Chapter 5

Pathway Analysis – Development of New Assay Methods for Glucuronate and *myo*-inositol oxygenase

TS Moon, S-H Yoon, M-J Tsang Mui Ching, A Lanza and KJ Prather. **Enzymatic assay of D-glucuronate using uronate dehydrogenase**, *Anal. Biochem.* 392, 183-185 (2009)

Abstract

D-Glucuronate is a key metabolite in the process of detoxification of xenobiotics and in a recently constructed synthetic pathway to produce D-glucaric acid, a "top-value added chemical" from biomass. A simple and specific assay of D-glucuronate would be useful for studying these processes, but existing assays are either time-consuming or non-specific. Using uronate dehydrogenase cloned from *Agrobacterium tumefaciens* and the oxidation reaction of D-glucuronate with concomitant reduction of NAD^+ cofactor, we developed an assay for D-glucuronate with a detection limit of 5 μM. This method was shown to be more suitable for a system with many interfering compounds than previous methods and was also applied to assays for *myo*-inositol oxygenase activity.

Keywords: enzymatic assay; uronate dehydrogenase; D-glucuronate; *myo*-inositol oxygenase

5.1 Introduction

D-Glucuronate is a metabolite in the synthesis of vitamin C (Sarma and Sastry, 1957) and plays an important role in glucuronidation, a conjugation reaction of D-glucuronate with xenobiotics and endobiotics, allowing for their detoxification and elimination (Bock and Kohle, 2005). D-Glucuronate has also been known to be a conversion product of inositol, a constituent of membrane phospholipids and a precursor for cell signaling molecules in mammals (Charalampous and Lyras, 1957), and to be converted to D-glucaric acid via a three-step route in mammals or a uronate dehydrogenase-catalyzed step in bacteria (Wagner and Hollmann, 1976). Recently, we constructed a synthetic pathway with D-glucuronate as an intermediate to produce D-glucaric acid from D-glucose and have been interested in developing a simple and specific assay of D-glucuronate (Moon et al., 2009a). More importantly, the activity of MIOX, the rate-limiting step enzyme should be measured accurately to diagnose the system for improvement, but the orcinol-based MIOX assay reveals specificity issues (see Chapter 4). In addition, D-glucuronate units occur in several classes of glycosaminoglycans (Haerry et al., 1997), and assays for the D-glucuronate units have been the standard methods for estimating the polysaccharides containing D-glucuronate (Cesaretti et al., 2003). This chapter describes newly developed assay methods for D-glucuronate and MIOX activity, based on the use of uronate dehydrogenase cloned from *Agrobacterium tumefaciens* (Chapter 3) (Yoon et al., 2009).

5.2 Results

Colorimetric assays using orcinol (Charalampous and Lyras, 1957) or carbazole (Bitter and Muir, 1962) have been widely used to quantify D-glucuronate, but these methods lack specificity. Pentoses such as 2-deoxy-D-ribose and hexoses such as D-glucose also react with these reagents (Linster and Van Schaftingen, 2004). The linearity of the orcinol method was assessed using D-glucuronate standard solutions (Alfa Aesar, Ward Hill, MA) and measuring absorbance at 670 nm. A fairly good linear relationship ($y=1.833x+0.007$, $R^2=0.993$) was obtained with a detection limit of 20 μM D-glucuronate, which corresponds to an absorbance change of 0.04 units. However, as expected, D-glucose, a common hexose component in culture media, also gave a strong signal ($y=0.627x-0.012$, $R^2=0.991$). In addition, both colorimetric assays require harsh conditions including the use of concentrated acids with boiling steps. An HPLC assay has been developed (Poon et al., 1993), but this approach is usually time consuming and requires the appropriate instrumentation. Moreover, pretreatment of samples with boronic acid affinity gel was reported to be required to separate D-glucuronate from D-glucarate because these two compounds had almost the same retention time (Moon et al., 2009a; Poon et al., 1993). A new assay based on two purified enzymes, uronate isomerase and mannonate dehydrogenase, and oxidation of NADH has also been reported as an alternative (Linster and Van Schaftingen, 2004). This method is quite sensitive and specific to D-glucuronate, with reduced activity on D-galacturonate. However, a significant amount of work to purify the two enzymes is required.

We have recently cloned, purified, and characterized uronate dehydrogenase (Udh) from *Agrobacterium tumefaciens* str. C58 using a commercially-available protein

purification kit (Pro-Bond purification system, Invitrogen Corporation, Carlsbad, CA) (Yoon et al., 2009). Udh is specific to D-glucuronate and D-galacturonate but does not accept aldohexoses, aldopentoses, polyols, or other uronates as substrates (Chang and Feingold, 1969; Yoon et al., 2009). This enzyme was found to have a high rate constant (k_{cat} = 1.9 x 10^2 s^{-1} on D-glucuronate) with a low K_m (0.37 mM). Using this newly cloned Udh, we constructed a specific and simple method to assay for D-glucuronate. The linearity of our newly developed enzymatic assay was assessed using D-glucuronate standard solutions similar to those used to examine the orcinol assay (Fig. 5-1). In this method, NAD^+ and purified Udh were added to the sample, NADH generation was monitored at 340 nm, and the D-glucuronate concentration was calculated using the extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ for NADH. As shown in Fig. 5-1, a comparable linear relationship was observed, with a detection limit of 5 μM, corresponding to an absorbance change of 0.03 units. The linear range was 5 to 320 μM D-glucuronate, in which the upper limit was determined not by the activity of Udh but by the reliable measurement of the spectrophotometer (maximum reliable absorbance unit of 2). The corresponding calibration curve (absorbance at 340 nm vs. D-glucuronate concentration) has a slope of 6.26, which is 3.4 times higher than that of the calibration curve obtained by the orcinol assay, indicating the increased sensitivity of the enzymatic assay. The plot of NADH produced versus D-glucuronate supplied as substrate yields a slope of 1.0 and an intercept of 0.0 (Fig. 5-1), indicating that the NADH concentration as determined by absorbance readings at 340 nm can be used to directly determine the D-glucuronate concentration.

The effect of culture broth from two different *E. coli* strains containing D-glucose on the enzymatic assay was also investigated. A standard solution of D-glucuronate (160 μM) was added to distilled water or samples of culture supernatant after centrifugation to see whether medium components or cellular metabolites interfere with the enzymatic assay. Almost 100% recovery was achieved, showing no interference by the culture broth of the two typical laboratory *E. coli* strains (D-glucuronate measured (μM, mean±SD): 161±2 for distilled water; 163±2 for DH10B; 162±2 for BL21 Star™ (DE3)). HPLC analysis indicated that the assay samples contained ~0.4 mM D-glucose, which would result in at least 180% recovery with the orcinol assay as calculated from the calibration curves (y=1.833x+0.007 for D-glucuronate; y=0.627x-0.012 for D-glucose).

We applied this enzymatic assay to a system (Moon et al., 2009a) in which D-glucuronate is produced from D-glucose by *E. coli* BL21 Star™ (DE3) containing two recombinant genes encoding *myo*-inositol-1-phosphate synthase and *myo*-inositol oxygenase (BL21 Star™ (DE3)(pRSFD-IN-MI)). BL21 Star™ (DE3)(pRSFDuet-1) containing empty plasmid pRSFDuet-1 was used as the negative control. Cultures were grown at 30°C for 24 hrs in LB medium supplemented with 10 g/L D-glucose and induced with 0.5 mM IPTG. Culture supernatant was collected after centrifugation, and the D-glucuronate concentration was determined by either the enzymatic method or HPLC analysis as described previously (Moon et al., 2009a). Three enzymatic or HPLC measurements gave D-glucuronate concentrations of 791±13 or 804±32 μM, respectively, from the culture of BL21 Star™ (DE3)(pRSFD-IN-MI). A t-test determined that the difference in the average values is not statistically significant (95% confidence level). As

expected, no D-glucuronate was detected from the negative control although ~0.6 mM D-glucose was present in the assay samples, confirming that the enzymatic assay is not affected by this component.

We applied the enzymatic assay to the determination of *myo*-inositol oxygenase (MIOX) activity. Assays for MIOX activity have been performed using lysates or purified enzymes and *myo*-inositol, and the produced D-glucuronate has been determined by the orcinol method (Arner et al., 2001; Moon et al., 2009a; Reddy et al., 1981a). Once the MIOX enzyme is purified, there would be no interfering compounds in the sample, and the orcinol method could be used reliably to determine the MIOX activity. However, MIOX assays using lysates are sometimes preferred because sample preparation does not require extensive enzyme purification steps. Enzyme activities measured in lysates might also estimate more accurately the *in vivo* activity compared to that of the purified enzyme. As discussed above, the orcinol method results in a higher level of background signal because many interfering compounds (e.g., pentoses and hexoses) are likely to exist in the lysates. To study the feasibility of our enzymatic method for this application, we determined the conversion rates of *myo*-inositol by MIOX with different lysate amounts by enzymatically measuring the D-glucuronate produced (Fig. 5-2). A good linearity was observed especially when more than 0.1 mL of the lysate was used (R^2 = 0.9805 with only the last four data points). We also performed experiments to determine the specific activity of MIOX using the orcinol method (Arner et al., 2001; Reddy et al., 1981a) or the enzymatic method. The total protein concentration of lysates was determined using the Bradford method (Bradford, 1976), and six measurements gave

specific activities of 7.2±1.5 nmol/min/mg or 5.4±0.9 nmol/min/mg, respectively. Net activity is reported for the orcinol assay, in which the background signal measured in the absence of substrate (20 to 80% of the total signal in these samples) has been subtracted. While the difference in activities obtained using these two different methods was found to be statistically significant (95% confidence level of a *t*-test), the high and variable background of the orcinol assay makes it less reliable. The difference in standard deviation between the two data sets was found to be insignificant (at the 95% confidence level).

5.3 Discussion

Our interest in studying a synthetic pathway to produce D-glucuronic and D-glucaric acids (Moon et al., 2009a) triggered the development of a simple and specific assay for D-glucuronate using Udh. The MIOX enzyme step of the biosynthetic pathway was found to be rate limiting, implying that improving its activity and measuring it accurately is important for enhancing the productivity of these acids. We are now applying these enzymatic assays to our systems to determine D-glucuronate titer and MIOX activity. This application will be further demonstrated in Chapter 6. Our methods might be also useful to study other systems, including the biosynthetic pathway for vitamin C, glucuronidation processes, and specific quantification of glycosaminoglycans containing D-glucuronate units in the mixture of polysaccharides. D-Galacturonate is also a substrate of Udh (k_{cat} = 9.2 x 10 s^{-1}; K_m = 0.16 mM) and its presence would interfere with D-glucuronate measurement (Yoon et al., 2009). However, D-

galacturonate is absent in both our system and many others of interest (Linster and Van Schaftingen, 2004; Moon et al., 2009a). The current assay could also prove equally useful as a specific assay for D-galacturonate. The stability in solution (stable at 4°C for more than six months so far) and high activity of Udh as well as its relatively easy purification method provides researchers with an alternative method to study D-glucuronate-related metabolism.

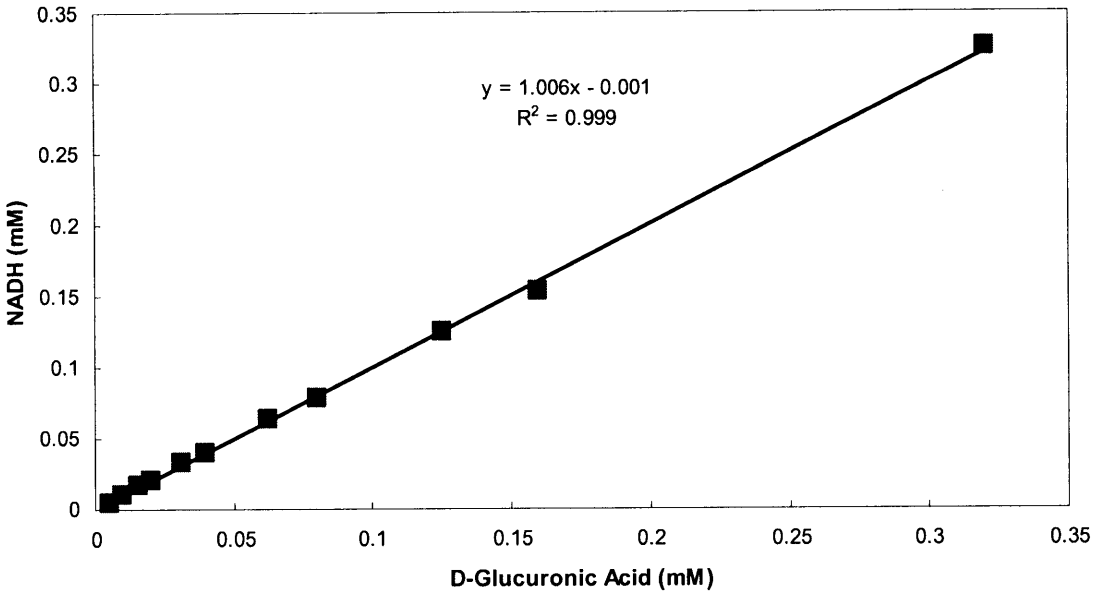


Fig. 5-1. Calibration curves for D-glucuronate by the enzymatic method using Udh. The reaction mixture (1ml) contained 0.8 mM NAD^+, 100 mM Tris-Cl (pH 8.0), 20 μl sample containing D-glucuronate, and 0.06 U (1U = 1 μmol/min) Udh prepared as described previously (Yoon et al., 2009) and was incubated at room temperature for 30 min. The absorbance increase was measured at 340 nm using a Beckman DU-800 spectrophotometer (Beckman Coulter, Fullerton, CA), and the D-glucuronate concentration was calculated using the extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ for NADH.

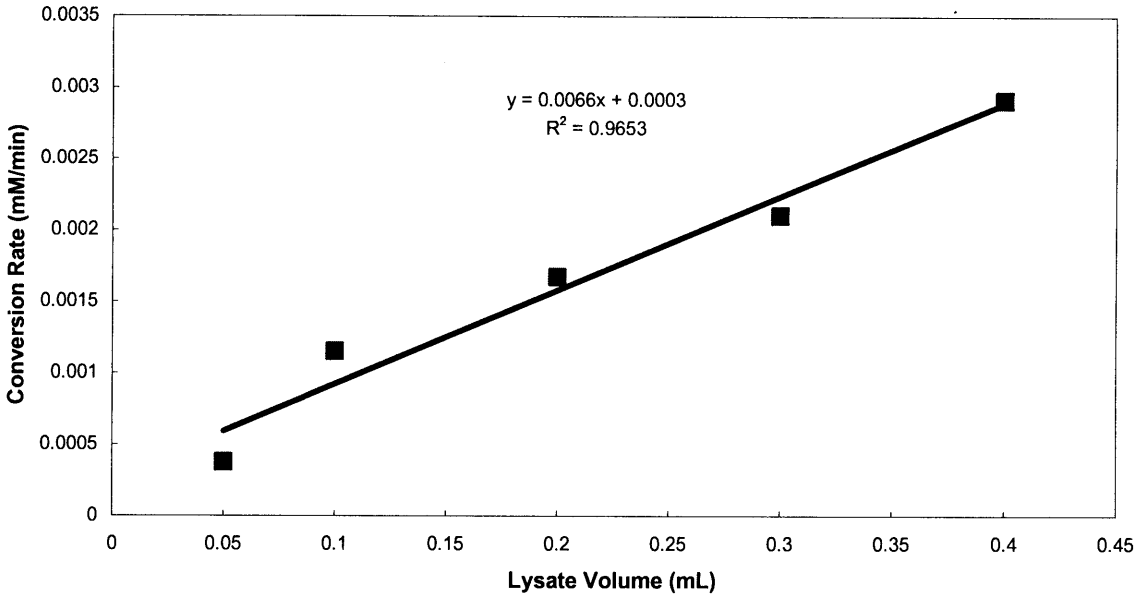


Fig. 5-2. Enzymatic assay of MIOX using Udh. MIOX was expressed from pTrc-MIOX (Moon et al., 2009a) in *E. coli* DH10B grown in LB medium with 0.1 mM IPTG at 30°C for 24 hrs. Lysates were prepared by resuspending cell pellets from 50 mL culture in 8 mL sodium phosphate buffer (100 mM, pH 7) with 1 mg/mL lysozyme. EDTA-free protease inhibitor cocktail tablets (Roche Applied Science, Indianapolis, IN) were added to the resuspension buffer according to the manufacturer's instructions. Cell solutions were lysed by sonication and the resulting solutions were centrifuged at 14,000 rpm at 4°C for 15 min to remove insolubles. The reaction buffer consisted of 77 mM sodium phosphate buffer (pH 7), 2 mM L-cysteine, 1 mM $FeSO_4$, and 60 mM *myo*-inositol. Samples were pre-incubated without substrate for 5 minutes at 30°C to activate the MIOX enzyme. Reactions were incubated for 1 hr at 30°C, and the D-glucuronate produced was quantified using Udh and 4 mM NAD^+.

Chapter 6

Pathway Improvement – Devices Engineering and Application to Synthetic Glucaric Acid Pathway

JE Dueber, GC Wu, GR Malmirchegini, TS Moon, CJ Petzold, AV Ullal, KJ Prather and JD Keasling. **Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux**, *Nat. Biotechnol.* 27, 753-759 (2009)

TS Moon, JE Dueber and KJ Prather. **Use of Modular, Synthetic Scaffolds for Improved Production of Glucaric Acid in Engineered *E. coli***, submitted to *Metab Eng.* (2009)

Abstract

The field of metabolic engineering has the potential to produce a wide variety of chemicals in both an inexpensive and ecologically-friendly manner. Heterologous expression of novel combinations of enzymes promises to provide new or improved synthetic routes towards a substantially increased diversity of small molecules. Recently, we constructed a synthetic pathway to produce D-glucaric acid, a molecule that has been deemed a "top-value added chemical" from biomass, starting from glucose. Limiting flux through the pathway is the second recombinant step, catalyzed by *myo*-inositol oxygenase (MIOX), whose activity is strongly influenced by the concentration of the *myo*-inositol substrate. To synthetically increase the effective concentration of *myo*-inositol, polypeptide scaffolds were built from protein-protein interaction domains to co-localize all three pathway enzymes in a designable complex as previously described (Dueber et al., 2009). Glucaric acid titer was found to be strongly affected by the number of scaffold interaction domains targeting upstream Ino1 enzymes, whereas the effect of increased numbers of MIOX-targeted domains was much less significant. We determined that the scaffolds directly increased the specific MIOX activity and that glucaric acid titers were strongly correlated with MIOX activity. Overall, we observed an approximately 5-fold improvement in product titers. These results further validate the utility of these synthetic scaffolds as a tool for metabolic engineering.

Keywords: glucaric acid; synthetic biology; metabolic pathway engineering; scaffold; modularity; colocalization.

