Bioretrosynthetic Construction of a Non-Natural Nucleoside Analog Biosynthetic Pathway

By

William Ross Birmingham

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Biochemistry

December, 2013

Nashville, Tennessee

Approved:

Brian O. Bachmann

Richard N. Armstrong

Tina M. Iverson

Lawrence J. Marnett

Carmelo J. Rizzo

To my wife, Annie, and my family,

for their continued love, support and encouragement.

ACKNOWLEDGEMENTS

I am appreciative for the positive influence many people have had on my life. I know that any measure of thanks stated here can't begin to do justice to the amount of gratitude that needs to be expressed, however, I can at least duly recognize many who have helped me along my way. It has been a privilege and greatly rewarding experience to have attended Vanderbilt University for my graduate studies. Without the support of the Departments of Biochemistry and Chemistry through an awarded Chemical Biology Interface Training Grant T32 GM065086, Organic Chemistry and General Chemistry Teaching Assistant positions and the D. Stanley and Ann T. Tarbell Endowment Fund, my research and this dissertation would not have been possible.

First and foremost, I would like to thank my wife for her constant support during my hectic years of graduate school. Annie's love, encouragement and cheerful nature made difficult research days seem not so bad in the grand scheme of things. Her smile and laugh always remind me how lucky I am to have her in my life. I cannot thank her enough for being understanding and patient beyond measure and for her bravery in following me to Nashville with her only connection being family friends who generously opened their home and made us feel like part of the family. For that, a very special thanks is also due to Keith and Kay Simmons for welcoming us to Nashville.

I owe a great deal of gratitude to my advisor, Dr. Brian O. Bachmann. This specific project that became the focus of my dissertation research was what initially drew me to Vanderbilt for graduate school. I thank him for allowing me to take up the reigns in this work on enzyme engineering and further my interests in biocatalysis. His scientific insight from a broad perspective of organic synthesis, biosynthesis and directed evolution has developed my scientific interests for future pursuits. Brian's guidance in making me a better researcher, writer, mentor and scientific thinker as well as his

excited interest in progressing this work and my career has been much appreciated over the past several years. In addition to Dr. Bachmann, I would also like to thank my Dissertation Committee members of Dr. Richard Armstrong, Dr. Tina Iverson, Dr. Lawrence Marnett and Dr. Carmello Rizzo for their valued discussions and for offering their time and energy in assisting my training as a graduate student.

I would additionally like to thank Dr. Iverson and her graduate students Dr. Tim Panosian and Crystal Starbird for their excellent collaborative work on the structural aspects related to my dissertation project. Tim and Chrystal determined the structures of wild-type and variant forms of phosphopentomutase that enabled a better understanding of the enzyme and the effects of mutagenesis. Tim also worked closely with Dr. David Nannemann, an alumnus of the Bachmann research group, on the biochemical characterization of phosphopentomutase prior to me joining the lab which allowed me to start directly with assay development and mutagenesis. The structural and mechanistic insights gained through their work on phosphopentomutase were invaluable for my directed evolution experiments. In David's primary research aims, he set the foundation for this biocatalytic project as a whole through his evolution of purine nucleoside phosphorylase for use in our proposed engineered biosynthetic pathway. For me, David was also a friend and graduate student mentor, actively answering my questions, discussing interesting enzyme engineering thoughts and showing me the ropes for designing protocols for experiments in directed evolution and biosynthesis assays. I am thankful for him passing along his knowledge and experience.

To the other members of the Bachmann group, thank you for creating an enjoyable research environment. Your camaraderie and sense of community over the years has been greatly appreciated. From organizing group activities (outside of lab even!) to continuing the search for Waldo, you all have made graduate school much more than just a research experience and I thank you for that.

I would be completely remiss if I did not thank two teachers in particular that first inspired and cultivated my interest in science. Pursuing a career in science has been a goal of mine ever since high school. I can attribute the major influence for that decision to my Advanced Placement Chemistry teacher, Mrs. Linda Pennington. She is truly a rare, inspirational teacher who not only made chemistry fun in learning, but also whetted my interest to continue studying. I left for college set on obtaining a degree in chemistry and had my first academic research experience under the mentoring of Dr. Rebecca Alexander at Wake Forest, studying the mechanism of *E. coli* methionyl-tRNA synthetase substrate recognition and catalysis. In the process, I found that my interests grew in the direction of biochemistry, specifically toward what I learned to be the growing field of synthetic biology. I wanted to research laboratory evolution of enzymes to catalyze new reactions and create new products, which ultimately lead me to join Dr. Bachmann's group at Vanderbilt. I really can't thank Mrs. P. and Dr. Alexander enough for shaping my interest in science and showing me how exciting and rewarding it can be.

Last, but never least, I would like to thank my family and friends. Thank you for everything. Your never ending love and support have made me who I am. I am grateful for everything from your encouraging my work to helping me pull away for a while to keep perspective. I could not have done this without you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter I

BIOCATALYST AND BIOSYNTHETIC PATHWAY ENGINEERING

Introduction

Biocatalysis is defined as the implementation of natural catalysts, such as enzymes or whole cells, to perform synthetic chemical reactions[1]. As a subset of synthetic biology, which in general aims to employ biological functions to perform a variety of designated tasks[2], these enzymes or cells are commonly engineered to be repurposed for applications and activities that they were not evolved to perform[1]. These new functional capacities are often instilled into the catalysts through directed evolution, a laboratory process that mimics Darwinian selection, whereby mutations are introduced into a gene sequence to allow the gain or loss of certain traits in the corresponding encoded protein. These mutations serve to evolve the enzyme toward a defined goal, which is guided along the way by careful experimental control of the process, hence the term directed evolution.

Biocatalysis offers an attractive alternative to chemical production processes. In many chemical routes, great attention and planning is invested in the optimization of synthetic schemes to enantioselectively produce a desired compound. However, traditional large scale synthetic methods are often met with low yields from incomplete conversion or formation of side products, requiring purification and generating high solvent waste volumes, all of which contribute to increased costs. Biocatalysts provide benefits from environmental and economic standpoints as they are made from renewable resources, are biodegradable and non-toxic, and the high reaction selectivity and partially aqueous reaction conditions can increase yields and reduce waste production and subsequent processing requirements. Additionally, reactions are

commonly performed at near ambient temperature, atmospheric pressure and relatively neutral pH and are therefore commonly safer and more easily managed than some corresponding chemical processes. For these reasons, both engineered and natural biocatalysts have been successfully applied to the synthesis of a variety of biofuels, drug intermediates, active pharmaceutical ingredients and both fine and commodity chemicals. However, as the majority of important synthetic compounds are not natural products, they cannot be created through native cellular metabolism. Instead, research efforts are invested in the development of individual enzymes and/or full biosynthetic pathways capable of forming these non-natural products.

For example, a small molecule transaminase was evolved to accept a considerably larger substrate through a process known as ‘substrate walking’ by engineering activity on successively larger substrates until reaching the desired starting material[3]. Further engineering by directed evolution increased activity ~40,000-fold and produced an enzyme that is able to replace the transition-metal based process of amination and asymmetric hydrogenation for production of the antidiabetic drug sitagliptin. The evolved enzyme was engineered to stereoselectively aminate a ketone substrate. The resulting biocatalytic process increased overall yield, productivity and enantioselectivity, reduced total waste volume and eliminated the need for transition-metal catalysts and is currently used in industrial scale production of sitagliptin[3, 4]. This one instance highlights many of the potential benefits of performing synthesis using biocatalysts.

In this dissertation, we present a biosynthetic pathway construction paradigm based on a model of natural pathway evolution and describe the evolution of multiple enzymes joined together to create an engineered biosynthetic pathway for production of the nucleoside analog reverse transcriptase inhibitor 2’,3’-dideoxyinosine (dideoxyinosine, ddI, didanosine). Dideoxyinosine is one of the many nucleoside analogs

prescribed as treatment for patients infected with human immunodeficiency virus (HIV). Currently, the World Health Organization estimates the price of dideoxyinosine to be US$ 243 per patient per year of treatment for low-income (less than US$ 1,025 gross national income per capita) and middle-income countries (US$ 1,026-4,035)[5]. These regions account for approximately 69% of the estimated 34 million people worldwide infected with this virus, and therefore this drug and other similar treatment options are unaffordable to the vast majority of those affected[6].

A major factor affecting the high cost of many of these nucleoside analog drugs is the synthetic process to manufacture the active pharmaceutical ingredient. For dideoxyinosine, manufacturing the active ingredient is approximately 66% of the direct costs, which is itself ~75% of the price per year of treatment, and these figures hold true for many similar nucleoside analogs as well[7]. The reason for the production process contributing such a high percentage of the final cost is due to the measures required for synthesis and purification of the active ingredient, primarily involved to retain the proper stereochemistry in the final product. There have been several strategies attempted to approach this hurdle, however, each have been met with different flaws. Natural nucleosides, such as inosine and adenosine, can be used as starting materials to produce dideoxyinosine and benefit from the appropriate stereochemistry already being in place in the initial compound. However, these materials are quite expensive even in bulk quantities and consequently drive up the cost of production. The other main production scheme involves a separate synthesis of the non-natural sugar activated at the C1 anomeric position, followed by attachment of the suitable nucleobase. In the case of dideoxynucleosides, the dideoxy-sugar moiety can be easily produced from glutamic acid, a very inexpensive bulk commodity chemical. On the other hand, racemic activation of the anomeric center by installation of a leaving group and subsequent displacement through addition of the nucleobase leads to the production of several

structural and regioisomers that ultimately reduce yield of the desired active nucleoside analog drug and increase costs through necessary added purification steps. A step in the direction of a potentially more efficient production method has incorporated a biocatalytic transformation to form the final product, however these methods still rely heavily on a chemical synthetic component to generate the substrate for the enzymatic reaction[8-12].

To reduce the synthetic burden in production, we propose to employ a short and simple chemical synthesis route to produce a non-natural sugar that then becomes a substrate for a sequence of enzymatic biotransformations to produce dideoxyinosine. Dideoxyribose, the proposed non-natural sugar, can be produced in three steps from the inexpensive starting material glutamic acid before a series of enzymes recruited from natural nucleoside biosynthesis catalyze the production of dideoxyinosine in a stereoselective manner (Figure 1-1). Taking advantage of the high stereoselectivity of enzymes enables the production of one properly activated C1 anomer of the dideoxysugar for subsequent addition of the nucleobase, and in turn could likely reduce sample processing and purification and may also increase the total yield of the active product. Combining these benefits, a biocatalytic or biosynthetic production route could directly affect the cost of production of many industrially relevant compounds, and in

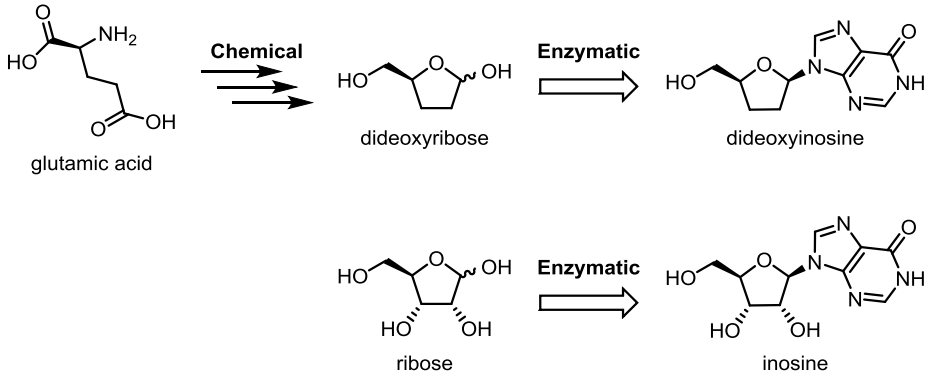

Figure 1-1. Proposed semisynthetic route from glutamic acid to dideoxyinosine and comparison of the non-natural compounds dideoxyribose and dideoxyinosine to the natural compounds ribose and inosine.

cases where manufacturing is a leading contributor to consumer pricing, such as this class of nucleoside analogs, their implementation could significantly reduce the final cost of the compound and increase overall availability.

In order to enable a biocatalytic process to be a relevant industrial method for manufacturing non-natural products, the productivity and turnover rate must be high enough to meet production demands. However, since naturally occurring enzymes are rarely capable of producing sufficient quantities of the desired non-natural product, they are frequently engineered to increase activity to meet required industrial scale production titers. Here, we describe the engineering of enzymes to catalyze a series of reactions on dideoxy-sugar substrates in the biosynthesis of the anti-HIV drug dideoxyinosine.

Enzyme Engineering for Biocatalysis

Although the use of biocatalysts provides many advantages over chemical synthetic routes, such as environmentally friendly conditions, high efficiency and extraordinary regio-, chemo- and enantio-selectivities[13], successful implementation of these processes are not without their difficulties. Most naturally occurring, or wild-type, enzymes have normally evolved over time to be quite substrate selective or maybe, at best, able to accept a small range of chemically and structurally similar compounds as substrates at reduced activity levels. While this is highly beneficial for living biological systems, it is frequently a hindrance to their use in biocatalytic applications since many desired transformations utilize non-natural compounds. Furthermore, the limited stability of enzymes to a narrow window of reaction conditions and a common requirement of expensive cofactors limits their use in large scale preparative reactions.

For these reasons, in order to utilize a biocatalytic component to replace a chemical step within a production process, a compromise must be made. In the past,

production conditions were tailored to the limitations of the enzyme. Now, enzymes are able to be engineered to fit the needs of the desired reaction and reaction conditions, such as the presence of organic co-solvents, non-neutral pH and very high substrate and product concentrations[1]. Typically, efforts are focused to increase the rate of catalysis on the target substrate under specific conditions through an iterative process of mutagenesis and screening for activity in an effort to evolve the enzymes for the new reaction parameters.

This practice of enzyme engineering frequently utilizes a two pronged approach, combining rational or targeted mutagenesis methods with random mutagenesis techniques to create libraries from which new biocatalysts with improved functions can be discovered through activity screening. As the first step, a progenitor enzyme must be identified for the reaction of interest. Part of the process to select a progenitor can involve biochemical characterization of few to many enzyme homologs and scaffolds to identify a variant that shows measurable turnover on the desired substrate. Selecting a progenitor panel may be as extensive as collecting all, or as many as possible, suitable enzymes (that is, those that catalyze the desired chemical transformation) that have been reported in recent literature, or may be more limited to only those that have been structurally characterized or possibly only variants that exhibit a degree of substrate promiscuity for natural and/or non-natural molecules. This last trait can be extremely valuable, as naturally present substrate promiscuity is thought to be important for engineering new activities[14].

In addition to demonstrating desirable activity, the availability of structural data for an enzyme of interest is highly beneficial. In order to perform targeted mutagenesis, some degree of knowledge of the enzyme active site must be determined, either via a crystal structure of the enzyme or through homology modeling with a related enzyme. This information can be used to identify first or second shell active site residues to target

for mutagenesis to analyze the respective contributions to activity through screening mutagenesis libraries[15, 16]. Mutations at sites directly contributing to substrate binding, the so-called first shell, or residues responsible for their positioning, the second shell, have been shown to provide the greatest changes in enantioselectivity, substrate selectivity and altered catalytic activity, all of which are highly valued traits of engineered enzymes for biocatalytic processes[17].

Following targeted mutagenesis approaches, techniques to introduce random mutations are then used to produce mutations throughout the gene. Because of the arbitrary addition of mutations at any point within the gene sequence, these methods can be used in the absence of structural data. Screening libraries created using these types of methods commonly results in mutations at positions that are unlikely to be predicted to have a beneficial effect on the desired activity. These sites frequently tend to be outside of the active site, accounting for why they may go unpredicted, as the active site normally comprises only a small portion of the sequence of the entire enzyme. Nonetheless, random mutagenesis methods seem to be the most commonly employed in biocatalyst development through directed evolution and are still very successful tools in enzyme engineering[18].

The use of directed evolution to evolve an enzyme for improved biocatalytic functions is a laboratory parallel to the natural selection of 'survival of the fittest,' but can be applied on the much shorter timescale of weeks to months. However, rather than survival advantage being the individual selection pressure to guide mutations as seen in nature, increases in the specific experimenter-defined activity conferred by the newly acquired mutation lead to selection to continue the enzymatic lineage. This activity may still provide a survival advantage, such as the degradation of a toxic molecule, but is often simply the most productive enzyme identified under experimental *in vitro* reaction conditions. The process of mutagenesis and selection can be iterated until the required

level of activity is gained, or until no further activity benefits can be imparted. An overview of mutagenesis methods and their application to enzyme engineering for biocatalytic process is provided below.

Targeted mutagenesis methods

After identifying the enzyme progenitor to be used in the directed evolution study, the first step is to create a pool of genetically diverse gene sequences. When structural data is available, mutagenesis commonly begins with a targeted approach to take advantage of known or predicted functional roles of specific substrate binding residues and active site architecture during the process of repurposing an enzyme for application in a biocatalytic process. Such targeted approaches, or 'smart libraries', reduce library size by targeting the points of gene diversity to known interactions between residues and specific functional groups of the substrate to reduce selectivity and improve substrate promiscuity and can possibly increase the likelihood of success[(19)].

One such method of introducing mutations in a targeted approach is saturation mutagenesis[(20)] (Figure 1-2). In this process, mutagenic primers containing a randomized sequence (commonly NNK or NNS, where N=A, T, C or G; K=T or G and S=C or G) at one or more codons in the primer are used to amplify a gene/plasmid of interest. The randomized codon sequence at specific locations creates a library of variants containing all 20 possible amino acids at the indicated position, giving full coverage of mutational analysis at the target residue. Since this approach is most effective when targeting residues known to contact the substrate, saturation mutagenesis is often used as an initial phase of enzyme engineering. O'Conner and coworkers applied saturation mutagenesis using the NNK codon at several positions that interacted with bound tryptamine in the active site of strictosidine synthase[(21)]. Screening of the resulting libraries provided an assortment of enzyme variants what were able to catalyze the

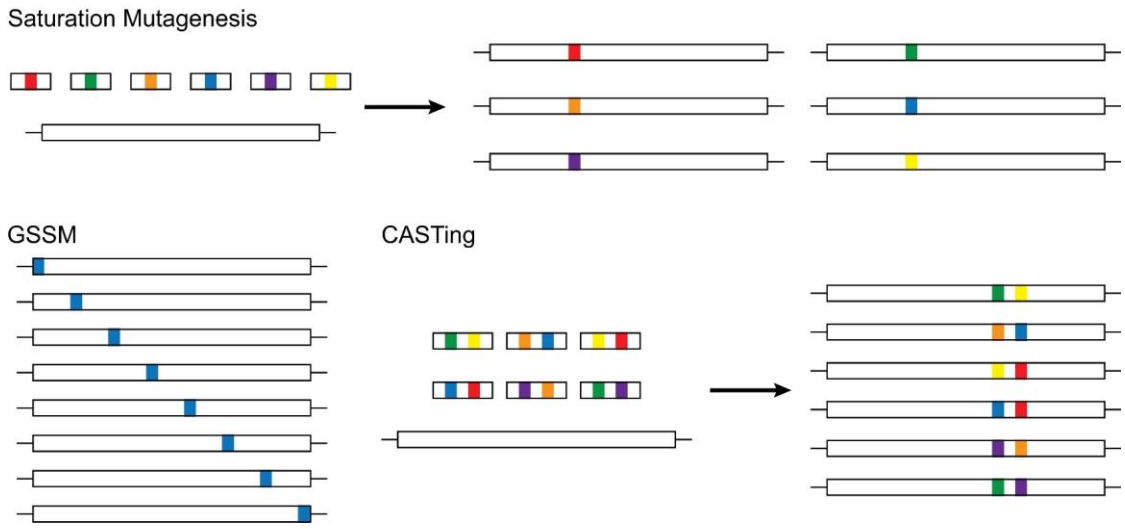

Figure 1-2. Methods for targeted gene mutagenesis. These methods allow for analysis of one or more specific positions in the enzyme through directly targeting locations of interest. GSSM: Gene Site Saturation Mutagenesis, CASTing: Combinatorial Active Site Testing.

Pictet-Spengler reaction using non-natural tryptamine analogs to produce a variety of new strictosidine analogs, several of which were halogenated and may be further derivatized through chemical strategies[21].

The basic premise of saturation mutagenesis can be applied in a variety of ways. For example, rather than targeting mutagenesis to a select few positions in the active site, every residue in a protein can be individually randomized in a process known as 'gene site saturation mutagenesis', or GSSM™ (trademark Diversa Co.) (Figure 1-2). While GSSM™ is not as commonly used as other methods, due to the generation of large libraries that are not focused for specifically targeted residues, one benefit is that this method can be used in the absence of structural data as all positions undergo mutational analysis by a randomized codon. GSSM™ was indeed shown to be an effective method to engineer an *R*-selective nitrilase for product formation in high enantiomeric excess and capable of catalyzing the hydrolysis reaction at substrate concentrations up to 3 M[22]. Under final conditions, a single mutant nitrilase variant identified through GSSM™ was capable of converting 3-hydroxyglutaryl nitrile to the

corresponding (*R*)-4-cyano-3-hydroxybutyric acid in 96 % yield at 98.5 % entiomeric excess in 15 hours[22].

Saturation mutagenesis can also be used iteratively, targeting a new site after identification of an advantageous mutation at a previous location. Reetz and coworkers combined iterative saturation mutagenesis with their previously published method known as the Combinatorial Active Site Test (CAST)[16] in engineering enantioselectivity of a lipase from *Pseudomonas aeruginosa* for a chiral ester[23]. CASTing involves the simultaneous saturation mutagenesis of multiple residues using one primer containing a randomized codon sequence at each of the target positions and can identify synergistic mutations where both residues are required to be mutated in order to be effective[16] (Figure 1-2). Demonstrating the value of these methods used together, the lipase was engineered over only two rounds of iterative CASTing to generate a variant with an enantioselective (E-value) factor of 594, which is 540-fold greater than the wild-type enzyme, after screening 10,000 total library members[23]. By comparison, a lipase variant generated using a combination of standard methods of error prone polymerase chain reaction (epPCR), saturation mutagenesis and DNA shuffling provided an E-value of only 51 after four rounds of mutagenesis with screening of approximately 50,000 transformants.

Another method of protein engineering involves applying a computational based protein sequence-activity relationship (ProSAR) algorithm to analyze experimental activity results from mutagenesis libraries. Functional data is collected from a library of enzyme variants containing multiple mutations, and a key aspect is that each mutation must be present in multiple separate clones[24]. The activity data is analyzed by the algorithm and each mutation is given a probability that the residue change is beneficial, neutral or detrimental to the desired activity. Each mutation is then sorted by the predicted effect, and all beneficial mutations are combined and incorporated into the

gene sequence to produce a new protein template. Codexis applied this ProSAR method in the evolution of a halohydrin dehalogenase to catalyze the conversion of ethyl (*S*)-4-chloro-3-hydroxybutyrate to ethyl (*R*)-4-cyano-3-hydroxybutyrate for use as a starting material in the production of the side chain of the cholesterol lowering drug atorvastatin[(25)]. Sequence diversity was generated using a combination of random mutagenesis, site saturation mutagenesis, gene shuffling and through analysis of the enzyme structure and of homologous enzyme sequences. The experimental data for activity of the variants under process conditions was then analyzed by the ProSAR algorithm after each round to suggest the most beneficial mutations that should be combined in the template for the following round of mutagenesis. The final variant after 18 rounds of evolution contained a total of 35 mutations and produced the desired compound at 99.5% purity, >99.9% enantiomeric excess and high yield at a productivity rate 4,000-fold greater than the wild-type enzyme[(25)] and is currently used in large scale industrial production of the atorvastatin side chain.

Random mutagenesis methods

A broad category of enzyme engineering approaches is composed of the random mutagenesis techniques. These methods arbitrarily incorporate mutations into the target sequence and allow very little control to the experimenter. Consequently, mutations can be identified throughout the entirety of the protein that have beneficial effects on the desired activity or trait. Additionally, these methods have the advantage of not requiring previously determined structural data, which in some cases can be unavailable or difficult to acquire. Because of this, random mutagenesis can be applied to any enzyme for directed evolution studies. However, it is commonly used as a complimentary approach after completing targeted mutagenesis guided by structural data for the

particular enzyme or based on homology modeling in order to further increase the desired activity[18].

The most commonly used of these random mutagenesis techniques is error prone PCR (epPCR)[18, 26], however, this can be performed in a few different modes (Figure 1-3). Early epPCR methods used a *Taq* polymerase lacking proofreading ability in combination with Mn^{2+} and unbalanced dNTP concentrations[27]. This composition increases the likelihood of mismatch pairing by the polymerase to increase the error rate during gene amplification and can generally be controlled by the concentrations of dNTP and Mn^{2+} added to the PCR sample. Currently, the most frequently used method of epPCR employs a DNA polymerase or combination of polymerases that have been engineered for lower extension fidelity to introduce mutations at random during DNA amplification cycles. Many of these polymerases also have reduced proofreading ability so that the incorporated mutations are not corrected during amplification. The Genemorph II kit from Stratagene Co. is one such example that is commercially available, which consists of a blend of a proprietary DNA polymerase mutant and an

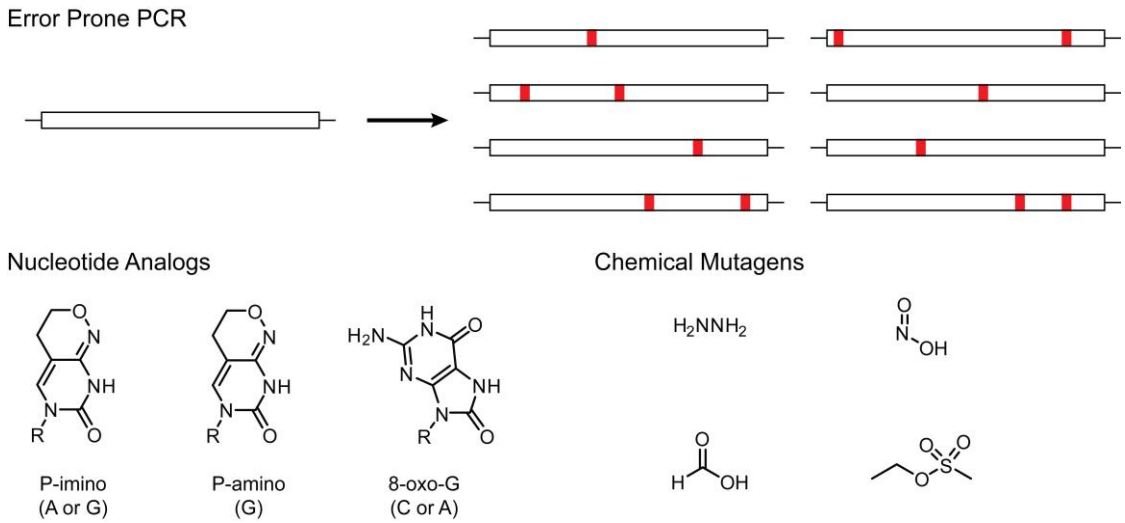

Figure 1-3. Methods for random gene mutagenesis. Each of these mutagenesis methods allows for incorporation of mutations into a gene sequence in an arbitrary manner, giving little to no preference for location or identity of the new nucleobase.

engineered *Taq* polymerase to provide a balanced mutation rate of transitions (purine to purine or pyrimidine to pyrimidine) and transversions (purine to pyrimidine or pyrimidine to purine).

In other random mutagenesis methods, nucleotide analogs are added into the PCR sample. After these non-canonical nucleotides are incorporated into the DNA sequence through PCR amplification, the non-natural nucleobases are capable of pairing with multiple canonical nucleotides to induce mutations during *in vivo* replication[28] (Figure 1-3). In a similar manner, exposing the DNA template to chemical mutagens, such as hydrazine, nitrous acid, formic acid or ethyl methane sulfonate (Figure 1-3), can chemically alter nucleobases to create new hydrogen bond donor and acceptor groups and allow mismatch pairing upon replication[29, 30]. Additionally, *E. coli* mutator strains and whole cell mutagenesis via UV exposure can be used to generate random mutations *in vivo*[31, 32].

Positions mutated via random mutagenesis occasionally become new sites to target for saturation mutagenesis[18, 33-35]. The hypothesis is that the process of random mutagenesis may pinpoint specific positions in the enzyme, termed 'hot sports'[36], which are sensitive to mutation and are responsible for improved activity. Targeting these sites for saturation mutagenesis allows a full analysis of amino acids at the position to determine the optimal residue identity for providing the highest level of desired activity. This complete characterization of identified hot spots allows targeted and random mutagenesis methods to work hand-in-hand to engineer the enzyme for the particular application.

Gene shuffling is another wide-ranging random mutagenesis technique that encompasses a number of methods, many with only slight variations. In general, the protocol enables the recombination of progenitor genes to create chimeric sequences with greater sequence diversity than that available through other random mutagenesis

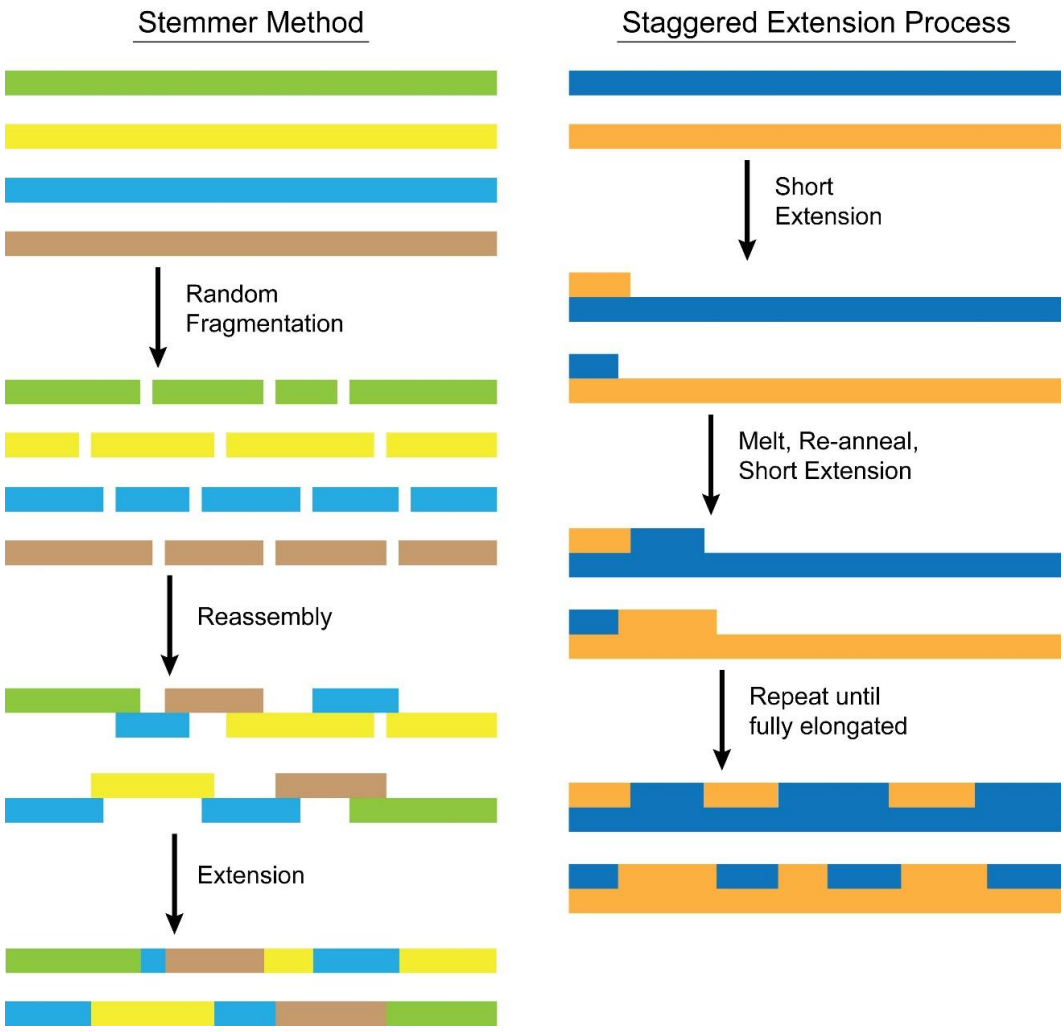

Figure 1-4. Methods for gene shuffling. Shuffling via the Stemmer method involves the recombination of fragmented homologous genes, while the StEP method relies on short DNA amplification cycles to allow partially extended genes to anneal with homologous templates to create multiple crossovers in the full length gene.

techniques (Figure 1-4). The sequences used may be either naturally occurring homologous genes and/or sequences containing beneficial mutations identified through other mutagenesis methods. The original DNA shuffling method was introduced by Willem P. C. Stemmer using a single gene with randomly identified point mutations[37] and was later expanded to demonstrate shuffling of a family of naturally occurring homologous genes from different species to provide an array of functional diversity, or sequence variation with proven activity[38]. In each of these methods, DNA sequences

were fragmented through mild treatment with DNase and the mixed oligonucleotides were reassembled through PCR extension. One particularly interesting application of DNA shuffling involved the recombination of thymidine kinase genes from herpes simplex virus types 1 and 2 to identify chimeras with increased ability to phosphorylate the nucleoside analog zidovudine, creating a metabolite toxic to the *E. coli* host[39]. Four rounds of shuffling created two clones that sensitized the bacteria to 32-fold and 16,000-fold lower zidovudine concentrations than the wild-type type 1 and type 2 thymidine kinase variants, respectively, indicating significant increases in activity on the non-natural substrate.

In newer variations of gene shuffling, recombination is facilitated in a variety of ways, some of which are seemingly inspired by Stemmer's initial process. In Staggered Extension Process (StEP), DNA amplification via PCR uses extremely short annealing and extension times, rather than fragmented DNA oligonucleotides, to create gene crossovers and chimeras using homologous templates[40] (Figure 1-4). In each cycle, the growing DNA fragments anneal to new template DNA based on homologous sequences and are partially extended before being melted apart. This cycle is repeated until full gene sequences have been created, and typically result in variants with sequence elements from multiple parental templates due to several gene crossovers. Other more recent methods allow recombination of two nonhomologous sequences by time-dependent exonuclease digestion followed by ligation to create chimeric genes. Such methods are termed incremental truncation for the creation of hybrid enzymes (ITCHY)[41] and sequence homology-independent protein recombination (SHIPREC)[42]. Another method, SCRATCHY, combines ITCHY and DNA shuffling to create a larger number of crossovers and therefore greater sequence diversity for screening[43].

Although it is most commonly used for improving activity on non-natural substrates or in non-natural reaction conditions, enzyme engineering has also been

demonstrated to have the capacity to reprogram an enzyme for an entirely different activity. Bromoperoxidase A2 from *Streptomyces aureofaciens,* an α/β-hydrolase fold family member, was converted to a lipase after comparison to *Bacillus subtilis* lipase A of the same enzyme fold[(33)]. Removal of the cap-like domain specific to the bromoperoxidase followed by remodeling the substrate binding site was sufficient to transplant the lipase activity and completely eliminate halogenation activity. Directed evolution and site directed mutagenesis were also used to further increase the rate of hydrolysis and substrate scope of the new enzyme, showing that not only substrate preference can be altered, but also enzyme chemistry can indeed be transformed through enzyme engineering[(33)].

While not widely used as a stand-alone technique, computational based methods have been employed to analyze large numbers of enzyme variants *in silico* in order to suggest functional combinations of active site residues to perform specific reactions (Figure 1-5). Additionally, major advantages of protein structure prediction and design algorithms, such as RosettaMatch[(44)], are the ability to identify and design catalytic sites in novel protein scaffolds and to introduce a desired function into locations where no function previously existed. The new enzyme possessing this designed activity can then serve as a progenitor for directed evolution to further optimize the designed function. For example, RosettaMatch was used to design a series of enzymes capable of catalyzing a retro-aldol cleavage of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (commonly called methodol) using a Schiff base mechanism[(45)] that was further evolved to increase activity[(46)] and subsequently lead to an enzyme variant with a completely reconstructed active site. The computationally designed catalytic lysine residue was relocated through random mutagenesis to a new and unpredicted position within the binding pocket resulting in >4,400-fold increased specific activity, nearing the catalytic efficiency of some natural enzymes[(47)].

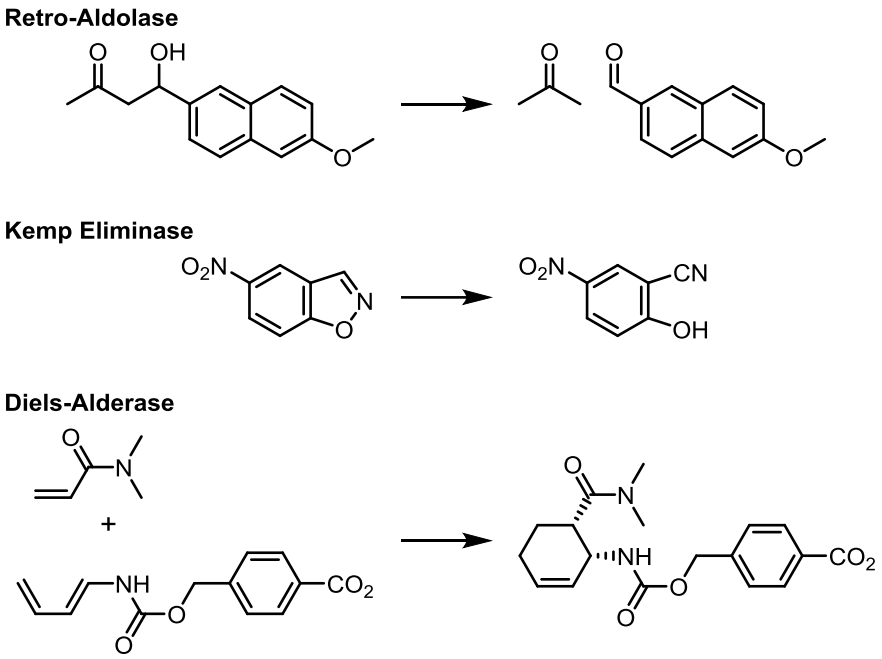

Figure 1-5. Reactions performed by computationally designed enzymes. Adapted from Nannemann *et al.*[18].

Furthermore, this same protein prediction tool has even been used to engineer non-natural enzymatic reactions. RosettaMatch was used to design a series of active sites on a number of scaffolds to catalyze a Kemp elimination, with several enzymes showing various degrees of detectable activity[48]. One candidate was further engineered through random mutagenesis and gene shuffling, resulting in a variant with >200-fold increased catalytic efficiency[49]. Similarly, RosettaMatch was also used to develop a Diels-Alderase scaffold that catalyzed the stereoselective cycloaddition of 4-carboxybenzyl *trans*-1,3-butadiene-1-carbamate and *N,N*-dimethylacrylamide. Iterative saturation mutagenesis of residues adjacent to those responsible for activating the substrates increased activity up to 100-fold[50].