6.1 Introduction

Synthetic biology is an evolving field involving the creation of new biological components and systems, such as enzymes, signaling molecules, and metabolic pathways (Benner and Sismour, 2005; Keasling, 2008; Leonard et al., 2008). Synthetic biologists seek to design and characterize interchangeable parts from which one can build devices and systems that can both help to understand natural biological systems and facilitate the creation of new biological "machines." Achievements in the field include rewiring signaling pathways (Park et al., 2003) as well as the development of microbes that can synthesize bulk chemicals (Nakamura and Whited, 2003), fuels (Atsumi et al., 2008), and drugs (Martin et al., 2003; Ro et al., 2006). In the latter examples, synthetic biology intersects directly with metabolic engineering in using enzymes as interchangeable parts for the construction or re-constitution of metabolic pathways. These pathways can be naturally existing, recruited from a heterologous organism, or they may be formed from novel combinations of enzymes to produce both natural compounds and products not yet observed in nature (Prather and Martin, 2008). Metabolic engineering has traditionally focused on the improvement of metabolic pathways for increased productivity. To this end, one focus of synthetic biology is to provide additional tools for producing high value compounds cheaply, efficiently, and cleanly (Arkin and Fletcher, 2006; Keasling, 2008; Tyo et al., 2007).

We recently constructed a synthetic pathway for the production of D-glucaric acid from D-glucose in *Escherichia coli* (Moon et al., 2009a). D-glucaric acid has been identified as a "top value-added chemical from biomass" (Werpy and Petersen, 2004),

and has been studied for therapeutic purposes including cholesterol reduction (Walaszek et al., 1996) and cancer chemotherapy (Singh and Gupta, 2003; Singh and Gupta, 2007). Its primary use is as a starting material for hydroxylated nylons (Werpy and Petersen, 2004). D-Glucaric acid is currently produced by chemical oxidation of glucose, a nonselective and expensive process using nitric acid as the oxidant. There is a known route for the production of D-glucaric acid from D-glucose in mammals; however, this is a lengthy pathway, consisting of more than ten conversion steps. Our synthetic pathway was assembled by recruiting enzyme activities from disparate sources into *Escherichia coli* (Moon et al., 2009a). Co-expression of the genes encoding *myo*-inositol-1-phosphate synthase (Ino1) from *Saccharomyces cerevisiae*, *myo*-inositol oxygenase (MIOX) from *Mus musculus* (mouse), and uronate dehydrogenase (Udh) from *Pseudomonas syringae* led to production of D-glucaric acid at titers of ~1 g/L. We next aimed to improve this level of productivity.

Other examples of the design and construction of synthetic pathways from the combination of heterologous enzymes have been recently reported (Atsumi et al., 2008; Martin et al., 2003; Nakamura and Whited, 2003; Niu et al., 2003); however, the main focus of metabolic engineering has been global optimization of metabolic flux (Stephanopoulos and Jensen, 2005). To this end, various approaches have been successfully implemented, including modulation of enzyme expression by varying the strengths of promoters and ribosome binding sites, control of mRNA processing by introducing tunable intergenic regions, and improvement of rate-limiting enzymes by directed evolution (Alper et al., 2005; Bloom et al., 2005; Pfleger et al., 2006; Pitera et

al., 2007; Stephanopoulos, 1999). Recently, an orthogonal, but compatible method for improving pathway efficiency was described (Dueber et al., 2009). In this method, pathway enzymes were colocalized using synthetic scaffolds built from protein-protein interaction domains that specifically bound corresponding ligands fused to the metabolic enzymes. By taking advantage of the modularity of these interaction domains, scaffold architectures were optimized to achieve a 77-fold improvement of mevalonate production at low expression levels of pathway enzymes (Dueber et al., 2009). In this same report, we were able to demonstrate a three-fold improvement in D-glucaric acid titer by co-localizing Ino1 and MIOX in a 1:1 ratio (Table 6-1), although the baseline of 0.6 g/L in the absence of scaffolding was somewhat lower than the titers previously achieved with different expression machinery (Moon et al., 2009a). Here we take advantage of the modular scaffold design to control enzyme stoichiometry at the synthetic complex in a targeted manner for further titer improvements.

Our interest in scaffolding the glucaric acid pathway is based on two prior observations. First, we observed that the activity of MIOX was lowest of the three enzymes in the recombinant system, more than two orders of magnitude lower than that of the most active enzyme (Udh) (Moon et al., 2009a). Second, we confirmed previous reports that high MIOX activity in *E. coli* is strongly influenced by exposure to high concentrations of *myo*-inositol, its substrate (Arner et al., 2004; Moon et al., 2009a). Based on these observations, we hypothesized that beyond merely reducing diffusion distance and transit time, recruitment of the pathway enzymes, particularly Ino1 and MIOX, to the synthetic scaffold could result in increased effective concentrations of the

myo-inositol substrate. This, in turn, could lead to increased MIOX activity and improved D-glucaric acid production. In this chapter, we report on further investigation of the effectiveness of these modular scaffolds to improve D-glucaric acid titers. We first examined, more fully, the effects of recruiting only Ino1 and MIOX to the scaffold that were previously reported (Dueber et al., 2009) in order to determine whether any impact on MIOX activity was observed. We next created synthetic scaffolds to co-localize all three enzymes on constructs that allowed the independent manipulation of scaffold and enzyme concentration. Finally, we varied the number of interaction domains targeting Ino1 and MIOX to modulate the effective concentration of *myo*-inositol at the synthetic complex and to improve glucaric acid titers.

6.2 Materials and Methods

6.2.1. Escherichia coli strains, plasmids and scaffold construction

Escherichia coli strains and plasmids used in this study are listed in Table 6-2. All molecular biology manipulations were carried out according to standard practices (Sambrook and Russell, 2001).

Scaffold devices consisting of GBD, SH3 and PDZ protein interaction domains (Dueber et al., 2009) were assembled using the BglBrick strategy with BamHI and BglII cohesive ends compatible for ligation (doi: 1721.1/46747). Basic parts were made such that they would be flanked on the 5' end by a BglII site and on the 3' end by BamHI and XhoI sites. Composite parts were then constructed by digesting the backbone vector with

BamHI and XhoI, and a 3' part was added as a BglII/XhoI-digested insert. The resultant parts could then be sub-cloned into the pWW306 and pWW308 expression plasmids (provided by Weston Whitaker of the Dueber Lab) carrying either a tetracycline-inducible (P_{TET}) or a lactose-inducible promoter (P_{lac}), respectively, upstream of a BglII/XhoI multi-cloning site. Both expression plasmids were constructed as a modification of pSB1A2, a plasmid obtained from the MIT Registry of Standard Biological Parts (http://partsregistry.org/Main_Page).

6.2.2 Synthesis of degenerate versions of SH3 domain

The sequence of the SH3 domain of mouse protein Crk (residues 134-191) was jumbled, optimized for *E. coli*, and diagnostic restriction enzyme sites incorporated using the online tool *Gene Design* at http://baderlab.bme.jhu.edu/gd/. This was done iteratively with some codon changes made by eye with alignment analysis in an attempt to maximize degeneracy of the five additional coding versions of the Crk SH3 domain. *Gene Design* returned oligonucleotide sequences that could be PCR assembled to produce the desired SH3-encoding product. These were made into basic BglBrick parts as described above.

6.2.3. Culture and analysis conditions for D-glucaric acid production

Cultures were grown in LB medium supplemented with 10 g/L D-glucose and induced at the exponential phase as indicated in the Results. An inoculum was prepared in LB medium, and 1 % (v/v) was used to inoculate 250-mL baffled flasks containing 50 mL of medium. The cultures were incubated at 30°C and 250 rpm for 2 days, and then

D-glucaric acid titer was analyzed using HPLC as described previously (Moon et al., 2009a).

6.2.4. Assay for MIOX activity

Assays for MIOX activity were performed using lysates as described previously (Moon et al., 2009b). Briefly, lysates were prepared by suspending cell pellets in sodium phosphate buffer (50 mM, pH 8.0) with 1 mg/mL lysozyme and EDTA-free protease inhibitor cocktail tablets (Roche Applied Science, Indianapolis, IN) and then by sonication, followed by centrifugation to remove insolubles. The total protein concentration of lysates was determined using the Bradford method (Bradford, 1976). The MIOX activity was measured by monitoring D-glucuronic acid produced when excess (60 mM) *myo*-inositol was incubated with lysates. For the determination of D-glucuronic acid produced, the reaction mixture also contained excess NAD^+ and purified Udh (Yoon et al., 2009), stoichiometrically generating NADH which can be determined by absorbance at 340 nm. Control reactions were established without *myo*-inositol to account for background.

6.3 Results

6.3.1. The effect of Ino1 and MIOX co-localization on D-glucaric acid titer and MIOX activity

Previous studies on the production of D-glucaric acid showed that the MIOX-catalyzed step is limiting and that MIOX activity is strongly influenced by the

concentration of *myo*-inositol, its substrate (Moon et al., 2009a). To improve this bottleneck step, we designed a simple scaffold that co-recruits Ino1 and MIOX in a 1:1 ratio (Table 6-1) (Dueber et al., 2009). The synthetic scaffold contains one SH3 domain and one PDZ domain that can interact with a C-terminal ligand fused to Ino1 and MIOX respectively. The genes encoding this scaffold and the two pathway enzymes were placed under *lac* promoter control, resulting in pJD706, a ColE1 origin-based plasmid, the *udh* gene was cloned into the IPTG-inducible plasmid pSTV28, resulting in pSTV28-ATudh, a p15A origin-based plasmid, and the production of D-glucaric acid was tested (Table 6-1). For simplicity, Udh was not designed to be co-recruited in this test system; Udh activity was more than 2 orders of magnitude higher than that of Ino1 and MIOX (Moon et al., 2009a) so that the Udh-catalyzed step was expected to proceed quickly even without colocalization. By co-localizing Ino1 and MIOX, we achieved 3-fold increase in D-glucaric acid titer (compared to the control without scaffold expression) at 0.2 mM IPTG concentration. At 0.4 mM IPTG, the titers from both MDP13 and MDP14 were lower than 0.4 g/L (data not shown), implying that metabolic burden effects (Bentley et al., 1990; Birnbaum and Bailey, 1991; Jones et al., 2000) might limit the production level. To determine whether this titer improvement was accompanied by changes in the most limiting step, we measured MIOX activity in lysate samples. Activity in the scaffolded system was 19.0±0.9 nmol/min/mg, more than 25% higher than the scaffold-free system activity of 15.0±1.3 nmol/min/mg ($p=0.013$) (Table 6-1). Given this result, we inquired whether increased MIOX activities could be observed by changing the scaffold architecture.

6.3.2. Constructing various synthetic scaffold devices

The previous results suggested that MIOX activity was indeed enhanced by co-localization, but that system was limited by the inability to independently control scaffold and enzyme expression levels. Additionally, only one scaffold architecture of a 1:1 recruiting domain ratio was tested, which may not be optimal for maximizing the local *myo*-inositol concentration. Here, we took advantage of the modular design of the synthetic scaffolds to vary the number of interaction domains of Ino1 and MIOX for control over relative enzyme stoichiometry. In particular, our hypothesis predicts improved titers to be achievable for architectures with higher numbers of Ino1-recruiting Src homology 3 (SH3) domains on the scaffold that would produce higher concentrations of *myo*-inositol at the complex. The increased numbers of SH3 domains should result in increased MIOX activation with a concomitant increase in D-glucaric acid production. However, if increasing the number of MIOX-recruiting PSD95/DlgA/Zo-1 (PDZ) domains results in substantial titer improvement, the scaffold effect is more likely due to shorter transit time of substrates and higher effective concentrations of MIOX, balancing the overall flux. Of course, these two mechanisms are not mutually exclusive.

For independent expression control of the glucaric acid synthetic pathway and the scaffolds, we used orthogonal IPTG-inducible (P_{Lac}) and tetracycline-inducible (P_{TET}) promoters, respectively, to drive expression. We used a two plasmid expression platform: the enzymes for biosynthesis of glucaric acid were cloned into a p15A-based medium copy plasmid, and the scaffolds were cloned into a ColE1-based high copy plasmid (Table 6-2). Scaffolds were constructed by assembling three protein-protein

interaction domains linked together by nine-residue glycine-serine linkers predicted to be flexible and unstructured. The GTPase binding domain (GBD), SH3, and PDZ domains were used to target Udh, Ino1, and MIOX, respectively. As the first test, a matrix of nine constructs was designed: $GBD_aSH3_cPDZ_b$ where a=1 and b and c are 1, 2, or 4 (Table 6-2). Because Udh activity was more than 2 orders of magnitude higher than that of Ino1 and MIOX (Moon et al., 2009a), all the scaffolds were designed to contain only one GBD domain (a=1).

To determine the most favorable expression levels, D-glucaric acid production by $GBD_1SH3_4PDZ_4$ was tested at different aTc and IPTG concentrations (Fig. 6-1). The highest titers were observed with IPTG and aTc concentrations of 0.05 mM and 108 nM, respectively, which led to ~5-fold titer improvement (compared to scaffold-free control). It is well-known that over-expression of recombinant enzymes can have a deleterious effect on cell growth in an unpredictable manner and that, ultimately, expression levels that are too high actually lead to decreases in productivity (Bentley et al., 1990; Birnbaum and Bailey, 1991; Jones et al., 2000; Kane and Hartley, 1988; Moon et al., 2009a; Tyo et al., 2009). This metabolic burden effect may be the reason that productivity decreases with increasing IPTG concentration. However, an expression level that is too low might reduce the encounter frequency between scaffolds and enzymes, diluting the pathway enzymes on the scaffolds and sequestering enzymes from their substrates. The sequestering effect (Burack and Shaw, 2000; Levchenko et al., 2000) is especially probable when the scaffold expression level is high (215 nM aTc).

6.3.3. Testing various synthetic scaffold devices and demonstrating that increased MIOX activity plays an important role in titer improvement

Using the previously identified IPTG concentration of 0.05 mM, we tested the matrix of scaffold constructs JT1 to JT9 at several scaffold inducer concentrations (215, 108, and 54 nM aTc) (Fig. 6-2). The results indicate that D-glucaric acid titer primarily depends on the number of Ino1-recruiting SH3 domains, not on the number of MIOX-recruiting PDZ domains. This observation supports the hypothesis that enhancement of MIOX activation by the *myo*-inositol substrate is responsible for the majority of the scaffolding's beneficial effect, since recruiting additional Ino1 enzymes should increase the local *myo*-inositol concentration while leaving the number of localized MIOX enzymes unchanged. The highest titers were achieved at moderate scaffold induction levels, consistent with the expectation that too high of a scaffold to enzyme ratio would result in a small number of recruited enzymes per scaffold molecule, resulting in small titer improvement with increasing number of SH3 domains targeting Ino1.

Given that the number of PDZ domains targeting MIOX does not affect the titer considerably, we constructed additional scaffolds where the number of PDZ domains was held constant at two and the number of SH3 domains was varied up to 8 (Fig. 6-3). Because the probability of recombination increases as the number of identical repeat sequences is increased, we built this new class of scaffolds with degenerate coding sequences for the SH3 domain with consideration of *E. coli* preferred codon usage. To verify that it is the interaction activities of the domains and not the codon usage that determines the scaffolding effect, two different constructs for $GBD_1SH3_2PDZ_2$ and

$GBD_1SH3_4PDZ_2$ were compared: one with the same SH3 sequences (JT5 and JT8, respectively), and the other with the degenerate sequences (JTK4 and JTK6, respectively). There were no statistically significant differences in the titers observed between constructs with identical and degenerate nucleotide sequences (Table 6-3). Increasing the number of SH3 domains past four repeats was not accompanied by an increase in titer (Fig. 6-3). In fact, there was a decrease in titers at the higher concentrations of scaffold inducer (81 and 108 nM aTc). There are several potential explanations for this decrease: 1. Sequestering effects, where the enzymes bind separate scaffold molecules and leave many unoccupied sites, are expected to become more prevalent with an increasing number of SH3 domains. 2. The recruited enzymes (e.g., Ino1 and MIOX) may become oriented in less efficient ways with increasing domain number. 3. Longer scaffolds with higher domain numbers may exhibit reduced stability and/or solubility. Any sequestering effect could potentially be overcome by an increase in IPTG concentration to increase enzyme levels. To investigate this possibility, we tested D-glucaric acid production by $GBD_1SH3_cPDZ_2$ where c is 3, 6, and 8 at four different IPTG concentrations (Fig. 6-4). Higher IPTG concentrations did not lead to titer enhancement, suggesting that metabolic burden effects are affecting productivity. We should also note that in this system, the synthetic pathway is consuming glucose-6-phosphate, the entry molecule for carbon flux into endogenous metabolism. Thus, in addition to burden specifically associated with recombinant protein expression, negative effects on productivity may result from variations in the flux distribution at this node that result from competition between the glycolysis and pentose phosphate pathways and glucaric acid production.

For direct evidence that the scaffolds affect D-glucaric acid titer by improving MIOX activation, we measured MIOX activity from multiple systems with various scaffold architectures and compared the specific activities with the corresponding titers (Fig. 6-5). Given that (1) the D-glucaric acid titer is proportional to MIOX activity and (2) MIOX activity depends on the effective number of Ino1 molecules colocalized, we conclude that titer improvement by scaffolds in our systems is strongly affected by MIOX activation, most likely as the result of increased local *myo*-inositol concentration.

6.4 Discussion

Engineering metabolic pathways from novel combinations of enzymes promises to dramatically expand the potential number of products that can be made more cheaply and more "green" than with current chemical schemes. However, several challenges must be overcome. First, enzymes with the desired catalytic activity must be isolated or engineered with the required substrate specificity. For the glucaric acid pathway described in this study, three enzymes with desired activities and substrate specificity were isolated from three separate organism sources. Second, in order to achieve viable titers, enzymatic activities of most engineered pathways will have to be balanced. As a result, this has been an area of intense focus where enzyme expression levels are controlled or enzyme activities are improved via directed evolution (Alper et al., 2005; Bloom et al., 2005; Pfleger et al., 2006; Pitera et al., 2007; Stephanopoulos, 1999).

Impressive successes have been achieved for the *in vivo* production of multiple targets (Atsumi et al., 2008; Menzella et al., 2005; Nakamura and Whited, 2003; Ro et al., 2006); however, a great need exists for additional strategies that are orthogonal, but additive, to achieve further gains in production yields if metabolic engineering is to prove viable for a wider spectrum of compounds, especially bulk compounds. One such strategy may be the use of synthetic scaffolds to co-target metabolic enzymes to the same complex to increase the effective concentration of each pathway component. Recently, scaffolding was shown to be highly effective for improving titers of the mevalonate pathway while simultaneously lowering the enzyme expression levels required to achieve these titers (Dueber et al., 2009). As a result, the production cells grew considerably faster than the non-scaffolded pathway expressed to levels required to achieve comparable titers. In the present study, the modular architecture was utilized for the glucaric acid pathway to optimize the effective concentration of the intermediate *myo*-inositol since the bottleneck enzyme MIOX had been observed to be more active when produced in the presence of *myo*-inositol (Moon et al., 2009a).