Repurposing mutant libraries for new targets

With a large variety of methods available for use in directed evolution, it is not surprising that many enzyme engineering efforts utilize a combination of methods over

several rounds of evolution to meet the goals of each particular project[18]. In doing so, the large libraries of enzyme variants created to improve activity on a specific substrate may also be useful in a second enzyme engineering campaign for a new substrate. Taking advantage of the available large sequence diversity, these variants may serve as an initial progenitor library to identify a template enzyme for the new reaction. A series of ketoreductase enzymes were evolved in this way by Codexis, Inc. to catalyze the asymmetric reduction of ketones to chiral alcohols in a variety of pharmaceutical intermediates[1, 51]. Using a ketoreductase from *Lactobacillus kefir* as the template, a library of enzyme variants was first generated through random mutagenesis, gene shuffling and ProSAR analysis in an effort to increase enantioselectivity and productivity of the enzyme for reduction of tetrahydrothiophene-3-one to (*R*)-tetrahydrothiophene-3-ol, an intermediate in sulopenem-type β-lactam antibiotics[52] (Figure 1-6). One of the ketoreductase variants from this library became the starting point for evolving activity for production of (*S*)-1-(2,6-dichloro-3-fluorophenyl)-ethanol, a

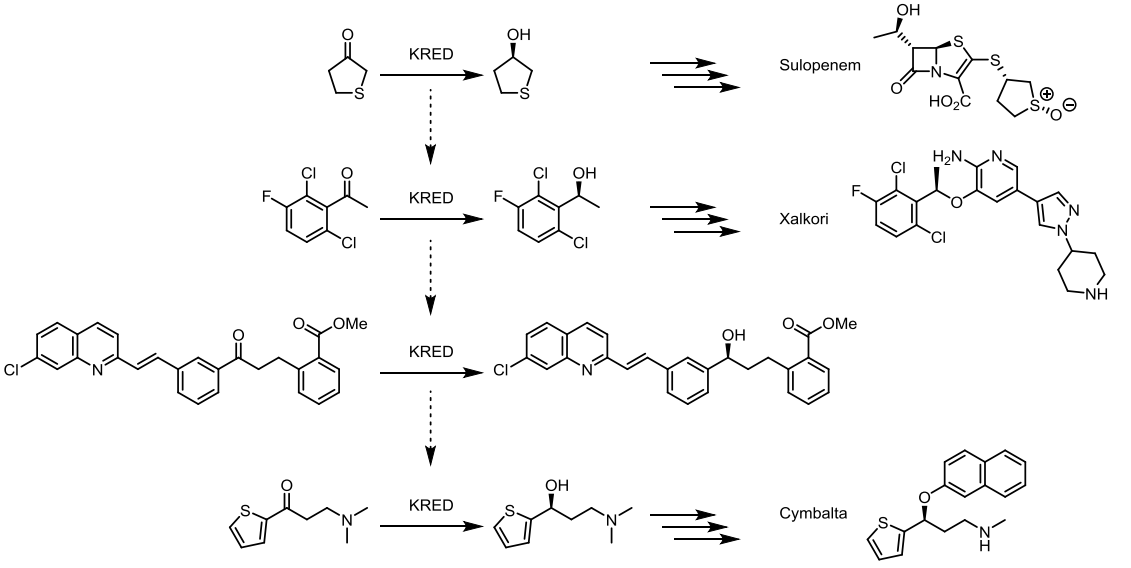

Figure 1-6. Series of engineered ketoreductases. For each of the reactions, the original template for engineering was generated during evolution for a different substrate in a previous directed evolution experiment.

raw material for production of crizotinib (Xalkori®), from the ketone substrate[1]. Similarly, one of the enzymes of this library was used as a progenitor for engineering activity for production of a montelukast (Singulair®) intermediate[53], during which time a starting point for evolution of a duloxetine (Cymbalta®) intermediate ketoreductase was identified[1]. In each of these cases, enzyme evolution was facilitated by using a non-natural progenitor enzyme that had increased stability for one set of process conditions and therefore accelerated the optimization process for the new substrate[1].

In a similar progression of enzyme engineering, Arnold and coworkers engineered a lineage of cytochrome P450 variants capable of converting alkanes of differing lengths to the corresponding alcohols using the substrate walking approach. A medium chain (C12-C18) fatty acid monooxygenase from *Bacillus megaterium* was used as the progenitor to evolve a new regioselective activity for the oxidation of octane to n-octanol[54]. The final variant from this study displayed hydroxylation activity on shorter chain alkanes down to propane, and was further evolved to enhance activity on this small hydrocarbon over several generations, successfully creating an extremely efficient propane monooxygenase[55, 56]. Eventually, a variant from this series was also evolved to produce ethanol by direct oxidation of ethane, providing an alternative method of producing this biofuel from petrochemical feedstocks[57].

In another impressive series, Turner and coworkers engineered a “toolbox” of *Aspergillus niger* monoamine oxidase variants for the production of enantiomerically pure chiral amines on a diverse range of racemic amine substrates[58] (Figure 1-7). Directed evolution initially began to develop variants capable of selectively oxidizing a variety of rather simple primary, secondary and some tertiary amines[31, 59, 60]. Coupling this catalysis with a nonselective chemical reduction enables the deracemization of these amine substrates through kinetic resolution. More recent evolution of these enzyme variants targeting residues in the active site and substrate channel has

expanded the substrate scope in several new variants capable of oxidizing significantly larger and more substituted complex amines[58, 61, 62]. Several of the chiral amine products that can now be accessed through these engineered monoamine oxidase enzymes are alkaloid natural products possessing interesting biological activities and intermediates for the synthesis of a variety of active pharmaceutical ingredients.

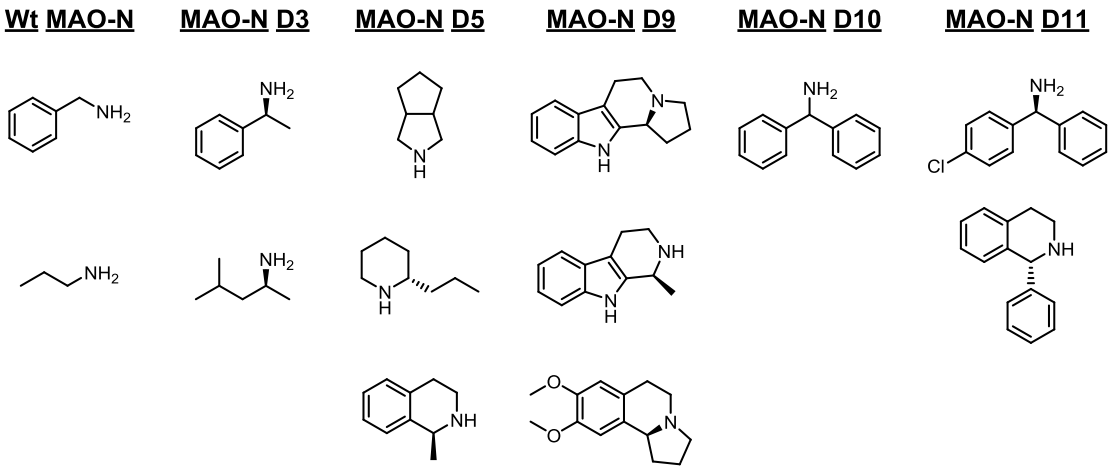

Figure 1-7. Toolbox of monoamine oxidase variants and substrates accessible by each variant. Adapted from Ghislieri *et al.*[58].

Pathway Construction

Biosynthetic pathways have the capability of increasing the complexity of enzymatically produced compounds in comparison to single enzyme biocatalytic reactions[63]. However, metabolic flux through natural pathways has been optimized through years of evolution to meet the minimal production requirements enabling organisms to grow and survive in a large variety of environmental conditions. Natural feedback inhibition mechanisms have evolved along with the pathway to regulate flux to maintain the concentration of pathway intermediates and final products in a narrow range to avoid toxicity[64]. In addition to preventing cytotoxic levels of certain metabolites, regulation also precludes pathways from diverting an unnecessary amount of resources

to one particular direction at the expense of others. While all of these traits are necessary for growth and healthy cell homeostasis, the associated restrictions limit the productivity of a pathway intended to be used for large scale production of compounds. In pathway engineering, the goal is to maximize production of the target compound rather than meet a certain metabolic need and these limitations must be circumvented for useful implementation.

Constructed biosynthetic pathways can be deemed 'integrated', 'hosted' or a combination of both in reference to how the pathway precursors are supplied[65]. Integrated biosynthetic pathways consume only natural metabolites produced through normal metabolism in the cell and can therefore be generated through biotransformations during fermentation of inexpensive and complex feedstocks. To contrast, hosted pathways are those where the host organism is only used for overexpression of the biosynthetic enzymes, and the necessary substrates are separately generated or obtained from commercial sources and added to the bioreactor for biocatalytic conversion. These pathways may be more applicable for production of completely non-natural compounds from non-natural substrates since the precursors would not be produced from central metabolism. In an approach that combines these two themes, some of the necessary substrates are natural cellular metabolites while others are added exogenously.

A classic approach to increase compound production is through metabolic engineering of the naturally producing organism. In this strategy, enzymes are either deliberately overexpressed to increase the levels of pathway precursors or knocked out to decrease consumption of intermediates by divergent metabolic pathways (Figure 1-8a). Examples of this innate biosynthesis of the natural product include production of β-lactams and other antibiotics[66, 67] as well as naturally occurring amino acids[68, 69]. Although this method has a proven success record for a large variety of compounds[70],

not all producing organisms and natural products are amenable to these strategies of improving innate biosynthetic production. In some cases, the naturally producing organism may not be cooperative to genetic manipulation, culturing in a laboratory setting or tolerating up-regulated expression of biosynthetic enzymes and therefore other strategies must be employed.

Instead, an often used approach is to move the pathway genes into a more readily culturable and genetically malleable heterologous host to optimize production (Figure 1-8b). In this non-innate biosynthesis strategy, biosynthetic genes can be separated from the inherent regulatory mechanisms that control metabolic flux in the original organism resulting in increased production of the natural products[(71)]. A recent application of this method is in the production of polyhydroxybutyrate in *Escherichia coli* expressing several genes from *Cupriavidus necator*[(72)]. Heterologous expression of three biosynthetic genes placed under an inducible promoter permitted accumulation of polyhydroxybutyrate up to 85% of dry cell weight in minimal media, which is comparable to the 90% production observed in the native host. Transfer of the pathway to *E. coli*, however, enabled production of the biopolymer in a much faster growing host organism that could be lysed more easily for faster and more efficient compound isolation[(73)].

Similar to the limitations of metabolic engineering for innate biosynthesis, this heterologous or non-innate biosynthesis method also has unpredictable challenges that may result in low production titers in the new organism. Poor performance may be the result of disproportionate expression or activity of biosynthetic enzymes, suboptimal levels of pathway precursors or required cofactors, formation of toxic intermediates or products, overall metabolic burden on the host cell or a combination of these and other factors[(74)]. Genes transferred from one organism to another are often found to be functionally expressed at lower levels due to either different codon bias in the heterologous host or consolidation of enzyme to inclusion bodies of insoluble protein. In

either case, the result is significantly lower enzymatic activity and, therefore, reduced total product formation. A synthetic version of the gene with codon usage optimized for the bias of the new organism is a commonly used strategy to improve expression of these heterologous genes[75] and fusion of the enzyme to maltose binding protein (MBP)[76], glutathione S-transferase (GST)[77], thioredoxin[78] or small ubiquitin-like modifier (SUMO)[79] can also aid in improving soluble expression. Additionally, production of the biosynthetic pathway may be limited by a single bottleneck enzyme that exhibits low activity or expression relative to the other enzymes. While low expression can be resolved through gene optimization, low activity of the enzyme may be corrected through other methods.

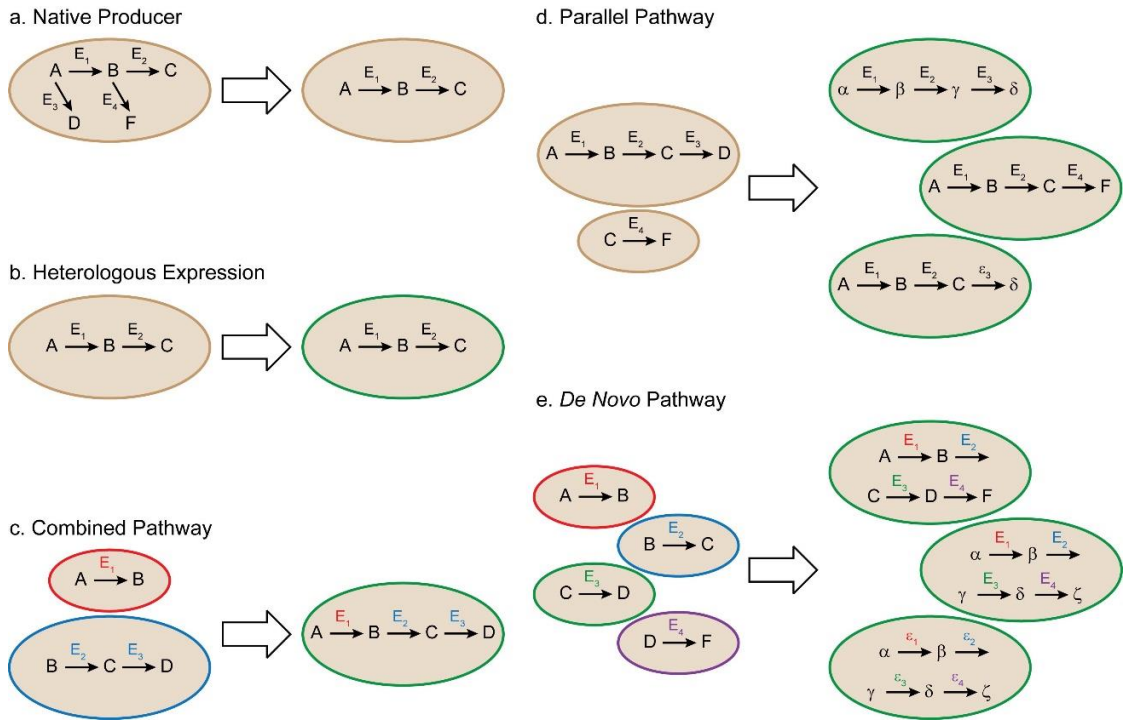

Figure 1-8. Pathway construction strategies. Natural enzymes (E_n) are collected from one or multiple organisms (ovals on left) to create biosynthetic pathways in (**a**) an engineered native producer (beige oval on right) or (**b-e**) heterologous hosts (green ovals on right). Natural substrates are indicated by letters A, B, C, D, F. Engineered enzymes (ε_n) and non-natural substrates or products are noted as greek letters α, β, γ, δ, ζ. Adapted from Martin *et al.*[80].

The more common application of heterologous pathway expression involves the combination of genes from multiple source organisms. Often when a biosynthetic enzyme, or particular section of the pathway, shows limited activity in a heterologous host it can be replaced with a similar enzyme from a different organism to create a more productive combined pathway (Figure 1-8c). Demonstration of this pathway construction method is epitomized through the assembly of biosynthetic pathways for production of precursors to the antimalarial drug artemisinin by the Keasling group, where genes for precursor and secondary metabolite production were isolated from three organisms and heterologously express in a separate host organism (Figure 1-9). Initial production of the isoprenoid amorphadiene in *E. coli* was first demonstrated after simultaneous heterologous expression of the mevalonate-dependent isoprenoid precursor pathway from *Saccharomyces cerevisiae* with the amorphadiene synthase gene from *Artemisia annua*[(81)]. Use of this pathway created a larger pool of isopentenyl pyrophosphate and

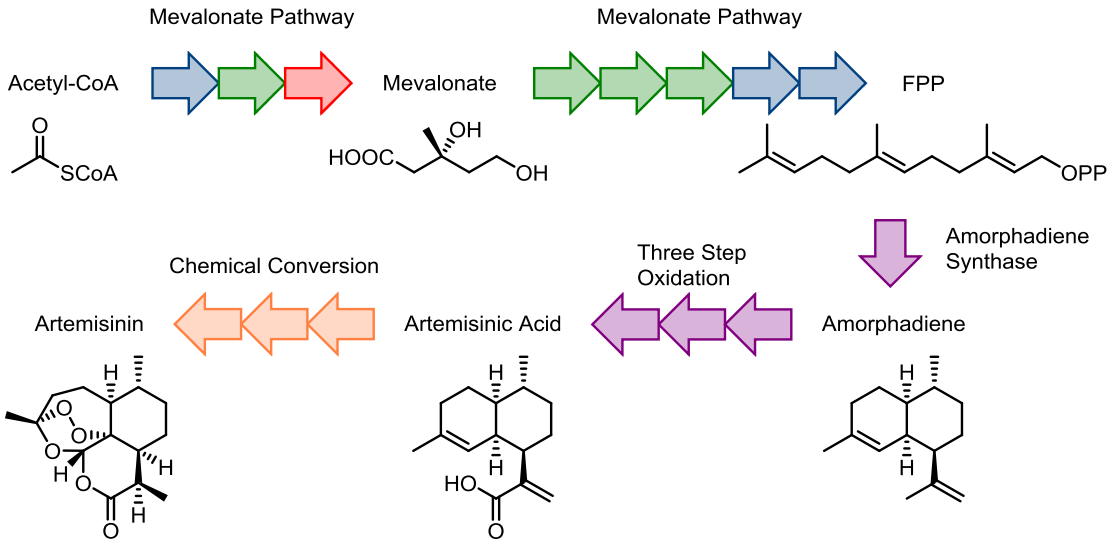

Figure 1-9. Heterologous construction of artemisinic acid biosynthetic pathway. Arrows indicate enzymatic steps between indicated pathway intermediates and products and are color coded to the source organism: *E. coli* (blue), *S. cerevisiae* (green), *S. aureus* (red), and *Artemisia annua* (purple). Chemical synthesis steps to produce artemisinin are indicated by orange arrows.

dimethylallyl pyrophosphate precursors than the host-provided mevalonate-independent, or deoxyxylulose 5-phosphate, pathway due to metabolic regulations in the native host. This larger precursor pool led to improved production of isoprenoid compounds, such as amorphadiene via the co-expressed amorphadiene synthase. Yeast strain engineering, through upregulating expression of genes involved in precursor production and preventing substrate flux through divergent isoprenoid pathways, further increased yields of amorphadiene. Addition of a cytochrome P450 monoxygenase isolated from the native producer *A. annua* performs the three step oxidation to allow direct production of artemisinic acid, a later stage biosynthetic intermediate of artemisinin, at 100 mg/L[(82)]. Most recently, production of artemisinic acid was vastly enhanced via additional pathway optimization and coexpression of the cognate cytochrome P450 reductase, an alcohol dehydrogenase and the artemisinic aldehyde dehydrogenase from *A. annua* with the optimized *S. cerevisiae* strain to yield 25 g/L artemisinic acid, a production level possibly high enough to allow industrial scale preparation of the antimalarial drug artemisinin after efficient chemical conversion to the active ingredient[(83)]. Likewise, a similar use of heterologous expression and host metabolic engineering was used to create *E. coli* capable of producing taxadiene, the first committed intermediate in production of the anticancer drug paclitaxel first isolated from the Pacific yew tree *Taxus brevifolia*[(84)].

To extend biosynthetic production to a wider variety of products and applications, existing natural biosynthetic pathways can be hijacked to produce natural product analogs or other non-natural products. General strategies for production in these methods typically exploit promiscuous properties of enzymes to catalyze reactions on non-natural substrates (creating parallel pathways) or are gained through the addition and/or subtraction of tailoring enzymes to create branches from the typical biosynthetic pathway[(80, 85)] (Figure 1-8d). For example, novel polyhydroxyalkanoates containing thioether linkages in the side chains were produced after feeding propylthiooctanoic acid

or propylthiohexanoic acid to a polyhydroxyalkanoate synthase knockout strain of *Cupriavidus necator* heterologously expressing a broadly specific polyhydroxyalkanoate synthase from *Pseudomonas mendocina*[(86)]. Additionally, naturally produced analogs of macbecin were produced after manipulating the gene cluster responsible for producing the natural product, thereby producing a large variety of macbecin analogs with some still retaining a high level of activity as Hsp90 inhibitors[(87)].

Many industrially useful compounds cannot be created through known biochemical routes, creating the need for the ability to design non-natural metabolic pathways. Merging the approaches of combinatorial pathway construction and use of promiscuous enzymes for diverse production creates a new strategy of *de novo* pathway design. In this scheme, a series of unrelated enzymes is proposed to form a non-natural pathway for production of a valuable compound (Figure 1-8e). At a conceptual level, this eliminates the necessity of relying on the predetermined natural pathways observed in organisms. Instead, *de novo* pathway design permits the organization of enzymes into new biosynthetic pathways not found in nature, capitalizing on enzyme promiscuity and the substantial diversity of enzymes and enzymatic reactions to establish biosynthetic pathways[(85)]. Such schemes can be utilized to create a more manageable biosynthetic pathway (*i.e.*, a shorter, more direct pathway or a replacement for a pathway that is not easily heterologously expressed) for natural products, but can also be used to design pathways for the production of non-natural products for which a natural biosynthetic pathway does not exist[(80)]. As a replacement for a more circuitous but natural metabolic route, Moon *et al.* constructed a synthetic pathway for production of glucaric acid from glucose in *E. coli*, combining five unrelated enzymes from three source organisms[(88)]. This *de novo* pathway is significantly shorter than the 10 step pathway that converts glucose or galactose to glucaric acid observed in mammals, and therefore provides a more direct production route to this value-added chemical.

Despite the potential of these *de novo* pathways to allow biosynthetic production of a seemingly immense number of compounds, only two cases exist for their use in biosynthesis of molecules not found in nature (Figure 1-10). The first example was by Frost and coworkers, where dehydrogenase, dehydratase and decarboxylase enzymes from *Pseudomonas fragi*, *Pseudomonas putida* and *E. coli* were combined in order to generate D- and L-1,2,4-butanetriol from D-xylose or L-arabinose, respectively[89]. This non-natural polyol serves as a precursor for D, L-1,2,4-butanetriol trinitrate, a more stable replacement for nitroglycerin-based explosives. The second case provides production of 1,4-butanediol, a high volume commodity chemical used to manufacture polymers, from

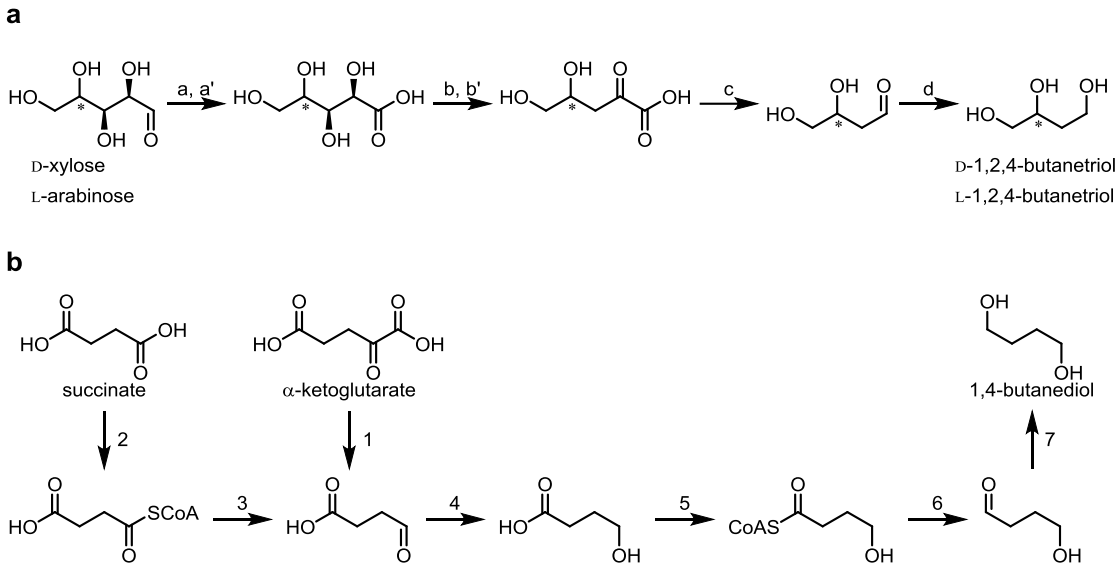

Figure 1-10. *De novo* pathways for production of non-natural products. (**a**) D- and L-1,2,4-butanetriol are produced from D-xylose and L-arabinose, respectively. (**b**) 1,4-butanediol is generated from citric acid cycle intermediates succinate and α-ketoglutarate. Enzymes are indicated by letters or numbers. a: D-xylose dehydrogenase (*P. fragi*), a': L-arabinoate dehydrogenase (*P. fragi*), b: D-xylonate dehydratase (*E. coli*), b': L-arabinoate dehydratase (*P. fragi*), c: benzylformate decarboxylase (*P. putida*), d: dehydrogenase. 1: α-ketoglutarate decarboxylase, 2: succinyl-CoA synthetase, 3: CoA-dependent succinate semialdehyde dehydrogenase, 4: 4-hydroxybutyrate dehydrogenase, 5: 4-hydroxybutyryl-CoA transferase, 6: 4-hydroxybutyryl-CoA reductase, 7: alcohol dehydrogenase. Figures adapted from Niu *et al.*[89] and Yim *et al.*[90].

a variety of pure and mixed sugar sources derived from biomass[90]. In this work, potential biosynthetic pathways were computationally predicted for production of 1,4-butanediol from *E coli* central metabolites, and then evaluated and ranked based on a number of factors including theoretical yield, pathway steps and thermodynamic favorability. The final implemented pathway directed carbon flux to produce succinyl semialdehyde, which was sequentially reduced over several steps to yield the desired 1,4-butanediol in a yield of 18 g/L. While each of these studies employed a combination of wild-type enzymes from the host organism and heterologous sources for production of these non-natural polyol compounds, these examples demonstrate the vastly different biosynthetic possibilities available via *de novo* pathway design. Completely different synthetic schemes were proven to be quite successful for relatively efficient production of two very similar compounds, differing only by the presence or absence of the C2 hydroxyl group.

While the majority of the pathway construction examples provided above involved recruitment of natural enzymes from either one or multiple source organisms to meet the activity needs of the pathway, intrinsic characteristics of some natural enzymes that limit pathway production titers—such as creating a bottleneck in pathway flux, low activity on non-natural substrates, low production of non-natural compounds or formation undesirable byproducts—may be unavoidable for wild-type enzymes. For this reason, another beneficial and increasingly successful option is improving enzymatic activity through enzyme engineering. This approach can be applied to create custom-made enzymes with individually tailored traits to meet specific needs where naturally occurring enzymes are insufficient for providing the desired activity[80].

For example, Ran and Frost engineered 2-keto-3-deoxy-6-phosphogalactonate aldolase for altered substrate selectivity as a way to circumvent a bottleneck in the shikimate biosynthetic pathway[91]. The evolved enzyme catalyzes the aldol

condensation of pyruvate and D-erythrose 4-phosphate to form 3deoxy-D-arabino-heptulosonic acid 7-phosphate (DHAP), which is naturally produced by DHAP synthase from phosphoenolpyruvate and D-erythrose 4-phosphate. This change in substrate selection to an abundant metabolite removes the dependence on phosphoenolpyruvate availability, which is limited due to competition between multiple metabolic pathways, allowing production of the downstream metabolite 3-dehydroshikimic acid in yields up to 19 g/L. This compound is the direct precursor to shikimic acid which can be used to produce the antiinfluenza drug Tamiflu, among other chemicals of commercial interest.

The addition of engineered enzymes into existing biosynthetic pathways can also expand production capabilities in these parallel pathways to form products for which no naturally occurring biosynthetic route exists (Figure 1-8d). Heterologous expression of *Bacillus subtilis* threonine dehydratase and a glutamate dehydrogenase evolved for production of L-homoalanine from 2-ketobutyrate permitted the biosynthetic production of this pharmaceutically relevant non-natural amino acid in a threonine overproducing strain of *E. coli*[92]. Similarly, the natural metabolic pathway responsible for producing branched-chain amino acids was directed through an evolved 2-isopropylmalate synthase and several downstream enzymes engineered for altered substrate specificity to produce 3-methyl-1-pentanol and other novel long chain alcohols via an engineered non-natural metabolic pathway[93]. Related work from the same lab resulted in the production of a wider variety of alcohols via reductions of naturally occuring α-ketoacids through the Ehrlich pathway[94]. In both of these highlighted cases, natural biosynthetic pathways were extended using engineered enzymes in order to produce non-natural products.

To date, there is only one instance where multiple engineered enzymes have been concatenated into a *de novo* biosynthetic pathway for production of a non-natural

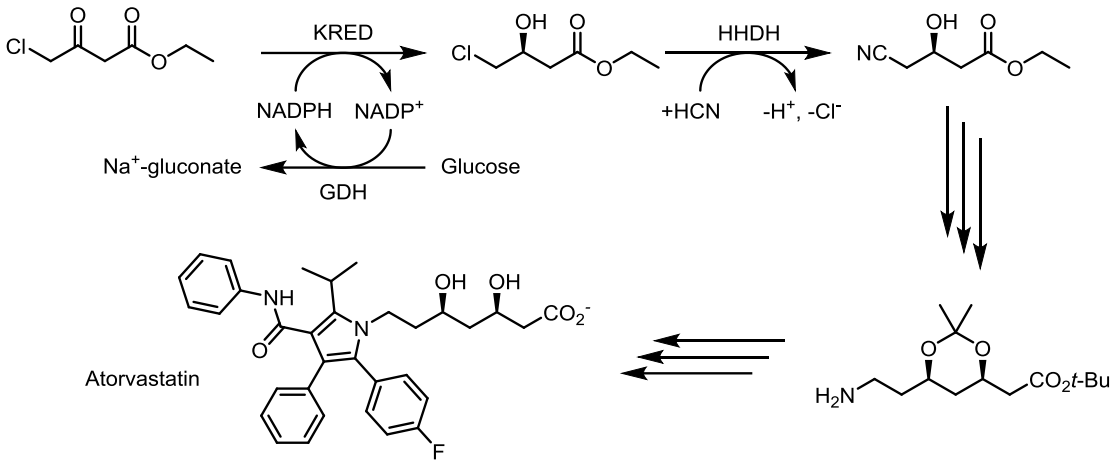

Figure 1-11. Production of atorvastatin side chain intermediate *(R)*-4-cyano-3-hydroxybutyrate through an engineered *de novo* non-natural biosynthetic pathway. The semisynthetic pathway continues through multiple chemical and one biocatalytic step to produce the full side chain for production of Atorvastatin.

molecule. The halohydrin dehalogenase previously mentioned that was evolved for production of an intermediate of the atorvastatin side chain was used in tandem with an engineered ketoreductase and glucose dehydrogenase to create a two-step, three enzyme biocatalytic system[(95)] (Figure 1-11). Each of these three enzymes have been engineered for activity on either a new substrate or in new reaction conditions to fit the required process parameters. The first step involves the reduction of ethyl-4-chloroacetoacetate by the ketoreductase to yield the *(S)*-ethyl-4-chloro-3-hydroxybutyrate product. The presence of glucose and glucose dehydrogenase in the reaction allows for *in situ* regeneration of the NADPH cofactor required for ketoreductase activity. Finally, the engineered halohydrin dehalogenase catalyzes the replacement of the chloro substituent with a cyano group to provide the desired ethyl *(R)*-4-cyano-3-hydroxybutyrate. Implementation of this biosynthetic pathway on large scale *in vitro* reactions has allowed the production of this intermediate at a rate of >100 tons per year[(1)].

Theories of Biosynthetic Pathway Evolution

Despite the individual successes of pathway engineering and enzyme evolution to produce a large variety of valuable products, *de novo* pathway design methods and enzyme engineering techniques must be combined in order to gain access a greater diversity of compounds and fully capitalize on the synthetic potential of biological systems. This type of strategy would eliminate the dependence on naturally occurring pathways and can even remove the limitation to naturally occurring substrates to enable the design of an entirely new biosynthetic route for production of either natural or non-natural compounds. However, the major obstacles in delineating a non-natural biosynthetic pathway is determining a reasonable arrangement of starting material, biotransformations and intermediates that enable the production of the desired compound.

Organization of data collected on metabolic pathways and individual enzymes from functional activity assays and genome sequencing into databases such as Swiss-Prot[(96)], BRENDA[(97)], KEGG[(98)] and MetaCyc[(99)] provide a large pool of information on natural enzymes that can be used for pathway construction, including details on biotransformations that the enzymes can catalyze. The massive amount of enzyme characterization and genome sequencing data collected has allowed researchers to develop algorithms in an effort to computationally predict sequential transformations to propose biosynthetic pathways for a particular molecule, such as the Biochemical Network Integrated Computational Explorer (BNICE)[(100)], the Pathway Prediction System of the University of Minnesota Biocatalysis and Biodegradation Database (UM-BBD)[(101)], Retro-Biosynthesis Tool (ReBiT)[(85)] and DESHARKY[(102)] along with many other recently reviewed computational pathway design tools[(103)]. Essentially, these prediction tools use specifically defined biotransformation guidelines based on assigned enzyme classification number and/or catalyzed reaction (*i.e.* functional group transformation) to

suggest a variety of metabolic routes connecting two compounds. Pathways can be ranked on biosynthetic potential based on a series of criteria including pathway length, thermodynamic favorability, host compatibility and predicted flux to name a few. While many of these pathway design tools have high potential for facilitating biosynthetic pathway design and engineering, unfortunately their regular implementation and success has been stifled by the inherent complexity of biological systems[104]. Additionally, many of these tools are limited to predictions involving known naturally occurring molecules (*i.e.* central metabolism) and/or xenobiotic degradation pathways[105], and are therefore less than ideal for constructing *de novo* biosynthetic pathways to produce non-natural compounds. A potentially useful alternative strategy for designing a new *de novo* biosynthetic pathway may be based on one of the multiple hypotheses described for natural pathway evolution, some of which are described below. In theory, researchers can attempt to construct a *de novo* non-natural biosynthetic pathway through mimicking one of these proposed process, or through combining aspects of the different hypotheses for early natural pathway evolution.

Although the theories on pathway evolution have several organizational ideas that separate them into unique models, all appear to have many fundamental aspects in common. Each models presupposes a specific set of environmental conditions, primarily the presence and availability of particular metabolic compounds, which form the basis of the individual theories. The common feature of each pathway evolution theory is the widely accepted mechanism of individual enzyme recruitment as the result of a gene duplication event[106] followed by functional divergence through gradual change in substrate specificity through a process termed neofunctionalization[107]. Originally outlined by Whaley in 1969[108] and later fully developed by Ycas[109] and Jensen[110] into what is known as the patchwork hypothesis, the small genome of an early organism is expected to consist primarily of genes encoding enzymes with broad substrate

specificity, allowing reactivity on a wide range of different but related substrates. This central proposed characteristic of ancestral enzymes would provide a maximal range of catalytic capability within the cell with only a limited number of genes and required genome size[110]. The low specificity of these enzymes creates a 'biochemical leakiness' that enables the production of new metabolites in low amounts, facilitating recruitment and eventual specialization for use in new pathways[110].

In this patchwork process, the gene for one of these unspecialized promiscuous enzymes is duplicated within the genome, creating a second and potentially expendable copy. Rather than natural selection strictly encouraging the loss of the spare copy, if the side metabolites produced by this enzyme are advantageous, the duplicated gene with be selectively maintained to preserve the metabolic benefit[110]. Slow and random accumulation of mutations within one of the duplicates could prove to be even more advantageous, eventually permitting a new predominant activity within the same original enzyme scaffold, and continued acquisition of mutations can generate a new specialized enzyme with higher substrate specificity that improves metabolic efficiency[108, 110] (Figure 1-12). In this way, recruitment and specialization of multiple genes in a linear sequence are able to form a complete biosynthetic pathway that is metabolically beneficial to the organism.

In terms of the directionality of constructing the series of enzymes into a pathway, there are formally two approaches: recruiting and incorporating enzymes into the pathway in the forward direction—that is, in the order of biosynthesis from precursor to final metabolite—or proceeding in the reverse direction, meaning opposite the direction of biosynthesis, from final product backwards to simple precursors. Each of these schemes have been proposed as models for natural biosynthetic pathway evolution. The Granick hypothesis suggests the development of biosynthetic pathways in the forward direction, toward the creation of complicated biological molecules for simple

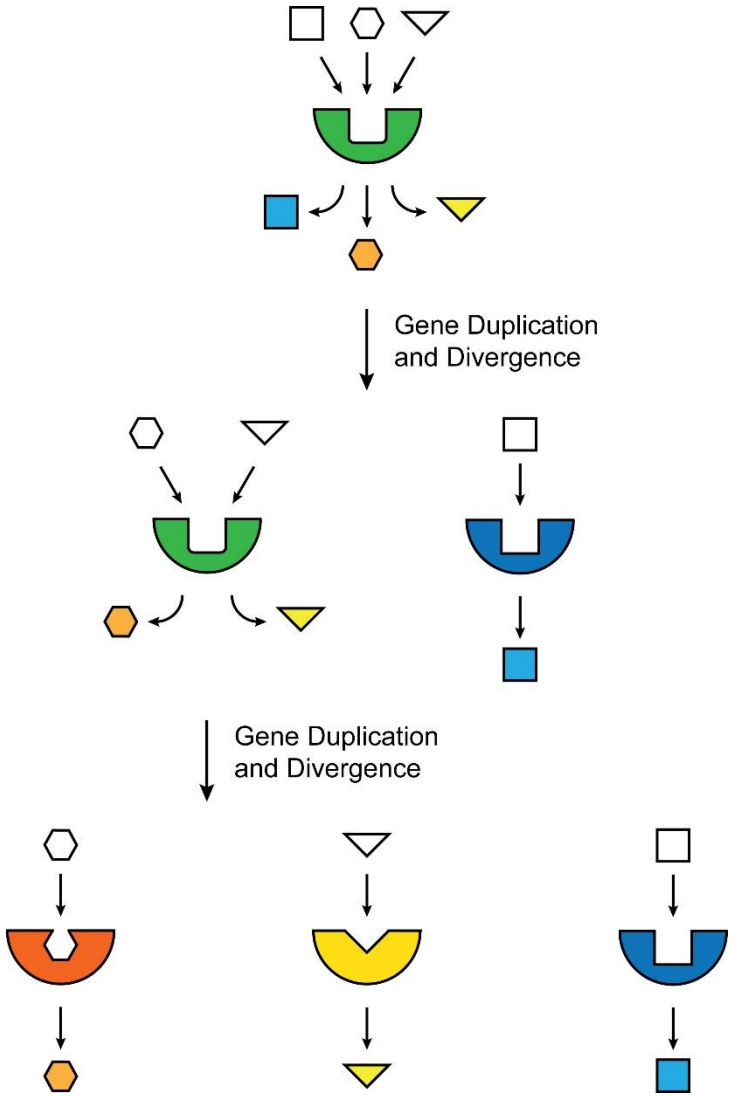

Figure 1-12. The patchwork model of gene duplication and enzyme functional divergence. From one progenitor enzyme with a wide substrate scope (green), multiple specialized enzymes can be evolved through gene duplication and new specialization. Adapted from Fondi *et al.*[111].

precursors[112] (Figure 1-13, left). The rationale for this model is that simple compounds and metabolites likely predated the presence of complex products, and therefore the enzymes catalyzing reactions in the early stages of biosynthesis must have evolved before those catalyzing the later steps. In this model, each intermediate within the pathway is assumed to be beneficial to the cell in some capacity, as otherwise a

selective pressure for maintaining or extending the pathway would not be present, and the simultaneous evolution of all enzymes to comprise the pathway is implausible[113].

While the absolute requirement of beneficial intermediates has limited the widespread acceptance of the forward evolution model for anabolic biosynthetic pathways, it is duly plausible for catabolic pathways to enable an organism to gain more energy through the breakdown of complex molecules into simple products[114]. Elements of this model have also been preserved in the shell hypothesis more recently proposed by Morowitz[115], in which layers of reaction networks with increasing complexity are built as shells extending outward from a central metabolic core of enzymatic reactions, describing the evolution of metabolism as a whole. Morowitz proposes the core series of reactions to be those within the reductive tricarboxylic acid cycle and argues that initially, in the very early stages of reactions, this metabolic cycle proceeded in the absence of enzymes. Also included in this central shell are the pathways of glycolysis and fatty acid biosynthesis.

Expansion into each new metabolic shell in this model was the result of a novel gateway activity that allowed new biochemical reactions to be incorporated into the metabolic repertoire. These new activities are added to the outer shells so that the core metabolic elements do not change, therefore preserving the components of central metabolism. The proposed gateway from the central shell of the tricarboxylic acid cycle is the introduction of nitrogen through the conversion of keto acids to amino acids. Therefore, this second shell includes many of biosynthetic pathways that produce a majority of the naturally occurring amino acids. Stepping outward to the third shell is the result of incorporating sulfur into molecules, namely the amino acids cysteine and methionine. Access to the fourth shell is through gaining activities for ring closure reactions and production of nitrogen containing heterocycles, such as purines, pyrimidines and those found in cofactors. In this way, each layer is dependent on all

previous layers, and the impetus for evolution through expansion to new shells is through the production of a molecule that provides a benefit to catalysis in a more central shell[115].

With the general delineation of the shells of metabolic pathways, certain themes can be described to characterize each level[115]. The central shell is primarily concerned with energetics, producing an amphiphile for eventual use as an encapsulating device via membrane formation and finally, providing precursors to biosynthesis. The second shell initiates production of amino acids, allowing catalysis via the introduction of chirality to cellular metabolites, while incorporation of sulfur in the third shell introduces the basis

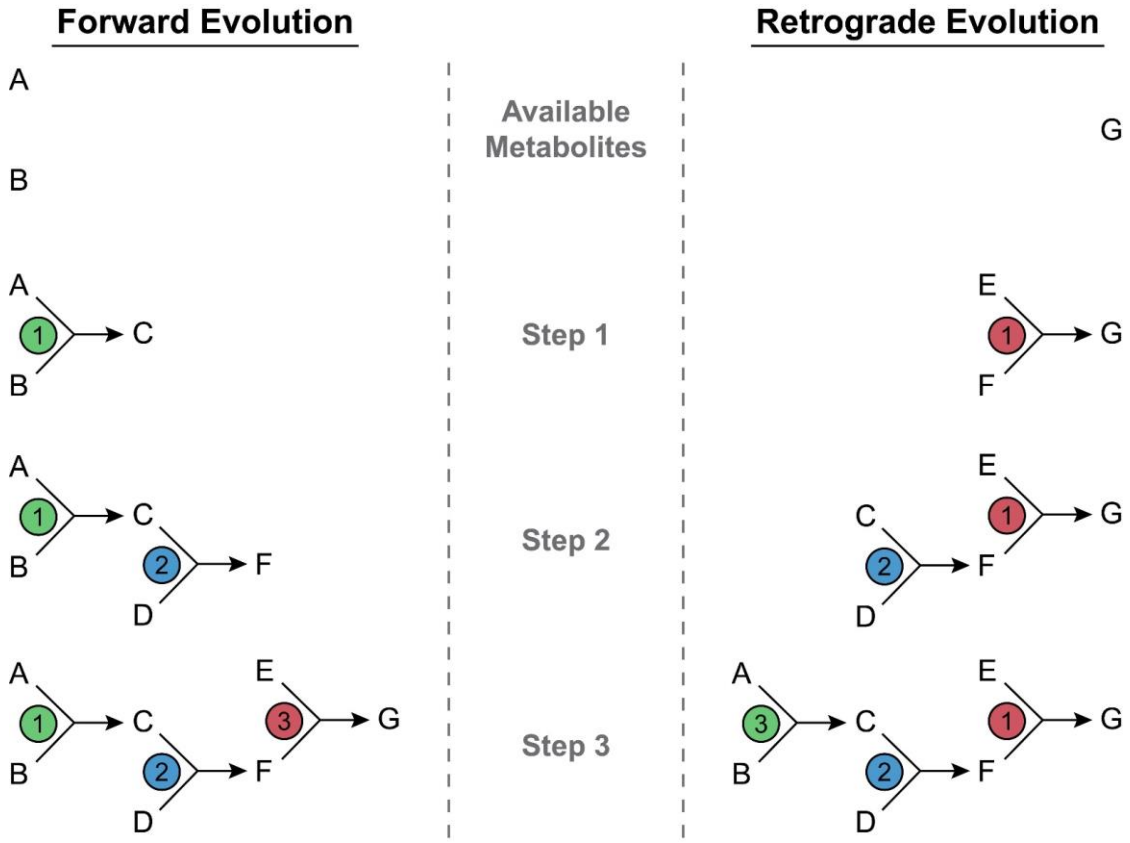

Figure 1-13. Forward and retrograde pathway evolution schemes. Forward pathway evolution creates more complex molecules from simple precursors. Retrograde pathway evolution enables continued production of one complex molecule through extending the pathway precursor backward to simple intermediates. Adapted from Rison and Thorton[116].

for higher protein structures allowed by cysteine disulfide bonds and methionine initiation. The fourth shell allows production of molecules such as cofactors to aid catalysis in the lower shells in addition to creating nucleobases that are used in the later shells dealing with creating templates and providing coding mechanisms.

The final model describing pathway evolution is for reverse, or retrograde, pathway evolution. The hypothesis of retrograde evolution was first explained by Horowitz in 1945, and suggests gene duplication and evolution of biosynthetic pathways in a step-wise manner in the reverse order to that of biosynthesis, meaning the last enzymatic reaction was the first to evolve, followed by the penultimate enzyme, and so on to the beginning of the pathway[117] (Figure 1-13, right). Central to this theory is the idea that distinct intermediates along the pathway cannot be assumed to provide any benefit to the organism and therefore do not provide a selective advantage. Instead, the individual reactions of a biosynthetic pathway are only valuable when viewed from the perspective of production of the final product, and therefore creation of the pathway is driven by improvements to continuously form this metabolite.

For example, an organism is presumed to be heterotrophic for a particular compound, G (Figure 1-13, right). As the environmental supply of G decreases due to consumption by the organism, a metabolic strain is created. This strain can be alleviated through evolution of an enzyme that is able to produce G from available precursors, E and F. Any organism that has gained this ability through the generation of a mutant enzyme will have a selective growth advantage over other organisms, especially after the levels of G have been completely exhausted. Additionally, under these conditions, a mutation to revert back to the original genetic strain, those incapable of producing G from E and F, would prove to be lethal, ensuring that this new enzymatic activity is maintained and propagated to future generations. Eventually as one of these precursors, say compound F, becomes limiting for growth, another retro-extension of the pathway

would be gained through a second event of gene duplication and mutation that would allow production of F from precursors C and D. These retro-extensions can continue until the pathway is connected back to abundant compounds, such as A and B, possibly produced in primary metabolism, and in this process the selective pressure is placed on the capacity of enzyme evolution to continue metabolite production rather than the chance occurrence of novel activity gain[117]. In this model, pathway intermediates are assumed to be readily available, either from a nutrient rich environment or possibly through side activities of endogenous enzymes. This comprises the main criticism of this model as many biosynthetic pathway intermediates are unstable due to chemical lability, so accumulation into a sufficiently large precursor pool is unlikely[107, 110]. Therefore, the retrograde pathway evolution hypothesis may be limited to compounds that can be produced directly from molecules available in pre-biotic Earth, such as those resulting from Miller-Urey type experiments[118, 119].

Pathway Design Through Bioretrosynthesis

The pathway evolution schemes described above provide a number of hypothetical approaches that are available to use as inspiration for *de novo* construction of a non-natural biosynthetic pathway. In perhaps all cases, neofunctionalization via the patchwork hypothesis will be a predominant aspect of pathway engineering. Enzyme engineering and directed evolution are, by definition, processes for laboratory controlled enzyme recruitment through gene duplication and diversification that are analogous to those invoked as the basis of the patchwork model. The major difference is that the selection pressure for enzyme evolution in nature is organism survival, whereas in the laboratory setting the researcher defined quota for high yielding generation of a particular product is the mechanism by which suitable activity is judged.

The choice, then, in applying a proposed pathway evolution scheme for creation of a non-natural pathway is between the forward evolution or retrograde evolution hypotheses, as the shell model proposes a process of evolution for metabolism as a whole with forward evolution being the primarily force for individual pathway formation[115]. Applying the forward evolution theme, directed evolution would begin with the enzyme catalyzing the first step of the pathway, continuing forward recruiting and evolving enzymatic activity step-by-step toward the end of the pathway (Figure 1-14, top). Similar to the caveat of forward evolution in a natural pathway in that all pathway intermediates must provide a selective advantage to the host, constructing a non-natural pathway in this manner would require the development some kind of activity assay for every individual enzymatic step in order to observe the appearance and/or improvement in the desired activity. As each biosynthetic step presumably performs a different kind of biotransformation on the substrate, a unique selection or screen would need to be developed to determine activity for each enzyme evolved in the pathway (Figure 1-14, gray boxes). Design of a robust and reliable assay can be quite time consuming and in some cases may ultimately be the rate determining step for analyzing a library of enzyme variants for improved activity, especially when several assays need be developed for construction of a non-natural biosynthetic pathway. Additionally, it is also often necessary to go back and reoptimize select enzymatic reactions within the pathway to optimal flux, balance concentration of intermediates and enzyme expression levels or remove feedback inhibition once downstream reactions have been concatenated to the biosynthetic sequence[120].

The alternative is to apply the principles of retrograde evolution[117] to evolve the biosynthetic pathway starting with the product forming (ultimate) enzyme and progress in the reverse direction toward the beginning of the pathway. This process would entail recruiting and evolving appropriate enzymes to catalyze reactions beginning from

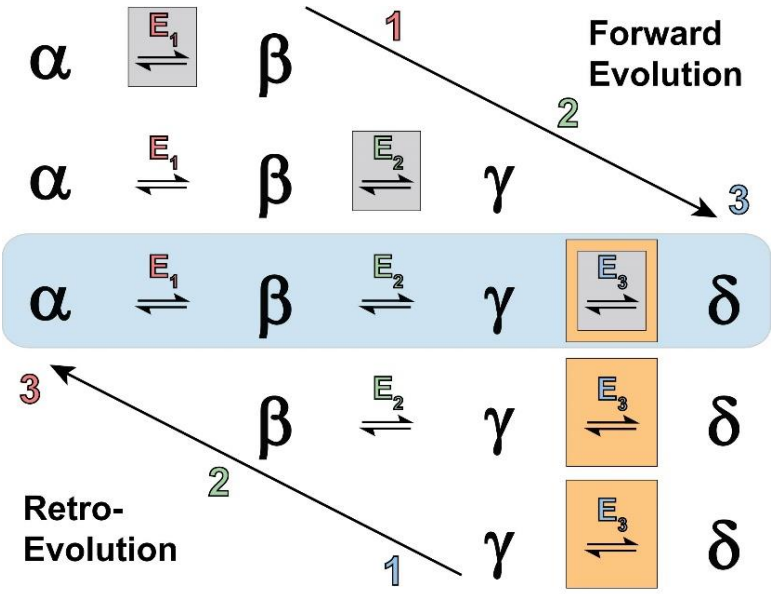

Figure 1-14. Pathway construction via forward and retrograde evolution schemes. Greek letters represent pathway substrates and products and enzymes are represented by the letter E. The order of evolution is indicated by the arrow and number colored the same as the corresponding enzyme at each step. Activity screening assays are indicated by gray and orange boxes.

anticipated pathway intermediates (Figure 1-14, bottom). Laboratory application of this strategy has been termed 'bioretrosynthesis' and is proposed to be a potentially generally applicable approach for non-natural pathway evolution by screening enzymatic activity in increasingly tandem assays[(105)]. Constructing a pathway in this way could allow a single, product forming assay to be used as the sole selection criterion for evolving the entire pathway (Figure 1-14, orange boxes), much like how a natural pathway would evolve for production of the beneficial final product in the Horowitz hypothesis[(117)]. Theoretically, this may greatly simplify the assay design component associated with directed evolution experiments by allowing a single assay to be used for screening enzyme variant libraries at each evolved step. As an added benefit, selecting for final product formation in the tandem assays required in bioretrosynthesis may ease the eventual tuning of pathway flux as each biosynthetic enzyme is engineered for increased production in the presence of all downstream enzymes. This progression

begins to link the activities of each enzyme together as a biosynthetic unit and facilitates pathway optimization throughout the building process. Instead of combining individually optimized pieces, enzymatic activity is evolved via comparison to the activity of the product forming enzyme, and therefore must only be improved so that it is no longer the rate limiting step rather than reaching maximal activity levels.

One key aspect of constructing a *de novo* biosynthetic pathway that is not encountered in the retrograde evolution of pathways in nature is that a deliberate series of biotransformations must be prearranged for conversion of a simple starting material to the desired final product, and is especially important for pathways utilizing non-natural substrates. Enzyme recruitment in a natural pathway relies on chance events of gene duplication and acquiring mutations in order to extend the pathway, however, these would not be sufficient for targeted high yielding production of compounds, especially when many would not provide a survival or growth benefit to the producing organism. Instead, following along with the inspiration from retrograde evolution, another element to bioretrosynthesis is the planning of a putative biosynthetic pathway in a strategy analogous to the conceptual approach of retrosynthesis used in synthetic organic chemistry[105, 121]. In a process that may simplify the task of designing a new biosynthetic pathway, the large collection of known enzymatic transformations, rather than the body of known chemical transformations, is considered in order to propose individual biosynthetic steps. Beginning from the final product, the researcher proposes a precursor or set of precursors that can undergo a particular enzyme catalyzed biotransformation to yield the target compound. These precursors can also then be the subject of an additional bioretrosynthetic step through identifying precursors for the precursors via another enzymatic activity, and so on until an appropriate simple starting material for the biosynthetic pathway has been reached[105].

Application of bioretrosynthesis for construction of a *de novo* biosynthetic pathway has several requirements[105]. Much like all directed evolution experiments, genes encoding potential progenitor enzymes for all steps of the pathway must be available. While this condition may be simple and rather obvious in nature, the consideration is made more complex when designing pathways catalyzing reactions on non-natural compounds. In many cases, a series of related enzymes from different genetic sources may need to be identified, collected and preliminarily tested for activity to determine a suitable progenitor enzyme for future engineering. Second, starting materials and intermediates must be readily available for use in assays. These materials may be of commercial or in-house synthetic origin, but as these compounds will likely not be generated initially at high yields during pathway evolution, stocks must be kept in-hand for screening and characterization of enzyme variants. The final major requirement is that of a robust, sensitive and reproducible assay for screening enzyme variant libraries. Assays employed should be capable of facilitating at least a medium throughput capacity and must be sensitive enough to detect low turnover and small increases in activity gained through evolution. Assay sensitivity also becomes a necessity as formation of the final product, and therefore detectable signal, may likely diminish as the tandem assays are extended into longer biosynthetic pathways.

Bioretrosynthetic design of a dideoxyinosine biosynthetic pathway

A biosynthetic pathway for production of dideoxyinosine has been planned using the retrosynthetic aspect of the bioretrosynthesis strategy. Biotransformations were proposed for the non-natural pathway based on similarity of the non-natural product to natural molecules and known enzymatic reactions. The high similarity between dideoxyinosine and the naturally occurring nucleoside inosine allowed for potential enzymes and biosynthetic pathways to be identified based on the production routes of

the known nucleosides described in the BRENDA and KEGG databases. Multiple pathways of varying length and complexity can be proposed using these natural pathways as a model (Figure 1-15), and each of which will be described.

Nucleosides can be generated through two very different biosynthetic pathways. The *de novo* pathway enzymatically builds the nucleobase component step by step onto the activated sugar substrate phosphoribosyl pyrophosphate (PRPP) consuming amino acids, folate, ATP and CO_2 to form the nucleotide. The produced mononucleotide can then be dephosphorylated to yield the nucleoside (Figure 1-15, blue arrows). Nucleoside production through the full *de novo* pathway consists of 13-15 enzymes for purines and 9-14 enzymes for pyrimidines, making biosynthesis through this route is a very expensive task from the perspective of energy and metabolites required. Emulating this biosynthetic pathway for dideoxyinosine production beginning with a simple sugar precursor and continuing through the *de novo* route would take an extraordinary research effort. Production of non-natural nucleosides through this pathway may require the evolution of approximately a dozen sequential enzymes to complete the system. Primarily for this reason, following the *de novo* biosynthetic route is not a realistic option for production of a non-natural nucleoside analog.

Instead of proceeding through the entire *de novo* route, stepwise construction of the nucleobase can be circumvented by creating nucleosides or nucleotides through the nucleobase salvage pathway. Salvage enzymes exist so that an organism can reclaim high value sources of sugar and nucleobases[(122)], among other metabolites, from the environment as a way to bypass the biosynthesis process from scratch. In addition to the *de novo* pathway, PRPP is similarly a substrate for hypoxanthine phosphoribosyl transferase (HPRT) to generate inosine monophosphate through the addition of hypoxanthine. Again, the nucleotide formed in this way can yield the unphosphorylated nucleoside through the activity of a nucleotidase. Inosine can also be formed directly by

purine nucleoside phosphorylase (PNP) from ribose 1-phosphate and hypoxanthine. The potential dideoxysugar substrates for these enzymes, if used to produce dideoxyinosine, would be 5-phospho-2,3-dideoxyribose 1-pyrophosphate and dideoxyinosine monophosphate for the HPRT and nucleotidase reactions, respectively, or 2,3-dideoxyribose 1-phosphate for PNP. Each of these three enzymes are very well

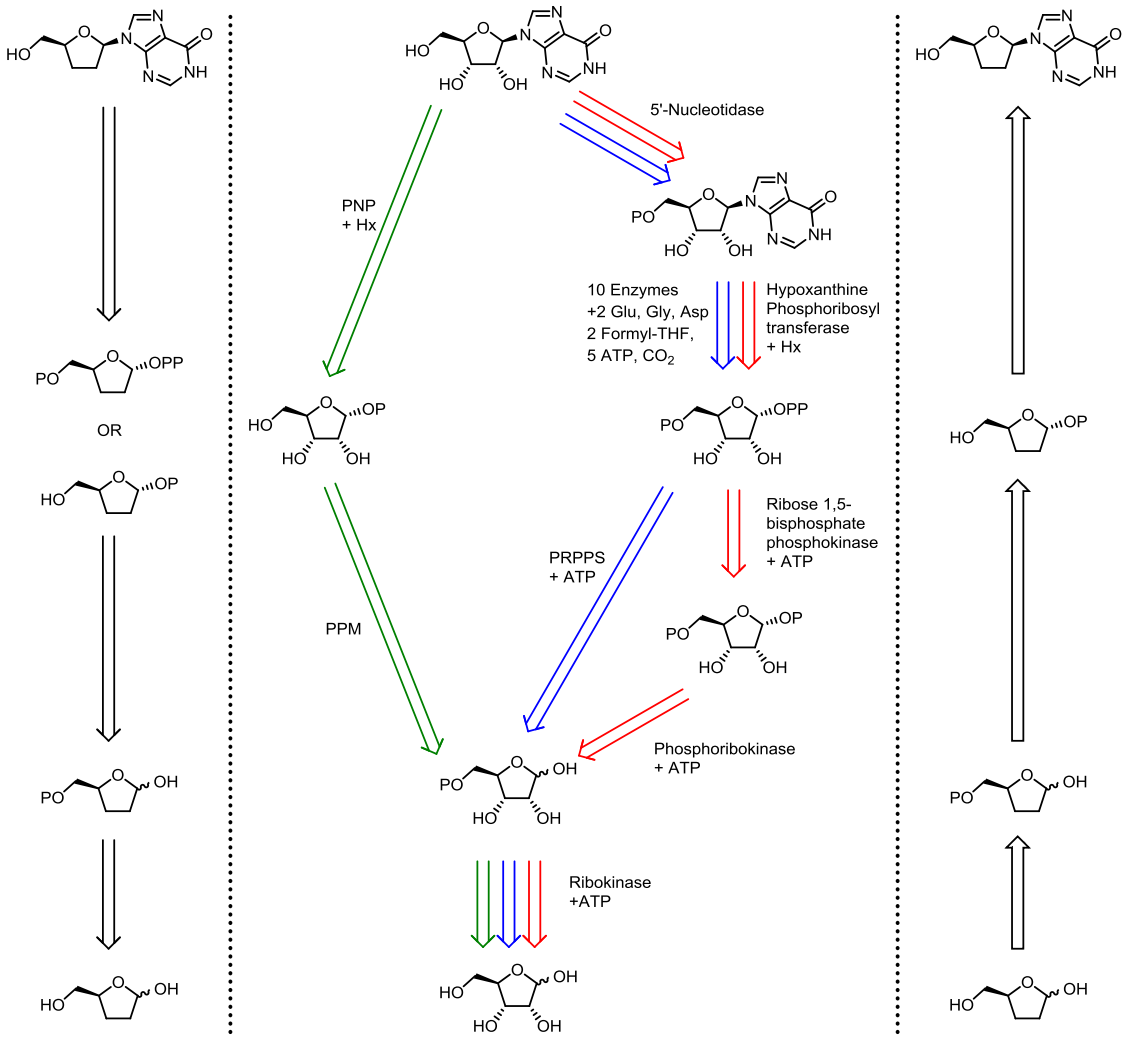

Figure 1-15. Bioretrosynthetic analysis of inosine biosynthesis routes to identify potential dideoxyinosine biosynthetic pathways. Natural substrates and intermediates in inosine biosynthesis are indicated in the center, with corresponding dideoxy sugar substrates on the left. The series of substrates required for the chosen pathway consisting of ribokinase, phosphopentomutase and purine nucleoside phosphorylase (green arrows) are shown on the right. PNP: purine nucleoside phosphorylase, Hx: hypoxanthine, PRPPS: phosphoribosyl pyrophosphate synthetase, PPM: phosphopentomutase.

studied from both a biochemical and structural standpoint, which would be of great help in a subsequent directed evolution experiment for improving activity on the potential non-natural substrates.

Formation of the PRPP precursor, or the non-natural dideoxy sugar analog, could occur in a single or dual step process. PRPP can be formed directly from ribose 5-phosphate in a one-step pyrophosphorylation reaction catalyzed by phosphoribosyl pyrophosphate synthetase (PRPPS). Additionally, PRPP can be produced through ribose 1,5-bisphosphate phosphokinase catalyzed phosphorylation of ribose 1,5-bisphosphate, which is generated by phosphoribokinase from ribose 5-phosphate. While these three enzymes may be envisioned to catalyze similar reactions on the dideoxy substrates, not all may be very likely candidates. Ribose 1,5-bisphosphate phosphokinase and phosphoribokinase are very minimally studied enzymes, and only the former has been cloned for overexpression and characterization just relatively recently[(123)]. PRPPS enzymes from several species, on the other hand, have been structurally and biochemically characterized, including activity on multiple substrates as well as determining the effects of mutagenesis at certain positions[(124-126)] and seem to be the more suitable candidate enzyme for production of PRPP through this pathway. The analogous non-natural substrates for these enzymes would be dideoxyribose 5-phosphate for PRPPS and phosphoribokinase and dideoxyribose 1,5-bisphosphate for ribose 1,5-bisphosphate phosphokinase.

Although the previously mentioned phosphoribokinase can phosphorylate ribose 5-phosphate at the C1 position, prior to my studies no enzyme is known to catalyze this reaction on ribose directly. The only enzyme known to perform a similar reaction is S-methyl-5-thioribose kinase, which phosphorylates the ribose-like substrate 5-methylthioribose at the C1 hydroxyl group as part of the methionine salvage pathway in bacteria[(127)]. However, this enzyme has not been shown to catalyze the similar

reaction with ribose, and as a result may require additional engineering to accept the differences in the 5-hydroxy sugar substrate as well as the removal of the hydroxyl groups at the C2 and C3 positions. Therefore, when planning the biosynthetic pathway prior to beginning enzyme evolution, the ribose 1-phosphate substrate for PNP could only be generated by phosphopentomutase (PPM) through isomerization of ribose 5-phosphate. While very little structural data was available for PPM enzymes at the beginning of this project, activity on dideoxyribose 5-phosphate, the analogous non-natural substrate for production of dideoxyinosine, has been demonstrated in the homolog from *Bacillus*[(128)], making this a strong potential candidate for inclusion in the dideoxyinosine biosynthetic pathway.

The ribose 5-phosphate precursor that feeds into all of the described pathways is formed through phosphorylation of the abundant metabolite ribose by ribokinase. Ribokinase homologs from many organisms have been extensively characterized, and therefore provide a large knowledgebase to use as reference while planning and performing directed evolution experiments. For this enzyme, then, the desirable non-natural substrate for eventual production of dideoxyinosine would be the simple sugar analog 2,3-dideoxyribose, leading to production of dideoxyribose 5-phosphate.

Considering these multiple routes, the pathway progressing through PRPP via the nucleobase salvage system requires 4-5 enzymes to generate dideoxyinosine from dideoxyribose (Figure 1-15, blue and red arrows) while the pathway that utilizes PPM consists of only three enzymes (Figure 1-15, green arrows). This shorter, PPM-catalyzed route also requires only one equivalent of ATP, versus the 2-3 needed to reach PRPP, which is an important consideration for future use of the pathway either as a hosted biosynthetic pathway or in an *in vitro* setting, as meeting the necessary precursor supply can often be challenging. Additionally, as each of these enzymes catalyze reactions on ribosyl substrates, catalytic activity on any of the proposed dideoxyribosyl molecules

would presumably be quite low. For this reason, enzyme engineering is expected to be necessary for each enzyme to be used in the dideoxyinosine biosynthetic pathway in order to increase activity on the corresponding non-natural substrate. While improved methods in directed evolution have increased the number of apparent successes and applications of biocatalysis, there is still a certain level of risk associated with enzyme engineering in not meeting the desired goals. Following the shortest and most direct pathway to the final product that proceeds through well studied enzymes are strategies to minimize these risks. Therefore, the biosynthetic pathway consisting of ribokinase (RK), phosphopentomutase (PPM) and purine nucleoside phosphorylase (PNP) was selected as the most desirable route for production of dideoxyinosine from dideoxyribose (Figure 1-15, green arrows). In this pathway, RK will phosphorylate dideoxyribose to form dideoxyribose 5-phosphate. This compound is then isomerized by PPM to create dideoxyribose 1-phosphate. The biosynthetic pathway is completed though PNP catalyzed addition of hypoxanthine to dideoxyribose 1-phosphate to form the desired non-natural nucleoside analog dideoxyinosine.

The sugar analog pathway precursor dideoxyribose can be chemically synthesized from glutamic acid in an extension of the procedure provided by Okabe *et al.*[(129)]. Briefly, glutamic acid can be cyclized in the presence of $NaNO_2$ under acidic conditions to provide *(S)*-γ-butyrolactone-γ-carboxylic acid. Cyclization under these conditions allows a retention of the stereochemistry at the gamma position due to double inversion, which is important for appropriate production of the dideoxy analog of D-ribose. Additionally, the very low cost of glutamic acid permits an economical option for synthesis of the substrate. A short sequence of reduction reactions on the carboxylic acid and the lactone moieties yield the final dideoxyribose substrate that can be introduced into the enzymatic portion of the proposed semisynthetic pathway (Figure 1-16).

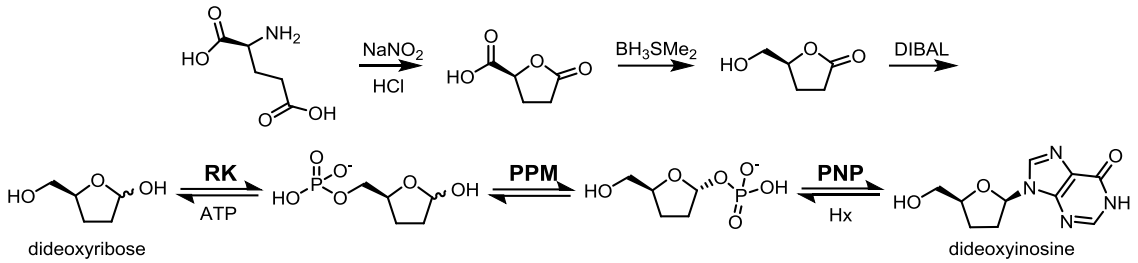

Figure 1-16. Proposed semi-synthetic route for production of dideoxyinosine. Glutamic acid is used to chemically synthesize the non-natural sugar analog dideoxyribose to serve as the biosynthetic pathway precursor.

Bioretrosynthetic construction of dideoxyinosine biosynthetic pathway

The bioretrosynthesis strategy not only provides a retrosynthetic platform for planning and designing a biosynthetic pathway, but also borrows the theme of retrograde evolution[(117)] in that the pathway is constructed and optimized in reverse. Enzyme engineering for the non-natural substrates begins with the product forming enzyme, and then continues in a reverse step-wise manner toward the first enzyme in the pathway. Following this course, only one screening or selection method for the terminal enzyme activity is needed for evolving all enzymes within the biosynthetic pathway. In the proposed RK-PPM-PNP pathway, all assays will focus on activity of PNP in individual or tandem assays measuring the consumption/production of hypoxanthine or production of the final nucleoside (Figure 1-17).

To begin optimization, the wild-type human variant of PNP (hPNP) was engineered to catalyze the phosphorolysis of dideoxyinosine into dideoxyribose 1-phosphate and hypoxanthine[(130)]. After analyzing published cocrystal structures with the natural substrate and analogs, the active site residue Tyr88 was predicted to modulate selectivity for ribosyl substrates. Based on these structures, the computational prediction software RosettaLigand[(131)] was used to calculate binding energies of transition state complexes for inosine and dideoxyinosine in single mutants of hPNP at

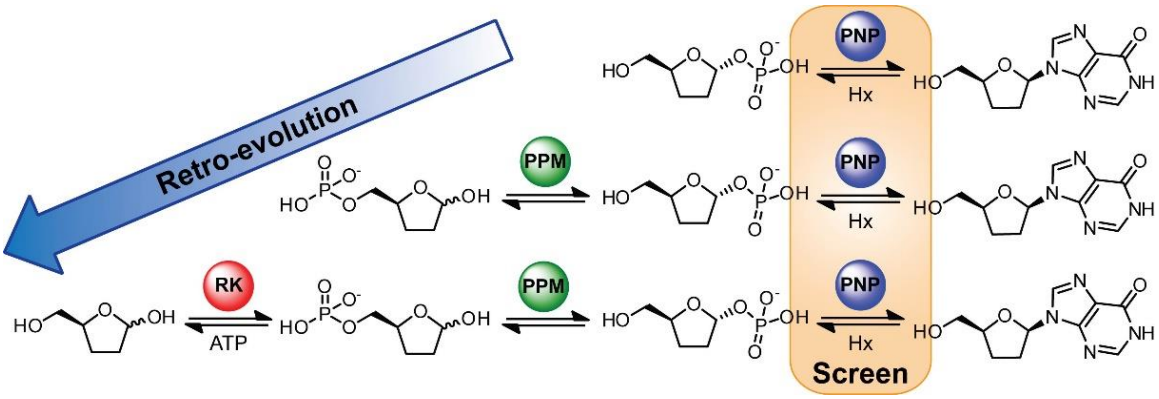

Figure 1-17. The single screen requirement of the bioretrosynthetic pathway construction strategy. Activity of each enzyme, PNP, PPM and RK can be determined through measuring activity of the terminal enzymatic step.

position 88, and predicted a Tyr88Phe mutation to improve PNP activity on dideoxyinosine. Simultaneous biochemical characterization of the mutants verified Tyr88Phe as the most beneficial residue at the position, providing a 9-fold improvement in k_{cat} and more than 2-fold decrease in K_M. Directed evolution by random mutagenesis was then used over three rounds to further improve activity, yielding an enzyme with approximately 22-fold increased catalytic efficiency over the wild-type PNP and 36-fold increased activity in cell lysate[(130)].

Having gained these large improvements in dideoxyinosine catalysis as the ultimate step of the pathway, the next phase of the pathway engineering is to evolve the penultimate enzyme, phosphopentomutase, to create the required dideoxyribose 1-phosphate from a dideoxyribose 5-phosphate precursor. The final step of pathway construction would then be the engineering of the antepenultimate enzyme, ribokinase, to form dideoxyribose 5-phosphate from the sugar analog dideoxyribose in an equivalent reaction to the naturally catalyzed phosphorylation of ribose. Successful evolution of these three enzymes would create the first biosynthetic pathway for production of a nucleoside analog and would serve as a demonstration of the potential of bioretrosynthesis as a pathway construction and optimization paradigm.

Dissertation Statement

While very few *de novo* biosynthetic pathways have been created for production of non-natural molecules, there is only one example where all biosynthetic enzymes were engineered for increased activity[(95)]. These pathways are rare developments as there is often a great deal of difficulty in planning, constructing and optimizing the entire pathway by hand since no general methods exist to simplify the pathway construction process. The research in this dissertation focuses on developing a potentially widely applicable strategy for *de novo* design and construction of non-natural biosynthetic pathways and is demonstrated through the creation of a pathway for production of the nucleoside analog dideoxyinosine.

A semi-synthetic pathway to produce dideoxyinosine consisting of a short chemical synthesis of the sugar analog dideoxyribose followed by a three enzyme biosynthetic sequence has been proposed using the principles of a bioretrosynthetic pathway design strategy. Application of a biosynthetic method for production of therapeutically useful nucleoside analog drugs from inexpensive starting materials has the potential to directly affect the cost of treatment, as a large percentage of treatment cost is associated with production of the nucleoside analog itself. Successful demonstration and extensive optimization of this biosynthetic pathway may provide a foundational process possessing economic advantages over the traditional chemical synthesis routes beginning with expensive natural nucleoside precursors, and may provide a more affordable and world-wide accessible treatment option.

Enzyme directed evolution is an essential aspect of many biocatalyst and biosynthetic pathway engineering studies, as turnover rates routinely need to be improved for application on a large scale. This is especially true when non-natural substrates and/or non-native reaction conditions must be used during the manufacturing process. Chapters II and III describe the directed evolution of phosphopentomutase

utilizing structure based targeted mutagenesis and whole gene random mutagenesis approaches, respectively, as methods to increase activity of the enzyme on the non-natural substrate dideoxyribose 5-phosphate.

Ribokinase is projected to be the leading candidate for the proposed reaction in the designed non-natural biosynthetic pathway to maintain continuity with natural nucleoside biosynthetic routes in this analogous pathway and also since ribose is a very similar natural metabolite to the non-natural substrate. However, it is equally possible that other sugar or small molecule kinase enzymes can fill this role in the proposed biosynthetic pathway for dideoxyinosine as phosphorylation of an alcohol group is a very common enzymatically catalyzed reaction in metabolism. For this reason, a panel of potential kinase progenitor enzymes was identified and screened for the desired activity and is presented in Chapter IV.

Chapter V describes the retrosynthetic construction of the biosynthetic pathway, comparing production of inosine and dideoxyinosine using either wild-type or evolved enzymes *in vitro*. The biosynthetic enzymes are formed into progressively longer tandem pathways leading to the demonstration of dideoxyinosine production through the full pathway beginning with dideoxyribose. Additionally, initial results in engineering ribokinase through site directed mutagenesis are presented.

The work presented in this dissertation demonstrates the ‘bioretrosynthesis’ strategy for *de novo* biosynthetic pathway construction and optimization through the creation of a biosynthetic pathway for the nucleoside analog drug dideoxyinosine. This method is a laboratory application of the theory of retrograde evolution for the creation of a non-natural pathway. With the main practical feature being the necessity of only one screening assay for evolution of all enzymes within the biosynthetic pathway, this approach can simplify the process of pathway evolution by greatly reducing the time and

effort required for assay design and should ultimately provide a widely applicable strategy for creating *de novo* biosynthetic pathways to produce non-natural compounds.

References

1. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, 485, 185-194.

2. Endy, D. Foundations for engineering biology. *Nature* **2005**, 438, 449-453.

3. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.

4. Desai, A. A. Sitagliptin manufacture: A compelling tale of green chemistry, process intensification, and industrial asymmetric catalysis. *Angew. Chem. Int. Ed. Engl.* **2011**, 50, 1974-1976.

5. WHO. Transaction prices for antiretroviral medicines from 2009 to 2012. *World Health Organization Global Price Reporitng Mechanism Summary Report* **2012**, 1-20.

6. WHO. HIV/AIDS Fact Sheet No 360. *World Health Organization* **2013**.

7. Pinheiro, E., Vasan, A., Kim, J. Y., Lee, E., Guimier, J. M., and Perriens, J. Examining the production costs of antiretroviral drugs. *AIDS* **2006**, 20, 1745-1752.

8. Kaminski, P. A., Dacher, P., Dugue, L., and Pochet, S. In vivo reshaping the catalytic site of nucleoside 2'-deoxyribosyltransferase for dideoxy- and didehydronucleosides via a single amino acid substitution. *J. Biol. Chem.* **2008**, 283, 20053-20059.

9. Prasad, A. K., Trikha, S., and Parmar, V. S. Nucleoside synthesis mediated by glycosyl transferring enzymes. *Bioorg. Chem.* **1999**, 27, 135-154.

10. Rogert, M. C., Trelles, J. A., Porro, S., Lewkowicz, E. S., and Iribarren, A. M. Microbial synthesis of antiviral nucleosides using *Escherichia coli* BL21 as biocatalyst. *Biocatal. Biotransformation* **2002**, 20, 347-351.

11. Shirae, H., Kobayashi, K., Shiragami, H., Irie, Y., Yasuda, N., and Yokozeki, K. Production of 2',3'-dideoxyadenosine and 2',3'-dideoxyinosine from 2',3'-dideoxyuridine and the corresponding purine bases by resting cells of *Escherichia coli* AJ 2595. *Appl. Environ. Microbiol.* **1989**, 55, 419-424.

12. Utagawa, T. Enzymatic preparation of nucleoside antibiotics. *J. Mol. Catal., B Enzym.* **1999**, 6, 215-222.

13. Hudlicky, T., and Reed, J. W. Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. *Chem. Soc. Rev.* **2009**, 38, 3117-3132.