It was reported that MIOX harbors a coupled dinuclear iron cluster which is perturbed by *myo*-inositol binding (Brown et al., 2006; Xing et al., 2006c). Interestingly, MIOX with the mixed-valent (II/III) diiron cluster is the catalytically active form, instead of MIOX (II/II) or MIOX (III/III) (Xing et al., 2006a). It was suggested that *myo*-inositol binding conditions the diiron cluster for activation of oxygen, the other substrate of MIOX (Xing et al., 2006a; Xing et al., 2006b; Xing et al., 2006c). The authors showed that in the presence of saturating *myo*-inositol and limiting oxygen, MIOX (II/II) with

myo-inositol bound (MIOX (II/II)•MI) was converted into MIOX (II/III)•MI as a stable product with a low yield of D-glucuronic acid; in contrast, MIOX (II/III) with *myo*-inositol bound (MIOX (II/III)•MI) reacted with limiting oxygen to stoichiometrically generate D-glucuronic acid with regeneration of MIOX (II/III)•MI. Given those findings and our results, we speculate that the increased local *myo*-inositol concentration near the scaffolded enzymes leads to an increase in the fraction of MIOX in the active (II/III) state, followed by activation of oxygen by MIOX (II/III)•MI and conversion of *myo*-inositol to D-glucuronic acid, with concomitant titer improvement (see the suggested mechanism scheme below). We also previously observed a dependence on oxidation state in non-scaffolded cultures, where D-glucaric acid titers were sensitive to aeration in shake flasks (Moon et al., 2009a).

MIOX (II/II) + MI → MIOX (II/II)•MI → MIOX (II/III)•MI → MIOX (II/III)•MI•O_2 → MIOX (II/III)•MI•activated O_2 → D-glucuronic acid + MIOX (II/III)•MI regenerated → Repeated reactions as long as MI concentation remains high enough

One obstacle we had to overcome was the construction of robust scaffolds despite high numbers of interaction domain repeats. The ultimate goal of synthetic biology is the creation of systems that behave in a programmable and robust manner like electronic circuits; however, the complexity of the cell makes this challenging. In our system, we have found it difficult to build constructs with more than four repeats of an identical gene sequence. We have found the use of degenerate codons encoding the desired protein sequence to be an effective strategy for constructing high numbers of repeats. In this

study, we constructed six distinct versions of the gene encoding the SH3 domain (including the natural mouse sequence) and used them in a random order to progressively build constructs as high as eight SH3 domain repeats without any noticeable recombination events.

We were unable to achieve as large an improvement in product titers as previously achieved for mevalonate biosynthesis (77-fold), a fact that was previously highlighted as a potential limitation of this system (DeLisa and Conrado, 2009). We point out several significant differences between the mevalonate system and the synthetic glucaric acid pathway. Firstly, the bottleneck enzymatic step in the mevalonate biosynthetic pathway results in an accumulation of a toxic intermediate (HMG-CoA). Thus, testing the matrix of scaffolds for improved titers was equivalent to screening for improved flux of this bottleneck step and reduction of toxicity to the *E. coli* production host. Unfortunately, these simultaneous effects can not be easily deconvoluted. For the glucaric acid production pathway, the bottleneck is MIOX activity that is improved with increased substrate concentrations, but there is no indication of intermediate toxicity. Consequently, the scaffolds that worked best were the ones predicted to increase the effective concentration of *myo*-inositol at the resultant complexes and not necessarily those that formed complexes with structures resulting in improved, balanced fluxes of each enzymatic step. Secondly, in the glucaric acid system, the increase in activity of the rate-limiting MIOX enzyme as the result of co-localization with Ino1 affects both the characteristic diffusion and characteristic reaction times, whereas co-recruitment of the enzymes in the mevalonate pathway should not affect the underlying kinetics. Again,

because of the nature of the enhancement, we can not deconvolute these two effects. Lastly, it should also be emphasized that we do not have the ability to predict the structural properties of the scaffolded complexes. These are likely to be complicated by the oligomerization state of the pathway enzymes. For example, a tetrameric enzyme will contain four scaffold-recruiting peptide ligands. Thus, the final scaffolded complex could be quite large and the optimal mevalonate biosynthetic complex is likely to differ considerably from the optimal glucaric acid biosynthetic complex.

In this study, we were able to treat metabolic enzymes as modular parts that can be combined in novel combinations to produce the high-value product glucaric acid. Similarly, we were able to take advantage of the functional and physical modularity of the metazoan protein-protein interaction domains to build synthetic scaffolds for the directed purpose of forming a synthetic complex with increased local concentrations of intermediate *myo*-inositol. Further, we were able to change the relative product titers depending on the number of enzymes producing *myo*-inositol expected to be recruited to the complex, consistent with a titration in this local metabolite concentration. In this manner, we used modularity as an engineering strategy similar to how it has facilitated evolution of new signaling connections in living cells (Pawson and Nash, 2003). Thus, these modular strategies should prove to be generalizable towards other engineered pathways for improvement of production yields in addition to the gains made by conventional strategies.

Table 6-1. Improved production of D-glucaric acid and increased MIOX activity from the clone MDP14 containing the two enzyme-recruiting scaffold. The gene encoding this scaffold and the two pathway enzymes (Ino1 and MIOX) were placed under *lac* promoter control, resulting in pJD706, a ColE1 origin-based plasmid. The *udh* gene was cloned into the IPTG-inducible plasmid pSTV28, resulting in pSTV28-ATudh, a p15A origin-based plasmid. MDP13 is the scaffold-free control containing pSTV28-ATudh and pJD705 (containing only the genes encoding Ino1 and MIOX, but no scaffold gene). At 0.05 mM IPTG, the improvement was not significant, but 3-fold improvement in titer and more than 25% increase in MIOX activity were observed at 0.2 mM IPTG. Data are the averages and standard deviations of three replicates.

Strains	D-Glucaric acid (g/L)		Specific Activity of MIOX (nmol/min/mg)	
	IPTG (mM)		IPTG (mM)	
	0.05	0.2	0.05	0.2
MDP13	0.44±0.05	0.57±0.10	8.3±0.5	15.0±1.3
MDP14	0.59±0.10	1.71±0.20	9.0±0.6	19.0±0.9

Table 6-2. *E. coli* strains and plasmids. All production strains were made by transforming BL21 Star™ (DE3) (F^- *ompT* *hsd*S_B ($r_B^-m_B^-$) *gal dcm rne*131 (DE3), Invitrogen Corporation, Carlsbad, CA) with pJD727 (Ino1, MIOX, and Udh under *lac* promoter control; p15A origin) and the scaffold plasmid as indicated.

Strains	Scaffold plasmid[a]	No. of GBD domain (for Udh)	No. of SH3 domain (for Ino1)	No. of PDZ domain (for MIOX)
JT1	pJD757	1	1	1
JT2	pJD758	1	1	2
JT3	pJD759	1	1	4
JT4	pJD760	1	2	1
JT5	pJD761	1	2	2
JT6	pJD762	1	2	4
JT7	pJD763	1	4	1
JT8	pJD764	1	4	2
JT9	pJD765	1	4	4
JT10	pWW306[b]	0	0	0
JTK1	pJD788[c]	1	6	2
JTK2	pJD789[c]	1	8	2
JTK3	pJD790[c]	1	3	2
JTK4	pJD791[c]	1	2	2
JTK5	pJD824	1	0	2
JTK6	pJD825[c]	1	4	2

a Scaffold plasmids are under pTET control and contain ColE1 origin.
b pWW306 is a control plasmid containing no scaffold.
c These plasmids contain degenerate sequences for SH3 domains.

Table 6-3. Comparison of titers achieved using scaffolds with the same SH3 sequences (JT5 and JT8) versus scaffolds with degenerate coding sequences (JTK4 and JTK6), with 0.05 mM IPTG. Numbers are the averages ± standard deviations, in g/L.

aTc (nM)	JT5 ($GBD_1SH3_2PDZ_2$)	JTK4 ($GBD_1SH3_2PDZ_2$)	JT8 ($GBD_1SH3_4PDZ_2$)	JTK6 ($GBD_1SH3_4PDZ_2$)
108	1.43±0.01	1.22±0.31	2.27±0.23	2.28±0.35
81	1.21±0.28	1.20±0.02	Not measured	Not measured
54	1.31±0.12	1.21±0.12	2.20±0.08	2.37±0.30
27	0.89±0.10	1.02±0.13	Not measured	Not measured

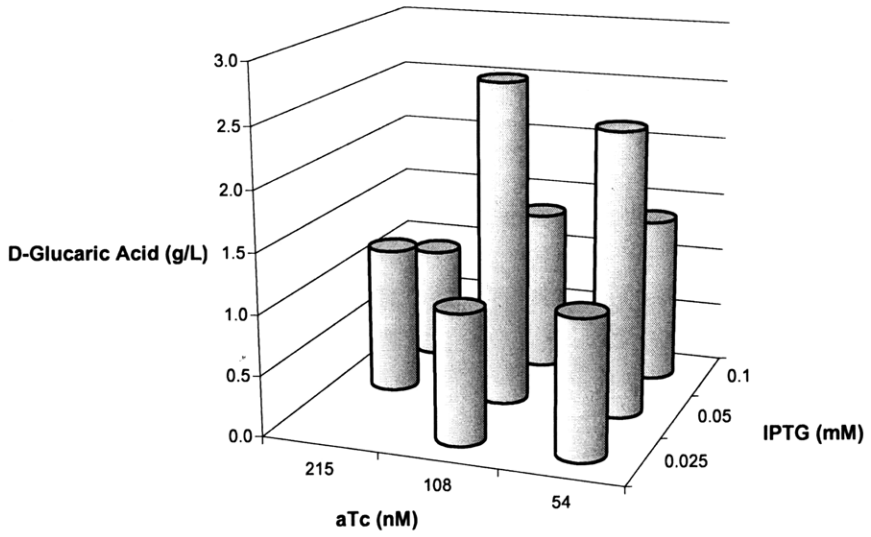


Fig. 6-1. Effect of induction levels (IPTG for pathway enzyme expression; aTc for scaffold expression) on the production of D-glucaric acid using the scaffold $GBD_1SH3_4PDZ_4$ (JT9). For the 0.025 mM IPTG/215 nM aTc condition, no experiment was performed. Data are the averages of three replicates and the standard deviations are not higher than 23% of the averages.

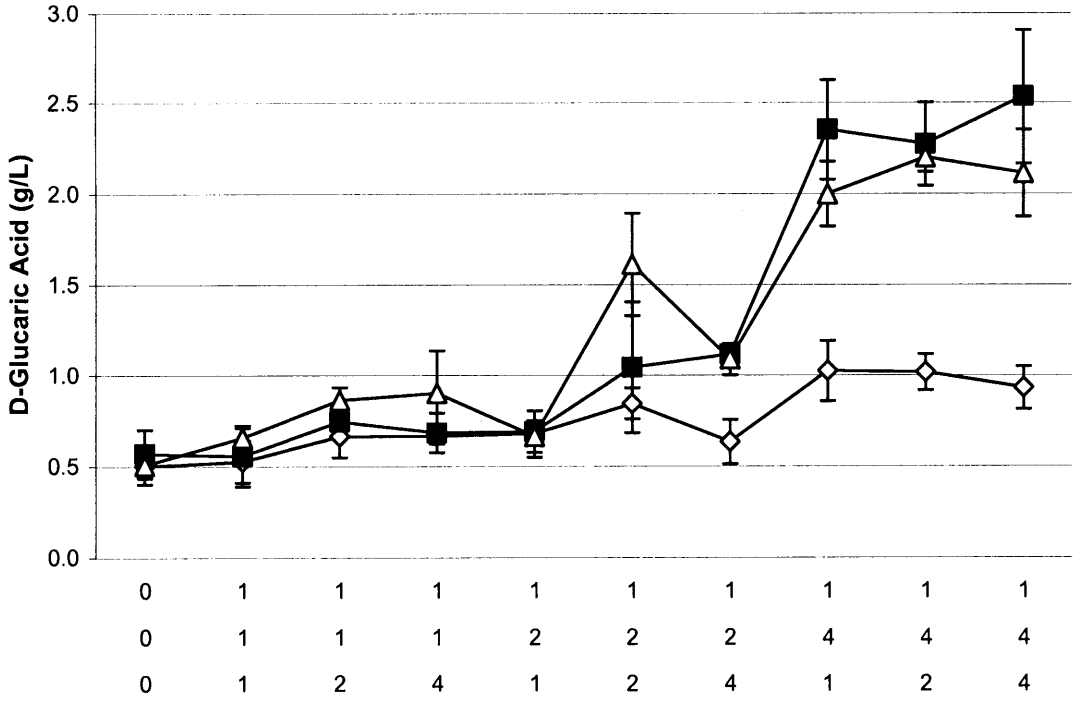


Fig. 6-2. Effect of various scaffold architectures (constructs JT1 to JT9) on the production of D-glucaric acid at 0.05 mM IPTG and three different aTc concentrations as indicated. Δ = 54 nM aTc; ■ = 108 nM aTc; ◊ = 215 nM aTc.

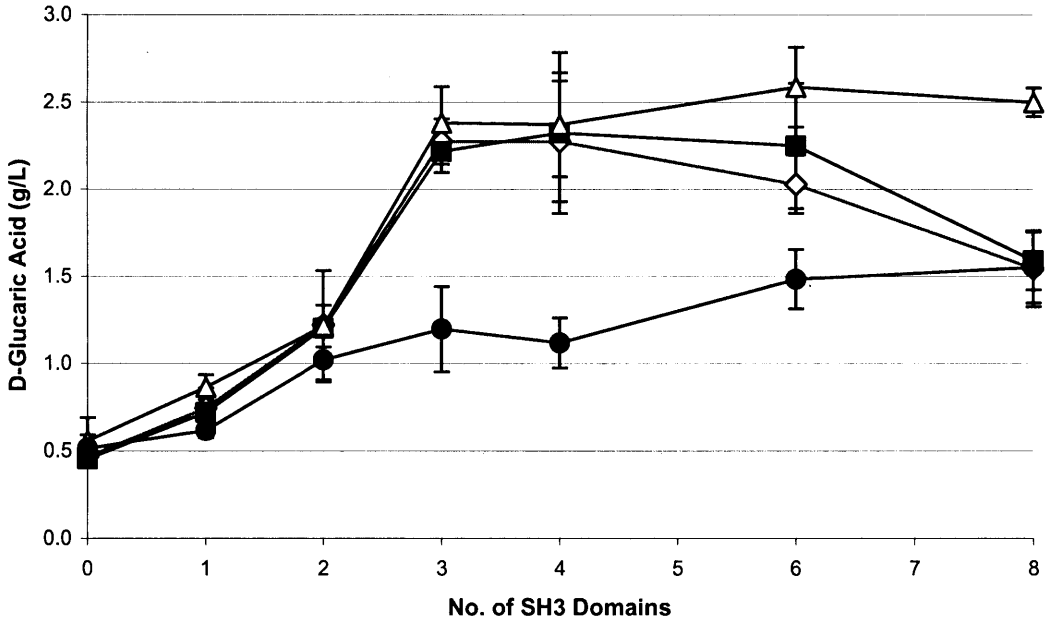


Fig. 6-3. Effect of the number of Ino1-recruiting SH3 domains on D-glucaric acid titer with 0.05 mM IPTG in $GBD_1SH3_cPDZ_2$, where c is 0, 1, 2, 3, 4, 6, and 8. Experiments were performed at four different aTc concentrations as indicated. The scaffolds with degenerate coding sequences were used for c = 2, 3, 4, 6, and 8. Lines are drawn to indicate trends. ● = 27 nM aTc; Δ = 54 nM aTc; ■ = 81 nM aTc; ◊ = 108 nM aTc.

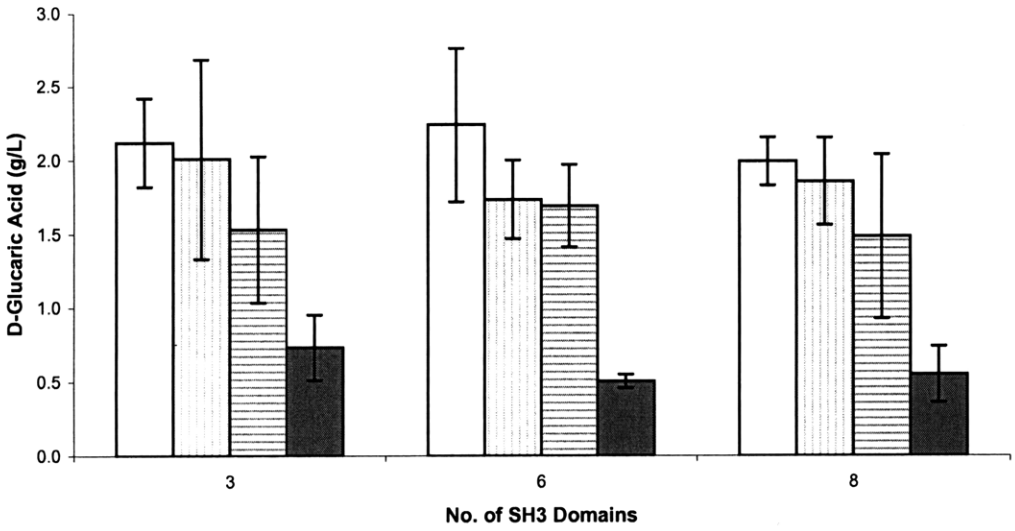


Fig. 6-4. Effect of IPTG concentration on D-glucaric acid titer at 54 nM aTc in $GBD_1SH3_cPDZ_2$, where c is 3, 6, and 8. Experiments were performed at four different inducer concentrations: from left to right, 0.05, 0.075, 0.1, and 0.2 mM IPTG. Data are the averages and standard deviations of two replicates.

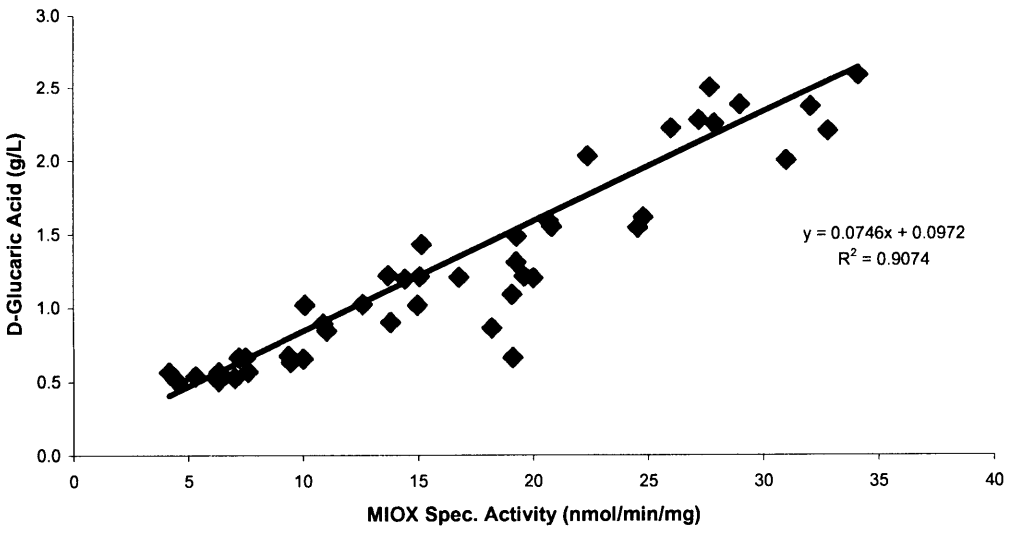


Fig. 6-5. Correlation between D-glucaric acid titer and MIOX activity across various scaffold architectures. Data are from JT1-10 at 0.05 mM IPTG, and 54 or 215 nM aTc; and JTK1-4, JT5, and JT10 at 0.05 mM IPTG and 108, 81, 54, or 27 nM aTc. Data points represent the average values of MIOX activity and D-glucaric acid titer. The standard deviations are not higher than 57% (MIOX activity) and 37% (D-glucaric acid titer) of the averages.

Chapter 7

Part Engineering – Engineering Substrate Specificity to Create Glucose 6-Oxidases

SM. Lippow*, TS Moon*, S Basu, S-H Yoon, X Li, B Chapman, K Robison, D Lipovšek and KJ Prather. **Engineering enzyme specificity using computational design of a defined-sequence library**, to be submitted (2009).

* These authors contributed equally to the work.

Abstract

Computational protein design has shown its potential to efficiently create proteins with improved properties, including protein stability and antibody affinity. However, the computational redesign of enzymes has proven to be far more challenging because of poor understanding of the catalytic process and inability to accurately predict the interactions at the transition states. Thus, experimental approaches like directed evolution have been the preferred method to create new enzymes. Advances in DNA synthesis technology, however, have made possible the rapid creation of genetic material from scratch, giving synthetic biologists opportunities to combine the benefits of the computational and experimental approaches. Here, a computational design and selection of glucose oxidase were reported. To this end, we computationally searched the sequence space of candidate mutations, combinatorially assembled the selected sequences, and experimentally screened the created library. We present computational and experimental results, including the methods used to generate the novel mutants and the structure-activity relationship of the newly created glucose oxidases.

Keywords: glucose oxidases; galactose oxidases; DNA synthesis technology; computational protein design; substrate specificity change; mutagenesis; library screening; protein characterization.

7.1 Introduction

Synthetic organic chemists approach a synthesis scheme for a compound, based on the structure of the target compound. This retrosynthetic design is a powerful tool, giving organic chemists a lot of flexibility: a wide range of target products and a large set of possible synthetic steps. Despite this flexibility, chemical synthesis is not efficient when the target compounds are complex and have many functional groups. To synthesize these challenging compounds, biochemists and metabolic engineers have recruited enzymes and have shown the potential of these biocatalysts (Atsumi et al., 2008; Martin et al., 2003; Moon et al., 2009a; Nakamura and Whited, 2003; Niu et al., 2003). However, most biosynthetic approaches have been made by utilizing "forward biosynthesis," a substrate-focused design of biological pathways, limiting the range of the target products. Retrobiosynthetic approach, a product-targeted design of biological pathways, has been proposed as an alternative strategy to exploit the diversity of enzyme-catalyzed reactions (Prather and Martin, 2008).

For the production of D-glucaric acid, I have designed an alternative synthetic pathway, entirely based on retrosynthetic scheme (PW2 in Fig. 1-1). The first step involves oxidation of the C-6 hydroxyl group on glucose, but no glucose oxidase in nature has been found to act on this hydroxyl group on glucose. Sun *et al.* attempted to generate glucose 6-oxidase (GlcOx) from galactose 6-oxidase (GaOx) by error prone PCR, but no improvement in activity on glucose was observed (Sun et al., 2002; Sun et al., 2001). Arnold and coworkers then screened multiple rounds of combinatorial libraries which were constructed by saturation mutagenesis, and a GaOx mutant (M-

RQW) with a low GlcOx activity was selected (Sun et al., 2002). However, the activity level is too low for the practical application to PW2.

Such a semi-rational approach attempted by Sun *et al.*, which targets smaller number of residues to mutate based on structural information (Sun et al., 2002), would decrease greatly the library size, but this approach still neglects pairwise and higher-order interrelationship between residues. Thus, their sequential search by saturation mutagenesis might lead to local minimum. The main challenge can be attributed to the difficulty of screening a library generated by combinatorial randomization. For example, a library generated by combinatorial randomization of only six residues has a theoretical size of 6.4×10^7, which is too large to manually screen. This daunting task, however, can be accomplished by computational screening (Dahiyat, 1999; Hayes et al., 2002). This type of computational screening allows multiple mutations to be identified simultaneously, which is beneficial especially when mutation effect is not an additive one.

Despite advances in computational screening (Moore and Maranas, 2004; Saven, 2002; Wong et al., 2007), computational redesign of enzyme activity is still difficult because there is no accurate model to represent the high-energy transition states (Chen et al., 2009). The simplified models for protein design assume a rigid protein backbone, a rotamer library for side-chain conformation, and a pairwise energy function (Georgiev et al., 2008; Gordon et al., 1999; Lovell et al., 2000; Ponder and Richards, 1987; Vizcarra and Mayo, 2005). Nonetheless, the computational protein design has shown its potential

to improve target properties and generate enzyme activity (Bolon and Mayo, 2001; Chen et al., 2009; Lippow et al., 2007). However, most libraries generated so far were from the use of degenerate codons or mutagenesis and recombination, making it difficult to substantially correlate protein sequence/structure with activity.

Here, computational enzyme design and experimental screening for activity are combined to create GlcOx mutants. For the computational part, structure-based computational design was attempted. Availability of crystal structure and substrate binding model for GaOx (Baron et al., 1994; Ito et al., 1994; Rogers et al., 2007) is prerequisite for this computational modeling. However, imperfect understanding of the catalytically relevant interactions at the transition state requires complementary methods. To this end, an experimental selection method was introduced, and the defined-sequence library (Table 7-1) designed by computation was screened. In addition, the connection between these two approaches was made possible by DNA synthesis technology which led to library creation with high-fidelity, defined sequences. Moreover, sequence-structure-activity relationship was elucidated from analysis of this library and characterization of newly created GlcOx mutants.

7.2 Materials and Methods

7.2.1 Screening Glucose Oxidase Libraries

Computational enzyme design and DNA synthesis for the defined sequence library construction were performed at Codon Devices, and the detailed methods will be

described in the manuscript under preparation (SM. Lippow, TS Moon, S Basu, S-H Yoon, X Li, B Chapman, K Robison, D Lipovšek and KJ Prather. Engineering Enzyme Specificity Using Computational Design of a Defined-Sequence Library). Briefly, we developed a method to assemble high-fidelity libraries of defined amino acid sequences that encode high-order correlated information, such as that obtained from structure-based computational design. In our defined-sequence libraries, only amino-acid mutations predicted from computational design are included, and only designed amino-acid combinations are included. The wild-type gene was split into N- and C-terminal constant regions and an internal variable region which contains all twelve designed positions. The variable region of the gene was constructed by PCR-based assembly of partially overlapping oligonucleotides (Tian et al., 2004). Single colonies from the designed or combinatorial library (Table 7-1 and 7-2) were inoculated into deep well microplates (well volume, 2 ml; VWR Scientific Products, Edison, NJ) containing 1 ml LB and 0.1 mg/ml carbenicillin. The parent plates were incubated overnight at 30°C and used to create a 10% (v/v) glycerol stock of the clones. The parent plates were duplicated by inoculating cells to new deep well microplates containing 0.2 ml LB-ampicillin (100 μg/ml), and the daughter plates were incubated for 12 hrs at 30°C and 250 rpm. 60 μl of the cultures were transferred to new deep well microplates containing 0.6 ml LB-ampicillin (100 μg/ml) and 0.5 mM IPTG, and cells were grown for 12 hrs at 30°C and 250 rpm. The cultures were centrifuged for 15 min at 3000 x g, and the cells were resuspended in 0.4 ml sodium phosphate buffer (0.1M, pH6.8) containing 0.8 mM $CuSO_4$ and 0.5 mM polymyxin B sulfate (EMD Chemicals, San Diego, CA). The plates were then incubated overnight at 4°C, and the cell lysates were mixed with Horseradish

peroxidase (HRP, 20 U/ml, Sigma Aldrich, St. Louis, Mo), ABTS (2.8 mg/ml, 2,2′-azino-bis[3-ethylbenzothiazoline-6-sulfonic acid], Sigma Aldrich, St. Louis, Mo), and glucose (0.25M, Mallinckrodt Baker, Phillipsburg, NJ). Positive clones were selected after 24 hr incubation at room temperature (for the 1st round of screening), 6 hr (for the 2nd round of screening), and 0.5 hr (for the 3rd round of screening). The selected clones were streaked on LB agar plates supplemented with 100 μg/ml ampicillin, and the resulting single colonies were used for further screening (the 2nd and/or 3rd round), plasmid purification, and DNA sequencing.

7.2.2 Strains, growth media, and plasmids

E. coli strain DH10B [F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*Δ139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L (Str^R) *nup*G] was used for all molecular biology manipulations and as host for the library screening. BL21 Star™ (DE3) [F⁻ *omp*T *hsd*S_B ($r_B^- m_B^-$) *gal dcm rne*131 (DE3)] was used as host for expression of pET21b (Novagen, Madison, WI). Competent cells of both strains were purchased from Invitrogen Corporation (Carlsbad, CA). Cultures were propagated in Luria-Bertani (LB) broth. LB (Miller) medium was prepared from dehydrated powder according to manufacturer's instructions (BD Biosciences, San Jose, CA). All molecular biology manipulations were performed according to standard practices (Sambrook and Russell, 2001). Plasmids containing genes encoding the wild type GaOx and M-RQW version (Sun et al., 2002; Sun et al., 2001) were obtained from Prof. Frances Arnold at Caltech.

7.2.3 Protein Purification and Characterization

The selected clones were sequenced (Table 7-3), and the DNA sequences of DES3-1 and DES-8 are found to be identical. The resulting unique clones were sub-cloned into the 6xHis-tag-containing pET21b using the *Not*I and *Xho*I sites. The primers used for the PCR reactions are as follows: 5′-GTACATATAAGCGGCCGCCTCAGCACCTATCG-3′ and 5′-GATACAACTTCTCGAGCTGAGTAACGCGAATCG-3′. The *Not*I and *Xho*I restriction sites are underlined. The resulting plasmids were confirmed by sequencing and used for transformation into BL21 Star™ (DE3). Transformed cells were spread on LB-ampicillin (100 μg/ml) agar plates and grown overnight at 30°C. Single colony was picked, and an inoculum was prepared in LB-ampicillin (100 μg/ml) medium by growing for 16 hrs at 30°C and 250 rpm. The culture was transferred to two 250-mL baffled flasks containing 50 ml of LB-ampicillin (100 μg/ml) medium, grown until the OD_{600} (optical density at 600 nm) reached 0.6, and induced by adding IPTG. The cells were grown for an additional 4 hrs and then centrifuged for 15 min at 4°C and 5000 x g. Protein purification was performed at 4°C using the harvested cells from the two 50 ml cultures and the ProBond™ Purification System as described by the manufacturer (Invitrogen Corporation, Carlsbad, CA). EDTA-free protease inhibitor cocktail tablets (Roche Applied Science, Indianapolis, IN) were added to the resuspension buffer containing 1 mg/ml lysozyme. After eluting the protein, diafiltration was performed using phosphate buffered saline (PBS, pH 7.4) and Amicon® Ultra-15 (Millipore, Billerica, MA) according to the manufacturer's instructions. The retentate containing the protein was incubated with 0.1 mM $CuSO_4$ overnight at 4°C and used for kinetic measurements.

SDS-PAGE was performed as described by Sambrook and Russell (Sambrook and Russell, 2001). The purified protein ran as a single band with the expected molecular weight of 70 kDa. Protein concentrations were measured by using the Sandwich ELISA. As the ELISA standard, *Pseudomonas syringae* uronate dehydrogenase containing T7 tag and 6xHis-tag was prepared as described by Yoon *et al.* (Yoon et al., 2009). ELISAs were performed on Costar® 96 well plate (Corning Incorporated, Corning, NY) coated with 50 µl of 4 µg/ml monoclonal anti-T7 tag antibody (Sigma Aldrich, St. Louis, Mo) and blocked with 0.5% (w/v) BSA-PBS buffer. The standards or samples containing T7 tag and 6xHis-tag were incubated for 1 hr at room temperature, and then monoclonal anti-polyHistidine, Peroxidase conjugate (Sigma Aldrich, St. Louis, Mo) was added. After 1 hr incubation at room temperature, 1-Step™ Ultra TMB-ELISA (Pierce, Rockford, IL) was used as a chromogenic substrate, the reaction was stopped by adding 2M sulfuric acid, and the absorbance was measured at 450 nm using a Varioskan Flash Spectral Scanning Multimode reader (Thermo Fisher Scientific, Waltham, MA).

The activity of glucose oxidases on glucose (4 to 300 mM) and galactose (2 to 70 mM) was determined in 0.1 M sodium phosphate buffer (pH 6.8) using the HRP-ABTS coupled assay (Baron et al., 1994). The absorbance change rate was monitored at 405 nm using a spectrophotometer (Beckman Coulter, Fullerton, CA), and the extinction coefficient of 36.8 $mM^{-1}cm^{-1}$ was used to calculate the specific activity (Sigma Quality Control Test Procedure, http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/General_Information/hrp_assay_

abts.Par.0001.File.tmp/hrp_assay_abts.pdf). For glucose oxidase system in which a lag time was observed (DES3-10 and M-RQW) (see Chaper 8 for details), the maximal rate was used to calculate the specific rate. Our initial data show that the rate-substrate concentration curves do not exhibit the typical saturation behavior up to the practical solubility limit of glucose (0.8 M) and galactose (0.4 M) for rate determination, implying that the K_m value is on the order of molar and is practically impossible to measure. Instead, the k_{cat}/K_m ratio was determined from the slope of the specific activity-substrate concentration curve with the assumption that K_m is much greater than the substrate concentration (glucose: 4 to 300 mM; galactose: 2 to 70 mM).

7.3 Results and Discussion

GaOx was selected as an enzyme engineering target for the first step of PW2 based on the following reasons: structural similarity of galactose to glucose and availability of its crystal structure and substrate binding model (Baron et al., 1994; Ito et al., 1994; Rogers et al., 2007). The Arnold group evolved GaOx to find mutant M-RQW, which has an activity on glucose (Sun et al., 2002). They failed to find the GlcOx by random point mutagenesis; sequential saturation mutagenesis of five amino acid residues generated GlcOx with a relatively low activity. To reduce the library size and find GaOx mutants with higher activities on D-glucose, molecular modeling-based mutagenesis has been performed. A model of the complex between wild type GaOx and galactose was generated by docking galactose into a crystal structure of the unbound enzyme (Fig. 7-1). Twelve active-site and second-shell positions involved in substrate recognition were

simultaneously redesigned around glucose in the active site, while preserving all remaining positions including five active-site side chains crucial for the conversion of alcohol to aldehyde. The resulting unique sequences with low predicted energy of the enzyme–substrate complex were narrowed down based on criteria including favorable enzyme–substrate binding interaction, enzyme stability, and electrostatic components of these energies, and 10,752 sequences were suggested as GlcOx candidates (Table 7-1). Compared to the number of sequences required for saturation mutagenesis of the twelve residues ($4x10^{15}$), 10,752 is a significantly smaller number. However, it is still a huge number for site directed mutagenesis.

In 2006, Codon Devices synthesized a 35,000 bp DNA strand, and synthetic biologists started to think that rapid and cheap creation of genetic material from scratch would be possible in near future. However, to my knowledge, no attempt to create a library of genes with specified sequences has been made. In this joint project, Codon Devices has created the library of 10,752 DNA sequences. In addition, a combinatorial library, created by using degenerate codons for the selected amino acid residues at the twelve residues, was constructed (Table 7-2). For a comparison, 11,266 clones (from the combinatorial library with a theoretical size of $1.1x10^{10}$) were screened by using the horseradish peroxidase-based method. One clone was found to have detectable glucose oxidase activity; however, the glucose oxidase activity is much lower than that of M-RQW, and its galactose oxidase activity was almost lost. In contrast, 402 clones were found to be active on glucose from the designed library, confirming that the computationally designed "synthetic" library samples sequence space more efficiently

(Table 7-3). The best clones were sequenced, sub-cloned into the 6xHis-tag-containing pET21b, and further studied after protein purification (Table 7-4). Three mutations (Q326E, Y329K/R, and R330K), two of which M-RQW does not share, were found to be important for activity on glucose, resulting in glucose oxidase mutants with higher activities than that of M-RQW (Table 7-4). W290F is found in only one of the top clones (and M-RQW), showing that saturation mutagenesis done by Arnold group (Sun et al., 2002) led to local minima in the sequence–function landscape. Keeping P463 and W290 intact was found to be important for activity on galactose.

The sequential search done by Sun *et al.* (Sun et al., 2002) is shown to hit local minimum. As discussed in Chapter 8, Trp290 plays an import role in stabilizing the O radical on Tyr272 and restricting entry of other molecules (e.g. water molecules), keeping the enzyme active. Replacing Trp290 with Phe, a smaller residue, would decrease K_m, but this beneficial effect could be cancelled out by the large decrease in k_{cat} (see Chapter 8 for more discussion). In addition, Des3-2, the best clone, is only three-residues different from the wild type (Table 7-4), indicating the risk of local minimum traps for the sequential approaches. Moreover, R330K, the common mutation for all eleven clones (and M-RQW) is only found in 48% of the top 380 hits (Table 7-3), further implying the risk of local minimum traps. Thus, our multiple simultaneous mutation simulation and screening is proved to more efficiently search sequence space than sequential search by saturation mutation.

Preliminary result shows that Schrodinger docking score, a computed score based on the extent of ligand-protein (glucose-GlcOx in this work) interaction, correlates linearly with enzyme activity of the top enzymes (data not shown). In addition, sequence-activity analysis of 382 hits (out of the 402 clones, Table 7-3) clearly indicates amino acid enrichment particularly at the important positions (Q326E, Y329K/R, and R330K) during the designed library screening (data will be available in the manuscript under preparation). To my knowledge, this type of sequence-activity analysis has never been performed, and our approach combining structure-based computational enzyme design, DNA synthesis technology, library screening, and activity characterization will pave a new way for enzyme engineering. In addition, this research may provide some insights to enzyme evolution processes (Tracewell and Arnold, 2009). Chapter 8 extends this sequence-activity analysis to a mechanism level to explain the lag phenomenon from the mutants containing W290F.