14. Khersonsky, O., Roodveldt, C., and Tawfik, D. S. Enzyme promiscuity: Evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **2006**, 10, 498-508.

15. Jaeckel, C., Kast, P., and Hilvert, D. Protein design by directed evolution, In *Annu. Rev. Biophys.* **2008**, 37, pp 153-173.

16. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. X., and Vogel, A. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* **2005**, 44, 4192-4196.

17. Morley, K. L., and Kazlauskas, R. J. Improving enzyme properties: When are closer mutations better? *Trends Biotechnol.* **2005**, 23, 231-237.

18. Nannemann, D. P., Birmingham, W. R., Scism, R. A., and Bachmann, B. O. Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med. Chem.* **2011**, 3, 803-819.

19. Shivange, A. V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* **2009**, 13, 19-25.

20. Miyazaki, K., and Arnold, F. H. Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *J. Mol. Evol.* **1999**, 49, 716-720.

21. Bernhardt, P., McCoy, E., and O'Connor, S. E. Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **2007**, 14, 888-897.

22. DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z. L., Tomlinson, G., Huang, H. J., Tan, X. Q., Bibbs, L., Chen, P., Kretz, K., and Burk, M. J. Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.* **2003**, 125, 11476-11477.

23. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., and Bocola, M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **2010**, 132, 9144-9152.

24. Fox, R. Directed molecular evolution by machine learning and the influence of nonlinear interactions. *J. Theor. Biol.* **2005**, 234, 187-199.

25. Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., and Huisman, G. W. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, 25, 338-344.

26. Wong, T. S., Zhurina, D., and Schwaneberg, U. The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* **2006**, 9, 271-288.

27. Wong, T. S., Roccatano, D., Zacharias, M., and Schwaneberg, U. A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* **2006**, 355, 858-871.

28. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **1996**, 255, 589-603.

29. Lai, Y. P., Huang, J., Wang, L. F., Li, J., and Wu, Z. R. A new approach to random mutagenesis in vitro. *Biotechnol. Bioeng.* **2004**, 86, 622-627.

30. Myers, R. M., Lerman, L. S., and Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. *Science* **1985**, 229, 242-247.

31. Alexeeva, M., Enright, A., Dawson, M. J., Mahmoudian, M., and Turner, N. J. Deracemization of alpha-methylbenzylamine using an enzyme obtained by in vitro evolution. *Angew. Chem. Int. Ed. Engl.* **2002**, 41, 3177-+.

32. Balashov, S., and Humayun, M. Z. Specificity of spontaneous mutations induced in mutA mutator cells. *Mutat. Res.* **2004**, 548, 9-18.

33. Chen, B., Cai, Z., Wu, W., Huang, Y. L., Pleiss, J., and Lin, Z. L. Morphing activity between structurally similar enzymes: From heme-free bromoperoxidase to lipase. *Biochemistry* **2009**, 48, 11496-11504.

34. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B. W., Reetz, M. T., and Jaeger, K. E. Directed evolution of an enantioselective lipase. *Chem. Biol.* **2000**, 7, 709-718.

35. Reetz, M. T. Controlling the enantioselectivity of enzymes by directed evolution: Practical and theoretical ramifications. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 5716-5722.

36. Reetz, M. T. Application of directed evolution in the development of enantioselective enzymes. *Pure Appl. Chem.* **2000**, 72, 1615-1622.

37. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994**, 370, 389-391.

38. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **1998**, 391, 288-291.

39. Christians, F. C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W. P. C. Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* **1999**, 17, 259-264.

40. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A., and Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **1998**, 16, 258-261.

41. Ostermeier, M., Shim, J. H., and Benkovic, S. J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **1999**, 17, 1205-1209.

42. Sieber, V., Martinez, C. A., and Arnold, F. H. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **2001**, 19, 456-460.

43. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D., and Benkovic, S. J. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 11248-11253.

44. Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A., Rothlisberger, D., and Baker, D. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **2006**, 15, 2785-2794.

45. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., III, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. *De novo* computational design of retro-aldol enzymes. *Science* **2008**, 319, 1387-1391.

46. Althoff, E. A., Wang, L., Jiang, L., Giger, L., Lassila, J. K., Wang, Z., Smith, M., Hari, S., Kast, P., Herschlag, D., Hilvert, D., and Baker, D. Robust design and optimization of retroaldol enzymes. *Protein Sci.* **2012**, 21, 717-726.

47. Giger, A., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N., and Hilvert, D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **2013**, 9, 494-U449.

48. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, 453, 190-U194.

49. Khersonsky, O., Rothlisberger, D., Dym, O., Albeck, S., Jackson, C. J., Baker, D., and Tawfik, D. S. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **2010**, 396, 1025-1042.

50. Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., Clair, J. L. S., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. Computational sesign of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010**, 329, 309-313.

51. Huisman, G. W., Liang, J., and Krebber, A. Practical chiral alcohol manufacture using ketoreductases. *Curr. Opin. Chem. Biol.* **2010**, 14, 122-129.

52. Liang, J., Mundorff, E., Voladri, R., Jenne, S., Gilson, L., Conway, A., Krebber, A., Wong, J., Huisman, G., Truesdell, S., and Lalonde, J. Highly enantioselective reduction of a small heterocyclic ketone: Biocatalytic reduction of tetrahydrothiophene-3-one to the corresponding (R)-alcohol. *Org. Process. Res. Dev.* **2010**, 14, 188-192.

53. Liang, J., Lalonde, J., Borup, B., Mitchell, V., Mundorff, E., Trinh, N., Kochrekar, D. A., Cherat, R. N., and Pai, G. G. Development of a biocatalytic process as an alternative to the (-)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org. Process. Res. Dev.* **2010**, 14, 193-198.

54. Glieder, A., Farinas, E. T., and Arnold, F. H. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* **2002**, 20, 1135-1139.

55. Fasan, R., Chen, M. M., Crook, N. C., and Arnold, F. H. Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew. Chem. Int. Ed. Engl.* **2007**, 46, 8414-8418.

56. Peters, M. W., Meinhold, P., Glieder, A., and Arnold, F. H. Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* **2003**, 125, 13442-13450.

57. Meinhold, P., Peters, M. W., Chen, M. M. Y., Takahashi, K., and Arnold, F. H. Direct conversion of ethane to ethanol by engineered cytochrome P450BM3. *Chembiochem* **2005**, 6, 1765-1768.

58. Ghislieri, D., Green, A. P., Pontini, M., Willies, S. C., Rowles, I., Frank, A., Grogan, G., and Turner, N. J. Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products. *J. Am. Chem. Soc.* **2013**, 135, 10863-10869.

59. Carr, R., Alexeeva, M., Enright, A., Eve, T. S. C., Dawson, M. J., and Turner, N. J. Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angew. Chem. Int. Ed. Engl.* **2003**, 42, 4807-4810.

60. Dunsmore, C. J., Carr, R., Fleming, T., and Turner, N. J. A chemo-enzymatic route to enantiomerically pure cyclic tertiary amines. *J. Am. Chem. Soc.* **2006**, 128, 2224-2225.

61. Koehler, V., Bailey, K. R., Znabet, A., Raftery, J., Helliwell, M., and Turner, N. J. Enantioselective Biocatalytic Oxidative Desymmetrization of Substituted Pyrrolidines. *Angew. Chem. Int. Ed. Engl.* **2010**, 49, 2182-2184.

62. Rowles, I., Malone, K. J., Etchells, L. L., Willies, S. C., and Turner, N. J. Directed evolution of the enzyme monoamine oxidase (MAO-N): Highly efficient chemo-enzymatic deracemisation of the alkaloid (+/-)-Crispine A. *Chemcatchem* **2012**, 4, 1259-1261.

63. Eriksen, D. T., Lian, J., and Zhao, H. Protein design for pathway engineering. *J. Struct. Biol.* **2013**.

64. Jackson, J., Williams, L. S., and Umbarger, H. E. Regulation of synthesis of the branched-chain amino acids and cognate aminoacyl-transfer ribonucleic acid synthetases of *Escherichia coli* - A common regulatory element. *J. Bacteriol.* **1974**, 120, 1380-1386.

65. Dalby, P. A., Baganz, F., Lye, G. J., and Ward, J. M. Protein and pathway engineering in biocatalysis. *Chim. Oggi* **2009**, 27, 18-+.

66. Li, R. F., and Townsend, C. A. Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in *Streptomyces clavuligerus*. *Metab. Eng.* **2006**, 8, 240-252.

67. Rokem, J. S., Lantz, A. E., and Nielsen, J. Systems biology of antibiotic production by microorganisms. *Nat. Prod. Rep.* **2007**, 24, 1262-1287.

68. Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U., and Lee, S. Y. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* **2007**, 3.

69. Luetke-Eversloh, T., Santos, C. N. S., and Stephanopoulos, G. Perspectives of biotechnological production of L-tyrosine and its applications. *Appl. Microbiol. Biotechnol.* **2007**, 77, 751-762.

70. Lee, S. Y., Kim, H. U., Park, J. H., Park, J. M., and Kim, T. Y. Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov. Today* **2009**, 14, 78-88.

71. Na, D., Kim, T. Y., and Lee, S. Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* **2010**, 13, 363-370.

72. Kang, Z., Wang, Q., Zhang, H., and Qi, Q. Construction of a stress-induced system in *Escherichia coli* for efficient polyhydroxyalkanoates production. *Appl. Microbiol. Biotechnol.* **2008**, 79, 203-208.

73. Li, R., Zhang, H., and Qi, Q. The production of polyhydroxyalkanoates in recombinant *Escherichia coli*. *Bioresour. Technol.* **2007**, 98, 2313-2320.

74. Du, J., Shao, Z., and Zhao, H. Engineering microbial factories for synthesis of value-added products. *J. Ind. Microbiol. Biotechnol.* **2011**, 38, 873-890.

75. Yin, E., Le, Y., Pei, J., Shao, W., and Yang, Q. High-level expression of the xylanase from *Thermomyces lanuginosus* in *Escherichia coli*. *World J. Microbiol. Biotechnol.* **2008**, 24, 275-280.

76. Di Guan, C., Li, P., Riggs, P. D., and Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **1988**, 67, 21-30.

77. Smith, D. B., and Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **1988**, 67, 31-40.

78. Lavallie, E. R., Diblasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *Escherichia coli* cytoplasm. *Biotechnology (N.Y.)* **1993**, 11, 187-193.

79. Panavas, T., Sanders, C., and Butt, T. R. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems, In *Methods Mol. Biol.* **2009**, 497, (Ulrich, H. D., Ed.), pp 303-317.

80. Martin, C. H., Nielsen, D. R., Solomon, K. V., and Prather, K. L. Synthetic metabolism: Engineering biology at the protein and pathway scales. *Chem. Biol.* **2009**, 16, 277-286.

81. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **2003**, 21, 796-802.

82. Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, 440, 940-943.

83. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013**, 496, 528-532.

84. Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., and Stephanopoulos, G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **2010**, 330, 70-74.

85. Prather, K. L., and Martin, C. H. *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **2008**, 19, 468-474.

86. Ewering, C., Lutke-Eversloh, T., Luftmann, H., and Steinbuchel, A. Identification of novel sulfur-containing bacterial polyesters: biosynthesis of poly(3-hydroxy-S-propyl-ω-thioalkanoates) containing thioether linkages in the side chains. *Microbiology* **2002**, 148, 1397-1406.

87. Zhang, M.-Q., Gaisser, S., Nur-E-Alam, M., Sheehan, L. S., Vousden, W. A., Gaitatzis, N., Peck, G., Coates, N. J., Moso, S. J., Radzom, M., Foster, T. A., Sheridan, R. M., Gregory, M. A., Roe, S. M., Prodromou, C., Pearl, L., Boyd, S. M., Wilkinson, B., and Martin, C. J. Optimizing natural products by biosynthetic engineering: Discovery of nonquinone Hsp90 inhibitors. *J. Med. Chem.* **2008**, 51, 5494-5497.

88. Moon, T. S., Yoon, S.-H., Lanza, A. M., Roy-Mayhew, J. D., and Prather, K. L. J. Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* **2009**, 75, 589-595.

89. Niu, W., Molefe, M. N., and Frost, J. W. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **2003**, 125, 12998-12999.

90. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **2011**, 7, 445-452.

91. Ran, N., and Frost, J. W. 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.* **2007**, 129, 6130-6139.

92. Zhang, K., Li, H., Cho, K. M., and Liao, J. C. Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 107, 6234-6239.

93. Zhang, K., Sawaya, M. R., Eisenberg, D. S., and Liao, J. C. Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, 105, 20653-20658.

94. Atsumi, S., Hanai, T., and Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, 451, 86-U13.

95. Ma, S. K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G. W., and Sheldon, R. A. A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem.* **2010**, 12, 81.

96. Apweiler, R., Martin, M. J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Alpi, E., Antunes, R., Arganiska, J., Casanova, E. B., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chan, W. M., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dimmer, E., Fazzini, F., Gane, P., Fedotov, A., Castro, L. G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Jacobsen, J., Jones, R., Legge, D., Liu, W., Luo, J., MacDougall, A., Mutowo, P., Nightingale, A., Orchard, S., Patient, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Sawford, T., Sehra, H., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Corbett, M., Donnelly, M., van Rensburg, P., Goujon, M., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Binz, P.-A., Blatter, M.-C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., de Castro, E., Cerutti, L., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., James, J., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lew, J., Lieberherr, D., Martin, X., Masson, P., Morgat, A., Neto, T., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.-L., Zerara, M., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Huang, H., Kukreja, A., Laiho, K., McGarvey, P., Natale, D. A., Natarajan, T. G., Roberts, N. V., Suzek, B. E., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L.-S., Yerramalla, M. S., Zhang, J.,

and UniProt, C. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **2013**, 41, D43-D47.

97. Schomburg, I., Chang, A., Placzek, S., Soehngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., and Schomburg, D. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Res.* **2013**, 41, D764-D772.

98. Kanehisa, M., and Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, 28, 27-30.

99. Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Paley, S., Popescu, L., Pujar, A., Shearer, A. G., Zhang, P., and Karp, P. D. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **2010**, 38, D473-D479.

100. Hatzimanikatis, V., Li, C. H., Ionita, J. A., Henry, C. S., Jankowski, M. D., and Broadbelt, L. J. Exploring the diversity of complex metabolic networks. *Bioinformatics* **2005**, 21, 1603-1609.

101. Ellis, L. B. M., Gao, J., Fenner, K., and Wackett, L. P. The University of Minnesota pathway prediction system: Predicting metabolic logic. *Nucleic Acids Res.* **2008**, 36, W427-W432.

102. Rodrigo, G., Carrera, J., Prather, K. J., and Jaramillo, A. DESHARKY: Automatic design of metabolic pathways for optimal cell growth. *Bioinformatics* **2008**, 24, 2554-2556.

103. Medema, M. H., van Raaphorst, R., Takano, E., and Breitling, R. Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* **2012**, 10, 191-202.

104. Kwok, R. Five hard truths for synthetic biology. *Nature* **2010**, 463, 288-290.

105. Bachmann, B. O. Biosynthesis: Is it time to go retro? *Nat. Chem. Biol.* **2010**, 6, 390-393.

106. Lewis, E. B. Pseudoallelism and gene evolution. *Cold Spring Harbor Symp. Quant. Biol.* **1951**, 16, 159-174.

107. Fani, R., and Fondi, M. Origin and evolution of metabolic pathways. *Phys. Life Rev.* **2009**, 6, 23-52.

108. Waley, S. G. Some aspects of evolution of metabolic pathways. *Comp. Biochem. Physiol.* **1969**, 30, 1-11.

109. Ycas, M. On earlier states of the biochemical system. *J. Theor. Biol.* **1974**, 44, 145-160.

110. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **1976**, 30, 409-425.

111. Fondi, M., Emiliani, G., and Fani, R. Origin and evolution of operons and metabolic pathways. *Res. Microbiol.* **2009**, 160, 502-512.

112. Granick, S. Speculations on the origins and evolution of photosynthesis. *Ann. NY Acad. Sci.* **1957**, 69, 292-308.

113. Lazcano, A., and Miller, S. L. The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time. *Cell* **1996**, 85, 793-798.

114. Ullrich, A., Rohrschneider, M., Scheuermann, G., Stadler, P. F., and Flamm, C. *In silico* evolution of early metabolism. *Artif. Life* **2011**, 17, 87-108.

115. Morowitz, H. J. A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **1999**, 4, 39-53.

116. Rison, S. C. G., and Thornton, J. M. Pathway evolution, structurally speaking. *Curr. Opin. Struct. Biol.* **2002**, 12, 374-382.

117. Horowitz, N. H. On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U.S.A.* **1945**, 31, 153-157.

118. Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A., and Bada, J. L. The Miller volcanic spark discharge experiment. *Science* **2008**, 322, 404-404.

119. Miller, S. L., and C., U.-H. Organic compound synthesis on the primitive earth. *Science* **1959**, 130, 245-251.

120. Carothers, J. M., Goler, J. A., and Keasling, J. D. Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.* **2009**, 20, 498-503.

121. Corey, E. J. The logic of chemical synthesis - Multistep synthesis of complex natural carbogenic molecules. *Angew. Chem. Int. Ed. Engl.* **1991**, 30, 455-465.

122. Murray, A. W. Biological significance of purine salvage. *Annu. Rev. Biochem.* **1971**, 40, 811-&.

123. Hove-Jensen, B., Rosenkrantz, T. J., Haldimann, A., and Wanner, B. L. *Escherichia coli* phnN, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): Dual role in phosphonate degradation and NAD biosynthesis pathways. *J. Bacteriol.* **2003**, 185, 2793-2801.

124. Hove-Jensen, B., Bentsen, A. K. K., and Harlow, K. W. Catalytic residues Lys197 and Arg199 of *Bacillus subtilis* phosphoribosyl diphosphate synthase - Alanine-scanning mutagenesis of the flexible catalytic loop. *FEBS J.* **2005**, 272, 3631-3639.

125. Li, S., Lu, Y., Peng, B., and Ding, J. Crystal structure of human phosphoribosylpyrophosphate synthetase 1 reveals a novel allosteric site. *Biochem. J.* **2007**, 401, 39-47.

126. Parry, R. J., Burns, M. R., Skae, P. N., Hoyt, J. C., and Pal, B. Carbocyclic analogues of D-ribose-5-phosphate: Synthesis and behavior with 5-phosphoribosyl alpha-1-pyrophosphate synthetases. *Bioorg. Med. Chem.* **1996**, 4, 1077-1088.

127. Ku, S.-Y., Yip, P., Cornell, K. A., Riscoe, M. K., Behr, J.-B., Guillerm, G., and Howell, P. L. Structures of 5-methylthioribose kinase reveal substrate specificity and unusual mode of nucleotide binding. *J. Biol. Chem.* **2007**, 282, 22195-22206.

128. Hamamoto, T., Noguchi, T., and Midorikawa, Y. Phosphopentomutase of *Bacillus stearothermophilus* TH6-2: The enzyme and its gene ppm. *Biosci. Biotechnol. Biochem.* **1998**, 62, 1103-1108.

129. Okabe, M., Sun, R. C., Tam, S. Y. K., Todaro, L. J., and Coffen, D. L. Synthesis of the dideoxynucleosides ddC and CNT from glutamic acid, ribonolactone, and pyrimidine bases. *J. Org. Chem.* **1988**, 53, 4780-4786.

130. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

131. Meiler, J., and Baker, D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins* **2006**, 65, 538-548.

Chapter II

TARGETED SATURATION MUTAGENESIS OF *BACILLUS CEREUS* PHOSPHOPENTOMUTASE ACTIVE SITE RESIDUES

Introduction

Nucleoside analog drugs have been and continue to be prevalent forms of treatment with over 30 compounds having received FDA approval for use as antivirals or antiretrovirals (most notably in drug combination regimens for HIV treatment[1]), and are additionally proving their worth as cancer therapeutics[2, 3]. These drugs, however, are frequently quite expensive to manufacture chemically, where in some cases expenses of producing the active ingredient reach up to 99% of direct costs[4]. The cost applied to synthesis of these compounds is commonly a result of expensive starting materials and low yield due to significant formation of side products.

Chemical synthesis of nucleoside analogs often proceeds through a sugar intermediate that is activated at the C1 position by means of a halide or acetate leaving group in a racemic mixture. Due to this lack of stereocontrol, subsequent addition of a nucleobase creates an anomeric mixture of products leading to a loss of material due to improper addition. This is compounded by the lack of nucleobase regioselectivity, where a variety of structural isomers may be formed depending on the specific nitrogen on nucleobase that takes part in the nucleophilic addition. Instead, a biocatalytic approach could offer a more selective method of preparing the activated sugar or sugar analog and also may be used to catalyze the stereospecific addition of a nucleobase through synthesis in a tandem enzymatic system. As a large percentage of the therapeutic price of these drugs is tied to synthesis of the active ingredient, a more efficient and economical approach to synthesizing pharmaceutically relevant nucleoside analogs

could immediately alleviate some of these production costs by providing a direct route to a stereospecific product with fewer side reactions and wasteful byproducts.

We are interested in developing a biosynthetic process as a possible complementary method for producing nucleoside analogs, specifically 2',3'-dideoxyinosine (ddI, didanosine), a nucleoside analog reverse transcriptase inhibitor (NRTI) prescribed as treatment for HIV. Dideoxyinosine is considered a representative compound in the NRTI class of nucleoside analog therapeutics, as most members can be classified as dideoxynucleosides[1]. For this reason, a biosynthetic process to create dideoxyinosine may have broader applications for enzymatic synthesis of other pharmaceutically relevant nucleoside analogs as well.

Our approach to designing this biocatalytic route to dideoxyinosine, specifically an engineered multi-enzyme biosynthetic pathway, has been termed 'bioretrosynthesis'[5] and has been outlined in greater detail in Chapter I. In short, this pathway construction paradigm is to optimize enzymes for activity on non-natural substrates in the reverse order of biosynthesis, starting with the product forming enzyme, and selecting for final product formation. Applying this approach, our lab has previously engineered a human purine nucleoside phosphorylase (hPNP) variant that demonstrated 22-fold improved catalytic efficiency in the phosphorolysis of dideoxyinosine to the corresponding dideoxyribose 1-phosphate[6]. This enzyme, although optimized for the degradation of dideoxyinosine, was also improved in the direction of dideoxyinosine synthesis from dideoxyribose 1-phosphate and hypoxanthine, and could therefore be used as a foundation to identify and begin evolving an enzyme capable of producing dideoxyribose 1-phosphate.

A prerequisite to the enzyme engineering process is the identification of a progenitor enzyme that performs the desired transformation to use as an initial template for mutagenesis. Although not absolutely necessary, the progenitor should preferably

also show some promiscuous activity on the substrate of interest. The likelihood of success increases if the goal of enzyme evolution is to improve activity on the target substrate rather than impart a completely new capability[7]. A new biocatalyst with improved properties can then be generated from the existing enzyme through complementary methods in targeted and random mutagenesis. With the aid of structural data, active site residues can be rationally selected for targeted mutagenesis in an effort to modify enzyme selectivity for a desired substrate. Whole gene random mutagenesis through directed evolution can then be applied to discover beneficial mutations in an untargeted and more impartial manner.

Phosphopentomutases (PPMs) catalyze the interconversion of α-D-ribose 5-phosphate (ribose 5-phosphate) and α-D-ribose 1-phosphate (ribose 1-phosphate), creating a metabolic link between glycolysis and RNA metabolism[8]. In this pathway, ribose 1-phosphate serves as a precursor for nucleoside biosynthesis, accepting a nucleobase addition at the activated C1 position by a purine or pyrimidine nucleoside phosphorylase to form the respective nucleosides. The biocatalytic potential of PPM and nucleoside phosphorylase can be envisioned in a parallel series of reactions with other sugars or sugar analogs, taking advantage of the activation step performed by PPM in order to form nucleoside analogs, and has indeed been utilized in some biocatalysis studies[9-14].

As a biosynthetic retro-extension from the previously engineered PNP[6], we have endeavored to evolve a PPM variant for use in this non-natural biosynthetic pathway (Figure 2-1). In a tandem biocatalytic reaction to produce dideoxyinosine, the initial substrate for PPM would be 2,3-dideoxyribose 5-phosphate (dideoxyribose 5-phosphate). This would undergo an isomerization by PPM to form dideoxyribose 1-phosphate which then becomes a substrate for the previously evolved hPNP variant to catalyze the addition of hypoxanthine and form dideoxyinosine through displacing the

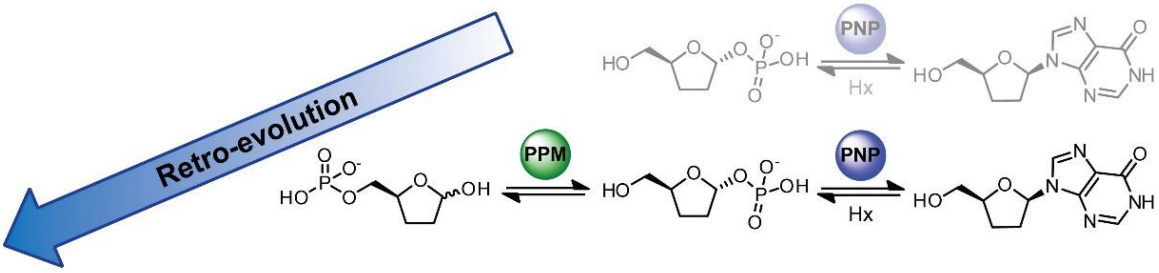

Figure 2-1. Retro-extension of the dideoxyinosine biosynthetic pathway to phosphopentomutase.

phosphate group. As dideoxyribose 5-phosphate is not a natural substrate for PPM the enzyme must be engineered to increase the level of desired activity.

Phosphopentomutase from *Bacillus cereus* was identified as a suitable progenitor enzyme for evolving activity on dideoxyribose 5-phosphate. As a preface to beginning evolution of PPM, we sought to further characterize this enzyme. Sequence and structural analysis classifies prokaryotic PPMs as members of the alkaline phosphatase superfamily of metalloenzymes, which includes cofactor-independent phosphoglycerate mutase, nucleotide pyrophosphatase/phosphodiesterase, phosphatases and sulfatases[(15-17)]. While most enzymes of this class function as hydrolases, PPM[(18)] and the cofactor-independent phosphoglycerate mutase[(19, 20)] catalyze phosphomutase reactions. PPM, however, has recently been characterized[(21)] as entering the general alkaline phosphatase mechanism at an altered point to undergo a unique intermolecular phosphate transfer reaction rather than an intramolecular phosphate transfer analogous to that observed in the cofactor-independent phosphoglycerate mutase[(19, 20)]. In this mechanism, active PPM is phosphorylated at Thr85, and during catalysis creates ribose 1,5-bisphosphate and a dephosphorylated enzyme as catalytic intermediates. A second phosphate transfer from the substrate back to Thr85 produces ribose 1-phosphate and the re-phosphorylated active PPM[(21)] (Figure

2-2, blue arrow). While undergoing the same overall cycle observed in the standard alkaline phosphatase mechanism (Figure 2-2, green arrow), PPMs have adapted to proceed through catalysis beginning from an altered entry point in order to perform this vital mutase reaction that connects the central metabolic pathways of glycolysis and ribonucleoside biosynthesis.

In this study, we have determined and analyzed cocrystal structures of wild-type and PPM variants with natural or non-natural substrates bound to indicate active site residues for targeted saturation mutagenesis. First shell active site residues were

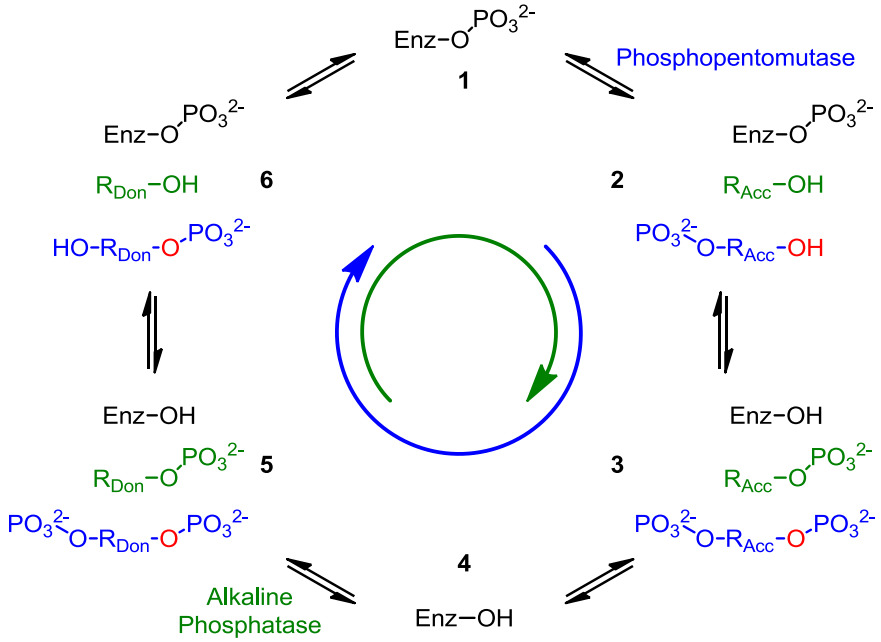

Figure 2-2. Comparison of PPM and alkaline phosphatase catalytic cycles. The PPM catalyzed intermolecular phosphate transfer follows the reaction cycle indicated in blue in order of states **1** through **6**. The phosphorylated enzyme **1** binds a phosphorylated substrate then transfers the phosphate group to the acceptor hydroxyl on the same substrate (indicated in red) to create a bisphosphate substrate and a dephosphorylated enzyme as intermediates **2-3**. Subsequent steps transfer a phosphate from the donor substrate back to the enzyme resulting in a singly phosphorylated substrate and the active phosphorylated enzyme **4-6**. The alkaline phosphatase cycle (shown in green) begins with an unphosphorylated enzyme **4** that binds a phosphorylated substrate and proceeds through an intermediate where the enzyme is phosphophorylated **1** before transferring the phosphate to the acceptor substrate **6-3**. Adapted from Panosian *et al.*[21].

selected for mutagenesis based on apparent interactions and close proximity to the bound substrates. The activity screening program for this targeted mutagenesis approach was designed to identify mutations that reduced natural substrate selectivity by comparing the activity of each clone on both substrates, ribose 5-phosphate and dideoxyribose 5-phosphate. The Ser154Gly and Val158Leu PPM variants were found to have substantially increased substrate selectivity over the wild-type enzyme, 49-fold and 881-fold, respectively. These two variants were unique in their new active site and activity characteristics as a result of each mutation, so both became templates for directed evolution to further increase activity.

Methods

PPM mutant library generation

Site-directed and saturation mutagenesis were performed using the QuikChange II mutagenesis kit (Stratagene) with either the wild-type PPM template (Genbank Accession Number Q818Z9.1) (for positions Ser154, Val158 and Ile195) or the corresponding Ser154Gly template (positions Val158 and Ile195) in pET28a+ vector[(22)]. PCR samples were prepared as recommended in the kit manufacturer's manual using 50 ng template and 125 ng forward and reverse primer and amplified through 16 cycles with 1 minute annealing at 55°C and 2 minute/kb extension at 68°C. *DpnI* restriction endonuclease was used to digest the template plasmid DNA prior to purification of the mutant plasmid by QIAquick PCR Purification Kit (QIAgen, Inc.) and subsequent transformation into BL21(*DE3*) cells (Invitrogen). Primers used for site directed mutagenesis and saturation mutagenesis are provided in Table 2-1.

Table 2-1. Primers used in site directed and saturation mutagenesis of PPM. Mutations in each sequence are underlined. N=A, T, C or G. K=G or T. M=C or A.

Primer Name	Nucleotide Sequence
S154A for	CAGGCTCTTTAATCGTTTATACTGCCGCTGATAGCGTATTGCAAATTGCAGC
S154A rev	GCTGCAATTTGCAATACGCTATCAGCGGCAGTATAAACGATTAAAGAGCCTG
S154X for	GGAAACAGGCTCTTTAATCGTTTATACTNNKGCTGATAGCGTATTGC
S154X rev	GCAATACGCTATCAGCMNNAGTATAAACGATTAAAGAGCCTGTTTCC
V158X for	GCTGATAGCNNKTTGCAAATTGCAGCACACGAAGAAGTAGTGCCAC
V158X rev	GTGGCACTACTTCTTCGTGTGCTGCAATTTGCAAMNNGCTATCAGC
I195X for	GGTAGGTCGTGTTNNKGCTCGTCCATTCGTTGGTGAACCTG
I195X rev	CAGGTTCACCAACGAATGGACGAGCMNNAACACGACCTACC

Library growth and screening

Individual colonies were picked to round bottomed 96-well plates and grown in 100 µL LB medium with 50 µg/mL kanamycin for 24 h at 37°C with shaking at 200 rpm. Plates held 4 wells of negative control (*E. coli* with empty pET28a+ vector), 4 wells of positive control (vector with template gene) and 88 wells of the mutant library. For saturation mutagenesis libraries (176 clones), plates were replicated to 100 µL and 150 uL LB medium for ribose 5-phosphate and dideoxyribose 5-phosphate, respectively, (wild-type PPM template) or 50 µL and 150 µL LB medium for ribose 5-phosphate and dideoxyribose 5-phosphate (Ser154Gly PPM template) and grown for an additional 24 h. The 50 µL cultures of Ser154Gly PPM were induced by addition of 50 µL LB medium containing 50 µg/mL kanamycin and 2 mM IPTG and grown for an additional 24 h. All cells were collected by centrifugation at 1600 rcf and the broth was removed by inversion before storing the plates at -80°C until ready for assay.

PPM activity was determined in tandem assays with purine nucleoside phosphorylase and hypoxanthine consumption was measured by xanthine oxidase at

endpoints using a typical hypoxanthine detection assay[23]. Thawed cell pellets were resuspended in 200 µL (for ribose 5-phosphate assay) or 100 µL (for dideoxyribose 5-phosphate assay) lysis mix containing 0.1 mM $MnCl_2$, 0.25 mg/mL egg white lysozyme (Sigma) and DNase I (Sigma) in 25 mM Tris-HCl, pH 8 and incubated for 10 min at 25°C before undergoing a single freeze/thaw cycle at -80°C to 37°C. After centrifugation, 25 µL of the clarified cell lysate was transferred to a 96-well flat bottomed plate and 55 µL of an assay mix was added to initiate the reaction. Final concentrations of components in 80 µL reactions for the ribose 5-phosphate activity screen were 0.1 mM $MnCl_2$, 5 µM PNP, 1 µM glucose 1,6-bisphosphate, 600 µM hypoxanthine and 1 mM ribose 5-phosphate. Final concentrations of components in 80 µL reactions for the dideoxyribose 5-phosphate screen were 0.1 mM $MnCl_2$, 10 µM hPNP-46D6, 1 µM glucose 1,6-bisphosphate, 600 µM hypoxanthine and 2 mM dideoxyribose 5-phosphate. Assays were incubated at room temperature for 12 min (ribose 5-phosphate) or from 50 - 80 min (dideoxyribose 5-phosphate) before quenching by addition of either 30 µL 1 M HCl (ribose 5-phosphate) or 30.5 µL 1M NaOH (dideoxyribose 5-phosphate). After a minimum of 30 min, 30 µL 1 M NaOH or 29.5 µL 1 M HCl was added to neutralize the reaction before addition of 35 µL of a developing solution containing 0.2% Triton X-100, 7.5 mM iodonitrotetrazolium chloride and xanthine oxidase in 25 mM Tris-HCl, pH 8. Hypoxanthine consumption was determined by measuring maximal absorbance of the colored formazan at 546 nm and normalized to percent activity in comparison to the positive and negative controls. Hits from each primary screen were regrown from fresh transformants and assayed again in duplicate to validate activity of the top ~45 hits. Top 4 - 5 hits were assayed in a tertiary screen with extended incubation with ribose 5-phosphate (25 min) and shorter incubation with dideoxyribose 5-phosphate (40 - 70 min) to better determine activity of each clone on the two substrates.

Enzyme expression and purification

Plasmids containing wild-type or variant PPM, PNP or hPNP-46D6[6] were transformed into *E. coli* BL21(*DE3*) and grown at 37°C in LB medium supplemented with 50 µg/mL kanamycin and induced with 1 mM IPTG for 3 - 6 h after OD_{600} had reached 0.5 - 0.6. Cell pellets were resuspended in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 7.4) and disrupted by passing through a French Pressure cell (1400 psi). The clarified lysate was applied to HisTrapFFcrude Nickel affinity column (GE Healthcare, Inc.) and washed at 10% Buffer B (Buffer A with 500mM imidazole). Protein was eluted by a linear gradient from 10% Buffer B to 60% Buffer B, before a step up to 100% Buffer B to fully elute the column. The purified enzyme was concentrated, desalted and exchanged into 25 mM Tris-HCl, pH 8 before storage at -80°C. All enzyme concentrations were determined using the BCA Protein Assay Kit (Thermo Scientific, Inc.). Xanthine oxidase was purified from raw bovine milk using previously reported protocols[24].

PPM kinetics assays

The activities of wild-type and variant PPMs were measured in tandem assays with either PNP or hPNP-46D6. Ribose 1-phosphate formed by PPM was subsequently consumed by a catalytic excess of PNP in the presence of hypoxanthine to produce inosine. Similarly, production of dideoxyribose 1-phosphate via PPM activity was converted to dideoxyinosine in the presence of hypoxanthine and a catalytic excess of hPNP-46D6. Inosine or dideoxyinosine produced in the assay was separated from other reaction components using a Luna Phenyl-Hexyl column (4.6 X 250 mm, Phenomenex) and an isocratic flow of 1.0 mL/min of 10 mM ammonium acetate in 95% water:5% acetonitrile, pH 6. A Thermopal autosampler was used to inject 10 µL of the sample for analysis. Nucleosides were analyzed on a TSQ Quantum Access triple quadrupole

electrospray ionization-LC/MS (Thermo, Inc.) using selected reaction monitoring fragmentation to the free nucleobase (inosine $[M+H]^+$ 269 m/z and dideoxyinosine $[M+H]^+$ 237 m/z transition to hypoxanthine $[M+H]^+$ 137 m/z) with 2-deoxyguanosine as the internal standard ($[M+H]^+$ 268 m/z to guanine $[M+H]^+$ 152 m/z). Nitrogen was used for both the auxiliary and sheath gases and was set to 45 units and 30 units, respectively. The following instrument parameters were used: source voltage 4.5kV; vaporizer temperature 0 °C; capillary temperature 270 °C; tube lens 101 V; skimmer offset -5 V; collision energy -10 V. Data acquisition and analysis were conducted with Thermo Xcalibur software, version 2.1.

All reactions were performed in 100 µL volumes in 96-well plates. Wild-type or variant PPM was activated at a concentration 10-fold higher than that used in the assay by incubation for 10 min at room temperature in 25 mM Tris-HCl and 0.1 mM $MnCl_2$ with either 5 µM (wild-type PPM) or 10 µM (variant PPM) glucose 1,6-bisphosphate then held at 4°C until assayed. Biochemical assays for PPM activity on ribose 5-phosphate contained 0.1 mM $MnCl_2$, 5 µM PNP, 600 µM hypoxanthine and 0 - 500 µM, 0 - 1000 µM or 0 - 4000 µM ribose 5-phosphate in 25 mM Tris-HCl, pH 8. Assays for PPM activity on dideoxyribose 5-phosphate contained 0.1 mM $MnCl_2$, 10 µM hPNP-46D6, 600 µM hypoxanthine and 0 - 5000 µM dideoxyribose 5-phosphate in 25 mM Tris-HCl, pH 8. PPM concentrations ranged from 0.02 - 1 µM for ribose 5-phosphate assays and 0.15 - 1 µM for dideoxyribose 5-phosphate assays. Reactions were initiated by addition of 10 µL of the sugar 5-phosphate substrate to 90 µL mix containing all other components and were incubated for 2 - 8 min at room temperature before being quenched by addition of 5 µL 2 M NaOH. After 30 min, 5 µL of 2 M HCl/1 M $CaCl_2$ was added to neutralize the mixture and the assay plate was centrifuged to pellet the precipitates. A 40 µL aliquot of each sample was combined with 10 µL of 50 µM 2-deoxyguanosine internal standard to prepare the sample for LC/MS analysis. Inosine and didanosine formation was quantified

by relative peak area of analyte to a 10 µM 2-deoxyguanosine internal standard in comparison to a standard curve made using authentic inosine (Acros Organics) and didanosine (3B Pharmachem (Wuhan) International Co. Ltd.). Retention times were approximately 5 min for inosine, 6.2 min for 2-deoxyguanosine and 14 min for didanosine. Kinetic constants were determined by fitting the data to the Michaelis-Menten equation using non-linear regression analysis in GraphPad Prism version 5.01.

Crystallization, data collection, and structure determination of wild-type and variant PPM

Purification and preparation of the PPM variants for crystallography was performed as previously described[22]. In short, crystals of wild-type PPM, the Ser154Ala variant, and the Ser154Gly variant were all grown using the hanging-drop vapor diffusion method by combining 1 µL of protein solution and 1 µL reservoir solution and incubating these over the reservoir solution at 18 °C while crystals of the Val158Leu variant grew after combining 2 µL protein solution and 2 µL reservoir solution. Crystals of wild-type PPM (12 mg/ml in 1 mM $MnCl_2$, 25 mM Tris-HCl, pH 7.4) grew over a reservoir solution containing 13% polyethylene glycol 3350, 50 mM $MnCl_2$, 75 mM NH_4CH_3COO, and 100 mM Bis-Tris, pH 5.5. Crystals of the Ser154Ala variant (12 mg/ml in 1 mM $MnCl_2$, 25 mM Tris-HCl, pH 7.4) grew over a reservoir solution containing 17% polyethylene glycol 3350, 50 mM $MnCl_2$, 50 mM NH_4CH_3COO, and 100 mM Bis-Tris, pH 5.5. Crystals of the S154G variant (12 mg/ml in 1 mM $MnCl_2$, 25 mM Tris-HCl, pH 7.4) grew over a reservoir solution containing 17% polyethylene glycol 3350, 50 mM $MnCl_2$, 50 mM NH_4CH_3COO, and 100 mM Bis-Tris, pH 5.5. Crystals of the V158L variant (12 mg/ml in 1 mM $MnCl_2$, 25 mM Tris-HCl, pH 7.4) grew over a reservoir solution containing 20% polyethylene glycol 3350, 50 mM $MnCl_2$, 50 mM NH_4CH_3COO, and 100 mM Bis-Tris, pH 5.45. Co-crystals of the Ser154Ala or Ser154Gly variants with ribose 5-phosphate were

prepared by soaking fully formed crystals in a solution that contained all components of the reservoir solution and 5 mM ribose 5-phosphate for 30 minutes. Co-crystals of the Ser154Ala or Ser154Gly variants with dideoxyribose 5-phosphate were prepared by soaking fully formed crystals in a solution that contained all components of the reservoir solution and 10 mM dideoxyribose 5-phosphate for 30 minutes. All crystals were cryoprotected with a solution that was 70% v/v reservoir solution and 30% glycerol before flash cooling by plunging in liquid nitrogen.

Data were collected at a temperature of 100 K using a wavelength of 0.979 Å and a MarMosaic225 CCD detector at the Advanced Photon Source (Argonne, IL) LS-CAT beamline 21-ID-G (Table 2-2). Data were processed and scaled using the HKL suite of programs[25]. Space group and unit cell information for all crystals is listed in Table 2-2. For wild-type PPM and the Ser154Ala and Ser154Gly variants, initial phases were determined by subjecting the structure of wild-type *B. cereus* PPM (PDB entry 3M8W[21]) to rigid body refinement using CNS[26]. For the Val158Leu variant, initial phases were obtained by molecular replacement with Molrep[27] of the CCP4 Suite[28], using PDB entry 3UN3[29]. The model was refined using iterative rounds of model building in COOT[30] and refinement in CNS[26] and Refmac5[28, 31] using Translation/Libration/Screw (TLS) refinement[32]. Figures 2-5, 2-6 and 2-7 were made using PyMOL[33].

Table 2-2. Data collection and refinement statistics for wild-type and variant PPM.

Protein	**Wild-type**	**S154A**	**S154A**	**S154A**	**S154G**	**S154G**	**S154G**	**V158L**
Ligand	ddR5P	None	R5P	ddR5P	none	R5P	ddR5P	none
PDB entry	4LRE	4LR7	4LR8	4LR9	4LRA	4LRF	4LRB	4LRC
Data collection								
Resolution (Å)	50-2.1	50-2.1	50-2.0	50-2.1	50-1.95	50-2.0	50-2.0	50-1.89
High resolution bin	2.18-2.1	2.18-2.1	2.01-2.0	2.18-2.1	1.98-1.95	2.07-2.0	2.07-2.0	1.92-1.89
Space group	P2_1	P2_1	P2_1	P2_1	P2_1	P2_1	P2_1	P2_1
Unit cell	a=91.5	a=91.3	a=91.2	a=92.3	a=91.1	a=90.9	a=91.2	a=91.7
	b=76.5	b=76.4	b=76.5	b=76.6	b=76.7	b=76.8	b=76.8	b=76.8
	c=107.3	c=108.1	c=107.0	c=107.2	c=106.8	c=107.5	c=106.5	c=107.2
	β=108.8	β =109.0	β =108.9	β =109.0	β =108.2	β =108.7	β =108.3	β =108.7
Total reflections	233,118	277,066	281,630	274,095	347,401	358,180	303,896	312,653
Unique reflections	79,568	81,695	90,848	81,866	98,816	94,933	93,620	102,942
[b]R_{sym} (%)	8.3 (28.3)[a]	9.4 (37.0)	8.0 (37.0)	8.7 (37.0)	7.5 (37.6)	8.7 (38.6)	6.7 (32.1)	5.9 (34.8)
I/σ	18.2 (3.2)	18.0 (3.3)	15.6 (2.2)	17.5 (3.1)	29.4 (4.7)	17.2 (3.2)	22.2 (4.3)	14.9 (3.4)
Completeness (%)	96.7 (88.7)	99.6 (99.4)	99.6 (85.6)	99.0 (95.6)	97.3 (97.3)	99.9 (99.9)	98.9 (98.4)	91.5 (92.7)
Refinement								
[c]R_{cryst} (%)	17.0	16.0	16.3	16.6	17.4	16.4	16.8	15.5
[d]R_{free} (%)	21.3	20.1	20.2	21.0	21.1	20.0	20.7	19.7

[a]Values in parentheses are for the highest resolution bin

[b]$R_{sym}=\sum |I_{obs}-I_{avg}|(100)/I_{avg}$,

[c]$R_{cryst}=[\sum||F_{obs}|-|F_{calc}|| (100)]/\sum|F_{obs}|$

[d]R_{free} is calculated using the same equation as R_{cryst} using a subset of reflections omitted from the refinement process.

Synthesis of 2,3-dideoxyribose 5-phosphate

(S)-γ-butyrolactone-γ-carboxylic acid (**2**)**.** Compound **2** was generated from L-glutamic acid **1** as generally described[34] with the following modifications. Briefly, a stirred solution of 53.3 g (363.6 mmol) L-glutamic acid suspended in 170 mL water was fixed with separatory funnels containing solutions of $NaNO_2$ (37.5 g, 545 mmol, 1.5 eq, in 100 mL water) and HCl (90.6 mL of 5.6 N, 508.6 mmol, 1.4 eq), which were added drop-wise simultaneously over a period 3 - 4 h. Reaction temperature was maintained between 15 - 20 °C during addition using an ice water bath. After complete addition, the reaction was warmed to room temperature and stirred overnight before removal of water by rotary evaporation and azeotropic dehydration using toluene (3 x 75 mL). The resulting solid was resuspended in 500 mL ethyl acetate and dried over anhydrous $NaSO_4$. After removing precipitates via filtration, AG50W-X4 resin (20 g, activated by washing successively with methanol and ethyl acetate) was added to the solution and stirred for 30 min to remove unreacted **1**. After filtration, solvent was removed by evaporation and the product recrystallized from dichloromethane at -20 °C (18.2 g, 139.9 mmol 39% yield). Product purity was confirmed by ^{1}H, ^{13}C NMR analysis and comparison to reported spectra.

(S)-γ-(hydroxymethyl)-γ-butyrolactone (**3**)**.** Butyrolactone **3** was produced via a procedure adapted from a previously described synthesis[34]. To a solution of 4.95 g **2** (38.1 mmol) in 25 mL anhydrous THF at 0 °C under Argon was added dropwise 4.38 mL BH_3SMe_2, 10 M in THF, (43.75 mmol, 1.15 eq) over 1 h. The solution was warmed to room temperature and stirred for an additional 2 h. Three dropwise additions of 2.25 mL methanol were performed to quench the reaction and remove unreacted BH_3SMe_2 prior to removal of solvents by rotary evaporation. Product **3** (3.8 g, 32.7 mmol 86 % yield) was sufficiently pure to proceed without further purification. Product purity was confirmed by ^{1}H, ^{13}C NMR analysis and comparison to reported spectra.

(S)-γ-(dibenzylphosphomethyl)-γ-butyrolactone (**4**)**.** Butyrolactone **3** (755 mg, 6.5 mmol) dissolved in 6 mL of anhydrous dichloromethane and 1.5 mL anhydrous acetonitrile at 0 °C was added to 38.5 mL tetrazole (3% wt/wt in acetonitrile, 13.0 mmol, 2 eq) under argon with stirring. A solution of 3.33 mL dibenzyl *N,N*-diisopropylphosphoramidite (9.1 mM, 1.4 eq) dissolved in 18.5 mL anhydrous dichloromethane was added dropwise and stirred for 2 h at 0 °C followed by 1.07 mL 7.3 M *tert*-butyl hydroperoxide in water (7.8 mmol 1.2 eq) and the reaction was stirred for an additional hour. The reaction was extracted with 20 mL half-saturated $NaHCO_3$, twice with 20 mL dichloromethane, dried over anhydrous $MgSO_4$ and concentrated by rotary evaporation. The crude products were purified by SiO_2 flash chromatography (2.5% methanol in dichloromethane, then 5% hexanes in ethyl acetate) to yield 1.49 g (4.0 mmol, 61%) compound **4**. TLC (methanol:dichloromethane, 2.5/97.5 v/v): R_f=0.13. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.35 (s, 10H), 5.1 - 4.99 (m, 4H), 4.64 - 4.55 (m, 1H), 4.17 - 4.1 (m, 1H), 4.0 - 3.95 (m, 1H), 2.5 - 4.4 (m, 2H), 2.3 - 2.15 (m, 1H), 2.04 - 1.93 (m, 1H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 176.7, 135.86 (d, $^3J_{cp}$ = 2.32 Hz), 135.8 (d, $^3J_{cp}$ = 2.47 Hz), 128.98 (d, J_{cp} = 5.36 Hz), 128.39 (d, J_{cp} = 3.75 Hz), 77.80 (d, $^3J_{cp}$ = 8.06 Hz), 69.95 (d, $^2J_{cp}$ = 5.63 Hz), 68.29 (d, $^2J_{cp}$ = 5.48 Hz), 28.3, 23.6. ^{31}P {^{1}H} NMR (200 MHz, $CDCl_3$): δ 0.26 (s). HRMS (*m/z*): $[M+H]^+$ calculated for $C_{19}H_{22}O_6P^+$, 377.1149; found 377.1158.

(S)-2,3-dideoxyribose 5-(di-O-benzyl)phosphate (**5**)**.** To a stirred solution of compound **4** (539 mg, 1.43 mmol) in 54 mL anhydrous dichloromethane under Argon at -78 °C was added dropwise diisobutylaluminium hydride (1.5 M in toluene, 3.8 mL, 5.7 mM, 4 eq) and the reaction was stirred for 2 h. Methanol (17.5 mL) was added to quench the reaction before warming to room temperature. Saturated potassium sodium tartrate (74 mL) was added and the mixture was stirred overnight, filtered through a fritted funnel and the resulting solid dissolved in water. The solution was extracted with

dichloromethane (3 x 25 mL), dried over $MgSO_4$ and concentrated by rotary evaporation. Compound **5** was purified by flash SiO_2 chromatography (100% ethyl acetate) to yield 350 mg (0.93 mmol, 65% yield). TLC (ethylacetate): R_f=0.38. ^{1}H NMR (400 MHz, $CDCl_3$, mixture of α and β anomers): δ 7.34 (s, 10H), 5.53 - 5.48 (m, 1H), 5.12 - 4.99 (m, 4H), 4.39-4.18 (m, 1H), 4.13-4.02 (m, 1H) 4.0-3.88 (m, 1H), 2.14 - 1.58 (m, 4H). ^{13}C NMR (100 MHz, $CDCl_3$, mixture of α and β anomers): δ 135.73 (m), 128.48 (d, J_{cp} = 3.7 Hz), 127.90 (d, J_{cp} = 3.9 Hz), 98.90, 98.86, 78.03 (d, $^3J_{cp}$ = 6.1 Hz), 76.20 (d, $^3J_{cp}$ = 7.6 Hz), 70.97 (d, $^3J_{cp}$ = 5.8 Hz), 69.43 (d, $^3J_{cp}$ = 5.5 Hz), 69.27 (m), 69.12 (d, J_{cp} = 6.0 Hz), 33.60, 32.58, 25.15, 24.88. ^{31}P {^{1}H} NMR (200 MHz, $CDCl_3$, mixture of α and β anomers): δ 0.63, 0.40. HRMS (*m/z*): $[M+H]^+$ calculated for $C_{19}H_{24}O_6P^+$, 379.1305; found 379.1305.

(S)-2,3-dideoxyribse 5-phosphate, disodium salt (**6**)**.** To a stirred solution of compound **5** (334 mg, 0.88 mmol) in 10 mL ethanol under a hydrogen balloon was added 10% palladium on activated carbon (38.3 mg). The reaction was stirred overnight. The resulting reaction was filtered through celite, concentrated by rotary evaporation, and then dissolved in 2 mL water. A solution of $NaHCO_3$ (148.3 mg, 1.8 mmol) in 1 mL water was added to neutralize the solution, giving 150 mg (62.0 mmol, 70% yield) compound **6,** which was used without further purification. When used in biochemical reactions, concentration of **6** was determined using ^{1}H NMR and comparison of integrals to an internal standard (2.5 mM of dimethylformamide) with using a relaxation delay of 15 sec. ^{1}H NMR (400 MHz, D_2O, mixture of α and β anomers): δ 5.58 - 5.48 (m, 1H, CH), 4.47 - 4.22 (m, 1H), 4.02 - 3.76 (m, 2H), 2.20 - 1.76 (m, 4H). ^{13}C NMR (100 MHz, D_2O, mixture of α and β anomers): δ 98.56, 98.30 (CH), 79.27 ($^3J_{cp}$ = 7.99 Hz, CH), 77.60 ($^3J_{cp}$ = 7.97 Hz, CH), 68.29 ($^2J_{cp}$ = 5.33 Hz, CH_2), 67.07 ($^2J_{cp}$ = 5.25 Hz, CH_2), 32.86, 32.28, 24.61, 24.58 (CH2-CH2). ^{31}P {^{1}H} NMR (200 MHz, D_2O, mixture of α and β anomers): δ

1.79, 1.65, 1.58. HRMS (*m/z*): $[M-H]^-$ calculated for $C_5H_{10}O_6P^-$, 197.0220; found 197.0209.

Results

Chemical synthesis of non-natural substrate 2,3-dideoxyribose 5-phosphate

2,3-Dideoxyribose 5-phosphate was synthesized from L-glutamic acid over several steps. Although 2,3-dideoxyribose has also been prepared from D-mannitol[35] and D-ribonolactone[36], the lower cost of the glutamic acid starting material and shorter synthesis sequence makes this the preferred route for synthesis of dideoxysugar substrates[34]. Nitrous acid created *in situ* from $NaNO_2$ and HCl was used to catalyze the diazotization of the amino group of glutamic acid (**1**) and subsequent cyclization to afford the (*S*)-γ-butyrolactone-γ-carboxylic acid (**2**) in 39% yield with complete retention of stereochemistry[34] (Figure 2-3). Compound **2** was reduced to (*S*)-γ-hydroxymethyl-γ-butyrolactone (**3**) using borane dimethyl sulfide complex in anhydrous THF[34], giving 86% yield of sufficient purity to proceed without further isolation by column chromatography.

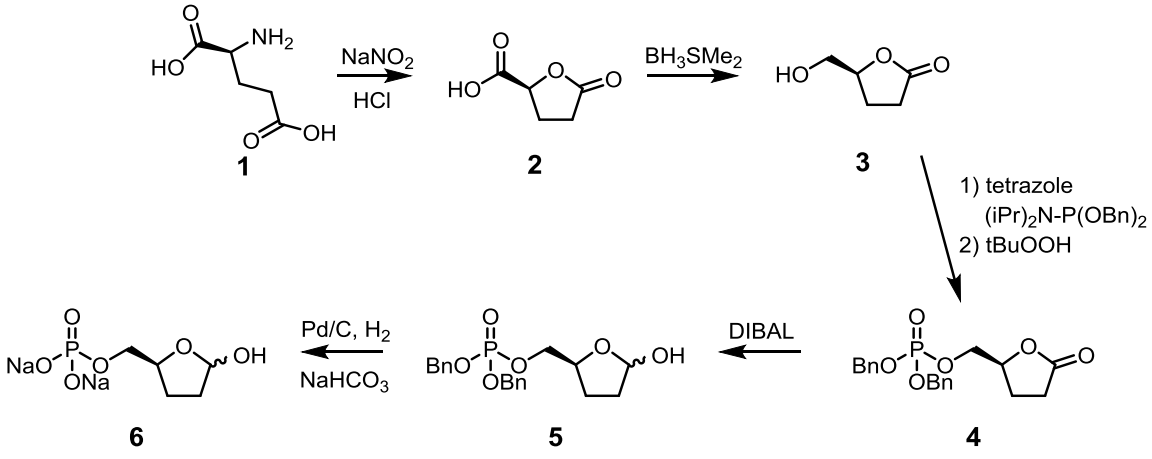

Figure 2-3. Synthesis of 2,3-dideoxyribose 5-phosphate.

The C5 hydroxyl of (*S*)-γ-hydroxymethyl-γ-butyrolactone (**3**) was phosphitylated and subsequently oxidized[37] to form the dibenzyl-protected phospholactone (**4**) with 61% yield over two steps. DIBAL reduction of the lactone lead to the corresponding dibenzyl-protected lactol (**5**) in 65% yield. Finally, hydrogenation[38] to remove the benzyl protecting groups followed by treatment with sodium bicarbonate to produce the desired 2,3-dideoxyribose 5-phosphate sodium salt (**6**) in 70% yield. This five step route from glutamic acid proved to be an effective method of producing dideoxyribose 5-phosphate in yields suitable to enable screening of the PPM mutagenesis libraries.

Selection of Bacillus cereus PPM progenitor enzyme

PPM activity on dideoxyribose 5-phosphate has been characterized for homologs from only two different species. The *E. coli* variant reportedly showed no activity toward this non-natural substrate when coupled with purine nucleoside phosphorylase and adenosine deaminase[9]. The enzyme from *Bacillus stearothermophilus*, however, reportedly demonstrated approximately 12% of the activity on the natural substrate, ribose 5-phosphate, when tested at expected saturating conditions in tandem with purine nucleoside phosphorylase II[10]. Although PPM from *Bacillus stearothermophilus* was not publicly available, a closely related enzyme from *Bacillus cereus* with 82% sequence identity and 92% sequence homology was identified. The gene was cloned from genomic DNA[22], expressed and tested for the desired activity on dideoxyribose 5-phosphate in tandem assays with hypoxanthine and the evolved hPNP-46D6 variant[6]. After confirming activity on dideoxyribose 5-phosphate (Figure 2-4), PPM from *B. cereus* became the progenitor enzyme for engineering. With an appropriately active template enzyme in hand, we next desired to evolve the enzyme for increased dideoxyribose 5-phosphate turnover.

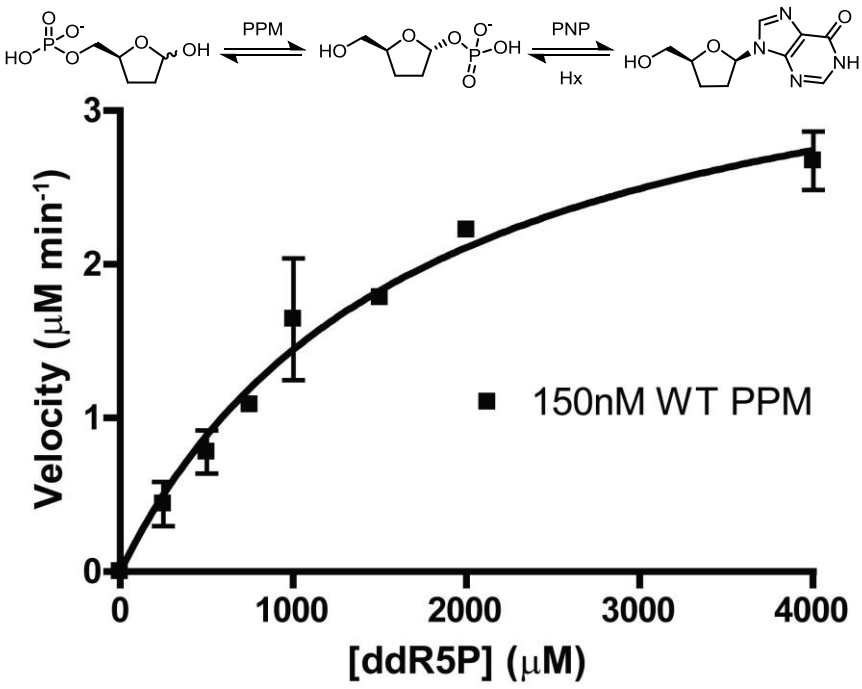

Figure 2-4. Michaelis-Menten plot of wild-type *B. cereus* PPM kinetics for dideoxyribose 5-phosphate. Initial velocities of turnover are plotted against substrate concentration. Data are the average of duplicate assays. The calculated K_M, V_{max} and k_{cat} are 1700 ± 300 μM, 3.9 ± 0.3 μM min^1 and 0.43 ± 0.04 s^{-1}.

Saturation mutagenesis of Ser154

Site directed saturation mutagenesis relies on the availability of structural data to identify residues of interest based on interactions with either natural or non-natural substrates. For this aspect of PPM evolution our aim was to target substrate interacting residues for mutagenesis in order to reduce selectivity for the natural substrate, however, no published crystal structures were readily available for PPM from *B. cereus* to aid in this process. We therefore determined costructures of PPM with the natural substrate ribose 5-phosphate at 1.8 Å resolution (PDB entry 3M8Z[21]) and the target non-natural substrate dideoxyribose 5-phosphate at 2.1 Å resolution (PDB entry 4LRE).

The cocrystal structures of each substrate in wild-type PPM shows similar binding positions of the phosphate moiety, however the ligand-associated electron

density of the sugar ring greatly differs between the natural and non-natural substrates (Figure 2-5a,b, Figure 2-6). The structure with ribose 5-phosphate shows well defined density for the substrate, indicating an organized and consistent binding orientation in the crystal. Conversely, only partial electron density for dideoxyribose 5-phosphate is observed for the unnatural ligand, suggesting that the dideoxysugar ring of the bound substrate is less structured, likely due to rotational disorder, and was therefore not optimally aligned with key catalytic residues. The Ser154 side chain Oγ that normally forms a hydrogen-bonding interaction with the sugar C3 hydroxyl of ribose 5-phosphate[21] does not similarly contribute to the orientation of dideoxyribose 5-phosphate (Figure 2-5a,b, Figure 2-6). Because dideoxyribose 5-phosphate lacks the C2 and C3 hydroxyl groups, the interaction of this substrate with Ser154 may be repulsive due to an unfavorable environment for the non-polar substrate created by the polar side chain. This unfavorable interaction may be contributing to the disorder of the dideoxysugar ring in binding of the non-natural substrate, decreasing the quality of electron density for dideoxyribose 5-phosphate observed in the structures.

Accordingly, the Ser154Ala mutant was created to decrease the active site polarity to potentially stabilize the more hydrophobic dideoxyribose ring and improve catalysis by removing the adverse interaction of the active site residue. Subsequent cocrystal structures of this variant indicated subjectively similar electron density for dideoxyribose 5-phosphate in both wild-type and Ser154Ala PPM (PDB entry 4LR9, Figure 2-5d). Additionally, no change was observed in the conformation of ribose 5-phosphate binding (PDB entry 4LR8, Figure 2-5c), but a loss of a hydrogen-bond between the Ser154Ala mutant and the C3 hydroxyl of ribose 5-phosphate indicates fewer substrate interactions specific for ribose 5-phosphate.

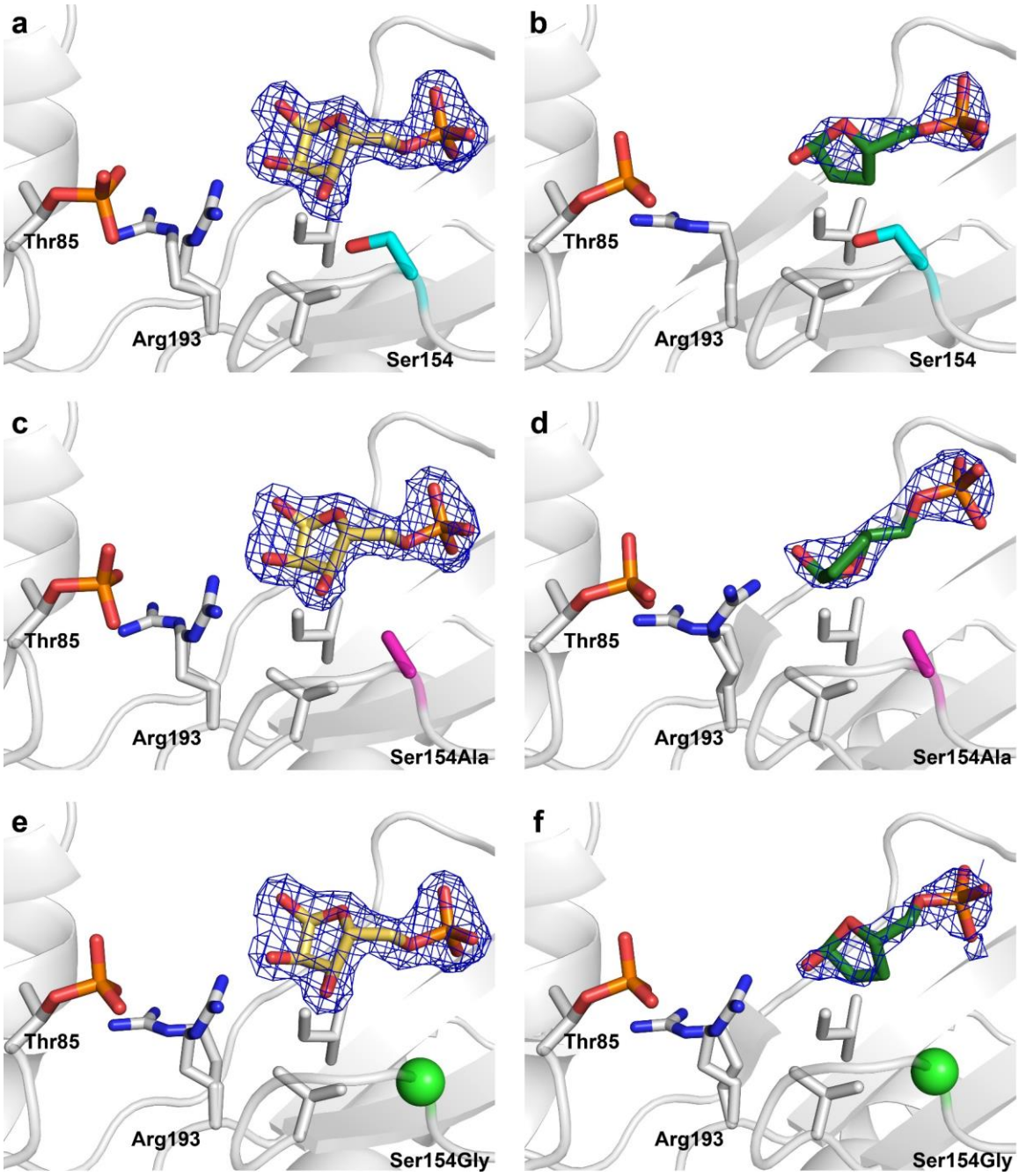

Figure 2-5. Substrate binding in PPM variants. R5P, left column, is shown in gold and ddR5P, right column, is shown in green. Costructure of wild-type PPM with (**a**) R5P (PDB entry 3M8Z[21]) and (**b**) ddR5P with S154 shown in cyan. Costructure of Ser154Ala PPM with (**c**) R5P and (**d**) ddR5P with Ser154Ala shown in pink. Costructure of Ser154Gly PPM with (**e**) R5P and (**f**) ddR5P with Ser154Gly Cα atom shown as a sphere in bright green.

Since Ser154 appeared to beneficially interact with the natural substrate, but made unfavorable contacts to the non-natural substrate, a saturation mutagenesis library was created targeting this position. The library was screened for activity on ribose 5-phosphate as well as dideoxyribose 5-phosphate to compare activity changes of each clone. Primary screen hits were validated in duplicate in a secondary assay, and sequencing of top hits identified Ser154Gly and Ser154Ala variants as having the greatest changes in desired activity. Kinetic characterization of these two top mutants, showed improved substrate selectivity in each enzyme, 49-fold in Ser154Gly and 70-fold in Ser154Ala, as determined by comparison of the ratio of catalytic efficiencies for both substrates between each PPM variant and the wild-type enzyme (Table 2-3). The glycine and alanine mutations provided 15-fold and 6-fold reduction in k_{cat} of ribose 5-phosphate, respectively, while only reducing dideoxyribose 5-phosphate turnover by 2-fold at most. Additionally, both enzymes showed improvements in substrate binding, where ribose 5-phosphate K_M value increased 3-fold in each variant and the K_M value for dideoxyribose 5-phosphate decreased 2-4-fold after removing the polar Ser154 side chain (Table 2-3).

The cocrystal structure of the Ser154Gly variant with dideoxyribose 5-phosphate was also associated with low electron density and appeared to still have significant rotational disorder of the sugar ring (PDB entry 4LRB). However, this rotational disorder was subjectively not as pronounced as it was for wild-type enzyme, with a percentage of the dideoxyribose 5-phosphate possibly stabilized in an orientation more similar to that seen in wild-type PPM. Similar to the Ser154Ala crystals, no change in the orientation of ribose 5-phosphate binding was observed in the Ser154Gly PPM variant (PDB entry 4LRF, Figure 2-5e,f). Although the Ser154Ala variant was slightly superior to the Ser154Gly variant in terms of selectivity and desired catalysis, the latter variant was chosen as the template for further evolution since the increased active site area could

Table 2-3. Kinetic Parameters of PPM variants for R5P and ddR5P substrates and comparison of changes in substrate selectivity.

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1}) ($x10^3$)	[k_{cat} / K_M (R5P)] / [k_{cat} / K_M (ddR5P)]	Fold Change
Wild-Type	R5P	10.4±0.2	40±3	260±20		
	ddR5P	0.43±0.04	1700±300	0.25±0.05	1028	1
S154A	R5P	1.7±0.1	130±20	13.1±2.2		
	ddR5P	0.39±0.02	440±60	0.89±0.13	14.8	70
S154G	R5P	0.71±0.02	130±10	5.5±0.5		
	ddR5P	0.21±0.02	800±200	0.26±0.07	20.8	49
V158L	R5P	0.42±0.02	1020±100	0.41±0.04		
	ddR5P	0.24±0.01	680±70	0.35±0.04	1.2	881
S154G/V158L	R5P	0.035±0.001	190±20	0.18±0.02		
	ddR5P	0.028±0.002	480±90	0.058±0.012	3.2	326

possibly accommodate later mutations that incorporated large side chains that may remodel the binding pocket and benefit dideoxyribose 5-phosphate catalysis.

Saturation mutagenesis of Val158 and Ile195

Evidence from the wild-type structures also suggested Val158 and Ile195 as potential targets for saturation mutagenesis. Each of these residues are within 4.5 Å of the bound substrates, so mutations at either position could possibly affect substrate binding through changes in steric interactions or polarity/hydrophobicity within the active site (Figure 2-6). Saturation mutagenesis libraries were created individually at each position on the wild-type and Ser154Gly templates and also tested against both substrates to compare activity changes of each clone. Both screens identified a Val158Leu mutation as the most beneficial on each template, creating the Val158Leu single mutant and the Ser154Gly/Val158Leu double mutant. No favorable mutation was observed at the Ile195 position on either template. Kinetic analysis showed that the Val158Leu mutation on the wild-type template resulted in a 25-fold increase in K_M as well as a 25-fold decrease in k_{cat} for ribose 5-phosphate, but a 2.5-fold decrease in K_M and only a slight decrease in k_{cat} for dideoxyribose 5-phosphate (Table 2-3). In total, this

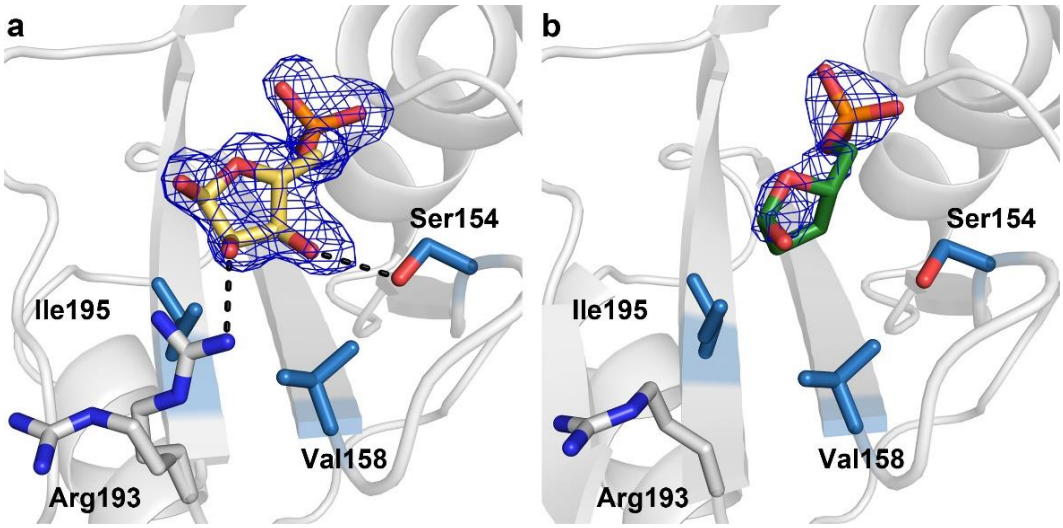

Figure 2-6. Additional first shell residues targeted for saturation mutagenesis. Distances between (**a**) R5P (PDB entry 3M8Z[21]) and (**b**) ddR5P and nearby residues Ser154, Val158 and Ile195 are highlighted in the wild-type PPM structures. Due to the close proximity of each residue to the bound substrates, both Val158 and Ile195 positions were also targeted individually for saturation mutagenesis on wild-type and Ser154Gly templates.

single mutation reduced ribose 5-phosphate catalytic efficiency 881-fold to only 1.2-fold greater than that of dideoxyribose 5-phosphate (Table 2-3). The Ser154Gly/Val158Leu double mutant also showed substantially improved dideoxyribose 5-phosphate substrate selectivity (>300-fold), however this benefit came at a severe loss in the rate of catalysis of both substrates (Table 2-3).

Superposition of the structures of wild-type (PDB entry 3M8Z[21]) and Val158Leu PPM (PDB entry 4LRC) identified that the active site mutation created steric hindrance between the Leu158 side chain and both substrate, ribose 5-phosphate at 2.5 Å, and Arg193, a substrate-orienting residue for ribose 5-phosphate[21], at 2.3 Å (Figure 2-7). These new putative unfavorable interactions prevent ribose 5-phosphate from binding in an orientation optimal for catalysis and would explain the dramatic increase in ribose 5-phosphate K_M observed in the Val158Leu variant (Table 2-3). Since Arg193 does not form a hydrogen-bonding interaction with dideoxyribose 5-phosphate (Figure 2-6b), the

Val158Leu mutation primarily affects binding of the natural substrate. This steric hindrace also explains why cocrystallization with both natural and non-natural substrate did not result in the appearance of interpretable electron density in 50 collected datasets.

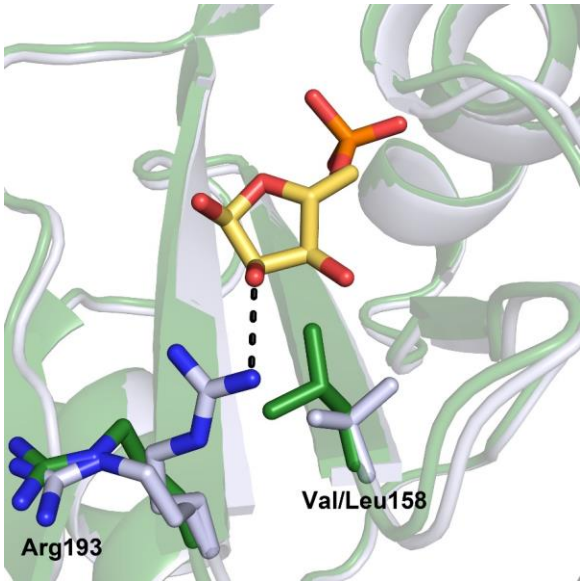

Figure 2-7. Overlay of the Val158Leu structure and wild-type PPM. The Val158Leu structure (green) is superimposed with the costructure of wild-type PPM (gray) with ribose 5-phosphate bound (gold, PDB entry 3M8Z[21]). A favorable hydrogen bond in the wild-type structure between Arg193 and ribose 5-phosphate induced upon substrate binding is indicated.