Table 7-1. The complete list of amino-acid sequences in the designed library.

Designed positions	290	324	326	329	330	333	334	383	405	406	441	463	Diversity
Wild type	W	D	Q	Y	R	N	H	C	Y	Q	F	P	
Assembly region	I	II						III			IV		
Designed amino-acid combinations	F	D	D	R	N	N	H	C	G	T	F	A	
	W	D	E	K	K	N	H	C	H	Q	F	D	
		D	E	R	K	N	H	C	S	Q	F	E	
		D	E	R	Q	N	H	C	S	R	F	P	
		D	E	R	S	N	H	C	Y	E	F	S	
		D	E	Y	Q	N	H	C	Y	K	H	D	
		D	K	K	E	N	H	C	Y	N	H	P	
		D	K	R	E	N	H	C	Y	Q	R	S	
		D	K	R	Q	N	H	C	Y	R			
		D	K	Y	E	N	H	C	Y	T			
		D	K	Y	Q	N	H	N	Y	Q			
		D	N	K	H	N	H	N	Y	T			
		D	N	R	D	N	H	S	Y	Q			
		D	N	R	H	N	H	S	Y	T			
		D	N	R	K	N	H						
		D	N	R	N	N	H						
		D	N	Y	H	N	H						
		D	Q	K	K	N	H						
		D	Q	K	Q	N	H						
		D	Q	N	Q	N	H						
		D	Q	R	K	N	H						
		D	Q	R	Q	N	H						
		D	Q	R	S	R	H						
		D	Q	Y	K	N	H						
		D	Q	Y	R	N	H						
		D	R	E	K	N	H						
		D	R	E	S	N	H						
		D	R	H	D	N	H						
		D	R	K	H	N	H						
		D	R	K	Y	N	H						
		D	R	N	K	N	H						
		D	R	R	D	N	H						
		D	R	R	H	N	H						
		D	R	R	K	N	H						
		D	R	R	M	N	H						
		D	R	R	N	N	H						
		D	R	R	Q	N	H						
		D	R	R	S	N	H						
		D	R	Y	D	N	H						
		D	R	Y	H	N	H						
		D	R	Y	K	N	H						
		D	R	Y	M	N	H						
		D	R	Y	Q	N	H						
		D	R	Y	S	N	H						
		D	S	R	K	N	H						
		G	R	Y	I	N	H						
		S	Q	R	D	R	E						
		S	R	R	S	N	H						
# combinations	2	48						14			8		10,752

Table 7-2. Comparison of the amino acids in the designed and combinatorial libraries.

Designed positions	290	324	326	329	330	333	334	383	405	406	441	463	Diversity
Wild type	W	D	Q	Y	R	N	H	C	Y	Q	F	P	
Assembly region	I	II						III			IV		
# designed combinations	2	48						14			8		10,752
Different amino acids in design	F	D	D	E	D	N	E	C	G	E	F	A	
	W	G	E	H	E	R	H	N	H	K	H	D	
		S	K	K	H			S	S	N	R	E	
			N	N	I				Y	Q		P	
			Q	R	K					R		S	
			R	Y	M					T			
			S		N								
					Q								
					R								
					S								
					Y								
# different amino acids in design	2	3	7	6	11	2	2	3	4	6	3	5	$1.2 \cdot 10^7$
Combinatorial library: best degenerate codon	TKS	RRC	VRS	NRS	NDS	MRC	SAS	WRC	NRC	VVS	YDC	BMW	
Different amino acids, or stop codon	C	D	D	C	C	H	D	C	C	A	C	A	
	F	G	E	D	D	N	E	N	D	D	F	D	
	L	N	G	E	E	R	H	S	G	E	H	E	
	W	S	H	G	F	S	Q	Y	H	G	L	H	
			K	H	G				N	H	R	P	
			N	K	H				R	K	Y	Q	
			Q	N	I				S	N		S	
			R	Q	K				Y	P		Y	
			S	R	L					Q		X	
				S	M					R			
				W	N					S			
				Y	Q					T			
				X	R								
					S								
					V								
					W								
					Y								
					X								
# different amino acids, or stop codon	4	4	9	13	18	4	4	4	8	12	6	9	$1.1 \cdot 10^{10}$
# total codons	4	4	12	16	24	4	4	4	8	18	6	12	$4.9 \cdot 10^{10}$

R=A,G; Y=C,T; M=A,C; K=G,T; S=C,G; W=A,T; B=C,G,T; D=A,G,T; H=A,C,T; V=A,C,G; N=A,C,G,T

Table 7-3. Experimental screening of the designed and combinatorial libraries.

	# screened	# hits*	Overall hit rate (%)
Combinatorial library	11,266	1 (weak)	0.009
Designed library, round 1	10,603	402 (varied)	4
		weak: 186	
Designed library, round 2	402	medium: 155	
		strong: 61	0.6
Designed library, round 3	61	best: 11	0.1

* Positive clones were selected after 24 hr incubation at room temperature (for the 1st round of screening), 6 hr (for the 2nd round of screening), and 0.5 hr (for the 3rd round of screening).

Table 7-4. Characterization of the top hits from the designed and combinatorial libraries[a].

													k_{cat} / K_m ($M^{-1}s^{-1}$)	
Position	290	324	326	329	330	333	334	383	405	406	441	463	On galactose	On glucose
Wild type	W	D	Q	Y	R	N	H	C	Y	Q	F	P	34,000 ± 5,000	Not detectable
M-RQW	F				K					T			91 ± 8	4.9 ± 0.7
Con1-1	F		N	D	K			Y	N	T	Y	A	Not measured	Not measured
Des3-1			E	R	K					T		A	460 ± 30	11.0 ± 0.6
Des3-2			E	K	K								1,100 ± 300	20.0 ± 3.0
Des3-3			E	K	K					T		A	550 ± 70	14.2 ± 1.1
Des3-4					K					T		S	172 ± 2	2.8 ± 0.4
Des3-5[b]				R	K								1,050 ± 150	1.8 ± 0.2
Des3-6			E	R	K					R			870 ± 180	5.8 ± 0.6
Des3-7			E	R	K					K			1,160 ± 40	4.8 ± 0.2
Des3-8[c]			E	R	K					T		A	460 ± 30	11.0 ± 0.6
Des3-9			E	R	K							S	210 ± 10	4.4 ± 0.4
Des3-10	F			K	K					R			14.8 ± 0.1	5.6 ± 0.8
Des3-11			E	K	K					T		S	130 ± 30	13.0 ± 1.1
Des3-12					K			N				A	152 ± 6	1.3 ± 0.1

[a] Changes from wild type are recoded.

[b] Des3-5 was not in the top 11 after the third round of screening, but was included for characterization.

[c] Identical to Des3-1.

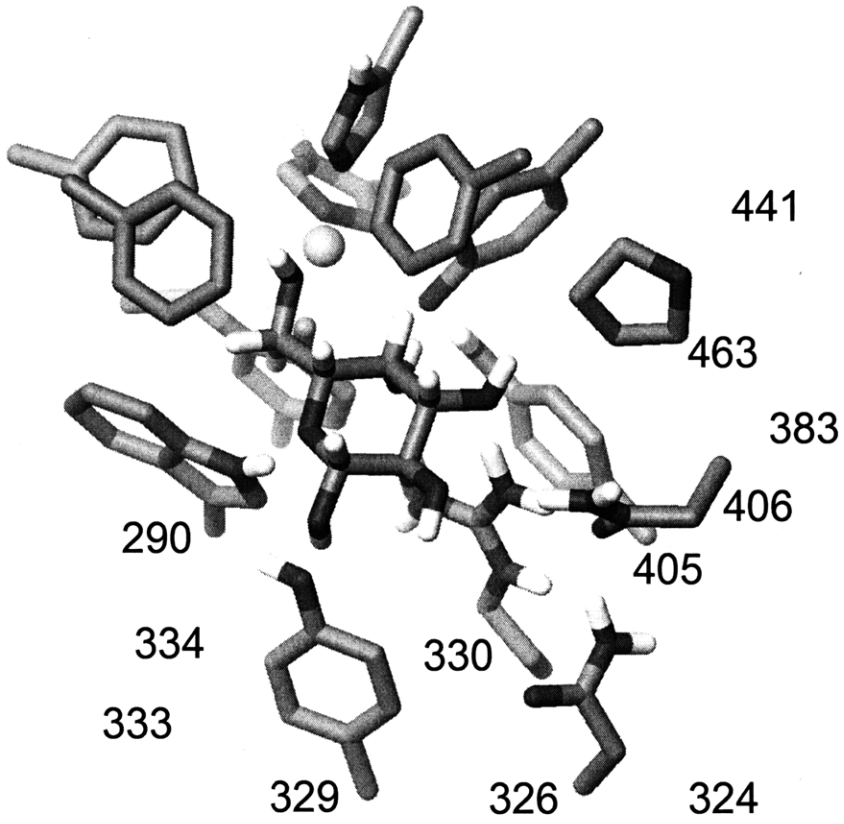


Fig. 7-1. A structural model of wild type galactose 6-oxidase with docked galactose.

Chapter 8

Part Characterization – Proposed Model Mechanism for Newly Created Glucose 6-Oxidases and Sensitivity Analysis

TS Moon*, D Nielsen* and KJ Prather. **Sensitivity Analysis of a Proposed Model Mechanism for Newly Created Glucose 6-Oxidases**, to be submitted (2009).

* These authors contributed equally to the work.

Abstract

Parametric sensitivity analysis of a proposed model mechanism can serve as a valuable diagnostic for studying the fundamental characteristics of an enzymatic reaction. As the first step towards understanding the kinetics of a series of newly engineered glucose 6-oxidases, the activity of which has never before been observed in nature, we have proposed a mechanistic kinetic model and performed a parametric sensitivity analysis. The mechanism of our model consists of two enzyme forms, namely 'less active' and 'activated,' and is shown to be consistent with experimental observations of prolonged periods of enzyme induction. The extended lag phase phenomenon was found to be well-described mechanistically by a slow rate of inter-conversion between the two enzyme forms (relative to the rate of product formation), and this prediction was further supported by our experimental results. The proposed enzyme model will serve as a blueprint with which to better understand the mechanistic behavior of newly generated glucose 6-oxidases.

Keywords: glucose oxidases; galactose oxidases; lag time; kinetic mechanism; sensitivity analysis; mutagenesis

8.1 Introduction

Galactose oxidase (GaOx) is a copper-containing enzyme that catalyzes the oxidation of primary alcohols to aldehydes with the concomitant reduction of molecular oxygen to hydrogen peroxide. This enzyme has been extensively studied for several decades and the determination of its crystal structure (Ito et al., 1991; Ito et al., 1994) has paved the way for an improved understanding of its kinetic mechanism (Baron et al., 1994; Humphreys et al., 2009; Whittaker et al., 1998). The copper at the active site is coordinated by two His and two Tyr residues, with a solvent serving as the fifth ligand of the square pyramid. One of these ligands, Tyr272, is the site of the O radical (Baron et al., 1994) which is stabilized by Tyr272-Cys228 cross-linking via a thioether bond. Meanwhile, Trp290 is believed to have an important role in stabilizing the radical by stacking interaction with Tyr272-Cys228 (delocalization of the unpaired spin density) and by entry restriction of other molecules to the active radical site (protection of the radical from solvents) (Saysell et al., 1997). Studies have shown that GaOx has three redox states: an activated form (Y•-Cu(II), Ox-GaOx) containing an oxidized complex and reacting with alcohols; an inactive form containing Cu(II) but no radical (Y-Cu(II), semi-Red-GaOx); and, finally, an oxygen-reactive form containing Cu(I) but no radical (Y-Cu(I), Red-GaOx). With consideration of these structural and redox characteristics, appropriate chemical mechanisms have been suggested (Humphreys et al., 2009; Whittaker et al., 1998).

GaOx is known to exhibit activity with a broad range of substrates including galactose, but it has almost no demonstrated activity on glucose, an isomer of galactose

(Sun et al., 2002; Sun et al., 2001). The inability to use glucose as a substrate is due to a steric clash between Tyr495 and the C4-hydroxyl group of glucose (Ito et al., 1994; Wachter and Branchaud, 1996). Sun *et al.* generated GaOx mutants with glucose activities, but the demonstrated activity levels remained quite low (Sun et al., 2002). Recently, we created glucose oxidases (GlcOx) from GaOx by computationally searching the sequence space of candidate mutations, combinatorially assembling the selected sequences, and experimentally screening the created library (Chapter 7). Our interest in constructing a synthetic pathway to produce a value-added compound glucaric acid triggered the creation of GlcOx variants which catalyze the first step of our designed pathway (Chapter 7). While characterizing these GlcOx to select the best enzyme for our pathway, I routinely observed that newly engineered proteins incorporating the W290F mutation experienced significant lag times during catalysis in addition to relatively high losses of GaOx activity.

Lag periods in enzyme activity have previously been observed in copper-containing tyrosinases (Palumbo et al., 1985; Xue et al., 2007), iron-containing lipoxygenases (Schilstra et al., 1993; Wang et al., 1993), and copper-containing copper,zinc superoxide dismutases (Winterbourn et al., 2002). Interestingly, I found that all of these enzymes contain metal ions and catalyze redox reactions involving molecular oxygen, but the lag was reported to be attributed to different factors (Gukasyan, 2002; Molina et al., 2007; Wang et al., 1993; Winterbourn et al., 2002). Tyrosinases, exhibiting two catalytic activities which yield quinones from monophenols via diphenols, were believed to incorporate an autocatalytic hydroxylase step (Gukasyan, 2002; Molina et al.,

2007). Two alternative models have been suggested to explain the lag period associated with lipoxygenases, including: 1) the product activation model in which the inactive enzyme form in the Fe (II) state is oxidized to the active Fe (III) form by the product; and, 2) the substrate inhibition model in which the substrate is displaced from the regulatory site by the product (Wang et al., 1993). Finally, sub-micromolar concentrations of transition metal ions have been found to interfere with copper,zinc superoxide dismutases, leading to an increased lag period (Winterbourn et al., 2002).

To better understand the nature of this observed lag phenomenon, I suggest a reaction mechanism to describe the function and activity of our newly engineered GlcOx. This mechanism was previously introduced to explain single substrate hysteretic enzyme cases in which two enzyme forms possessing different kinetic parameters lead to the occurrence of lag phenomena (Bates and Frieden, 1973). As the first step towards supporting the validity of this proposed mechanism for GlcOx, we have performed a parametric sensitivity analysis to investigate how the co-existence of multiple enzyme forms (less active and activated) may explain the observed occurrence of delayed enzyme kinetics. Sensitivity analysis of a proposed model mechanism can serve as a valuable diagnostic to study the fundamental characteristics of an enzymatic reaction. Whereas this technique does not allow us to directly assess the suitability of the proposed model structure, inferences regarding which parameters are most responsible for the prediction of an objective criterion may be deduced by identifying the parameters to which that output is most sensitive. Thus, the associated predictions can provide fundamental insight into the relative significance of the individual mechanisms which compose the

overall model. As will be demonstrated for the present case, this can be of particular value for enzyme models incorporating multiple, simultaneous phenomena.

8.2 Materials and Methods

8.2.1 Kinetic Studies

Three representative GlcOx (Des3-2, Des3-6, and Des3-7) mutants previously constructed that contain the intact Trp290 residue (Chapter 7) were compared with the W290F-containing mutants M-RQW (Sun et al., 2002) and Des3-10 (Chapter 7). The vector, which contains the GaOx mutant M-RQW, was kindly provided by Professor Frances Arnold at Caltech (Sun et al., 2002), and the M-RQW gene was sub-cloned into the 6xHis-tag-containing pET21b as described previously (Chapter 7). The kinetics of the purified GlcOx on glucose (220 mM) was monitored using the HRP-ABTS coupled assay (Baron et al., 1994) as described previously (Chapter 7). The absorbance change rate was determined at 405 nm using a spectrophotometer (Beckman Coulter, Fullerton, CA), and the extinction coefficient of 36.8 $mM^{-1}cm^{-1}$ was used to calculate the concentration of product formed (Chapter 7).

8.2.2 Numerical Methods

Solution of the selected model, as well as the associated dynamic sensitivity equations, was performed using Matlab® and the intrinsic ordinary differential equation solver *ode15s*. Parameter estimates for solution of the model of interest (Fig. 8-1) were obtained from Bates and Frieden (Bates and Frieden, 1973), and are listed in Table 8-1.

Lag times (Θ_{lag}) associated with each experimental data set were approximated by first evaluating the tangent of each $\Delta P/\Delta P_{max}$ (calculated as $\left(\frac{P-P_0}{P_f-P_0}\right)$) curve at the point where the slope approaches its steady state value. The tangent curve was then extrapolated to determine the value at which it intersects with the x-axis. This value was designated as Θ_{lag}. To improve the comparative analysis between experimental data and model predictions, t_{max} was selected such that the maximum lag time among all simulations, $\Theta_{lag,\,max}$, would be one-quarter of its value (or 0.25, which is similar in magnitude to the mean value associated with the experimental data sets). This provided a similar characteristic lag time between experiment and model.

8.2.3 Parametric Sensitivity Analysis

Although sensitivity coefficients may be estimated by numerous, previously described techniques, the 'Direct Method', originally described by Dickinson and Gelinas (Dickinson and Gelinas, 1976), was employed for the present study as it is considered to be of high computational rigor. According to the Direct Method, the differential equations which comprise a dynamic model may be supplemented with appropriate differential sensitivity equations. This allows the complete set to be solved simultaneously and yields predictions of both the state variables and the sensitivity coefficients, s_{ij} (Mauch et al., 1997; Patnaik, 1994), defined as the partial derivatives of the state variables (x) with respect to the associated parameters (p):

$$s_{ij} = \frac{\partial x_i}{\partial p_j} \qquad \text{(a)}$$

To generate the set of sensitivity equations, we first consider an enzyme kinetic model of the following general form:

$$\frac{dx}{dt} = g(x,t,p), \qquad y = h(x,p), \qquad x(t_0) = x_0 \tag{b}$$

where t is time, the state variables are given by x, the parameters by p, and relevant output, or measured, variables are represented by y. In this case, the differential sensitivity equations can be defined as:

$$\frac{\partial s_{ij}}{\partial t} = \sum_{k=1}^{n} \frac{\partial g_i}{\partial x_k} s_{kj} + \frac{\partial g_i}{\partial p_j} \tag{c}$$

whereas the sensitivities associated with the output responses may then be evaluated as:

$$\frac{\partial y_i}{\partial p_j} = \sum_{k=1}^{n} \frac{\partial h_i}{\partial x_k} s_{kj} + \frac{\partial h_i}{\partial p_j} \tag{d}$$

To facilitate comparison during analysis, scaling effects can be removed by normalizing the sensitivity coefficients as $\frac{p_j}{y_i}\frac{\partial y_i}{\partial p_j}$.