Discussion

Here we present the identification of phosphopentomutase from *Bacillus cereus* for use in the biotransformation of dideoxyribose 5-phosphate to dideoxyinosine in tandem experiments with an evolved purine nucleoside phosphorylase[6]. This enzyme may serve as a key step in the biosynthesis of dideoxyinosine from non-natural starting materials, however, in order to be useful in such a pathway, activity on and selectivity for dideoxyribose 5-phosphate must be increased to improve pathway productivity. Enzyme engineering can address this problem and to aid in the process of targeted rational mutagenesis, high resolution cocrystal structures of wild-type and variant PPM with the

natural and non-natural substrate have been determined to observe enzyme-substrate binding interactions[21].

PPM is a member of the alkaline phosphatase superfamily of enzymes and folds into two distinct domains. The core domain (residues 2-99 and 219-393) is organized into an alkaline phosphatase like fold, while the cap domain (residues 102-216) comprises a fold that is unique to prokaryotic PPMs, with the active site of the enzyme in a cleft created between the domains[21]. The cap domain primarily confers the particular activity of the enzyme and therefore largely determines the substrate selectivity[17]. Since the most dramatic effect in altering substrate activity can be gained by active site mutations[39], residues projecting into the active site from the cap domain of PPM were thought to provide a better starting point for engineering new substrate selectivity and were therefore targeted for mutational analysis.

We began PPM evolution with saturation mutagenesis guided by a series of crystal structures with natural and non-natural substrates bound. Costructures of the Ser154Ala active site mutant were determined and confirmed this position to interact with the natural substrate ribose 5-phosphate and also influence binding of the non-natural dideoxyribose 5-phosphate substrate. Additional structure evaluation and activity studies of enzyme variants identified Val158 as another modulator of substrate selectivity in *Bacillus cereus* PPM. The top hits determined from the saturation mutagenesis screens showed modest (49-fold in Ser154Gly PPM) to extreme (881-fold in Val158Leu PPM) changes in substrate preference compared to wild-type PPM (Table 2-3). The Ser154Gly mutation offered an enlarged active site that held the potential to be refit by later mutations to benefit dideoxyribose 5-phosphate binding. In a separate series, the rather conservative change in the Val158Leu single mutant triggered a drastic 25-fold change in both binding and turnover of ribose 5-phosphate, completely eliminating substrate selectivity to create a generalist enzyme (Table 2-3). The overlaid

comparison of wild-type PPM and the Val158Leu mutant, as well as the lack of success at cocrystallization with all substrates, suggests that the side chain extension by an additional methylene group is enough to sterically hinder ribose 5-phosphate binding, and could possibly also prevent proper movement of the substrate during catalysis. When the effects of Ser154Gly and Val158Leu were combined in the double mutant, the expanded binding pocket and removal of a strict hydrogen bond interaction provided by the Ser154Gly mutation appears to lessen the effect of the proposed steric hindrance of ribose 5-phosphate with the Val158Leu mutation. This can be observed by the difference in ribose 5-phosphate K_M of the Val158Leu variant compared to the similar K_M values between the Ser154Gly single mutant and the Ser154Gly/Val158Leu double mutant (Table 2-3). The total loss in activity observed in the double mutant indicates that each mutation can impart large selectivity changes on an individual basis, but cannot be combined for cooperative improvement.

Having two variants with mutually exclusive effects on substrate selectivity provided an opportunity to follow the evolution of the same enzyme with two different active site features: One variant with a larger active site and the other enzyme showing no preference between the two substrates. Accordingly, the Ser154Gly PPM and Val158Leu PPM variants were both selected to be progenitors for directed evolution by random mutagenesis. Continuing with both templates provides two enzymes with interesting and different characteristics to be improved upon through further mutagenesis, giving a higher chance of success to arrive at a suitable biocatalyst to use in the biosynthetic pathway for production of dideoxyinosine.

Conclusions

Saturation mutagenesis of the active site of *Bacillus cereus* PPM was successfully used to identify two residues responsible for regulating substrate selectivity

of the enzyme. Mutagenesis of these residues was exploited to engineer substantial changes in selectivity between the natural substrate, ribose 5-phosphate, and the target non-natural substrate, dideoxyribose 5-phosphate, resulting in a generalist enzyme showing no preference between the two compounds. Although not tested within the scope of this work, further optimization by mutagenesis at either positions 154 or 158 may be able to expand the utility of this enzyme to catalyze a similar isomerization reaction on other phosphorylated sugars or sugar analogs. Expanding the substrate allowance of this enzyme through further mutagenesis to accept sugar analogs with methyl-, fluoro- or azido-substitutions at the C2 or C3 positions, as well as activity on multiple heterocyclic furanose rings, may be of high value in biosynthetic applications toward multiple other nucleoside analogs containing these functionalities, many of which are pharmaceutical ingredients as well[(3)].

Acknowledgements

This work was contributed to by William R. Birmingham, Chrystal A. Starbird, Timothy D. Panosian, David P. Nannemann, Tina M. Iverson and Brian O. Bachmann. Bioretrosynthetic evolution of a didanosine biosynthetic pathway. W.R.B. and D.P.N. designed assays. W.R.B performed assays, screened all mutagenesis libraries, performed the kinetic characterization, expressed, purified and tested enzymes in the biosynthetic pathway studies. C.A.S. determined the crystal structure of the Val158Leu PPM variant. T.D.P. determined crystal structures of wild-type PPM and the Ser154Ala and Ser154Gly variants. D.P.N. established initial synthesis routes of dideoxyribose and dideoxyribose 5-phosphate. B.O.B supervised biochemical experiments and T.M.I. supervised X-ray crystallographic work. We thank V. Phelan for assistance with establishing the mass spectrometry based analysis methods, K. McCulloch for help in

crystallography data collection and to C. Goodwin for high resolution mass spectrometry data collection.

References

1. De Clercq, E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents* **2009**, 33, 307-320.

2. Egeblad, L., Welin, M., Flodin, S., Graslund, S., Wang, L., Balzarini, J., Eriksson, S., and Nordlund, P. Pan-pathway based interaction profiling of FDA-approved nucleoside and nucleobase analogs with enzymes of the human nucleotide metabolism. *PloS one* **2012**, 7, e37724.

3. Parker, W. B. Enzymology of purine and pyrimidine antimetabolites used in the treatment of cancer. *Chem. Rev.* **2009**, 109, 2880-2893.

4. Pinheiro, E., Vasan, A., Kim, J. Y., Lee, E., Guimier, J. M., and Perriens, J. Examining the production costs of antiretroviral drugs. *AIDS* **2006**, 20, 1745-1752.

5. Bachmann, B. O. Biosynthesis: Is it time to go retro? *Nat. Chem. Biol.* **2010**, 6, 390-393.

6. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

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

8. Tozzi, M. G., Camici, M., Mascia, L., Sgarrella, F., and Ipata, P. L. Pentose phosphates in nucleoside interconversion and catabolism. *FEBS J.* **2006**, 273, 1089-1101.

9. Barbas, C. F., and Wong, C. H. Overexpression and substrate-specificity studies of phosphodeoxyribomutase and thymidine phosphorylase. *Bioorg. Chem.* **1991**, 19, 261-269.

10. Hamamoto, T., Noguchi, T., and Midorikawa, Y. Phosphopentomutase of *Bacillus stearothermophilus* TH6-2: The enzyme and its gene ppm. *Biosci. Biotechnol. Biochem.* **1998**, 62, 1103-1108.

11. Horinouchi, N., Kawano, T., Sakai, T., Matsumoto, S., Sasaki, M., Mikami, Y., Ogawa, J., and Shimizu, S. Screening and characterization of a phosphopentomutase useful for enzymatic production of 2'-deoxyribonucleoside. *Nat. Biotechnol.* **2009**, 26, 75-82.

12. Horinouchi, N., Ogawa, J., Kawano, T., Sakai, T., Saito, K., Matsumoto, S., Sasaki, M., Mikami, Y., and Shimizu, S. One-pot microbial synthesis of 2 '-deoxyribonucleoside from glucose, acetaldehyde, and a nucleobase. *Biotechnol. Lett.* **2006**, 28, 877-881.

13. Horinouchi, N., Sakai, T., Kawano, T., Matsumoto, S., Sasaki, M., Hibi, M., Shima, J., Shimizu, S., and Ogawa, J. Construction of microbial platform for an energy-requiring bioprocess: Practical 2 '-deoxyribonucleoside production involving a C-C coupling reaction with high energy substrates. *Microb. Cell Fact.* **2012**, 11.

14. Taverna-Porro, M., Bouvier, L. A., Pereira, C. A., Montserrat, J. M., and Iribarren, A. M. Chemoenzymatic preparation of nucleosides from furanoses. *Tetrahedron Lett.* **2008**, 49, 2642-2645.

15. Galperin, M. Y., Bairoch, A., and Koonin, E. V. A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Sci.* **1998**, 7, 1829-1835.

16. Gijsbers, R., Ceulemans, H., Stalmans, W., and Bollen, M. Structural and catalytic similarities between nucleotide pyrophosphatases/phosphodiesterases and alkaline phosphatases. *J. Biol. Chem.* **2001**, 276, 1361-1368.

17. Zalatan, J. G., Fenn, T. D., Brunger, A. T., and Herschlag, D. Structural and functional comparisons of nucleotide pyrophosphatase/phosphodiesterase and alkaline phosphatase: Implications for mechanism and evolution. *Biochemistry* **2006**, 45, 9788-9803.

18. Hammer-Jespersen, K., and Munch-Petersen, A. Phosphodeoxyribomutase from *Escherichia coli*. *Eur. J. Biochem.* **1970**, 17, 397-407.

19. Breathnach, R., and Knowles, J. R. Phosphoglycerate mutase from wheat germ: studies with oxygen-18 labeled substrate, investigations of the phosphatase and phosphoryl transfer activities, and evidence for a phosphoryl-enzyme intermediate. *Biochemistry* **1977**, 16, 3054-3060.

20. Gatehouse, J. A., and Knowles, J. R. Phosphoglycerate mutase from wheat germ: Studies with isotopically labeled 3-phospho-D-glycerates showing that the catalyzed reaction is intramolecular. *Biochemistry* **1977**, 16, 3045-3050.

21. Panosian, T. D., Nannemann, D. P., Watkins, G. R., Phelan, V. V., McDonald, W. H., Wadzinski, B. E., Bachmann, B. O., and Iverson, T. M. *Bacillus cereus* phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle. *J. Biol. Chem.* **2011**, 286, 8043-8054.

22. Panosian, T. D., Nannemann, D. P., Bachmann, B. O., and Iverson, T. M. Crystallization and preliminary X-ray analysis of a phosphopentomutase from *Bacillus cereus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, 66, 811-814.

23. De Groot, H., De Groot, H., and Noll, T. Enzymic determination of inorganic phosphates, organic phosphates and phosphate-liberating enzymes by use of nucleoside phosphorylase-xanthine oxidase (dehydrogenase)-coupled reactions. *Biochem. J.* **1985**, 230, 255-260.

24. Ball, E. G. Xanthine oxidase: Purification and properties. *J. Biol. Chem.* **1939**, 128, 51-67.

25. Otwinowski, Z., and Minor, W. Processing of X-ray diffraction data collected in oscillation mode, In *Methods Enzymol.* **1997**, 276, pp 307-326.

26. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2007**, 2, 2728-2733.

27. Vagin, A., and Teplyakov, A. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, 30, 1022-1025.

28. Bailey, S. The CCP4 Suite - Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, 50, 760-763.

29. Iverson, T. M., Panosian, T. D., Birmingham, W. R., Nannemann, D. P., and Bachmann, B. O. Molecular differences between a mutase and a phosphatase: investigations of the activation step in *Bacillus cereus* phosphopentomutase. *Biochemistry* **2012**, 51, 1964-1975.

30. Emsley, P., and Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, 60, 2126-2132.

31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, 53, 240-255.

32. Winn, M. D., Isupov, M. N., and Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2001**, 57, 122-133.

33. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, USA.

34. Okabe, M., Sun, R. C., Tam, S. Y. K., Todaro, L. J., and Coffen, D. L. Synthesis of the dideoxynucleosides ddC and CNT from glutamic acid, ribonolactone, and pyrimidine bases. *J. Org. Chem.* **1988**, 53, 4780-4786.

35. Takano, S., Goto, E., Hirama, M., and Ogasawara, K. An alternative synthesis of (S)-(+)-γ-hydroxymethyl-γ-butyrolactone from (D)-(+)-mannitol. *Heterocycles* **1981**, 16, 951-954.

36. Camps, P., Cardellach, J., Font, J., Ortuno, R. M., and Ponsati, O. Studies on structurally simple alpha,beta-butenolides-II: (-)-(S)-γ-hydroxymethyl-α,β-butenolide and derivatives from D-ribonolactone efficient synthesis of (-)-ranunculin. *Tetrahedron* **1982**, 38, 2395-2402.

37. Graham, S. M., and Pope, S. C. Selective phosphitylation of the primary hydroxyl group in unprotected carbohydrates and nucleosides. *Org. Lett.* **1999**, 1, 733-736.

38. Humphries, M. J., and Ramsden, C. A. A fresh AIR synthesis. *Synthesis* **1999**, 985-992.

39. Morley, K. L., and Kazlauskas, R. J. Improving enzyme properties: When are closer mutations better? *Trends Biotechnol.* **2005**, 23, 231-237.

Chapter III

DIRECTED EVOLUTION OF PHOSPHOPENTOMUTASE BY WHOLE GENE RANDOM MUTAGENESIS

Introduction

Directed evolution has become a prevailing tool within the array of methods in protein engineering. Fundamentally, it is a direct laboratory application of Darwinian evolution, harnessing the process of natural selection to gain or increase experimenter-defined traits in target enzymes[(1)]. The genetically diverse libraries screened in directed evolution experiments are created through a process of random mutagenesis. Common methods of generating this population include error-prone PCR, DNA shuffling, chemical mutagenesis or the use of *E. coli* mutator strains. Iterative application of these techniques applied through multiple generations leads to an accumulation of codon and amino acid mutations that result in improvements in characteristics determined by the screening or selection method. Desired new traits may be increased stability at higher temperatures; reduced sensitivity to organic solvents; higher chemo-, regio- or enantio-selectivity toward a particular substrate; increased activity on a non-natural substrate; reduction of unwanted side reactions; improvements in solubility or expression or any combination of these. The characteristics necessary for implementation of the enzyme in a biocatalytic process (i.e. industrial process conditions) determine the goals of the directed evolution project at hand, and may differ for each application

Complementary to the targeted mutagenesis approach used in Chapter II, random mutagenesis does not rely on prior knowledge of enzyme structure, mechanism or active site architecture in order to improve the desired characteristic. For this reason,

methods in directed evolution can also be successfully applied to systems that are minimally characterized. Although it is a purely stochastic method, incorporating mutations arbitrarily by position and nucleobase shift into a gene, some control can be granted to the researcher through carefully designed experimental protocols. Mutation frequency can be titered to fit a desired range of mutations per gene by adjusting template concentration and PCR amplification cycles. Furthermore, rather than amplifying the entire gene, a segment of the sequence can be subjected to the mutagenic PCR conditions in order to target mutagenesis to a small portion of the full gene. By these methods, beneficial mutations can be identified throughout the protein sequence that may otherwise have been difficult to predict.

In this chapter, PPM variants with beneficial mutations identified through targeted active site saturation mutagenesis were subjected to further engineering by directed evolution. The Ser154Gly and Val158Leu templates each provided a unique characteristic that allowed two interesting avenues of evolution to be pursued through random mutagenesis in order to develop the best catalyst for selective activity on dideoxyribose 5-phosphate. One round each of error-prone PCR and recombination of mutations were used to generate libraries of PPM variants, leading to a final mutant possessing four mutations, 3-fold higher turnover of dideoxyribose 5-phosphate in cell lysate and 710-fold improved substrate selectivity.

Methods

PPM mutant library generation

Error prone PCR (epPCR) was performed using the GeneMorph II Random Mutagenesis kit (Stratagene) using 25 replication cycles and 7 - 13 ng/µL template plasmid. epPCR products were gel purified, digested with *NheI*/*XhoI* (New England Biolabs) and gel purified again before ligation into purified pET28a+ that had been

similarly digested and treated with alkaline phosphatase (New England Biolabs). Ligations were prepared overnight at 4°C using T4 DNA ligase (Promega) and contained 3:1 insert:vector ratio. Pellet Paint Co-Precipitant (Novagen) was used to concentrate DNA before transformation. Libraries generated from 10 of 8 ng/µL template were used to create the library of variants for screening, with each gene possessing 1.5 - 2.5 mutations per gene on average in 10 randomly sequenced clones from each library. Random recombination of epPCR mutations was performed using the QuikChange Multi Site-Directed Mutagenesis kit (Stratagene) where PCR samples contained a single template and a forward primer for each mutation not present at a position on the template, allowing any additional mutations to be incorporated at random in any combination. For example, the 500F6 sample contained primers for the Thr81Ile, Thr81Asn, Phe101Leu and Ile238Ile mutations (See Figure 3-4 for the complete series of mutations). PCR samples and thermal cycling for recombination of mutations were performed following the recommended protocol by the kit manufacture's manual. Primers used are listed in Table 3-1. Sample preparation for cloning protocols was performed as recommended in kit manuals.

Table 3-1. Primers used in random mutagenesis and recombination of PPM. Mutations in each sequence are underlined. N=A, T, C or G. K=G or T. M=C or A.

Primer Name	Nucleotide Sequence
epPCR for	CATGGGCAGCAGCCATCATCATCATCATCACAGC
epPCR rev	GTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCG
T81I for	GCAAGAGAAATCTAT<u>T</u>TGGTAAAGATACAATGACAG
T81N for	GCAAGAGAAATCTA<u>A</u>TGGTAAAGATACAATGACAG
F101L for	ATTGATACACCA<u>C</u>TCCAAGTGTTCCCAGAAGG
M190K for	TTAGATGAGAAATACA<u>A</u>GGTAGGTCGTGTTATTGC
I238I for	GACTACGATGTAATTGCTAT<u>A</u>GGTAAAATCTCTG
P361P for	CGGACAAGAGTTACC<u>T</u>CTTCGTCAAACATTTGC

Library growth and screening

Colonies were picked to 96-well round bottom plates containing 100 µL LB medium with 50 µg/mL kanamycin and grown for 24 h at 37°C with shaking at 200 rpm. Plates held 4 wells of negative control (*E. coli* with empty pET28a+ vector), 4 wells of positive control (vector with template gene) and 88 wells of the mutant library. For epPCR and recombination libraries (1056 and 88 clones, respectively), initial 100 µL overnight plates were copied to 50 µL LB medium with kanamycin and grown for an additional 24 h before collecting the cells. All cells were collected by centrifugation at 1600 rcf and the broth was removed by inversion before storing the plates at -80°C until ready for assay. PPM activity was determined in a tandem assay with purine nucleoside phosphorylase and hypoxanthine consumption was measured by xanthine oxidase at endpoints using a typical hypoxanthine detection assay[(2)].

Thawed cell pellets were resuspended in 100 - 200 µL lysis mix containing 0.1 mM $MnCl_2$, 0.25 mg/mL egg white lysozyme (Sigma) and DNase I (Sigma) in 25 mM Tris-HCl, pH 8 and incubated for 10 min at 25°C before undergoing a single freeze/thaw cycle at -80°C to 37°C. After centrifugation, 25 µL of the clarified cell lysate was transferred to a 96-well flat bottom plate and 55 µL of an assay mix was added to initiate the reaction. Final concentrations of components in 80 µL reactions for the dideoxyribose 5-phosphate screen were 0.1 mM $MnCl_2$, 10 µM hPNP-46D6, 1 µM glucose 1,6-bisphosphate, 600 µM hypoxanthine and 1-2 mM dideoxyribose 5-phosphate. Assays were incubated at room temperature for 40 - 70 min before quenching using 30.5 µL 1 M NaOH. After a minimum of 30 min, the solution was neutralized using 29.5 µL 1 M HCl before addition of 35 µL developing solution containing 0.2% Triton X-100, 7.5 mM iodonitrotetrazolium chloride and xanthine oxidase in 25 mM Tris-HCl, pH 8. Hypoxanthine consumption was determined by measuring absorbance of the colored formazan at 546 nm and normalized to percent

activity in comparison to the positive and negative controls. The top ~45 hits from each primary screen were regrown from fresh transformants and retested in duplicate to validate activity. The best 4 - 9 hits after validation were screened in a tertiary assay in duplicate under the same conditions as above with 500, 1000 and 2000 µM dideoxyribose 5-phosphate to select the top hit from each round.

Libraries containing the top hits from each round were freshly prepared to directly compare activity changes through the rounds of evolution. Plates containing 50 µL cultures of colonies hosting empty pET28a+, wild-type PPM, Ser154Gly, 12D2, 500F7, 2G8, Val158Leu, 650G11, 500F6 and 4H11 were lysed in 150 µL (for dideoxyribose 5-phosphate) or 200 µL (for ribose 5-phosphate) lysis mix and assayed with 1 mM ribose 5-phosphate or dideoxyribose 5-phosphate as described above. Reactions were incubated for 10-90 min (ribose 5-phosphate) or 45 - 60 min (dideoxyribose 5-phosphate) before being quenched with 30.5 µL 1 M NaOH then neutralized and developed as described above. Turnover was normalized by incubation length then to activity of wild-type PPM for final comparison.

Enzyme expression and purification

Plasmids containing wild-type or variant PPM, PNP, hPNP-46D6[(3)] were transformed into *E. coli* BL21(DE3) and grown at 37°C in LB medium supplemented with 50 µg/mL kanamycin and induced with 1 mM IPTG for 3 - 6 h after OD_{600} had reached 0.5 - 0.6. Cell pellets were resuspended in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM Imidazole, pH 7.4) and disrupted by passing through a French Pressure cell (1400 psi). The clarified lysate was applied to HisTrapFFcrude Nickel affinity column (GE Healthcare, Inc.) and washed at 10% Buffer B (Buffer A with 500mM imidazole). Protein was eluted by a linear gradient from 10% Buffer B to 60% Buffer B, before a step up to 100% Buffer B to fully elute the column. The purified enzyme was concentrated,

desalted and exchanged into 25 mM Tris-HCl, pH 8 before storage at -80°C. All enzyme concentrations were determined using the BCA Protein Assay Kit (Thermo Scientific, Inc.). Xanthine oxidase was purified from raw bovine milk using previously reported protocols[4].

PPM kinetics assays

The activity of wild-type and variant PPMs was measured in a tandem assay with either PNP or hPNP-46D6. Ribose 1-phosphate formed by PPM was subsequently consumed by a catalytic excess of PNP in the presence of hypoxanthine to produce inosine. Similarly, production of dideoxyribose 1-phosphate via PPM activity was converted to dideoxyinosine in the presence of hypoxanthine and a catalytic excess of hPNP-46D6. Inosine or dideoxyinosine produced in the assay was separated from other reaction components using a Luna Phenyl-Hexyl column (4.6 X 250 mm, Phenomenex) and an isocratic flow of 1.0 mL/min of 10 mM ammonium acetate in 95% water:5% acetonitrile, pH 6. A Thermopal autosampler was used to inject 10 µL of the sample for analysis. Nucleosides were analyzed on a TSQ Quantum Access triple quadrupole electrospray ionization-LC/MS (Thermo, Inc.) using selected reaction monitoring fragmentation to the free nucleobase (inosine $[M+H]^+$ 269 m/z and dideoxyinosine $[M+H]^+$ 237 m/z transition to hypoxanthine $[M+H]^+$ 137 m/z) with 2-deoxyguanosine as the internal standard ($[M+H]^+$ 268 m/z to guanine $[M+H]^+$ 152 m/z). Nitrogen was used for both the auxiliary and sheath gases and was set to 45 units and 30 units, respectively. The following instrument parameters were used: source voltage 4.5kV; vaporizer temperature 0 °C; capillary temperature 270 °C; tube lens 101 V; skimmer offset -5 V; collision energy -10 V. Data acquisition and analysis were conducted with Thermo Xcalibur software, version 2.1.

All reactions were performed in 100 µL volumes in 96-well plates. Wild-type or variant PPM was activated at a concentration 10-fold higher than that used in the assay by incubation for 10 min at room temperature in 25 mM Tris-HCl and 0.1 mM $MnCl_2$ with either 5 µM (wild-type PPM) or 10 µM (variant PPM) glucose 1,6-bisphosphate then held at 4°C until assayed. Biochemical assays for PPM activity on ribose 5-phosphate contained 0.1 mM $MnCl_2$, 5 µM PNP, 600 µM hypoxanthine and 0 - 1000 µM or 0 - 4000 µM ribose 5-phosphate in 25 mM Tris-HCl, pH 8. Assays for PPM activity on dideoxyribose 5-phosphate contained 0.1 mM $MnCl_2$, 10 µM hPNP-46D6, 600 µM hypoxanthine and 0-5000 µM dideoxyribose 5-phosphate in 25 mM Tris-HCl, pH 8. PPM concentrations ranged from 0.02-0.25 µM for ribose 5-phosphate assays and 0.25 µM for dideoxyribose 5-phosphate assays. Reactions were initiated by 10 µL addition of the sugar 5-phosphate substrate to 90 µL mix containing all other components and were incubated for 2-8 min at room temperature before being quenched by addition of 5 µL 2 M NaOH. After 30 min, 5 µL of 2 M HCl/1 M $CaCl_2$ was added to neutralize the mixture and the assay plate was centrifuged to pellet the precipitates. A 40 µL aliquot of each sample was combined with 10 µL of 50 µM 2-deoxyguanosine internal standard to prepare the sample for LC/MS analysis. Inosine and dideoxyinosine formation was quantified by relative peak area of analyte to a 10 µM 2-deoxyguanosine internal standard in comparison to a standard curve made using authentic inosine (Acros Organics) and dideoxyinosine (3B Pharmachem (Wuhan) International Co. Ltd.). Retention times were approximately 5 min for inosine, 6.2 min for 2-deoxyguanosine and 14 min for dideoxyinosine.

Crystallization, data collection, and structure determination of wild-type and variant PPM

Purification and preparation of the PPM variants for crystallography was performed as previously described[(5)]. Crystals of the 4H11 variant grew after combining

2 µL protein solution and 2 µL reservoir solution. Crystals of the 4H11 variant (20 mg/ml enzyme in 1 mM $MnCl_2$, 25 mM Tris-HCl, pH 7.4) grew over a reservoir solution containing 25% polyethylene glycol 3350, 50 mM $(NH_4)_2SO_4$, and 100 mM Bis-Tris, pH 5.45. All crystals were cryoprotected with a solution that was 70% v/v reservoir solution and 30% glycerol before flash cooling by plunging into liquid nitrogen.

Data were collected at a temperature of 100 K using a wavelength of 0.979 Å and a MarMosaic225 CCD detector at the Advanced Photon Source (Argonne, IL) LS-CAT beamline 21-ID-G (Table 3-2). Data were processed and scaled using the HKL suite of programs[6]. Space group and unit cell information for all crystals is listed in Table 3-2. For the 4H11 variant, initial phases were obtained by molecular replacement with Molrep[7] of the CCP4 Suite[8], using PDB entry 3M8Z. The model was refined using iterative rounds of model building in COOT[9] and refinement in CNS[10] and Refmac5[8, 11] using Translation/Libration/Screw (TLS) refinement[12]. Figures 3-5, 3-6 and 3-7 were made using PyMOL[13].

Table 3-2. Data collection and refinement statistics for wild-type and variant PPM.

Protein	**4H11**
Ligand	none
PDB entry	4LRD
Data collection	
Resolution (Å)	50-1.78
High resolution bin	1.81-1.78
Space group	$P2_1$
Unit cell	a=39.2
	b=60.5
	c=78.9
	β =98.5
Total reflections	108,153
Unique reflections	34,906
[b]R_{sym} (%)	7.3 (25.3)
I/σ	15.3 (3.3)
Completeness (%)	99.5 (99.7)
Refinement	
[c]R_{cryst} (%)	13.9
[d]R_{free} (%)	17.9

[a]Values in parentheses are for the highest resolution bin
[b]$R_{sym}=\sum |I_{obs}-I_{avg}|(100)/I_{avg}$,
[c]$R_{cryst}=[\sum||F_{obs}|-|F_{calc}||\ (100)]/\sum|F_{obs}|$
[d]R_{free} is calculated using the same equation as R_{cryst} using a subset of reflections omitted from the refinement process.

Results

Optimization of epPCR mutagenesis conditions

Error-prone PCR has become the most common method of generating a random mutagenesis library[(14)]. Although several variations of epPCR have been developed differing in reaction conditions used to create the random mutations (described in Chapter I), this work used the error prone polymerase Mutazyme II from Stratagene to generate the mutant pool for activity screening. The mutagenesis and screening process is outlined in Figure 3-1. In our application, Mutazyme II was used to create a genetically diverse library of PPM variants using plasmids contianing the Ser154Gly and Val158Leu PPM variants as templates. The mutant gene pool was subsequently ligated to a pET28a expression vector and transformed into *E. coli* BL21(*DE3*) to generate the

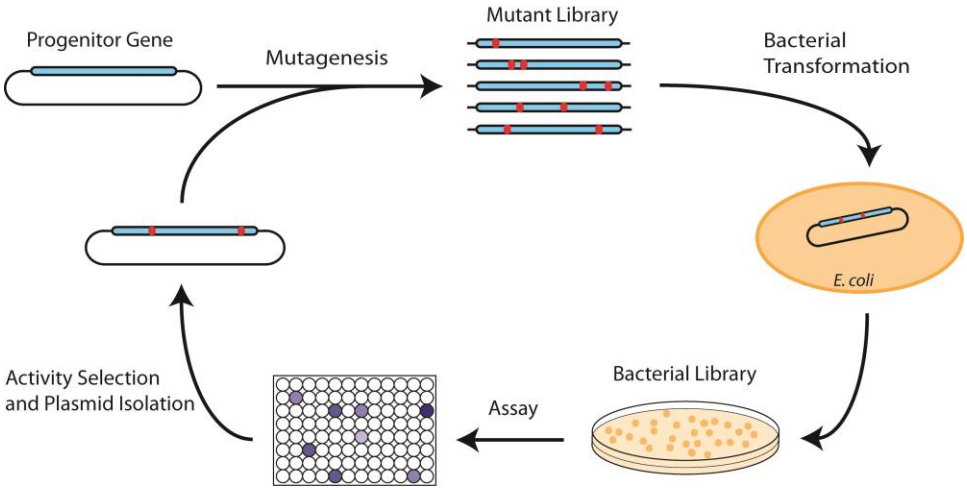

Figure 3-1. Iterative process of mutagenesis and screening used in directed evolution of PPM. A progenitor gene is mutagenized to create a pool of randomly mutated genes. After ligation into an expression vector and transformation into *E. coli*, clones are grown in 96-well plates, lysed and assayed for the desired activity. Secondary screens are used to validate activity and the top performing clone is chosen as the progenitor for a new round of mutagenesis.

bacterial library. Individual colonies were picked and grown in 96-well plates, lysed and assayed for increased turnover of dideoxyribose 5-phosphate in tandem assays with hPNP-46D6 and hypoxanthine. Clones showing improved activity in the primary screen were retested in duplicate in a secondary assay and again at varied substrate concentrations in a tertiary screen in order to identify the most improved variant from the round of mutagenesis. This clone then became the progenitor for a new round of mutagenesis to further increase activity.

PCR sample conditions for directed evolution must be optimized for the gene of interest in order to ensure a sufficiently assorted population of mutants within the library. The rate of mutation is easily titered by adjusting the template concentration and number of amplification cycles used in thermal cycling. Typically, mutation rates are adjusted to provide a dead rate (i.e. <20% of the progenitor activity in this study) of approximately 30-40%[(15)]. Small libraries of PPM variant were assayed to test epPCR parameters in order to arrive at the desirable conditions, which resulted in 45% inactive clones using

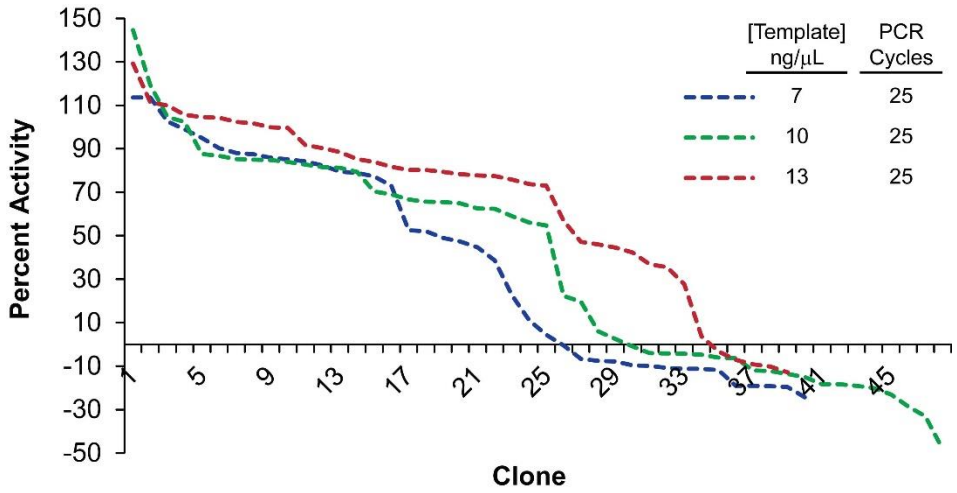

Figure 3-2. Error-prone PCR mutagenesis rates determined through testing sample conditions. Conditions in green were selected for use in directed evolution of PPM as they provided a desirable range of inactive colonies.

500 ng PPM template per 50 µL reaction and 25 replication cycles (Figure 3-2). Sequencing a small subset of randomly selected variants within this library indicated an average of 1.5 -2 nucleobase mutations per gene and ranging from zero to five mutations in the 10 clones.

Random mutagenesis and recombination of PPM variants

Since each of the two PPM active site mutations provided large substrate selectivity changes through seemingly distinct and potentially competitive mechanisms, the Ser154Gly and Val158Leu variants were separately used as templates for random mutagenesis by epPCR to further improve dideoxyribose 5-phosphate catalysis. Approximately 1000-member libraries for each template were screened for dideoxyribose 5-phosphate activity and the top two variants from the first generation epPCR libraries for each template provided 150 - 250% higher dideoxyribose 5-phosphate turnover in cell lysate than the respective progenitor (Figure 3-3). These new variants were 12D2 (Thr81Ile mutation and Ile238Ile codon change) and 500F7 (Phe101Leu) from the Ser154Gly template and 650G11 (Thr81Asn) and 500F6

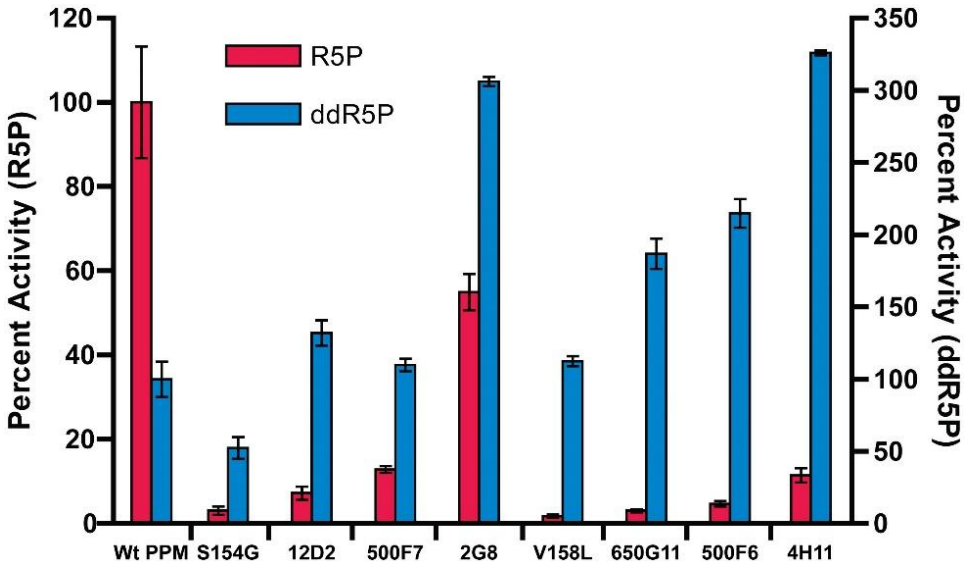

Figure 3-3. Substrate activity through generations of PPM evolution. Comparison of changes in R5P and ddR5P turnover rate per minute by PPM variants in cell lysate normalized to wild-type PPM. R5P activity initially shows substantial loss as each single mutant, but slowly regains activity through directed evolution. ddR5P activity steadily improves throughout directed evolution, reducing overall R5P selectivity in the later variants. Data are mean ± s.d. (n=3). Wild-type PPM turnover is 3.18 ± 0.52 μM min^{-1} for R5P and 0.486 ± 0.083 μM min^{-1} for ddR5P. (R5P, ribose 5-phosphate; ddR5P, dideoxyribose 5-phosphate).

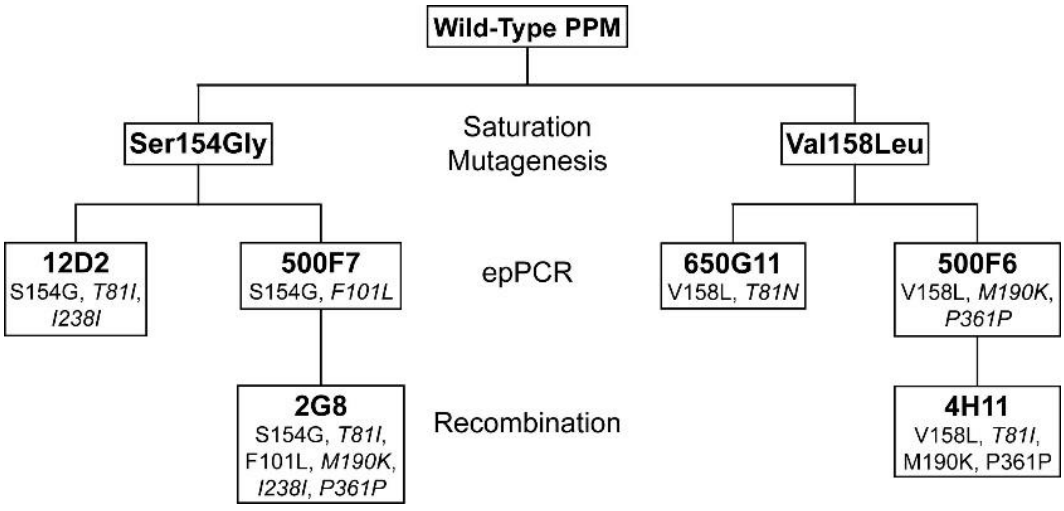

Figure 3-4. Lineage tree of PPM variants. Clone name is given in bold with mutations listed below. New mutations accumulated through the indicated method of mutagenesis are listed in italics.