8.3 Results

8.3.1 Proposed Mechanism

GaOx is known to exist in three different forms: an activated Ox-GaOx and two semi-reduced or reduced forms (semi-Red-GaOx or Red-GaOx) which require oxidation before becoming activated. However, to simplify our analysis, we have assumed a model mechanism (Bates and Frieden, 1973) consisting of only two enzyme forms: less active GlcOx (E, low affinity to substrate and low turnover rate) and activated GlcOx (E', high

affinity to substrate and high turnover rate) (Table 8-1 and Fig. 8-1). In addition, our experimental protocol was designed so as to eliminate oxygen limitations, allowing us to assume a single substrate model. Given that GaOx follows a ping-pong mechanism consisting of two sequential steps, where the first step (Ox-GaOx + alcohols → Red-GaOx + aldehydes) is known to be rate-limiting (Humphreys et al., 2009; Whittaker et al., 1998), the second step, which involves reaction with oxygen (Red-GaOx + O_2 → Ox-GaOx + H_2O_2), was omitted from our model mechanism for GlcOx. We have also included the following key assumptions: 1) a rapid equilibrium is achieved for the substrate-binding steps, 2) product formation steps are irreversible, and 3) no product inhibition occurs (Table 8-1 and Fig. 8-1). In contrast to a typical allosteric model (Monod et al., 1965), this mechanism is capable of explaining the extended lag time phenomenon that occurs when the inter-conversion steps between ES (or E) and E'S (or E') is slow relative to the product formation steps (Bates and Frieden, 1973). When substrate interacts with the enzyme, the enzyme-substrate complex (ES) concentration immediately reaches a high value, but the E'S concentration increases only gradually, resulting in a slower rate of product (*P*) formation. As the system then approaches the steady state, the E'S concentration reaches its maximum value with the concomitant product generation at the maximum rate.

8.3.2 Lag Time Phenomena and Sensitivity Analysis of the Proposed Model Mechanism

One practical utility of sensitivity coefficients is in their ability to describe the changes (both in magnitude and direction) that occur to the predicted values of the state

variables as associated with incremental perturbations in the parameter values. As with state variables of a dynamic model, sensitivity coefficients may also vary dynamically. Thus, it is important to characterize the dynamic sensitivity of a model to understand how, and under what temporal regions, each parameter may play a role in the overall predictions.

The presented model was solved using initial conditions of E and S of 1 and 1000 μM, respectively, whereas the initial values for all remaining state variables were taken to be 0. The initial conditions associated with each of the sensitivity coefficients were also appropriately determined using the initial values of the state variables. For simplification, the initial substrate concentration was assumed to be sufficiently excessive that its value should not change appreciably over the simulated range. For experiments, we used the initial substrate concentration of at least 220 mM, and the produced product concentration was never above 0.04 mM, justifying this assumption. As a result, $\frac{dS}{dt}$ was set to a value of zero for all times for the present study.

The effects of variations in *k2f* (and, appropriately, *k2r* as a constant equilibrium constant was maintained for this reaction step) on the overall model predictions was also investigated, and the results are compared in Fig. 8-2. As *k2f* was decreased, the predicted lag phase associated with product (P) synthesis was found to increase. As is further illustrated in Fig. 8-3, among the range of *k2f* values simulated, although initial rates of product formation were found to vary dramatically, each was found to approach the same steady-state value. This is similarly consistent with the results of Fig. 8-2,

wherein as t/t_{max} approaches its maximum value, each of the predicted curves arrives at the same steady-state slope.

The lag times associated with the experimental data and model simulations could also be directly approximated according to the method described above. These results are shown in Fig. 8-4, together with an illustration of this approach. Whereas a range of lag times between 0 and 0.40 was observed to be obtained experimentally, we find that a similar range could be generated by varying the value of *k2f* from its nominal value of 0.005 s^{-1} to as high as 5 s^{-1}. Interestingly, the lag times of M-RQW or Des3-10, both of which contain the W290F mutation, were higher than that of Des3-2, Des3-6 or Des3-7, which each contain the intact Trp290 (Fig. 8-4). Trp290 is known to play an important role in stabilizing the O radical of Tyr272 (Baron et al., 1994; Saysell et al., 1997). Thus, the mutation of this residue is expected to affect the inter-conversion steps between ES (or E) and E'S (or E'). Delayed accumulation of the activated enzyme forms would be expected to result in increased lag times (higher Θ_{lag} values).

The dynamic sensitivity of the selected model was subsequently characterized with respect to each of the parameters of Table 8-1. The predicted results with respect to *P*, the state variable of the highest interest to this study, are plotted in Fig. 8-5. As is clearly seen, the impact of the predictions for individual parameters is found to vary greatly with respect to the state of the system. For instance, although the model is initially quite sensitive to the parameters *k1f* and *k4f*, as the system proceeds towards its steady-state, the influence of *k1f* diminishes and approaches zero. Conversely, while the

model is initially predicted to be insensitive to both *k2f* and *k5f*, this condition quickly changes as these parameters become increasingly important to the prediction of *P*, particularly at the steady state. Similarly, predictions of *P* were also found to be increasingly sensitive to *k2r*, however in a direction opposite to that of *k2f* (indicative of their opposing reaction directions). Meanwhile, predicted values of *P* were found to be only mildly sensitive to both *k3f* and *k3r*. Finally, for all practical purposes, *P* was predicted to be insensitive to each of *k1r*, *k4r*, *k5r*, *k6f*, and *k6r* over the simulated range (data not shown).

From initialization ($t/t_{max} = 0$) through the preliminary stage of the reaction (i.e., the lag phase), predicted values of P were found to be rather sensitive to *k4f* and much less so to *k5f* (Fig. 8-5). This prediction suggests that, during this stage, most of the product formation occurred through *k4f* and was thus derived from *ES*, not *E'S*. Towards the steady-state, however, we have predicted the opposite to be true. As seen in Fig. 8-5, predicted values of P become increasingly sensitive to *k5f* and less so to *k4f* (though *k4f* still remains a significant parameter). Accordingly, this would suggest that, at steady-state (i.e., after the lag phase ends) more P is derived via the flux through *k5f* from E'S than is from ES. Overall, these predictions support the hypothesis that activation of the enzyme-substrate complex is an important step associated with the proposed mechanism and the function of this enzyme.

8.4 Discussion

Whereas the presented sensitivity analysis results are characteristic of the parameter values given in Table 8-1 (i.e., the location in the parameter space), the model and its sensitivity coefficients may also be solved for different sets of parameter values. Although we are free to explore the parameter space, we will confine our investigation to only physically relevant sets of parameters. From its initial value of 0.0001 s^{-1}, predictions of *P* remained almost completely insensitive to *k6f*. However, had *k6f* been 1000-times greater (i.e., estimated as 0.1 s^{-1}) we would have then found that *P* would be sensitive to *k6f* throughout the initial transient period, until approaching steady-state. From a mechanistic perspective, as *k6f* increases so does the rate of formation of *E'* which, in turn, can lead to increased rates of *E'S* formation (through *k3r*, see Fig. 8-1), and ultimately to *P* (by *k5f*). Thus, as the value of *k6f* increases, so does its importance relative to the formation of *P* (data not shown). On the other hand, if *k5f* were to be increased by a factor of 1000, for example, we would then find that predictions of *P* would be rendered essentially insensitive to the value of this parameter. In other words, *k5f* would become so large that small changes to its value would then have little overall effect on the flux through this step. Under such conditions it would also follow that the model would become increasingly sensitive to those reactions which supply the precursor *E'S*. Accordingly, our predictions suggest that the sensitivity of *P* to *k3r* and *k2f* would become greatly elevated under these conditions (data not shown).

No naturally-occurring glucose oxidase has yet been found which acts upon the C-6 hydroxyl group of glucose. Sun *et al.* generated glucose 6-oxidase (M-RQW) (Sun et al., 2002), but the kinetics of this mutant has never been studied intensively. We have

generated and kinetically analyzed a set of glucose-6-oxidases and found that the W290F mutation was highly correlated with an observable delay in the initiation of glucose oxidation. Whereas the exact cause of this lag phenomenon remains unknown, here we have proposed a mechanistic model to describe the enzyme kinetics and performed sensitivity analysis as the first step towards understanding this behavior. Our analyses have shown that the inter-conversion step between ES and E'S, which is responsible for activation of the enzyme-substrate complex, may be primarily responsible for the observed delay. Trp290 is known to protect the active site by restricting the entry of other molecules and by stabilizing the radical on Tyr272 (Baron et al., 1994; Saysell et al., 1997). When Trp290 is replaced by Phe, a smaller residue, we expect a more accessible and less stable radical site, resulting in a much slower accumulation of E'S (as predicted by a larger *k2r* or, alternatively, a smaller *k2f*). As illustrated in Fig. 8-2, the predicted results based on our proposed model mechanism are consistent with this explanation. It has also been observed that the W290H mutant displays almost the same value of K_m (on galactose) as that of the wild type GaOx, but that its k_{cat} value is more than three orders of magnitude lower than that of the wild type GaOx (Baron et al., 1994). The fact that the steady-state model predictions were found to be most sensitive to the parameters associated with enzyme turnover (i.e., *k4f* and *k5f* which are both related to k_{cat}) than those parameters associated with the equilibrium steps (i.e., *k1f*, *k1r*, *k3f*, and *k3r*, which are all related to K_m) might be connected to this experimental result. Similarly, the newly created GlcOx containing the W290F mutation (M-RQW and Des3-10) (Sun et al., 2002) showed a very low GaOx activity (Chapter 7). This phenomenon may likewise be explained by the co-existence of the two enzyme forms and the

increased destabilization of the E'S form by the W290F mutation, leading to slower turnover. Finally, the observation that by arbitrarily changing the equilibrium constant of Reaction 1 (Fig. 8-1, where K1 = *k1r/k1f*) from a nominal value of 250 to 1/250 μM imposed no significant effect to our predictions of the lag profile or the normalized sensitivity coefficients (data not shown) further supports our finding that the model predictions are less sensitive to the equilibrium steps associated with the substrate, enzyme, and substrate-enzyme complexes than the turnover steps.

Our interest in constructing a synthetic pathway to produce glucaric acid triggered the creation and characterization of GlcOx mutants (Chapter 7). As the first step toward understanding these new enzymes, we proposed a kinetic model and performed sensitivity analysis. With this model, we were able to explain the relationship between the lag phenomenon and one of the key amino acid residues for catalysis. Whereas detailed kinetic, spectroscopic and structural studies will be required to broaden our understanding of these newly created GlcOx, our current work would serve as the first blueprint to this end.

Nomenclature

Parameter	Units	Definition
kif, kir	s^{-1} or $\mu M^{-1} s^{-1}$	Forward and reverse rate constants associated with reaction step *i*
s_{ij}		Sensitivity coefficient associated with state variable *i* and parameter *j*
x		Model state variable
p		Model parameter
y		Model output or measured variable
t	s	Time
Θ_{lag}		Lag time
E	μM	Enzyme concentration, less active form
E'	μM	Enzyme concentration, activated form
S	μM	Substrate concentration
E'S	μM	Enzyme-substrate complex, activated form
ES	μM	Enzyme-substrate complex, less active form
P	μM	Product concentration

Subscripts	
0, f	Denote initial and final conditions, respectively
max	Denotes maximum value

Table 8-1. Parameter estimates used for solution of the model presented in Fig. 8-1.

Parameter	Estimate	Units
k1f	1	$\mu M^{-1} s^{-1}$
k1r	250	s^{-1}
k2f	0.05	s^{-1}
k2r	0.1	s^{-1}
k3f	50	s^{-1}
k3r	1	$\mu M^{-1} s^{-1}$
k4f	5	s^{-1}
k4r	0	$\mu M^{-1} s^{-1}$
k5f	50	s^{-1}
k5r	0	$\mu M^{-1} s^{-1}$
k6f	0.0001	s^{-1}
k6r	0.001	s^{-1}

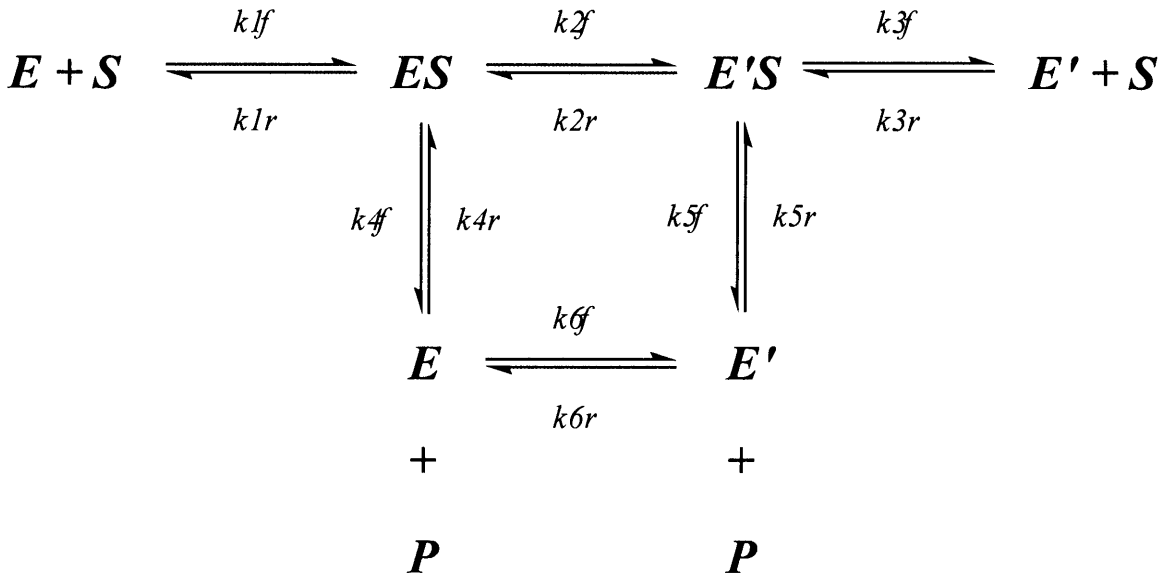


Fig. 8-1. A proposed model mechanism (Bates and Frieden, 1973) for GlcOx.

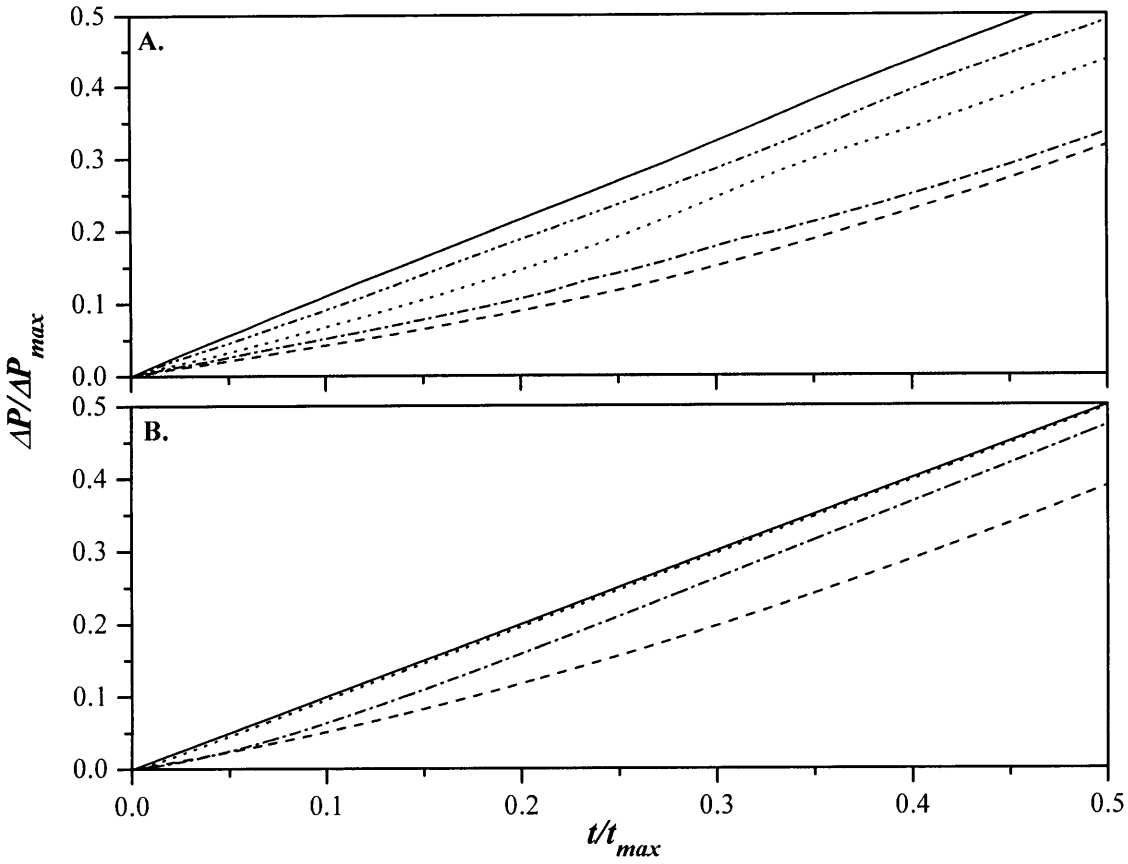


Fig. 8-2. Comparison of the normalized change in product concentration as a function of dimensionless time as obtained experimentally (A.) for Des3-10 (dashed), Des3-2 (dot-dashed), Des3-7 (solid), Des3-6 (dot-dot-dash) and M-RQW (dotted), and as predicted with the presented model (B.) for *k2f* values of 0.005 (dashed), 0.05 (dot-dashed), 0.5 (dotted), and 5 s^{-1} (solid).

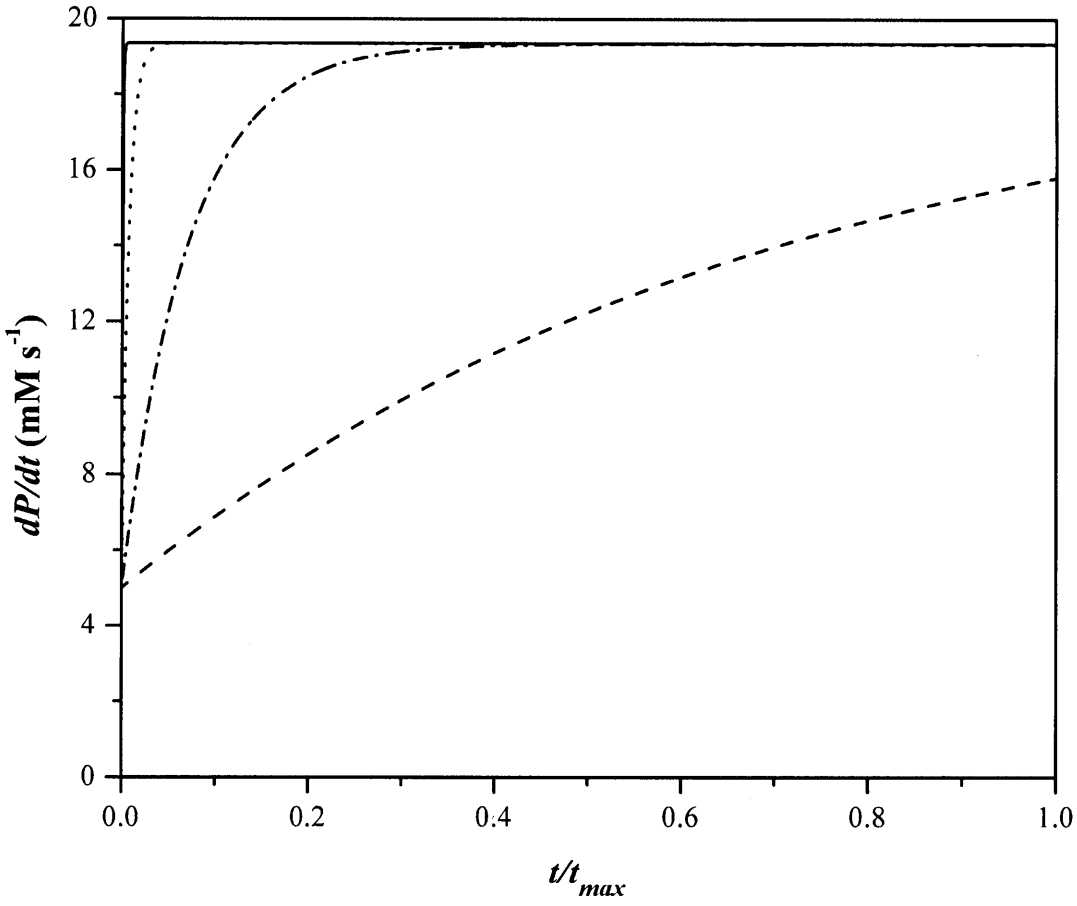


Fig. 8-3. The rate of product concentration (*P*) with respect to time predicted as a function of dimensionless time for *k2f* values of 0.005 (dashed), 0.05 (dot-dashed), 0.5 (dotted), and 5 s^{-1} (solid).

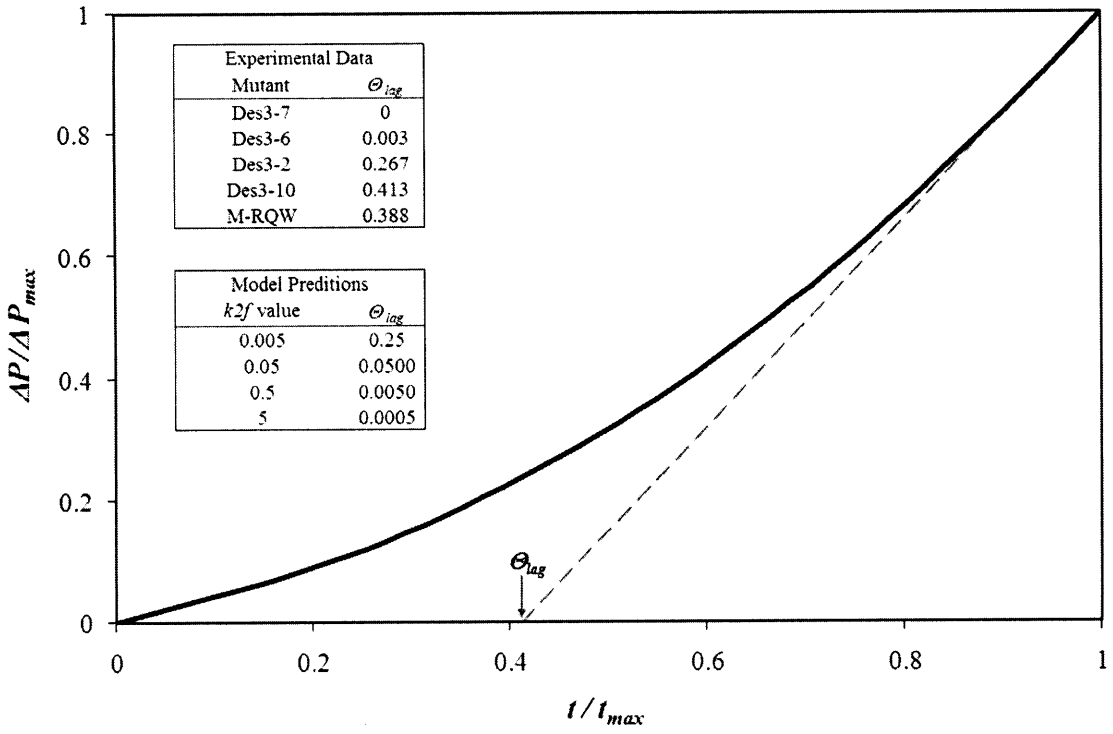


Fig. 8-4. Approximation of the lag time, Θ_{lag}, associated with mutant Des3-10. Inset tables summarize the experimental lag times associated with each of the studied mutants, as well as the lag times associated with the model predictions as approximated via the same procedure.

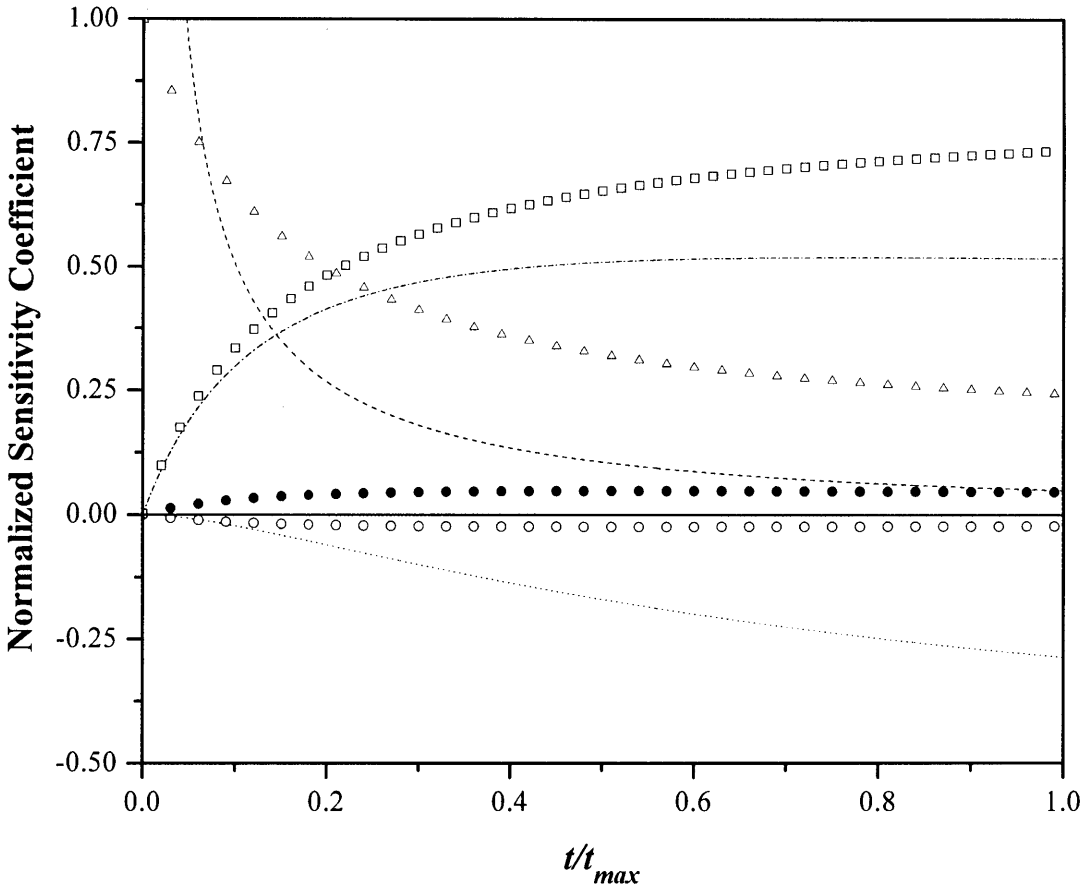


Fig. 8-5. Normalized sensitivity coefficients predicted as a function of dimensionless time for *k1f* (dashed line), *k2f* (dot-dash line), *k2r* (dotted line), *k3f* (open circles), *k3r* (solid circles), *k4f* (open triangles), and *k5f* (open squares).

Chapter 9

Conclusion and Recommendations

9.1 Summary

In this thesis, synthetic biology approaches were introduced to new pathway design and optimization problems. As a case study, glucaric acid pathways were designed and constructed, and parts/devices engineering was tried to improve the productivity of glucaric acid. Firstly, three enzymes were recruited from three different organisms to construct a synthetic pathway toward glucaric acid (Chapter 3 and 4). The newly constructed pathway showed flux imbalance due to the low activity of MIOX enzyme, and this conversion step was investigated to find that MIOX enzyme is activated by *myo*-inositol, its substrate. A simple and specific assay of D-glucuronate and MIOX activity was developed to better study this rate-limiting MIOX step (Chapter 5). Secondly, the synthetic glucaric acid pathway was equipped with engineered scaffold devices to improve pathway performance (Chapter 6). Various scaffold architectures were tested and 5-fold titer improvement was achieved. More importantly, the hypothesis that MIOX activity could be enhanced by co-localization of pathway enzymes was proven. Lastly, alternative pathways were designed by purely retrobiosynthetic approach. As the first step toward constructing one of these pathways (PW2 in Fig. 1-1), GlcOx variants were created from GaOx by combinational approach of computational design and experimental screening (Chapter 7 and 8). The newly created enzymes were characterized, and their kinetic mechanism was proposed and analyzed. Collectively, parts, devices, and chassis engineering were applied to a microbial chemical factory testbed, and this synthetic biology approach was proven to be effective for new pathway design and improvement.

9.2 Recommendations for Future Work

As synthetic biology is an evolving field, the above described projects also wait for several engineering approaches that can evolve the projects themselves. Firstly, the constructed glucaric acid pathway (PW1 in Fig. 1-1) could be further improved. As Voigt and coworkers demonstrated (Bayer et al., 2009), synthetic metagenomic approach would bring better enzyme parts for the pathway. Different sources of pathway enzymes were recruited and tested in the glucaric acid pathway, but better candidates (than currently recruited enzymes) were not found (personal communication with Sang-Hwal Yoon). However, these enzymes tested were not codon-optimized, and search was limited. Thus, systematic screening of candidate enzymes with the help of bioinformatics tools would lead to a bigger matrix of available enzymes and a higher productivity of glucaric acid. In addition, bioreactor optimization (e.g. aeration, glucose/*myo*-inositol feeding method, media composition, and induction condition) would result in higher titer and lower cost per gram of glucaric acid. Preliminary scale-up test using bioreactors (with 1L working volume) reproduced the productivity results from flask cultures. Given that preliminary flask culture showed that iron ion play an important role in production of glucaric acid in M9 minimal media, optimization of fermentation conditions including media composition would result in further productivity improvement and cost reduction. Furthermore, the overall flux could be increased by various ways including separate control of Ino1 and MIOX to find optimum expression levels, use of yeast chassis to increase expression of yeast enzyme Ino1 (thereby MIOX activity), global flux analysis

to find knockout targets and to achieve redox balance, and use of synthetic devices to better compete with endogenous pathways.

Secondly, the constructed pathway (PW1 in Fig. 1-1) could be extended and modified to produce the other molecules including galactaric acid and ascorbic acid. More imminently, the other glucaric acid pathways (PW2 ~ 5 in Fig. 1-1) should be constructed to show the potential of retrobiosynthesis. To this end, more parts engineering should be accompanied to generate new enzymes. At the same time, GlcOx could be further improved by error-prone PCR using the newly created GlcOx mutants as templates. Mechanism study on the GlcOx variants could be continued to better characterize these enzymes. Once new glucaric acid pathways are constructed, scaffold technology could be applied to the new pathways to minimize possible side reactions of unnatural intermediates and to prevent the evolved but still promiscuous enzymes from being used for the other conversions, generating toxic compounds and/or consuming key metabolites. The ultimate limit of retrobiosynthesis is enzyme availability, so advances in enzyme engineering will allow this approach to boom as synthetic organic chemistry flourished with rapid development of catalysts.

References

Addyman, R., et al., 1996. Urinary glucaric acid excretion in rheumatoid arthritis: influence of disease activity and disease modifying drugs. Ann Rheum Dis. 55, 478-81.

Adhikari, J., et al., 1987. Chloroplast as a locale of L-*myo*-inositol-1-phosphate synthase. Plant Physiol. . 85, 611-614.

Alper, H., et al., 2005. Tuning genetic control through promoter engineering. Proc Natl Acad Sci U S A. 102, 12678-83.

Amann, E., et al., 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene. 69, 301-15.

Arkin, A. P., Fletcher, D. A., 2006. Fast, cheap and somewhat in control. Genome Biol. 7, 114.

Arner, R. J., et al., 2004. Molecular cloning, expression, and characterization of *myo*-inositol oxygenase from mouse, rat, and human kidney. Biochem. Biophys. Res. Comm. 324, 1386-1392.

Arner, R. J., et al., 2001. *myo*-Inositol oxygenase: molecular cloning and expression of a unique enzyme that oxidizes *myo*-inositol and D-*chiro*-inositol. Biochem. J. 360, 313-320.

Arnold, F. H., et al., 1999. Directed evolution of mesophilic enzymes into their thermophilic counterparts. Ann N Y Acad Sci. 870, 400-3.

Arnold, F. H., Moore, J. C., 1997. Optimizing industrial enzymes by directed evolution. Adv Biochem Eng Biotechnol. 58, 1-14.

Arnold, F. H., et al., 2001. How enzymes adapt: lessons from directed evolution. Trends Biochem Sci. 26, 100-6.

Ashwell, A., et al., 1958. A new pathway of uronic acid metabolism. Biochim Biophys Acta. 30, 186-7.

Atsumi, S., et al., 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature. 451, 86-9.

Bailey, J. E., 1991. Toward a science of metabolic engineering. Science. 252, 1668-1675.

Barnett, J. E. G., et al., 1970. A colorimetric determination of inositol monophosphatases as an assay for D-glucose 6-phosphate-1L-*myo*-inositol 1-phosphate cyclase. Biochem. J. 119, 183-186.

Baron, A. J., et al., 1994. Structure and Mechanism of Galactose Oxidase - the Free-Radical Site. Journal of Biological Chemistry. 269, 25095-25105.

Bateman, D. F., et al., 1970. Purification and properties of uronate dehydrogenase from *Pseudomonas syringae*. Arch. Biochem. Biophys. 136, 97-105.

Bates, D. J., Frieden, C., 1973. Treatment of enzyme kinetic data. 3. The use of the full time course of a reaction, as examined by computer simulation, in defining enzyme mechanisms. J Biol Chem. 248, 7878-84.

Bayer, T. S., et al., 2009. Synthesis of methyl halides from biomass using engineered microbes. J Am Chem Soc. 131, 6508-15.

Benner, S. A., 2003. Synthetic biology: Act natural. Nature. 421, 118.

Benner, S. A., Sismour, A. M., 2005. Synthetic biology. Nat Rev Genet. 6, 533-543.

Bentley, W. E., et al., 1990. Plasmid-encoded protein: The principal factor in the metabolic burden associated with recombinant bacteria. Biotechnol. Bioeng. 35, 668-681.

Birnbaum, S., Bailey, J. E., 1991. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. Biotechnol. Bioeng. 37, 736-745.

Bitter, T., Muir, H. M., 1962. A modified uronic acid carbazole reaction. Anal. Biochem. 4, 330-334.

Bloom, J. D., et al., 2005. Evolving strategies for enzyme engineering. Curr Opin Struct Biol. 15, 447-52.

Bock, K. W., Kohle, C., 2005. UDP-glucuronosyltransferase 1A6: structural, functional, and regulatory aspects. Methods Enzymol. 400, 57-75.

Bolon, D. N., Mayo, S. L., 2001. Enzyme-like proteins by computational design. Proc Natl Acad Sci U S A. 98, 14274-9.

Bradford, M. M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Brown, P. M., et al., 2006. Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. Proc Natl Acad Sci U S A. 103, 15032-7.

Buell, C. R., et al., 2003. The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. Proc Natl Acad Sci U S A. 100, 10181-6.

Burack, W. R., Shaw, A. S., 2000. Signal transduction: hanging on a scaffold. Curr Opin Cell Biol. 12, 211-216.

Cello, J., et al., 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science. 297, 1016-8.

Cesaretti, M., et al., 2003. A 96-well assay for uronic acid carbazole reaction. Carbohyd. Polym. 54, 59-61.

Chang, Y. F., Feingold, D. S., 1969. Hexuronic acid dehydrogenase of Agrobacterium tumefaciens. J. Bacteriol. 99, 667-673.

Chang, Y. F., Feingold, D. S., 1970. D-glucaric acid and galactaric acid catabolism by *Agrobacterium tumefaciens*. J Bacteriol. 102, 85-96.

Charalampous, F. C., 1959. Biochemical studies on inositol. V. Purification and properties of the enzyme that cleaves inositol to D-glucuronic acid. J. Biol. Chem. 234, 220-227.

Charalampous, F. C., Lyras, C., 1957. Biochemical studies on inositol. IV. Conversion of inositol to glucuronic acid by rat kidney extracts. J. Biol. Chem. 228, 1-13.

Chen, C. Y., et al., 2009. Computational structure-based redesign of enzyme activity. Proc Natl Acad Sci U S A. 106, 3764-9.

Colombi, A., et al., 1983. Low-pH method for the enzymatic assay of D-glucaric acid in urine. Clin Chim Acta. 128, 337-47.

Cynkin, M. A., Ashwell, G., 1960. Uronic acid metabolism in bacteria. IV. Purification and properties of 2-keto-3-deoxy-D-gluconokinase in *Escherichia coli*. J Biol Chem. 235, 1576-9.