(Met190Lys and Pro361Pro) from the Val158Leu template (Figure 3-4). Kinetic characterization of the top clone in each library of the initial screen revealed increased turnover of both substrates, but further improvements in selectivity as a result of large changes in substrate K_M values in favor of dideoxyribose 5-phosphate (Table 3-3). While the Ser154Gly 12D2 variant showed 8.4-fold preference for ribose 5-phosphate, selectivity of the Val158Leu 650G11 variant was entirely reversed, actually favoring dideoxyribose 5-phosphate at 1.4-fold over the natural substrate ribose 5-phosphate after >1400-fold change in substrate selectivity (Table 3-3).

To maximize the total improvement from this round of evolution, each of these four hits were subjected to random recombination using the QuikChange Multi Site-Directed Mutagenesis kit from Stratagene. PCR samples contained a single template and a forward primer for each mutation not present at a position on the template. PCR sample conditions were adjusted according to the kit manual to allow

Table 3-3. Kinetic parameters of all PPM variants for R5P and ddR5P substrates and comparison of changes in substrate selectivity.

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1}) (x10^3)	[k_{cat} / K_M (R5P)] / [k_{cat} / K_M (ddR5P)]	Fold Change
Wild-Type	R5P	10.4±0.2	40±3	260±20		
	ddR5P	0.43±0.04	1700±300	0.25±0.05	1028	1
S154A	R5P	1.7±0.1	130±20	13.1±2.2		
	ddR5P	0.39±0.02	440±60	0.89±0.13	14.8	70
S154G	R5P	0.71±0.02	130±10	5.5±0.5		
	ddR5P	0.21±0.02	800±200	0.26±0.07	20.8	49
V158L	R5P	0.42±0.02	1020±100	0.41±0.04		
	ddR5P	0.24±0.01	680±70	0.35±0.04	1.2	881
S154G/V158L	R5P	0.035±0.001	190±20	0.18±0.02		
	ddR5P	0.028±0.002	480±90	0.058±0.012	3.2	326
12D2	R5P	1.08±0.03	180±10	6.0±0.4		
	ddR5P	0.27±0.02	380±70	0.71±0.14	8.4	122
650G11	R5P	1.06±0.04	2160±150	0.49±0.04		
	ddR5P	0.32±0.01	470±60	0.68±0.09	0.7	1426
2G8	R5P	5.4±0.1	230±10	23.5±1.1		
	ddR5P	0.38±0.01	450±60	0.84±0.11	27.8	37
4H11	R5P	4.1±0.1	2830±180	1.5±0.1		
	ddR5P	0.41±0.01	410±50	1.0±0.1	1.4	710

additional mutations contained on the mutagenic primers to be incorporated at random in any combination. Screening 88 clones from each of the four libraries and validating the top hits revealed clones 2G8 in the Ser154Gly lineage and 4H11 of the Val158Leu lineage as the most active recombinants in the screen, accumulating the mutations listed in Figure 3-4. In total, the 2G8 variant showed approximately 300% improved activity after gaining six mutations while the 4H11 variant demonstrated nearly 325% greater activity than wild-type PPM in cell lysate after accumulating four mutations (Figure 3-3). Additional rounds of epPCR on both the 2G8 and 4H11 templates using the same epPCR conditions and at a lower template concentration to increase the mutation rate failed to produce variants with improved turnover, so each were analyzed as the final PPM variants in the directed evolution study.

In tracking changes of ribose 5-phosphate in cell lysate through PPM evolution, initial saturation mutagenesis hits Ser154Gly and Val158Leu lost greater than 95% activity on the natural substrate due to the single active site mutations (Figure 3-3). As also observed in the purified enzyme kinetic parameters, successive rounds of random mutagenesis and recombination slowly recovered a portion of the lost activity, reaching 55% and 11% of the wild-type level for 2G8 and 4H11, respectively. Comparing the final two variants, 2G8 contained all of the mutations present in the 4H11 variant gathered through random mutagenesis (Thr81Ile, Met190Lys and Pro361Pro), along with two others unique to the Ser154Gly lineage (Phe101Leu and Ile238Ile) (Figure 3-4).

Overall, most catalytic parameters were very similar between the final variant from each lineage (Table 3-3). The 2G8 variant has a k_{cat} and K_M of 0.38 s^{-1} and 450 μM, respectively, for dideoxyribose 5-phosphate, which is a slight decrease in turnover rate but a 3.8-fold improvement in K_M from the wild-type enzyme. The 4H11 variant shows similar dideoxyribose 5-phosphate kinetic parameters with a k_{cat} of 0.41 s^{-1} and a K_M of 410 μM. Because random mutations were not anticipated to selectively improve

dideoxyribose 5-phosphate turnover, it was not surprising that each variant also showed increased turnover rate of ribose 5-phosphate over the single mutant progenitors, up to 5.43 s^{-1} for 2G8 and 4.11 s^{-1} for 4H11. The one notable difference between the 2G8 and 4H11 variants was in K_M values for ribose 5-phosphate. The 2G8 variant showed a much lower binding constant of 230 μM than the 4H11 variant at 2830 μM, which is 70-fold higher than the wild-type enzyme. This large difference in K_M values is presumably due to the putative steric effects caused by the original Val158Leu mutation. The contribution of this difference is reflected in the change in substrate selectivity of each enzyme, where wild-type PPM exhibits a >1000-fold preference for ribose 5-phosphate while selectivity in the best final variant, 4H11, was reduced by 710-fold so that ribose 5-phosphate is only favored 1.4-fold over dideoxyribose 5-phosphate (Table 3-3).

To identify the molecular basis for this greatly improvement in selectivity, we determined the structure of the 4H11 variant. The active site of PPM is located between two domains[(16)], and the 4H11 structure was associated with an interdomain rotation that resulted in a unique alignment of the substrate binding and catalytic residues at the active site. One consequence of the cap domain movement is highlighted in the inset. Asp156 has previously been shown to coordinate an active site Mn^{2+}, however, in the 4H11 structure, the location of this coordinating residue is shifted 4.6 Å away from the Mn^{2+} and the coordination sphere is instead completed by water molecules (Figure 3-5a).

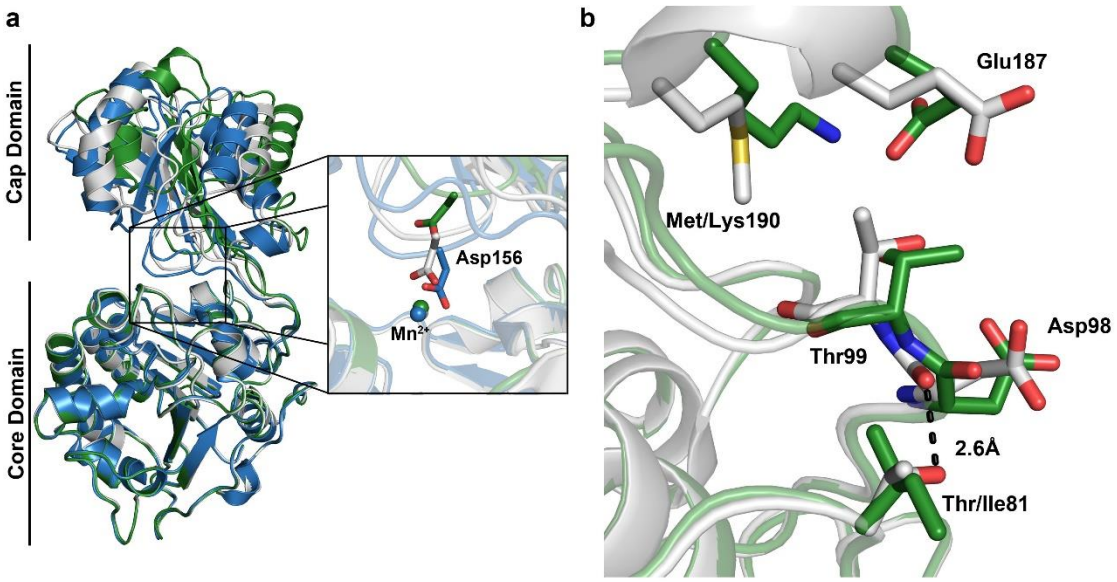

Figure 3-5. Structure comparison of wild-type and 4H11 PPM. (**a**) The core and cap domains of the 4H11 variant (green, PDB ID 4LRD) are related by a different interdomain angle than wild-type PPM in both its active, phosphorylated form (gray, PDB ID 3TWZ[17]) and the unphosphorylated and unactive form (blue, PDB ID 3TX0[17]). The insert indicates the extent of domain movement via the shift in Mn^{2+} coordinating residue Asp156. (**b**) The hinge region of 4H11 PPM (green, PDB ID 4LRD) appears to be destabilized as a result of the mutations at positions 81 and 190, contributing to the observed domain movement. The Thr81Ile mutation removes a hydrogen bond to the backbone of Asp98 in wild-type PPM (gray, PDB ID 3TZW[17]) and the Met190Lys mutation may allow for potential new interactions with Thr99 and Gly187.

Specifically contributing to the observed domain rotation, the mutations at positions 81 and 190 seem to destabilize important hinge region contacts that serve to modulate interdomain movement. The Thr81Ile mutation appears to remove a hydrogen bond interaction between the wild-type residue and the backbone amide of Asp98 (Figure 3-5b). The new unfavorable non-polar to polar interactions between these residues slightly displaces the beginning of this hinge region, which is propagated through the rest of the hinge. Furthermore, the position of the Met190Lys mutation seems to allow for potential new interactions with the side chains of Thr99 and Glu187, which is suggested, in part, by the altered orientations of these residues observed in the 4H11 structure (Figure 3-5b). However, the high B-factors observed for the Lys190 side chain in the 4H11 structure indicate that this residue does not form a completely stable

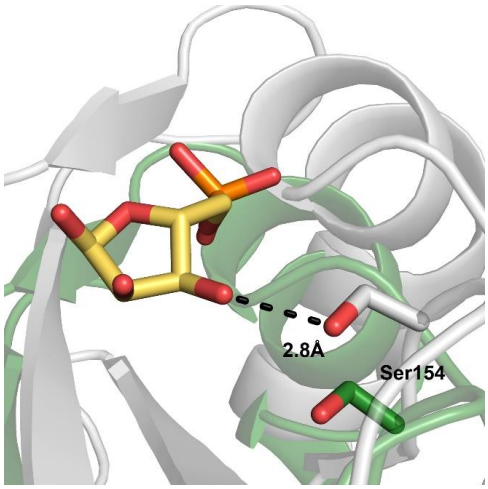

Figure 3-6. Repositioning of Ser154 after domain movement. Residue Ser154 in wild-type PPM (gray, PDB ID 3M8Z[16]) interacts with ribose 5-phosphate via a hydrogen bond at a distance of 2.8Å. This residue is shifted in the structure of the 4H11 variant (green, PDB ID 4LRD) as a result of the domain movement and is no longer able to interact with ribose 5-phosphate bound in the same position (4.3Å).

interaction with any of the neighboring residues, possibly further contributing to the proposed destabilization and flexibility of this region. Combined, these observations suggest that the 4H11 variant is associated with increased interdomain flexibility, which could partly explain the increased accommodation of the non-natural dideoxyribose 5-phosphate substrate. Cocrystallization with ribose 5-phosphate or dideoxyribose 5-phosphate did not result in the appearance of clear electron density in the active site. However, manual docking based on the position of each substrate in the wild-type enzyme showed that the domain twist resulted in Ser154, which hydrogen-bonds to ribose 5-phosphate in the wild-type enzyme, no longer being positioned for substrate interaction (Figure 3-6). Loss of this key interaction could further explain the observed increase in K_M for ribose 5-phosphate (Table 3-3).

Discussion

Identifying two active site mutations with mutually exclusive effects on substrate selectivity, the Ser154Gly and Val158Leu single mutant variants were both selected to be progenitors for further evolution by random mutagenesis. The particular active site qualities of each enzyme provided an opportunity to evaluate two divergent avenues of evolution through random mutagenesis: one following the Ser154Gly variant with a larger active site and the other following the generalist Val158Leu variant showing no preference between the natural and target non-natural substrates. Hits were found on both templates after gaining one or two mutations through random mutagenesis, with improvement in dideoxyinosine production in cell lysate up to 210% of wild-type PPM activity (Figure 3-3).

Because the random mutations in the top two hits from each library were an assortment of conservative, silent and drastic changes, we generated and screened a recombination library to identify further improvements from new beneficial combinations of the random mutations. The top performing clones were 2G8 in the Ser154Gly family and 4H11 in the Val158Leu line (mutations listed in Figure 3-4). These new variants showed an additional 275% and 150% increase in dideoxyinosine formation, respectively, over the template enzyme (Figure 3-3). Interestingly, the largest changes in activity, for both substrates, were observed after the mutagenesis round where beneficial mutations were randomly combined. This seems to suggest that several of the mutations act independently to provide new characteristics that are not mutually exclusive with respect to the other mutations. Overall, total dideoxyinosine production improvements of approximately 3-fold over wild-type PPM were seen in the two final mutants in cell lysate. These parameters are also comparable to the 3.4-fold and 3.9-fold increase in dideoxyribose 5-phosphate catalytic efficiency seen in the purified 2G8 and 4H11 variants, respectively (Table 3-3), indicating that the improvements in cell lysate are not

solely due to increased expression, but likely a combination of enhanced expression and kinetic parameters.

Accompanying changes in ribose 5-phosphate activity via purified enzyme kinetics and assays in cell lysate revealed substantial differences between the two final variants and the wild-type enzyme. Kinetic parameters indicate approximately a 700-fold change in substrate selectivity in 4H11 PPM (Table 3-3), resulting in only a 1.4-fold preference for ribose 5-phosphate over dideoxyribose 5-phosphate, thus preserving the bulk of the selectivity imparted by the original Val158Leu single mutation. This extensive change of selectivity is reflected in the small percentage of ribose 5-phosphate activity remaining in 4H11 in the cell lysate activity comparison (Figure 3-3). The Ser154Gly single mutant showed a 14.6-fold loss in ribose 5-phosphate turnover as the main contributing factor in the 47-fold change in substrate selectivity. However, through directed evolution, the k_{cat} for ribose 5-phosphate was restored to approximately half that of wild-type PPM and cell lysate showed 55% of the wild-type enzyme activity in the final 2G8 variant (Figure 3-3), reducing the substrate selectivity to only a 36-fold improvement (Table 3-3).

Unfortunately, these results contradicted the initial hope that the expanded binding pocket of the Ser154Gly variant might permit ensuing evolution to refit the active site more specifically to dideoxyribose 5-phosphate, but instead allowed mutations to generally improve activity on both substrates. Perhaps if ribose 5-phosphate activity was tested in a side-by-side assay throughout the directed evolution process, mutations on the Ser154Gly template directly affecting active site architecture and dideoxyribose 5-phosphate binding could have been identified by only selecting variants that presented further improvements in substrate selectivity. Though not necessarily a general gauge of predicting success, in this directed evolution study the library beginning with the Val158Leu general catalyst proved to be the more advantageous template for

engineering. The initial targeted saturation mutagenesis provided a solid foundation of engineered selectivity that was maintained through subsequent rounds of directed evolution, and, in the case of the 4H11 variant, appears to be integral in maintaining selectivity changes as directed evolution was used to improve catalysis on the non-natural substrate.

Modeling the location of each mutation onto wild-type PPM reveals that all three positions that underwent a beneficial residue change (Thr81, Phe101 and Met190) were located in or near the hinge region that connects the PPM core and cap domains rather than within the active site (Figure 3-7). Instead of directly affecting substrate interactions, these mutations might allow an improvement of a structural or dynamic capacity to generally enhance catalysis, which is partially supported by the domain movement and new interactions observed in the 4H11 structure (Figure 3-5). While speculative, it is possible that these mutations (i) increase the flexibility of the hinge, allowing a higher likelihood of domain closure to improve catalysis, (ii) help stabilize the closed form of the enzyme to provide more time for proper catalysis to occur or possibly a combination of both. Further crystallography studies may be able to help clarify these potential contributions, possibly by aiming to crystallize the 4H11 and 2G8 variants in the normally observed interdomain orientation or as well attempting to capture the enzyme in the closed form that is suggested to coincide with the enzymatic activity[17].

Interestingly, one top hit from each random mutagenesis library contained a mutation at Thr81, Thr81Ile (along with Ile238Ile) in 12D2 and Thr81Asn as the sole mutation in 650G11, and additionally a second variant in the top 5 clones identified in the Ser154Gly library held Thr81Ile as a lone mutation (not shown). Although the location of this residue is outside the substrate binding pocket (Figure 3-7), the fact that multiple PPM variants containing a mutation at this position indicates that it may have an impact on the active site structure that ultimately affects catalysis. Thr81 is near the

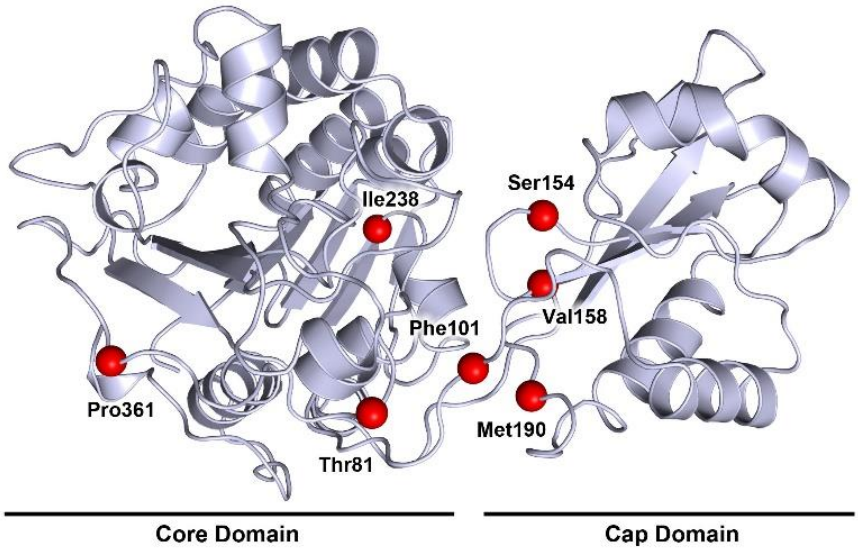

Figure 3-7. Positions of mutations mapped onto wild-type PPM with ribose 5-phosphate bound (PDB ID 3M8Z[16]). Positions 81, 238 and 361 are located in the core domain, 154 and 158 are located in the cap domain and 101 and 190 are in the hinge region connecting the two domains. Combinations of mutations specific to each PPM variant in the directed evolution process are provided in Figure 3-4.

phosphorylated Thr85 residue required for PPM catalysis[16, 17]. As stated above, the observed effect of the Thr81Ile mutation on the hydrogen bond network within the hinge region appears to contribute to the domain movement in the 4H11 crystal structure. As another possibility, mutations within this loop region may alter the presentation of phospho-Thr85 to incoming substrates for the phosphotransfer reactions. This may also explain why the Thr81Asn mutation, which is likely still able to form a hydrogen bond with the backbone of Asp98, was identified as another beneficial mutation at this location. Because mutations at this position were found after modifying the active site through saturation mutagenesis, small changes in the orientation of phospho-Thr85 may have been necessary to compensate for a slightly altered substrate binding orientation to realign the phosphate group for proper and efficient in-line attack during catalysis.

Conclusions

Directed evolution using whole gene epPCR and screening for activity in cell lysate successfully identified improved PPM variants through two rounds of mutagenesis in two separate libraries. The most improved mutant, 4H11 PPM, had a catalytic efficiency 4-fold greater than the wild-type enzyme, and substrate selectivity was shifted from >1000-fold in favor of ribose 5-phosphate to only a 1.4-fold preference. This large change in substrate selectivity would greatly aid in production of dideoxyinosine in an *in vivo* or cell lysate process by minimizing undesirable activity on the natural substrate.

Although attempts at improving activity through further rounds of epPCR after random recombination were not met with success, other mutagenesis methods may be more beneficial in identifying new mutations to increase activity on dideoxyribose 5-phosphate, and indeed applying different mutagenesis methods can yield different levels of success in enzyme engineering toward a particular goal[18]. Additionally, random mutagenesis methods are commonly used to identify sequence 'hotspots' that are later targeted directly for mutagenesis[14, 19, 20], and it is possible that some of the positions identified here through random mutagenesis are within such areas. Further mutagenesis studies targeting residues in the loop region around Thr81 as well as the hinge region near residues Phe101 or Met190, perhaps by saturation mutagenesis or focused random mutagenesis, could lead to further improvements in dideoxyribose 5-phosphate turnover to increase the productivity of the enzyme. However, as indicated in the variants identified through this mutagenesis program, changes in activity may not be selective for dideoxyribose 5-phosphate but rather be general for both the natural and non-natural substrate. A secondary assay for activity on ribose 5-phosphate may be required to selectively screen for improvements in non-natural substrate catalysis to ensure that a high degree of substrate selectivity is retained through the additional rounds of evolution. It should also be noted that in some instances where significant

enhancements of non-natural substrates were achieved, enzyme engineering was continued through 10-20 rounds of directed evolution using a variety of mutagenesis methods[21-23].

This tandem system consisting of the evolved 4H11 PPM and hPNP-46D6 represent a step toward the successful engineering of a non-natural biosynthetic pathway and a demonstration of bioretrosynthesis applied to pathway construction and optimization. Eventual implementation of a multistep biosynthetic pathway for the production of nucleoside analogs has great potential to directly affect the high cost of these drugs. Providing an economical alternative or at least a supplementary method of large volume production of these valuable pharmaceuticals could reduce the prohibitively high price of these drugs and make them more affordable to wider population in need of treatment.

Acknowledgements

This work was contributed to by William R. Birmingham, Chrystal A. Starbird, David P. Nannemann, Tina M. Iverson and Brian O. Bachmann. Bioretrosynthetic evolution of a didanosine biosynthetic pathway. W.R.B. and D.P.N. designed assays. W.R.B performed assays, screened all mutagenesis libraries, performed the kinetic characterization, expressed, purified and tested enzymes in the biosynthetic pathway studies. C.A.S. determined the crystal structures of the 4H11 PPM variant. D.P.N. established initial synthesis routes of dideoxyribose and dideoxyribose 5-phosphate. B.O.B supervised biochemical experiments and T.M.I. supervised X-ray crystallographic work. We thank V. Phelan for assistance with establishing the mass spectrometry based analysis methods and K. McCulloch for help in crystallography data collection.

References

1. Cobb, R. E., Sun, N., and Zhao, H. Directed evolution as a powerful synthetic biology tool. *Methods* **2013**, 60, 81-90.

2. De Groot, H., De Groot, H., and Noll, T. Enzymic determination of inorganic phosphates, organic phosphates and phosphate-liberating enzymes by use of nucleoside phosphorylase-xanthine oxidase (dehydrogenase)-coupled reactions. *Biochem. J.* **1985**, 230, 255-260.

3. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

4. Ball, E. G. Xanthine oxidase: Purification and properties. *J. Biol. Chem.* **1939**, 128, 51-67.

5. Panosian, T. D., Nannemann, D. P., Bachmann, B. O., and Iverson, T. M. Crystallization and preliminary X-ray analysis of a phosphopentomutase from *Bacillus cereus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, 66, 811-814.

6. Otwinowski, Z., and Minor, W. Processing of X-ray diffraction data collected in oscillation mode, In *Methods Enzymol.* **1997**, 276, pp 307-326.

7. Vagin, A., and Teplyakov, A. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, 30, 1022-1025.

8. Bailey, S. The CCP4 Suite - Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, 50, 760-763.

9. Emsley, P., and Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, 60, 2126-2132.

10. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2007**, 2, 2728-2733.

11. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, 53, 240-255.

12. Winn, M. D., Isupov, M. N., and Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2001**, 57, 122-133.

13. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, USA.

14. Nannemann, D. P., Birmingham, W. R., Scism, R. A., and Bachmann, B. O. Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med. Chem.* **2011**, 3, 803-819.

15. Cirino, P. C., Mayer, K. M., and Umeno, D. Generating mutant libraries using error-prone PCR, In *Methods in Molecular Biology: Directed Evolution Library Creation: Methods and Protocols*. **2003**, 231, (Arnold, F. H., and Georgiou, G., Eds.), pp 3-9, Humana Press, Totowa, N.J.

16. Panosian, T. D., Nannemann, D. P., Watkins, G. R., Phelan, V. V., McDonald, W. H., Wadzinski, B. E., Bachmann, B. O., and Iverson, T. M. *Bacillus cereus* phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle. *J. Biol. Chem.* **2011**, 286, 8043-8054.

17. Iverson, T. M., Panosian, T. D., Birmingham, W. R., Nannemann, D. P., and Bachmann, B. O. Molecular differences between a mutase and a phosphatase: investigations of the activation step in *Bacillus cereus* phosphopentomutase. *Biochemistry* **2012**, 51, 1964-1975.

18. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y., and Bocola, M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: Rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **2010**, 132, 9144-9152.

19. Chen, B., Cai, Z., Wu, W., Huang, Y. L., Pleiss, J., and Lin, Z. L. Morphing activity between structurally similar enzymes: From heme-free bromoperoxidase to lipase. *Biochemistry* **2009**, 48, 11496-11504.

20. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B. W., Reetz, M. T., and Jaeger, K. E. Directed evolution of an enantioselective lipase. *Chem. Biol.* **2000**, 7, 709-718.

21. Castle, L. A., Siehl, D. L., Gorton, R., Patten, P. A., Chen, Y. H., Bertain, S., Cho, H. J., Duck, N., Wong, J., Liu, D. L., and Lassner, M. W. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **2004**, 304, 1151-1154.

22. Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., and Huisman, G. W. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, 25, 338-344.

23. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.

Chapter IV

IDENTIFICATION OF DIDEOXYRIBOKINASE PROGENITOR ENZYME

Introduction

As the first step in the process of evolving an enzyme for activity on a non-natural substrate, properly identifying a suitable progenitor enzyme can ultimately determine the success of a biocatalyst development program. A prerequisite for selection is that the enzyme must catalyze the desired transformation, whether it be oxidation, reduction, phosphorylation, isomerization, hydrolysis, etc. Ideally, the enzyme would also catalyze the reaction on the non-natural substrate of interest. Even indications of low activity give a foundation to build upon, as iterative rounds of directed evolution can be used to amplify these weak activities into useful traits[(1)]. Additionally, enzymes from large families that perform a variety of reactions on diverse substrates tend to indicate a degree of 'evolvability' by demonstrating a high level of active site mutability that is both structurally and functionally accepted[(2)]. Laboratory methods of mutagenesis to impart new or enhanced functions can mimic this natural processes of diversification (i.e. natural selection) that has successfully expanded the enzyme family, permitting the generation of new biocatalysts with specifically tailored traits[(1)].

In addition to performing the desired biotransformation, structural data from refined crystal structures or homology models are also beneficial when selecting potential enzyme progenitors. Knowledge of the active site architecture and substrate binding will aid in determining interactions and roles of catalytic and substrate binding residues. This information can then be used to identify specific targets for functional analysis through methods such as saturation mutagenesis[(3-6)] without disturbing crucial residues involved in catalysis. Additional biochemical characterization in the form of

substrate acceptance is also a consideration when creating a library of prospective progenitor enzymes for a non-natural substrate, however, this data is not always available. Similar to how a diverse enzyme family can indicate the evolvability of an enzyme, awareness of low level promiscuity for additional substrates may prove to be useful data to obtain before pursuing a directed evolution campaign, as broad substrate acceptance in a wild-type enzyme may be a beneficial trait to indicate the flexibility and potential for engineering into a designer biocatalyst. On the other hand, enzymes known to exhibit particularly strict natural substrate specificity could be avoided if this information is known before-hand.

To continue the bioretrosynthetic construction process of a non-natural biosynthetic pathway for dideoxyinosine, we must develop a kinase enzyme with activity toward the non-natural sugar analog dideoxyribose to form the substrate for the evolved PPM (Figure 4-1). An efficient enzyme is necessary to produce dideoxyribose 5-phosphate to feed into the pathway and undergo successive biotransformations by the evolved PPM and PNP enzymes. This retro-extension expands the application of the model of retrograde evolution toward construction of a non-natural biosynthetic pathway by progressing to a simpler precursor substrate. Additionally, engineering kinase activity on the sugar analog will streamline the chemical synthetic component required to make

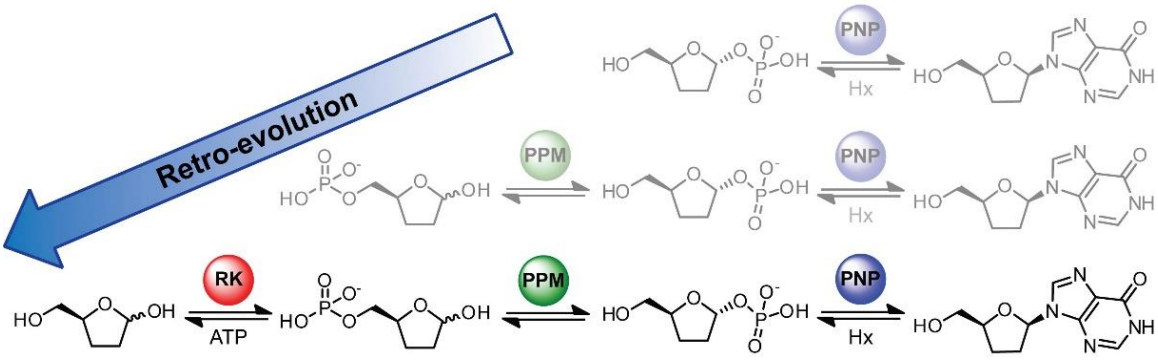

Figure 4-1. Retro-extension of the dideoxyinosine biosynthetic pathway to a kinase enzyme capable of phosphorylating dideoxyribose.

the non-natural sugar substrates needed for biosynthesis of dideoxyinosine, as dideoxyribose is much more easily tractable than the phosphorylated sugar analog.

Extending the biosynthetic pathway to include a kinase enzyme also added an additional complexity of requiring a necessary cofactor (ATP) to perform the reaction. Biocatalytic processes that consume stoichiometric amounts of biological cofactors (such as NAD(P)H, $NAD(P)^+$, ATP, etc) in order to catalyze the desired transformation are traditionally coupled to an enzymatic cofactor regeneration cycle. These cofactors are often more expensive than the desired product itself, so efficient strategies to recycle the used material are required for practical and cost effective application in industrial scale synthesis, usually by coupling the regeneration to consumption of inexpensive secondary substrates[7]. For example, to regenerate reducing equivalents of NAD(P)H, the desired reduction reaction can be coupled to the oxidation of formate[8], glucose[9] or isopropanol[10] by the respective dehydrogenase enzymes (Figure 4-2a). The oxidized form $NAD(P)^+$ can similarly be restored by reducing 2-ketoglutarate by glutamate dehydrogenase in the presence of ammonia[11] or by NADH oxidase catalyzed reduction of O_2[12] (Figure 4-2b). ATP can also be regenerated from ADP in several methods, using pyruvate kinase and the high energy phosphate donor phosphoenolpyruvate[13] or through a phosphate transfer from polyphosphate by polyphosphate kinase[14]. Additionally, in reactions where pyrophosphate is transferred from ATP to form AMP, adenylate kinase can be coupled to the activity of pyruvate kinase or polyphosphate kinase to convert AMP to ADP then then reform the active ATP cofactor[15] (Figure 4-2c).

In this work, we have identified a variety of kinase enzymes that fit the criteria above for potential use as the progenitor enzyme in engineering dideoxyribose kinase activity. To allow the greatest chance of success, the panel of enzymes selected possessed a variety of natural substrate specificities and were also from a diverse group of species. Of the ten plasmids received from other research groups and one from

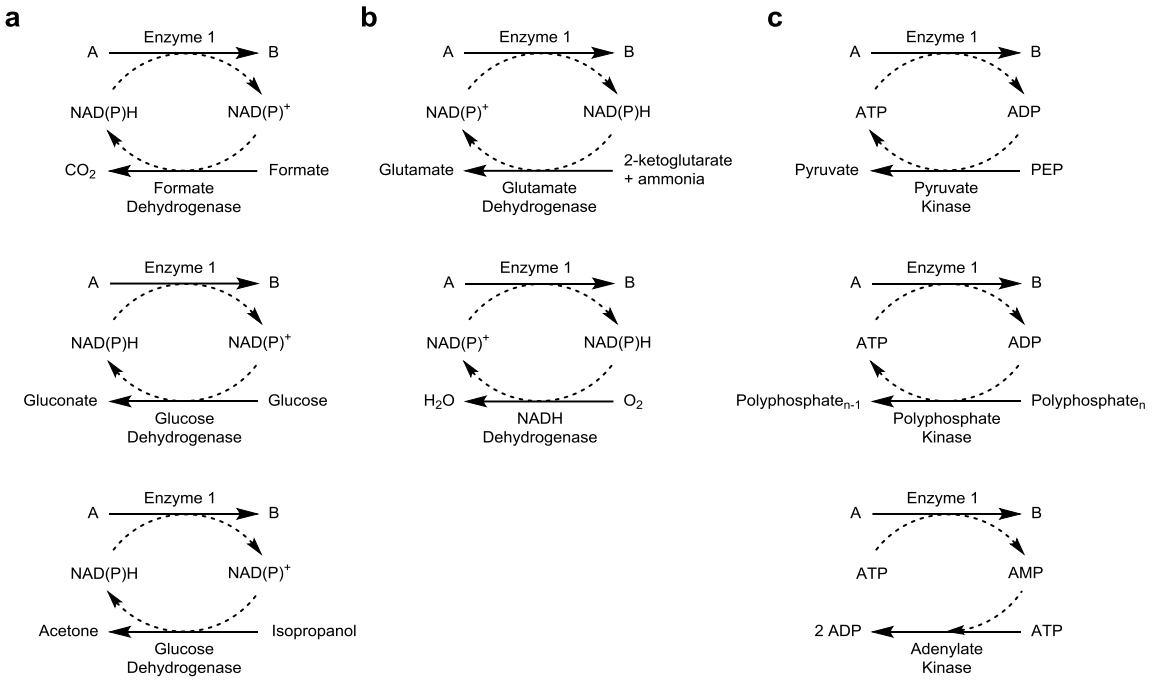

Figure 4-2. Examples of cofactor regeneration methods. In each instance, restoration of the required cofactor is provided by consumption of an inexpensive substrate through the coupled activity of another enzyme. Adapted from Xue and Woodley[17].

previous studies in our lab, five produced soluble His-tagged protein to be assayed for activity *in vitro*. Each were tested in the full pathway in tandem with the engineered PPM-4H11 and hPNP-46D6[16], measuring production of dideoxyinosine from dideoxyribose to identify the most productive variant to use as the template enzyme in rational and random mutagenesis studies to engineer an efficient dideoxyribokinase. Additionally, to counteract the stoichiometric requirement of ATP for phosphorylation of dideoxyribose by the kinase enzyme, the biosynthetic pathway was further extended to include an enzymatic ATP regeneration cycle, bringing the full system to a five enzyme pathway.

Methods

Synthesis of 2,3-dideoxyribose

(S)-γ-butyrolactone-γ-carboxylic acid (**2**)**.** Compound **2** was generated from L-glutamic acid **1** as generally described[18] with the following modifications. Briefly, a stirred solution of 53.3 g (363.6 mmol) L-glutamic acid suspended in 170 mL water was fixed with separatory funnels containing solutions of $NaNO_2$ (37.5 g, 545 mmol, 1.5 eq, in 100 mL water) and HCl (90.6 mL of 5.6 N, 508.6 mmol, 1.4 eq), which were added drop-wise simultaneously over a period 3 - 4 h. Reaction temperature was maintained between 15 - 20 °C during addition using an ice water bath. After complete addition, the reaction was warmed to room temperature and stirred overnight before removal of water by rotary evaporation and azeotropic dehydration using toluene (3 x 75 mL). The resulting solid was resuspended in 500 mL ethyl acetate and dried over anhydrous $NaSO_4$. After removing precipitates via filtration, AG50W-X4 resin (20 g, activated by washing successively with methanol and ethyl acetate) was added to the solution and stirred for 30 min to remove unreacted **1**. After filtration, solvent was removed by evaporation and the product recrystallized from dichloromethane at -20 °C for (18.2 g, 139.9 mmol 39% yield). Product purity was confirmed by ^{1}H, ^{13}C NMR analysis and comparison to reported spectra.

(S)-γ-(hydroxymethyl)-γ-butyrolactone (**3**)**.** Butyrolactone **3** was produced via a procedure adapted from a previously described synthesis[18]. To a solution of 4.95 g **2** (38.1 mmol) in 25 mL anhydrous THF at 0 °C under Argon was added dropwise 4.38 mL BH_3SMe_2, 10 M in THF, (43.75 mmol, 1.15 eq) over 1 h. The solution was warmed to room temperature and stirred for an additional 2 h. Three dropwise additions of 2.25 mL methanol were performed to quench the reaction and remove unreacted BH_3SMe_2 prior to removal of solvents by rotary evaporation. Product **3** (3.8g, 32.7 mmol 86 % yield)

was sufficiently pure to proceed without further purification. Product purity was confirmed by ^{1}H, ^{13}C NMR analysis and comparison to reported spectra.

(S)-2,3-dideoxyribose (**7**)**.** To a stirred solution of compound **3** (106 mg, 0.9 mmol) in 10.6 mL anhydrous dichloromethane at -78°C was added dropwise diisobutylaluminium hydride (1.22 mL, 1.5 M in toluene, 2 eq) and stirred for 2 hr. Methanol (5 mL) was added to quench the reaction and the solution was warmed to room temperature. The reaction was directly purified using flash SiO_2 chromatography (10% methanol in dichloromethane), yielding 95 mg (88%) compound **7** as an oil. TLC (methanol:dichloromethane, 10/90 v/v): R_f=0.25. ^{1}H NMR (400 MHz, D_2O, mixture of α and β anomers of furanose and pyranose forms): δ 5.56 - 4.92 (m, 1H) 4.33 - 3.97 (m, 1H), 3.82 - 3.37 (m, 2H), 2.17 - 1.55 (m, 4H). ^{13}C NMR (100 MHz, D_2O, mixture of α and β anomers of furanose and pyranose forms): δ 97.93, 97.55, 93.11, 92.89, 80.30, 78.35, 66.85, 66.27, 64.51, 63.95, 63.42, 63.07, 32.54, 32.00, 27.44, 26.99, 26.55, 26.47, 24.20, 24.12. HRMS (*m/z*): $[M+Na]^+$ calculated for $C_5H_{10}NaO_3^+$ 141.0522; found 141.0528.