Dahiyat, B. I., 1999. In silico design for protein stabilization. Curr Opin Biotechnol. 10, 387-90.
Daniel, R., 2004. The soil metagenome--a rich resource for the discovery of novel natural products. Curr Opin Biotechnol. 15, 199-204.
Dean-Johnson, M., Henry, S. A., 1989. Biosynthesis of inositol in yeast: primary structure of *myo*-inositol-1-phosphate synthase (EC 5.5.1.4) and functional analysis of its structural gene, the *INO1* locus. J. Biol. Chem. 264, 1274-1283.
DeLisa, M. P., Conrado, R. J., 2009. Synthetic metabolic pipelines. Nat Biotech. 27, 728-729.
Dickinson, R. P., Gelinas, R. J., 1976. Sensitivity Analysis of Ordinary Differential Equation Systems - Direct Method. Journal of Computational Physics. 21, 123-143.
Dueber, J. E., et al., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol. 27, 753-759.
Duff, K., 2002. Calcium-D-glucarate. Altern Med Rev. 7, 336-9.
Ellis, L. B., et al., 2006. The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. Nucleic Acids Res. 34, D517-21.
Elowitz, M. B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature. 403, 335-8.
Endy, D., 2005. Foundations for engineering biology. Nature. 438, 449-53.
Farmer, J. J., 3rd, Eagon, R. G., 1969. Aldohexuronic acid catabolism by a soil Aeromonas. J Bacteriol. 97, 97-106.
Gardner, T. S., et al., 2000. Construction of a genetic toggle switch in Escherichia coli. Nature. 403, 339-42.
Georgiev, I., et al., 2008. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. J Comput Chem. 29, 1527-42.
Goodner, B., et al., 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science. 294, 2323-8.
Gordon, D. B., et al., 1999. Energy functions for protein design. Curr Opin Struct Biol. 9, 509-13.
Gukasyan, G. S., 2002. Study of the kinetics of oxidation of monophenols by tyrosinase. The effect of reducers. Biochemistry (Mosc). 67, 277-80.
Haerry, T. E., et al., 1997. Defects in glucuronate biosynthesis disrupt Wingless signaling in Drosophila. Development. 124, 3055-3064.
Handelsman, J., 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev. 68, 669-85.
Hansen, C. A., et al., 1999. Synthesis of 1,2,3,4-tetrahydroxybenzene from D-glucose: exploiting *myo*-inositol as a precursor to aromatic chemicals. J. Am. Chem. Soc. 121, 3799-3800.
Hatzimanikatis, V., et al., 2005. Exploring the diversity of complex metabolic networks. Bioinformatics. 21, 1603-9.
Hayes, R. J., et al., 2002. Combining computational and experimental screening for rapid optimization of protein properties. Proceedings of the National Academy of Sciences of the United States of America. 99, 15926-15931.

Horton, D., Walaszek, Z., 1982. Conformations of the D-Glucarolactones and D-Glucaric acid in Solution. Carbohydrate Research. 105, 95-109.
Hou, B. K., et al., 2003. Microbial pathway prediction: a functional group approach. J Chem Inf Comput Sci. 43, 1051-7.
Hubbard, B. K., et al., 1998. Evolution of enzymatic activities in the enolase superfamily: characterization of the (D)-glucarate/galactarate catabolic pathway in *Escherichia coli*. Biochemistry. 37, 14369-14375.
Hugouvieux-Cotte-Pattat, N., Robert-Baudouy, J., 1987. Hexuronate catabolism in *Erwinia chrysanthemi*. J Bacteriol. 169, 1223-31.
Humphreys, K. J., et al., 2009. Galactose oxidase as a model for reactivity at a copper superoxide center. J Am Chem Soc. 131, 4657-63.
Ibert, M., et al., 2002. Determination of the side-products formed during the nitroxide-mediated bleach oxidation of glucose to glucaric acid. Carbohydr Res. 337, 1059-63.
Ishidate, M., et al., 1965. Biochemical studies on glucuronic acid and glucaric acid. 1. Quantitative chemical determination of D-glucaric acid in urine. Anal Biochem. 11, 176-89.
Ito, N., et al., 1991. Novel thioether bond revealed by a 1.7 A crystal structure of galactose oxidase. Nature. 350, 87-90.
Ito, N., et al., 1994. Crystal structure of a free radical enzyme, galactose oxidase. J Mol Biol. 238, 794-814.
Jiang, L., et al., 2008. De novo computational design of retro-aldol enzymes. Science. 319, 1387-1391.
Jones, K. L., et al., 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab. Eng. 2, 328-338.
Jung, K., et al., 1981. Improved determination of D-glucaric acid in urine. Clin Chem. 27, 422-6.
Kane, J. F., Hartley, D. L., 1988. Formation of recombinant protein inclusion bodies in Escherichia coli. Trends in Biotechnology. 6, 95-101.
Kang, Y., et al., 2004. Systematic mutagenesis of the *Escherichia coli* genome. J Bacteriol. 186, 4921-30.
Keasling, J. D., 2008. Synthetic biology for synthetic chemistry. ACS Chem Biol. 3, 64-76.
Khosla, C., Keasling, J. D., 2003. Metabolic engineering for drug discovery and development. Nat. Rev. Drug Discov. 2, 1019-1025.
Kiely, D. E., et al., 1994. Simple preparation of hydroxylated nylons - polyamides derived from aldaric acids. ACS Sym. Ser. 575, 149-158.
Kilgore, W. W., Starr, M. P., 1959a. Catabolism of galacturonic and glucuronic acids by *Erwinia carotovora*. J Biol Chem. 234, 2227-35.
Kilgore, W. W., Starr, M. P., 1959b. Uronate oxidation by phytopathogenic pseudomonads. Nature. 183, 1412-3.
Kuellmer, V., 2001. Ascorbic acid. Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons.
Laakso, E. I., et al., 1983. Determination of D-glucaric acid by high-performance liquid chromatography. J Chromatogr. 278, 406-11.

Leonard, E., et al., 2008. Engineering microbes with synthetic biology frameworks. Trends Biotechnol. 26, 674-81.
Levchenko, A., et al., 2000. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proc Natl Acad Sci U S A. 97, 5818-23.
Levskaya, A., et al., 2005. Synthetic biology: engineering Escherichia coli to see light. Nature. 438, 441-2.
Li, C., et al., 2004. Computational discovery of biochemical routes to specialty chemicals. Chem. Eng. Sci. 59, 5051-5060.
Linster, C. L., Van Schaftingen, E., 2004. A spectrophotometric assay of D-glucuronate based on Escherichia coli uronate isomerase and mannonate dehydrogenase. Protein Expr. Purif. 37, 352-360.
Lippow, S. M., et al., 2007. Computational design of antibody-affinity improvement beyond in vivo maturation. Nat Biotechnol. 25, 1171-6.
Lovell, S. C., et al., 2000. The penultimate rotamer library. Proteins. 40, 389-408.
MacNeil, M. L., et al., 1986. A semiautomated procedure for urinary D-glucaric acid using a centrifugal analyzer. J Anal Toxicol. 10, 15-7.
Marsh, C. A., 1963. Metabolism of d-glucuronolactone in mammalian systems. Identification of d-glucaric acid as a normal constituent of urine. Biochem J. 86, 77-86.
Marsh, C. A., 1985. An enzymatic determination of D-glucaric acid by conversion to pyruvate. Anal Biochem. 145, 266-72.
Marsh, C. A., 1986. Biosynthesis of D-glucaric acid in mammals: a free-radical mechanism? Carbohydr Res. 153, 119-31.
Martin, V. J., et al., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. 21, 796-802.
Mata-Gilsinger, M., Ritzenthaler, P., 1983. Physical mapping of the *exuT* and *uxaC* operators by use of exu plasmids and generation of deletion mutants *in vitro*. J Bacteriol. 155, 973-82.
Matsuhisa, A., et al., 1995. Inositol monophosphatase activity from the *Escherichia coli suhB* gene product. J. Bacteriol. 177, 200-205.
Mauch, K., et al., 1997. Dynamic sensitivity analysis for metabolic systems. Chemical Engineering Science. 52, 2589-2598.
Mavrovouniotis, M. L., 1991. Estimation of standard Gibbs energy changes of biotransformations. J. Biol. Chem. 266, 14440-14445.
Mc, R. R., Novelli, G. D., 1958. Glucuronate metabolism by *Aerobacter aerogenes*. Nature. 182, 1504-5.
Mc, R. R., et al., 1959. Alduronic acid metabolism by bacteria. J Bacteriol. 77, 212-6.
Meinhold, P., et al., 2005. Direct conversion of ethane to ethanol by engineered cytochrome P450 BM3. Chembiochem. 6, 1765-8.
Menzella, H. G., et al., 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat Biotechnol. 23, 1171-6.
Merbouh, N., et al., 2001. Facile nitroxide-mediated oxidations of D-glucose to D-glucaric acid. Carbohydr Res. 336, 75-8.

Mocarelli, P., et al., 1988. A new method for D-glucaric acid excretion measurement that is suitable for automated instruments. Clin Chem. 34, 2283-90.
Molina, F. G., et al., 2007. An approximate analytical solution to the lag period of monophenolase activity of tyrosinase. Int J Biochem Cell Biol. 39, 238-52.
Monod, J., et al., 1965. On the Nature of Allosteric Transitions: a Plausible Model. J Mol Biol. 12, 88-118.
Moon, T. S., et al., 2009a. Production of glucaric acid from a synthetic pathway in recombinant Escherichia coli. Appl. Environ. Microbiol. 75, 589-595.
Moon, T. S., et al., 2009b. Enzymatic assay of d-glucuronate using uronate dehydrogenase. Anal Biochem. 392, 183-5.
Moore, G. L., Maranas, C. D., 2004. Computational challenges in combinatorial library design for protein engineering. Aiche Journal. 50, 262-272.
Morales, V. M., et al., 1991. A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. Gene. 97, 39-47.
Nakamura, C. E., Whited, G. M., 2003. Metabolic engineering for the microbial production of 1,3-propanediol. Curr. Opin. Biotechnol. 14, 454-459.
Neidhardt, F. C., Curtiss, R., 1996. *Escherichia coli* and *Salmonella* : cellular and molecular biology. ASM Press, Washington, D.C.
Niu, W., et al., 2003. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. J. Am. Chem. Soc. 125, 12998-12999.
Palumbo, A., et al., 1985. Effect of metal ions on the kinetics of tyrosine oxidation catalysed by tyrosinase. Biochem J. 228, 647-51.
Park, S. H., et al., 2003. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. Science. 299, 1061-1064.
Patnaik, P. R., 1994. Dynamic Sensitivity of a Chemostat for a Microbial Reaction with Substrate and Product Inhibition. Applied Mathematical Modelling. 18, 620-627.
Pawson, T., Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science. 300, 445-52.
Payne, W. J., Mc, R. R., 1958. Glucuronate isomerase from *Serratia marcescens*. Biochim Biophys Acta. 29, 466-7.
Pfeifer, B. A., et al., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science. 291, 1790-2.
Pfleger, B. F., et al., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol. 24, 1027-32.
Pitera, D. J., et al., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metab Eng. 9, 193-207.
Ponder, J. W., Richards, F. M., 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol. 193, 775-91.
Poon, R., et al., 1993. HPLC determination of D-glucaric acid in human urine. J. Anal. Toxicol. 17, 146-150.
Portalier, R. C., et al., 1974. [Studies of mutations in the uronic isomerase and altronic oxidoreductase structural genes of *Escherichia coli* K 12 (author's transl)]. Mol Gen Genet. 128, 301-19.
Postma, P. W., et al., 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev. 57, 543-94.

Prather, K. L., Martin, C. H., 2008. De novo biosynthetic pathways: rational design of microbial chemical factories. Curr Opin Biotechnol. 19, 468-74.
Purnick, P. E., Weiss, R., 2009. The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol. 10, 410-22.
Ran, N., Frost, J. W., 2007. Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase. J Am Chem Soc. 129, 6130-9.
Reddy, C. C., et al., 1981a. *myo*-Inositol oxygenase from hog kidney. II. Catalytic properties of the homogeneous enzyme. J. Biol. Chem. 256, 8519-8524.
Reddy, C. C., et al., 1981b. *myo*-Inositol oxygenase from hog kidney. I. Purification and characterization of the oxygenase and of an enzyme containing the oxygenase and D-glucuronate reductase. J. Biol. Chem. 256, 8510-8518.
Ro, D. K., et al., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 440, 940-3.
Robert-Baudouy, J., et al., 1982. D-Mannonate and D-altronate dehydratases of *Escherichia coli* K12. Methods Enzymol. 90 Pt E, 288-94.
Robert-Baudouy, J. M., Stoeber, F. R., 1973. [Purification and properties of D-mannonate hydrolyase from *Escherichia coli* K12]. Biochim Biophys Acta. 309, 473-85.
Roberton, A. M., et al., 1980. Two genes affecting glucarate utilization in *Escherichia coli* K12. J Gen Microbiol. 117, 377-82.
Rogers, M. S., et al., 2007. The stacking tryptophan of galactose oxidase: a second-coordination sphere residue that has profound effects on tyrosyl radical behavior and enzyme catalysis. Biochemistry. 46, 4606-18.
Rondon, M. R., et al., 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol. 66, 2541-7.
Rothlisberger, D., et al., 2008. Kemp elimination catalysts by computational enzyme design. Nature. 453, 190-5.
Sambrook, J., Russell, D. W., 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Sarma, P. S., Sastry, K. S., 1957. Glucuronic acid, a precursor of ascorbic acid in Aspergillus niger. Nature. 179, 44-45.
Saven, J. G., 2002. Combinatorial protein design. Current Opinion in Structural Biology. 12, 453-458.
Saysell, C. G., et al., 1997. Properties of the Trp290His variant of Fusarium NRRL 2903 galactose oxidase: interactions of the GOase(semi) state with different buffers, its redox activity and ability to bind azide. Journal of Biological Inorganic Chemistry. 2, 702-709.
Schilstra, M. J., et al., 1993. Kinetic analysis of the induction period in lipoxygenase catalysis. Biochemistry. 32, 7686-91.
Simmons, C. J., et al., 1974. Urinary D-glucaric acid assay by an improved enzymatic procedure. Clin Chim Acta. 51, 47-51.
Singh, J., Gupta, K. P., 2003. Calcium glucarate prevents tumor formation in mouse skin. Biomed Environ Sci. 16, 9-16.

Singh, J., Gupta, K. P., 2007. Induction of apoptosis by calcium D-glucarate in 7,12-dimethyl benz [a] anthracene-exposed mouse skin. J Environ Pathol Toxicol Oncol. 26, 63-73.

Steinberg, K. K., et al., 1985. Assessment of 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure using a modified D-glucaric acid assay. J Toxicol Environ Health. 16, 743-52.

Stemmer, W. P., 1994a. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A. 91, 10747-51.

Stemmer, W. P., 1994b. Rapid evolution of a protein in vitro by DNA shuffling. Nature. 370, 389-91.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metab Eng. 1, 1-11.

Stephanopoulos, G., Jensen, K. L., 2005. Metabolic engineering: Developing new products and processes by constructing functioning biosynthetic pathways *in vivo*. AIChE Journal. 51, 3091-3093.

Sun, L. H., et al., 2002. Modification of galactose oxidase to introduce glucose 6-oxidase activity. Chembiochem. 3, 781-783.

Sun, L. H., et al., 2001. Expression and stabilization of galactose oxidase in Escherichia coli by directed evolution. Protein Engineering. 14, 699-704.

Tabor, J. J., et al., 2009. A synthetic genetic edge detection program. Cell. 137, 1272-81.

Tian, J., et al., 2004. Accurate multiplex gene synthesis from programmable DNA microchips. Nature. 432, 1050-4.

Tracewell, C. A., Arnold, F. H., 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol. 13, 3-9.

Tyo, K. E., et al., 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol. 27, 760-5.

Tyo, K. E., et al., 2007. Expanding the metabolic engineering toolbox: more options to engineer cells. Trends Biotechnol. 25, 132-7.

Vizcarra, C. L., Mayo, S. L., 2005. Electrostatics in computational protein design. Curr Opin Chem Biol. 9, 622-6.

Voigt, C. A., et al., 2002. Protein building blocks preserved by recombination. Nature Structural Biology. 9, 553-558.

Wachter, R. M., Branchaud, B. P., 1996. Molecular modeling studies on oxidation of hexopyranoses by galactose oxidase. An active site topology apparently designed to catalyze radical reactions, either concerted or stepwise. Journal of the American Chemical Society. 118, 2782-2789.

Wagner, G., Hollmann, S., 1976. Uronic acid dehydrogenase from *Pseudomonas syringae*. Purification and properties. Eur. J. Biochem. 61, 589-596.

Walaszek, Z., 1990. Potential use of D-glucaric acid derivatives in cancer prevention. Cancer Lett. 54, 1-8.

Walaszek, Z., et al., 1996. D-Glucaric acid content of various fruits and vegetables and cholesterol-lowering effects of dietary -glucarate in the rat. Nutrition Research. 16, 673-681.

Wang, Z. X., et al., 1993. Kinetic evaluation of substrate-dependent origin of the lag phase in soybean lipoxygenase-1 catalyzed reactions. Biochemistry. 32, 1500-9.

Werpy, T., Petersen, G., 2004. Top value added chemicals from biomass. Volume I: Results of screening for potential candidates from sugars and synthesis gas. N. R. E. L. (NREL) and P. N. N. L. (PNNL) (ed.).

Whittaker, M. M., et al., 1998. Kinetic isotope effects as probes of the mechanism of galactose oxidase. Biochemistry. 37, 8426-36.

Winterbourn, C. C., et al., 2002. Thiol oxidase activity of copper, zinc superoxide dismutase. J Biol Chem. 277, 1906-11.

Wong, T. S., et al., 2007. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. Environmental Microbiology. 9, 2645-2659.

Xing, G., et al., 2006a. Oxygen activation by a mixed-valent, diiron(II/III) cluster in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. Biochemistry. 45, 5402-12.

Xing, G., et al., 2006b. Evidence for C-H cleavage by an iron-superoxide complex in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. Proc Natl Acad Sci U S A. 103, 6130-5.

Xing, G., et al., 2006c. A coupled dinuclear iron cluster that is perturbed by substrate binding in myo-inositol oxygenase. Biochemistry. 45, 5393-401.

Xue, C. B., et al., 2007. Inhibition kinetics of cabbage butterfly (Pieris rapae L.) larvae phenoloxidase activity by 3-hydroxy-4-methoxybenzaldehyde thiosemicarbazone. Appl Biochem Biotechnol. 143, 101-14.

Yebra, M. J., et al., 2007. Identification of a gene cluster enabling *Lactobacillus casei* BL23 to utilize myo-inositol. Appl Environ Microbiol. 73, 3850-8.

Yeh, B. J., Lim, W. A., 2007. Synthetic biology: lessons from the history of synthetic organic chemistry. Nat. Chem. Biol. 3, 521-525.

Yoon, S. H., et al., 2009. Cloning and characterization of uronate dehydrogenases from two pseudomonads and Agrobacterium tumefaciens strain C58. J. Bacteriol. 191, 1565-1573.

Yoshida, K., et al., 2008. myo-Inositol Catabolism in *Bacillus subtilis*. J Biol Chem. 283, 10415-24.

Zajic, J. E., 1959. Hexuronic dehydrogenase of *Agrobacterium tumefaciens*. J Bacteriol. 78, 734-5.

Zanghellini, A., et al., 2006. New algorithms and an in silico benchmark for computational enzyme design. Protein Sci. 15, 2785-94.

Zha, W., et al., 2004. Rational pathway engineering of type I fatty acid synthase allows the biosynthesis of triacetic acid lactone from D-glucose in vivo. J Am Chem Soc. 126, 4534-5.