Enzyme expression and purification

Plasmids containing wild-type or variant PPM, PNP, hPNP-46D6[16], wild-type RK, adenylate kinase or pyruvate kinase[19] were transformed into *E. coli* BL21(*DE3*) and grown at 37°C in LB medium supplemented with 50 µg/mL kanamycin or 50 µg/mL streptomycin (RK) and induced with 1 mM IPTG for 3 - 6 h after OD_{600} had reached 0.5 - 0.6. Plasmids containing *Staphylococcus aureus* ribokinase[20], *Bacillus subtilis* fructokinase[21], *Enterococcus casseliflavus* glycerol kinase[22] and *Bacillus subtilis* hydroxyethylthiazole kinase[23] were transformed into *E. coli* BL21(*DE3*) and grown at 37°C in LB or 2xYT medium (hydroxyethylthiazole kinase only) supplemented with 50

µg/mL ampicillin and were induced with 0.5 - 2 mM IPTG for 20 hr at 16°C using the published protocols. *E. casseliflavus* glycerol kinase and *B. subtilis* hydroxyethylthiazole kinase were also cotransformed with pREP4 plasmid and media therefore contained 25 µg/mL kanamycin. Cell pellets were resuspended in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM Imidazole, pH 7.4) and disrupted by passing through a French Pressure cell. The clarified lysate was applied to HisTrap FF crude Ni-NTA affinity column (GE Healthcare) and washed at 10% Buffer B (Buffer A with 500mM imidazole). Protein was eluted by a linear gradient from 10% Buffer B to 60% Buffer B, before a step up to 100% Buffer B to fully elute the column. Purified enzymes were concentrated, desalted and exchanged into 5 mM $MgCl_2$ 25 mM Tris-HCl, pH 8 before storage at -80°C. Hydroxyethylthiazole kinase purification buffers also contained 2mM β-mercaptoethanol and both hydroxyethylthiazole kinase and fructokinase enzymes were exchanged into Tris-HCl and $MgCl_2$ buffer containing 2mM DTT before storage. All enzyme concentrations were determined using the BCA Protein Assay Kit (Thermo Scientific).

Characterization of dideoxyribose activity of kinase enzymes

Dideoxyinosine production by wild-type *E. coli* ribokinase[19], *S. aureus* ribokinase[20], *B. subtilis* fructokinase[21], *E. casseliflavus* glycerol kinase[22] or *B. subtilis* hydroxyethylthiazole kinase[23] was screened to determine the most active enzyme. Reactions contained 0.6 mM $MnCl_2$, 0.6 mM $MgCl_2$, 30 mM KCl, 10 µM hPNP-46D6, 10 µM 4H11 PPM, 100 µM above kinase, 5 µM adenylate kinase, 5 µM pyruvate kinase, 15 µM glucose 1,6-bisphosphate, 3 mM hypoxanthine, 1 mM ATP, 2 mM phosphoenolpyruvate and 1 mM dideoxyribose in 25 mM Tris, pH 8. Reactions were initiated by addition of dideoxyribose and incubated 22 h at room temperature. Aliquots of 250 µL were quenched by loading onto Oasis® HLB (3 mL, 60 mg) solid-phase extraction cartridges (Waters) preconditioned with 3 mL methanol and 3 mL water on a

vacuum manifold. Loaded cartridges were then washed with 1 mL water and nucleosides were eluted with 1.5 mL methanol[24]. The methanol fractions were evaporated to dryness and the remaining residue was reconstituted in 50 µL water. A 30 µL aliquot of each sample was combined with 7.5 µL of 50 µM 2-deoxyguanosine internal standard to prepare the sample for LC/MS analysis.

Dideoxyinosine produced in the assay was separated from other reaction components using a Luna Phenyl-Hexyl column (4.6 X 250 mm, Phenomenex) and an isocratic flow of 1.0 mL/min of 10 mM ammonium acetate in 95% water:5% acetonitrile, pH 6. A Thermopal autosampler was used to inject 10 µL of the sample for analysis. Nucleosides were analyzed on a TSQ Quantum Access triple quadrupole electrospray ionization-LC/MS (Thermo, Inc.) using selected reaction monitoring fragmentation to the free nucleobase (dideoxyinosine $[M+H]^+$ 237 m/z transition to hypoxanthine $[M+H]^+$ 137 m/z) with 2-deoxyguanosine as the internal standard ($[M+H]^+$ 268 m/z to guanine $[M+H]^+$ 152 m/z). Nitrogen was used for both the auxiliary and sheath gases and was set to 45 units and 30 units, respectively. The following instrument parameters were used: source voltage 4.5kV; vaporizer temperature 0 °C; capillary temperature 270 °C; tube lens 101 V; skimmer offset -5 V; collision energy -10 V. Data acquisition and analysis were conducted with Thermo Xcalibur software, version 2.1. Dideoxyinosine formation was quantified by relative peak area of analyte to internal standard in comparison to a standard curve made using authentic inosine (Acros Organics) and dideoxyinosine (3B Pharmachem (Wuhan) International Co. Ltd.). Retention times were approximately 5 min for inosine, 6.2 min for 2-deoxyguanosine and 13.5 min for dideoxyinosine. Turnover in the reaction mix without the sugar substrate incubated for the same time was subtracted from the final turnover measurements.

Inhibition of PPM by ATP, ADP and AMP

The extent of inhibition of wild-type PPM by ATP, ADP and AMP was measured in tandem assays with wild-type PNP. All reactions were performed in 100 µL on 96-well plates. Assays contained 0.1 mM $MnCl_2$, 10 nM PPM, 5 µM PNP, 0.5 µM glucose 1,6-bisphosphate, 600 µM hypoxanthine, 500 µM ribose 5-phosphate and 0 - 0.5 mM ATP, 0 - 0.5 mM ADP or 0 - 5 mM AMP in 25 mM Tris-HCl, pH 8. Reactions were initiated by addition of 5 µL ribose 5-phosphate to 95 µL containing all other components and were incubated for 20 min at room temperature before being quenched by addition of 5 µL 2 M NaOH. After 30 min, 5 µL of 2 M HCl was added to neutralize the mixture and 75 µL of the quenched assay was combined with 20 µL of 0.2% Triton X-100, 7.5 mM iodonitrotetrazolium chloride and xanthine oxidase in 25 mM Tris-HCl, pH 8. Hypoxanthine consumption was determined by the absorbance at 546 nm in comparison to a standard curve and normalized to turnover measured in the absence of adenosine nucleotide.

Results

Chemical synthesis of the non-natural sugar 2,3-dideoxyribose

The synthesis of (*S*)-γ-hydroxymethyl-γ-butyrolactone (**3**) from L-glutamic acid (**1**), described in Chapter II, creates to a branching point in the synthesis of 2,3-dideoxysugar substrates. One path can be followed to produce 2,3-dideoxyribose 5-phosphate, as previously described in Chapter II, and the other can lead to 2,3-dideoxyribose (**7,** Figure 4-3). A simple DIBAL reduction of the lactone (**3**) to the lactol[25] provides the dideoxyribose substrate. This reaction was sufficient to provide dideoxyribose in high enough yield for biochemical characterization of kinase progenitors in the *in vitro* non-natural biosynthetic pathway.

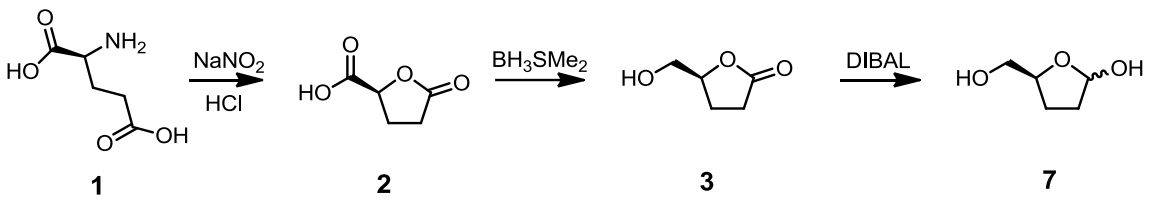

Figure 4-3. Synthesis of 2,3-dideoxyribose from glutamic acid.

Identification of potential kinase progenitors

An understanding of enzyme structure, active site design and substrate recognition are necessary for implementing a targeted mutagenesis approach for altering enzyme substrate specificity. Considering these elements, we searched the Protein Data Bank[(26)] for available structures of potential kinase progenitors. To limit the search, we focused on sugar and other small molecule alcohol kinases. These enzymes were more likely to have chemically similar and/or relatively comparable sized substrates to dideoxyribose, which may eventually provide a more beneficial foundation to evolve the desired non-natural activity[(27)]. Although not utilized in this work, accessing the structural and biochemical information available through the Enzyme Function Initiative (EFI)[(28)] is another potential option for choosing enzymes as part of the progenitor identification process. The Cores within the EFI work to provide structural and functional biochemical characterization of unknown enzymes discovered in genome sequencing projects and therefore could also be a valuable resource for this initial step.

Based on available published structures, we narrowed the list of progenitors to several enzyme classes with overall fairly extensive structural characterization, namely ribokinase, glycerol kinase, fructokinase, ketohexokinase, xylulokinase and hydroxyethylthiazole kinase. Fortunately, many of the structures were deposited as cocrystals with natural substrates and/or substrate analogs bound in the active sites, which provided additional information about residues involved in substrate binding and

orientation. Apparent substrate interacting residues are highlighted in Figure 4-4 in a representative costructure, as well as the phosphorylation reaction specific to each enzyme class. Having identified these six enzyme classes with sufficient structural data, we then expanded the search using the Braunschweig Enzyme Database (BRENDA)[29] to include genes that had been cloned into expression vectors but not crystallized in order to create a more diverse list of homologous enzymes to screen for phosphorylation of dideoxyribose. The total list included 27 unique enzyme variants from the classes given above and is provided in Table 4-1.

Of the 27 plasmids requested from other laboratories, 10 were provided for our test panel for dideoxyribose phosphorylation activity along with *E. coli* ribokinase which we had on hand from previous experiments[19]. However, several of these 10 were not able to be tested due to insoluble expression, low expression or absence of a His-tag sequence to facilitate purification. The final list of enzymes screened for activity were *E. coli* ribokinase[19], *S. aureus* ribokinase[20], *B. subtilis* fructokinase[21], *E. casseliflavus* glycerol kinase[22] and *B. subtilis* hydroxyethylthiazole kinase[23].

Figure 4-4. (Next page) Phosphorylation reaction and substrate binding interactions in kinase enzymes from the enzyme classes identified as potential progenitors for evolving dideoxyribokinase activity. Apparent interactions between the substrate and enzyme are indicated with dashed lines. All indicated distances are 4.0 Å or less. (**a**) *E. coli* ribokinase with ribose and β,γ-methyleneadenosine 5′-triphosphate (ADPCP) bound (PDB entry 1GQT[30]). (**b**) *E. casseliflavus* glycerol kinase with glycerol bound (PDB entry 1XUP[31]). (**c**) *B. subtilis* fructokinase with fructose and ADP bound (PDB entry 3LM9[21]) (**d**) Human ketohexokinase isoform C with fructose and adenosine 5′-(β,γ-imido)triphosphate (AMP-PNP) bound (PDB entry 3NBV[32]) (**e**) Human xylulokinase with xylose and ADP bound (PDB entry 4BC2[33]). (**f**) *B. subtilis* hydroxyethylthiazole kinase with 4-methyl-5-β-hydroxyethylthiazole bound (PDB entry 1C3Q[23]). The active site of this enzyme is at the interface of two subunits which are distinguished separately in gray and blue.

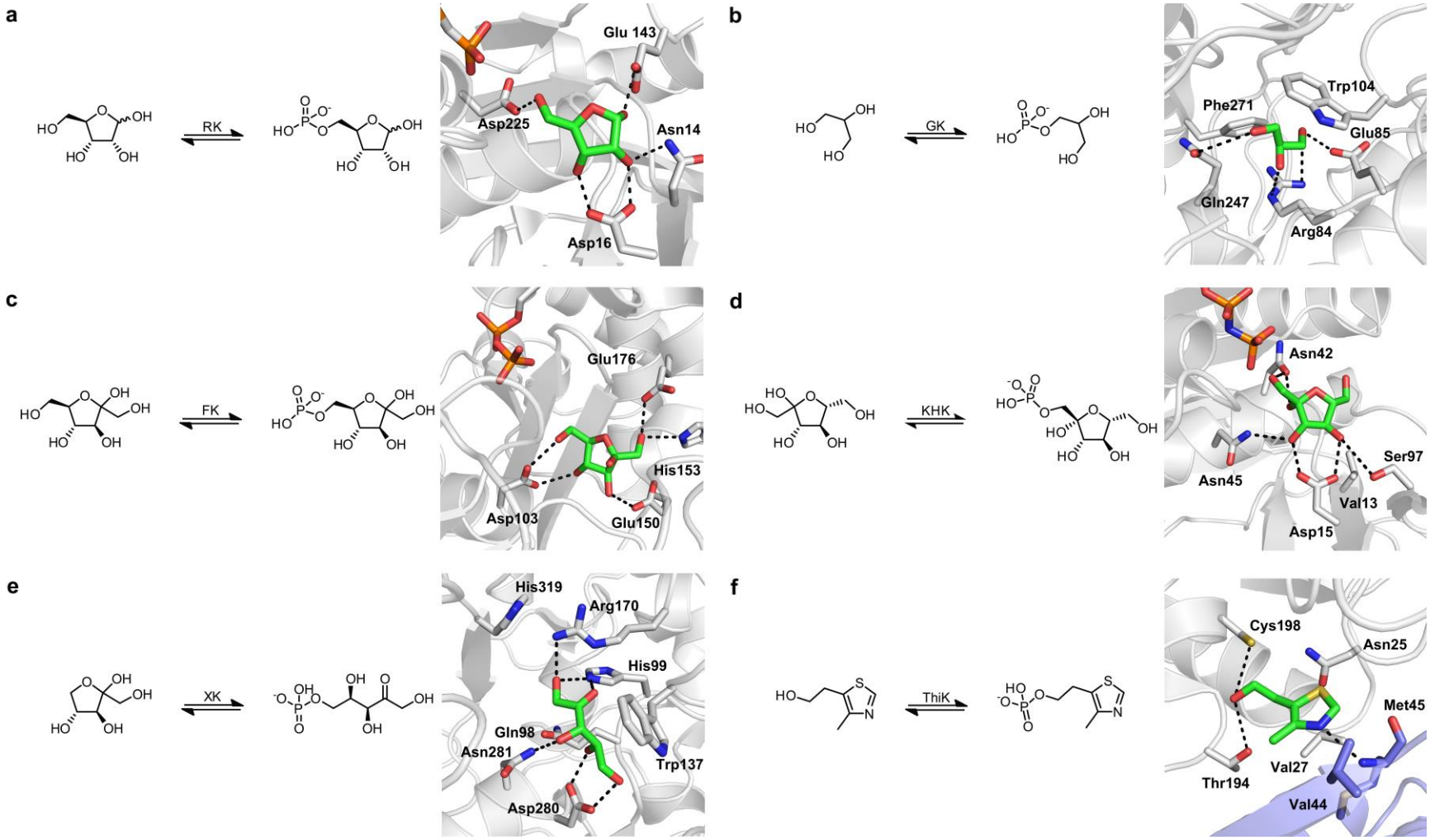
a
RK
Glu 143
Asp225
Asn14
Asp16
b
GK
Trp104
Phe271
Glu85
Gln247
Arg84
c
FK
Glu176
His153
Asp103
Glu150
d
KHK
Asn42
Ser97
Asn45
Val13
Asp15
e
XK
His319
Arg170
His99
Gln98
Asn281
Trp137
Asp280
f
ThiK
Cys198
Asn25
Met45
Val27
Thr194
Val44

Table 4-1. List of potential kinase progenitors found through searching the Protein Data Bank and BRENDA. Expression plasmids containing genes for all 27 enzymes listed were requested from the corresponding author of the respective publications. Ten plasmids were received, each denoted with an asterisk (*). ‡ denotes that a structure had been posted to the PDB.

Enzyme	Organism	Ref.
Ribokinase	*‡*E. coli*	(19)
	‡*H. orenii*	(34)
	*‡*S. aureus*	(35)
	‡Human	(36)
	Leishmania major	(37)
Deoxyribokinase (Putative)	*S. enterica serovar typhi*	(38)
Glycerol Kinase	‡*Thermococcus kodakaraensis*	(39)
	*‡*Plasmodium falciparum*	(40)
	*‡*Enterococcus casseliflavus*	(22)
	H. influenza	(41)
	T. brucei	(42)
	Thermus flavus	(43)
	Thermus aquaticus	(44)
Fructokinase	*‡*B. subtilis*	(21)
	‡*H. orenii*	(45)
	L. pseudomesenteroides	(46)
	Thermococcus litoralis	(47)
	Lycopersicon esculentum Mill.	(48)
Ketohexokinase	Human (isoforms *‡A and *‡C)	(49, 50)
Xylulokinase	**Kluyveromyces marxianus* NBRC1777	(51)
	‡Human	(52)
	Pichia stipitis	(53)
	**S. cerevisiae*	(54)
	Arabinopsis thaliana	(55)
	**Hansenula polymorpha*	(56)
Hydroxyethylthiazole Kinase	*‡*B. subtilis*	(23)

Screening progenitor enzymes for dideoxyribokinase activity

Each of the five enzymes listed above were assayed for *in vitro* production of dideoxyinosine from dideoxyribose. Activity was tested in the presence of the evolved PPM-4H11 and hPNP-46D6[16] variants in a tandem system and analyzed by HPLC/MS. Determining activity in tandem with these two previously engineered enzymes was found to be the most sensitive detection of kinase activity on dideoxyribose for this system, more so even than a typical ATP consumption assay[57] following the oxidation of NADH to NAD^+ at 340 nm in a tandem assay with pyruvate kinase and lactate dehydrogenase (data not shown). Assays were incubated for 22 h to allow for any small differences in activity to be more apparent in the measured turnover.

A low level of dideoxyinosine formation was detected in all five pathway series (Figure 4-5). Production was clearly highest in the two pathways that contained a ribokinase homolog, reaching 2.94 ± 0.05 µM with *E. coli* ribokinase and 2.76 ± 0.12 µM

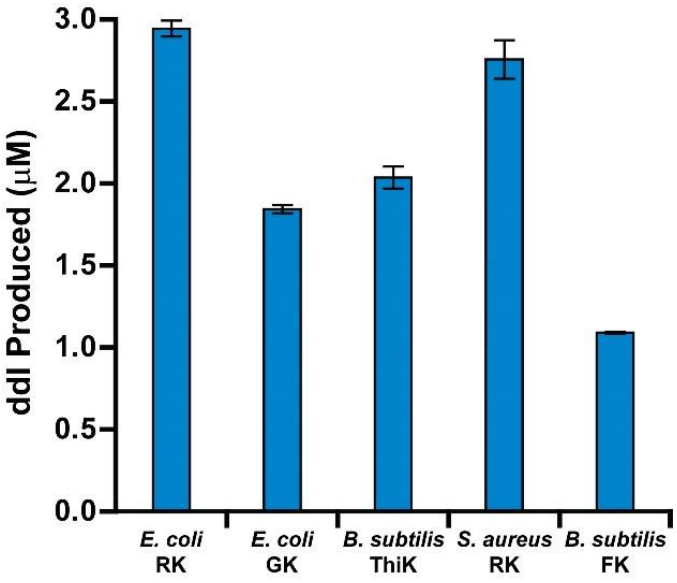

Figure 4-5. Production of didanosine from dideoxyribose by potential kinase progenitors. Kinases listed were tested in tandem with 4H11 PPM and hPNP-46D6[16] to identify the highest producing variant. Reactions were incubated at room temperature for 22 h. (ddI, dideoxyinosine; RK, ribokinase; GK, glycerol kinase; ThiK, hydroxyethylthiazole kinase; FK, fructokinase).

with the *S. aureus* enzyme. Although the two ribokinase enzymes showed comparable activity on dideoxyribose through similar production levels of dideoxyinosine, the ribokinase variant from *E. coli* was chosen to be the progenitor enzyme further evolution and use in the biosynthetic pathway for dideoxyinosine. The extensive biochemical and structural characterization available for this enzyme was thought to provide a firm foundation of functional insight that could be used to guide future mutagenesis studies.

ATP regeneration cycle

In constructing the proposed biosynthetic pathway, the retro-extension to ribokinase brought the necessary consideration of an ATP regeneration method. In addition to avoiding the costly requirement of stoichiometric levels of ATP for phosphorylation of dideoxyribose, ATP was also found to inhibit PPM, as did ADP and AMP, the products of enzymatic reactions and hydrolysis (Figure 4-6). For these

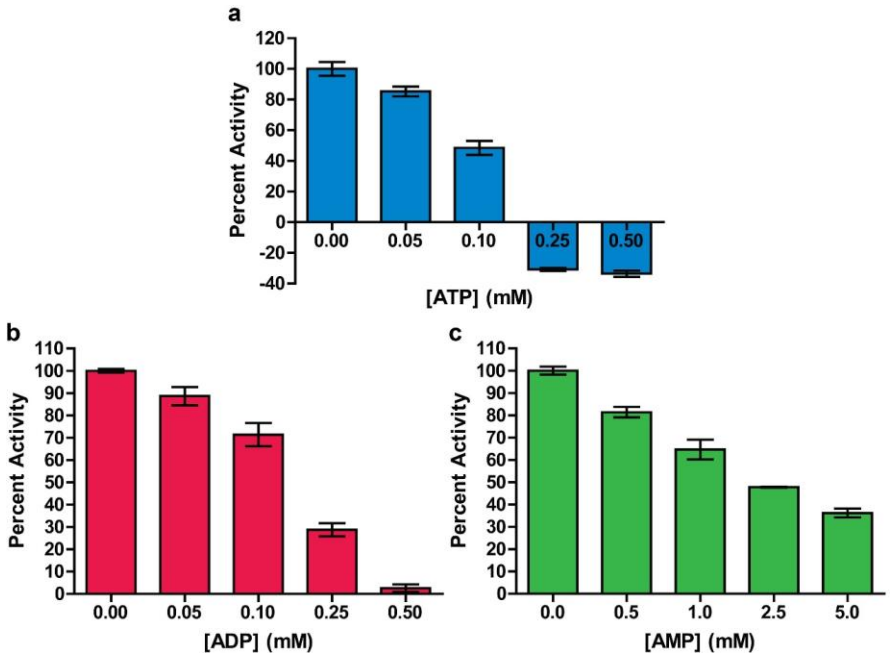

Figure 4-6. Inhibition of PPM by adenine nucleotides. Consumption of hypoxanthine by 10 nM PPM in tandem with PNP was determined in the presence of a range of (**a**) ATP, (**b**) ADP and (**c**) AMP concentrations to observe the effect on activity.

reasons, the concentration of ATP in the reaction must be kept relatively low and levels of inactive ADP and AMP nucleotides must also be minimized to prevent enzyme inhibition that would ultimately reduce biosynthetic pathway productivity. To address this need, a tandem enzymatic system of pyruvate kinase and adenylate kinase were introduced to the *in vitro* reaction along with phosphoenolpyruvate as the source of high energy phosphate. Together, these two enzymes recycle any ADP and AMP produced as a result of kinase activity and/or hydrolysis over the course of the long incubation time. Using this *in situ* ATP regeneration system, we reduced the concentration of ATP to sub-stoichiometric levels and compensated the loss by including phosphoenolpyruvate at levels high enough to maintain a steady active ATP concentration throughout the extended assay lengths. Including this cofactor regeneration system, the full system now consists of a five enzyme biosynthetic pathway (Figure 4-7).

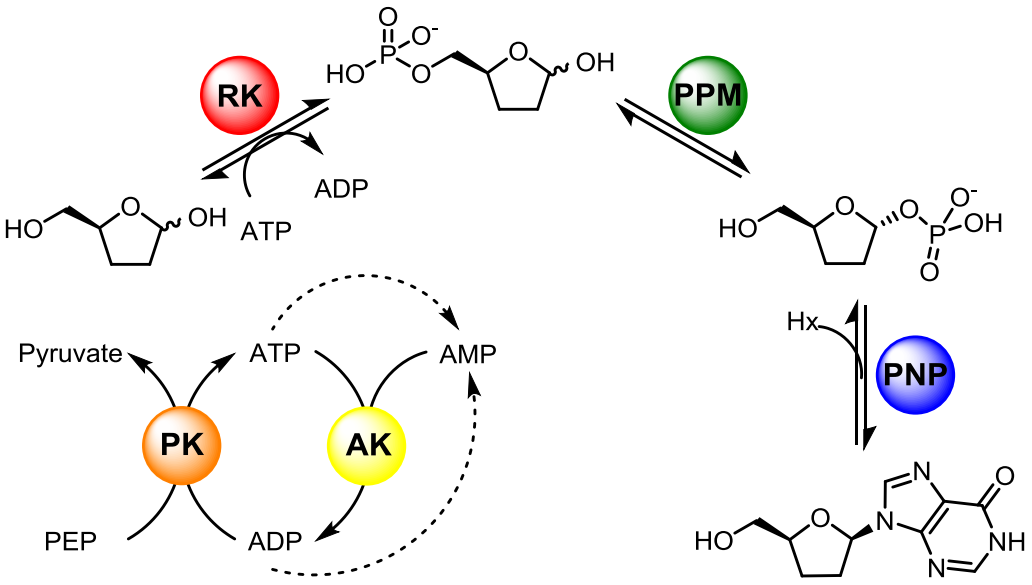

Figure 4-7. The five step dideoxyinosine biosynthetic pathway. Pyruvate kinase (PK) and adenylate kinase (AK) form the ATP regeneration cycle, while biosynthesis of ddI continues from dideoxyribose via RK, PPM and PNP catalysis. Expected products of degradation through non-enzymatic hydrolysis are indicated with dashed arrows.

Discussion

Identifying a proper progenitor enzyme is a critical aspect of engineering an effective biocatalyst. Ideally, a progenitor enzyme that naturally acts on a substrate as similar as possible to the desired non-natural substrate would provide the best place to begin[27]. For this reason, we set out to select a panel of enzyme classes that could be useful starting templates for evolving dideoxyribose phosphorylation activity. Aside from the chemical nature of the substrate, we considered the desired transformation (i.e. phosphorylation of a hydroxyl group) and the availability of structural data indicating active site residues involved in enzyme-substrate interactions as part of the analysis. This process led us to the ribokinase, glycerol kinase, fructokinase, ketohexokinase, xylulokinase and hydroxyethylthiazole kinase enzyme classes as potential progenitors (Figure 4-4).

After requesting all of the homologs from each class that were published as being cloned into expression vectors (Table 4-1), the list of progenitors was considerably shortened to five enzymes that were able to be expressed at a high level and purified by His-tag affinity chromatography. Each enzyme was individually tested in tandem with the evolved PPM and PNP variants in the presence of an ATP regeneration system. The assay measuring total production of dideoxyinosine from dideoxyribose by each kinase enzyme identified the two ribokinase variants (*E. coli* and *S. aureus*) as the most productive progenitors.

E. coli ribokinase has been reported to show low level activity on a variety of other naturally occurring D-pentose sugars. In their assay conditions, Chuvikovsky *et al.* detected activity on 2-deoxy-D-ribose, D-arabinose, D-xylose and D-fructose at 31, 0.74, 1.06 and 0.28% of the measured activity on ribose by the wild-type enzyme[58]. This noted promiscuous activity may prove to be a valuable quality in future studies to engineer activity. The *S. aureus* enzyme, on the other hand, has not been characterized

as having additional substrate allowance, however this enzyme in general has not been studied in great detail. Combined with the highest level of activity on dideoxyribose, the greater extent of biochemical and structural characterization of this enzyme lead to the selection of ribokinase from *E. coli* as the template for future evolution.

The chemical similarity between dideoxyribose and ribose (and dissimilarity between dideoxyribose and the natural substrates of the other enzymes) could perhaps explain why the ribokinase homologs demonstrated the highest activity of the kinase enzymes tested. The two substrates differ only by the presence or absence of the C2 and C3 hydroxyls, which appear to interact most directly with the active site residue Asp16 in *E. coli*[59] (Figure 4-4a). Although this interaction would likely be unfavorable toward dideoxyribose, perhaps dideoxyribose binds in a slightly altered orientation to minimize the interaction between the hydrophobic sugar and the negatively charged Asp16 residue.

Dideoxyribose possesses unique structural features when compared to the other natural substrates of each enzyme class. The non-natural sugar is somewhat larger than glycerol and not as planar as 4-methyl-5-β-hydroxyethylthiazole. The reduced activity on dideoxyribose observed in the enzymes responsible for phosphorylating these substrates, glycerol kinase and hydroxyethylthiazole kinase, respectively, could simply be the result of stearic hindrance in the active site. This would result in limited substrate binding due to residues preventing access to the binding pockets in the respective kinases. Mutagenesis targeted to specific active site residues with the preference for smaller amino acid side chains, such as glycine or alanine, may be able to open the active site and relieve some of the potentially unfavorable steric interactions and increase catalysis, if indeed this is the reason for the observed low activity. This strategy has previously been demonstrated to be quite successful in engineering activity into a transaminase where the active site was originally much too small to bind the ketone

substrate of interest, and targeted engineering to expand the active site was the first priority[60].

To contrast, the fructose sugar substrate of fructokinase is both larger and also more polar than the ribose substrate for ribokinase. Because of these traits, the substrate binding pocket of fructokinase has evolved accordingly to appropriately match the size and increase number of polar functional groups to accommodate fructose binding. These structural and electrostatic differences likely allow the smaller and much less polar dideoxyribose to bind in multiple different non-productive orientations in the active site. This would reduce the likelihood of acceptable binding and positioning of dideoxyribose for proper catalysis, causing fructokinase be rather inefficient in phosphorylating dideoxyribose as observed in this study.

Although not tested here, ketohexokinase could be an interesting progenitor to characterize for dideoxyribose kinase activity. This enzyme may also have minimal activity on dideoxyribose, comparable to that of fructokinase, as fructose is the natural substrate for this enzyme as well. On the other hand, ketohexokinase could possibly have a similar activity profile of the two ribokinase enzymes due to structural similarity between the three enzymes. Ketohexokinase isoforms A and C are both part of the ribokinase-like superfamily of enzymes[50], the same classification of both ribokinase enzymes tested in this study, while the fructokinase from *B. subtilis* is a member of the actin-like ATPase superfamily[21] along with the related enzymes glucokinase[61] and ATP-glucomannokinse[62]. It is possible that this ribokinase-like superfamily and the conserved active site architecture offers a more favorable scaffold for catalyzing the phosphorylation of dideoxyribose, and ketohexokinase may indeed provide another active template for evolution. It is also noteworthy that not all fructokinase enzymes belong to the actin-like ATPase superfamily. The structures of homologs from *Halothermothrix orenii* (PDB entry 3HJ6[45]) and *Xyella fastidiosa* (PDB entry 3LJS) are

classified as having a ribokinase-like fold and, if made available, could also be tested for the desired activity on dideoxyribose to further expand characterization of this superfamily for dideoxyribose phosphorylation.

Conclusions

Screening a panel of representative sugar and small molecule kinases led to the selection of *E. coli* ribokinase as the progenitor enzyme for engineering dideoxyribokinase activity. Although ribokinase from *S. aureus* was equivalently active toward dideoxyribose in the pathway production experiments, the wealth of structural and biochemical data available for the *E. coli* homolog was a key factor is selecting this enzyme as the template for use in the biosynthetic pathway and for further optimization in targeted and random mutagenesis studies.

Even though *E. coli* ribokinase was identified as the best wild-type enzyme from those tested, further studies investigating the effect of targeted mutagenesis of active site residues in select enzymes of the tested panel or of additional kinase enzymes may provide more beneficial kinase variants to use as template genes. Carefully targeted mutations within the substrate binding pocket of either glycerol kinase or hydroxyethylthiazole kinase may provide a necessary increase in active site volume to appropriately bind dideoxyribose and allow catalysis on the non-natural substrate. Additionally, mutagenesis of the active site Cys198 to aspartate in hydroxyethylthiazole kinase has been shown to increase natural substrate specific activity by 9-fold in the enzyme from *B. subtilis*[23], and could possibly affect the activity of this enzyme on dideoxyribose as well.

Expanding the collection of possible template enzymes to include other kinases within the ribokinase-like superfamily may also return new potential progenitors. Additionally, this exploration could yield a great deal of information regarding the

‘evolvability’ of the active site of this particular enzyme superfamily. Comparing sequence variation between enzymes in the ribokinase-like superfamily that show competent activity on dideoxyribose could suggest residues of interest for mutagenesis and new beneficial sequence combinations in an analysis similar to that performed as part of the ProSAR algorithm used to evolve a halohydrin dehalogenase for a non-natural substrate[63]. Progenitor expansion and characterization in this way could increase the aspect of rational design in this approach through providing a deeper biochemical understanding of the active site of ribokinase-like enzymes, and may ultimately lead to a more efficient dideoxyribokinase enzyme.

Acknowledgements

This work was contributed to by William R. Birmingham, David P. Nannemann and Brian O. Bachmann. Bioretrosynthetic evolution of a didanosine biosynthetic pathway. W.R.B. identified kinase progenitors, expressed, purified and screened the kinases for dideoxyribose phosphorylation. D.P.N. established initial synthesis routes of dideoxyribose. B.O.B. supervised biochemical experiments. We thank V. Phelan for assistance with establishing the mass spectrometry based analysis methods and to C. Goodwin for high resolution mass spectrometry data collection.

References

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

2. O'Loughlin, T. L., Patrick, W. M., and Matsumura, I. Natural history as a predictor of protein evolvability. *Protein Eng. Des. Sel.* **2006**, 19, 439-442.

3. Bernhardt, P., McCoy, E., and O'Connor, S. E. Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **2007**, 14, 888-897.

4. DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z. L., Tomlinson, G., Huang, H. J., Tan, X. Q., Bibbs, L., Chen, P., Kretz, K., and Burk, M. J. Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.* **2003**, 125, 11476-11477.

5. Jaeckel, C., Kast, P., and Hilvert, D. Protein design by directed evolution, In *Annu. Rev. Biophys.* **2008**, 37, pp 153-173.

6. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. X., and Vogel, A. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* **2005**, 44, 4192-4196.

7. Liu, W., and Wang, P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* **2007**, 25, 369-384.

8. Bommarius, A. S., Schwarm, M., and Drauz, K. Biocatalysis to amino acid-based chiral pharmaceuticals - Examples and perspectives. *J. Mol. Catal., B Enzym.* **1998**, 5, 1-11.

9. Patel, R. N. Enzymatic synthesis of chiral intermediates for drug development. *Adv. Synth. Catal.* **2001**, 343, 527-546.

10. Seisser, B., Lavandera, I., Faber, K., Spelberg, J. H. L., and Kroutil, W. Stereo-complementary two-step cascades using a two-enzyme system leading to enantiopure epoxides. *Adv. Synth. Catal.* **2007**, 349, 1399-1404.

11. Romisch, W., Eisenreich, W., Richter, G., and Bacher, A. Rapid one-pot synthesis of riboflavin isotopomers. *J. Org. Chem.* **2002**, 67, 8890-8894.

12. Riebel, B. R., Gibbs, P. R., Wellborn, W. B., and Bommarius, A. S. Cofactor Regeneration of both NAD+ from NADH and NADP+ from NADPH:NADH oxidase from *Lactobacillus sanfranciscensis*. *Adv. Synth. Catal.* **2003**, 345, 707-712.

13. Gross, A., Abril, O., Lewis, J. M., Geresh, S., and Whitesides, G. M. Practical synthesis of 5-phospho-D-ribosyl alpha-1-pyrophosphate (PRPP) - Enzymatic routes from ribose 5-phosphate or ribose. *J. Am. Chem. Soc.* **1983**, 105, 7428-7435.

14. Liu, Z. Y., Zhang, J. B., Chen, X., and Wang, P. G. Combined biosynthetic pathway for *de novo* production of UDP-galactose: Catalysis with multiple enzymes immobilized on agarose beads. *Chembiochem* **2002**, 3, 348-355.

15. Koeller, K. M., and Wong, C. H. Synthesis of complex carbohydrates and glycoconjugates: Enzyme-based and programmable one-pot strategies. *Chem. Rev.* **2000**, 100, 4465-4493.

16. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

17. Xue, R., and Woodley, J. M. Process technology for multi-enzymatic reaction systems. *Bioresour. Technol.* **2012**, 115, 183-195.

18. Okabe, M., Sun, R. C., Tam, S. Y. K., Todaro, L. J., and Coffen, D. L. Synthesis of the dideoxynucleosides ddC and CNT from glutamic acid, ribonolactone, and pyrimidine bases. *J. Org. Chem.* **1988**, 53, 4780-4786.

19. Scism, R. A., and Bachmann, B. O. Five-component cascade synthesis of nucleotide analogues in an engineered self-immobilized enzyme aggregate. *Chembiochem* **2010**, 11, 67-70.

20. Li, J., Wang, C., Wu, Y., Wu, M., Wang, L., Wang, Y., and Zang, J. Crystal structure of Sa239 reveals the structural basis for the activation of ribokinase by monovalent cations. *J. Struct. Biol.* **2012**, 177, 578-582.

21. Nocek, B., Stein, A. J., Jedrzejczak, R., Cuff, M. E., Li, H., Volkart, L., and Joachimiak, A. Structural studies of ROK fructokinase YdhR from *Bacillus subtilis*: Insights into substrate binding and fructose specificity. *J. Mol. Biol.* **2011**, 406, 325-342.

22. Charrier, V., Buckley, E., Parsonage, D., Galinier, A., Darbon, E., Jaquinod, M., Forest, E., Deutscher, J., and Claiborne, A. Cloning and sequencing of two Enterococcal *glpK* genes and regulation of the encoded glycerol kinases by phosphoenolpyruvate dependent, phosphotransferase system-catalyzed phosphorylation of a single histidyl residue. *J. Biol. Chem.* **1997**, 272, 14166-14174.

23. Campobasso, N., Mathews, II, Begley, T. P., and Ealick, S. E. Crystal structure of 4-methyl-5-β-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 angstrom resolution. *Biochemistry* **2000**, 39, 7868-7877.

24. Bezy, V., Morin, P., Couerbe, P., Leleu, G., and Agrofoglio, L. Simultaneous analysis of several antiretroviral nucleosides in rat-plasma by high-performance liquid chromatography with UV using acetic acid/hydroxylamine buffer - Test of this new volatile medium-pH for HPLC-ESI-MS/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2005**, 821, 132-143.

25. Winterfeldt, E. Applications of diisobutylaluminun hydride (DIBAH) and triisobutylaluminum (TIBA) as reducing agents in organic synthesis. *Synthesis* **1975**, 617-630.

26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28, 235-242.

27. Nannemann, D. P., Birmingham, W. R., Scism, R. A., and Bachmann, B. O. Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med. Chem.* **2011**, 3, 803-819.

28. Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., Dunaway-Mariano, D., Imker, H. J., Jacobson, M. P., Minor, W., Poulter, C. D., Raushel, F. M., Sali, A., Shoichet, B. K., and Sweedler, J. V. The Enzyme Function Initiative. *Biochemistry* **2011**, 50, 9950-9962.

29. Schomburg, I., Chang, A., Placzek, S., Soehngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., and Schomburg, D. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Res.* **2013**, 41, D764-D772.

30. Andersson, C. E., and Mowbray, S. L. Activation of ribokinase by monovalent cations. *J. Mol. Biol.* **2002**, 315, 409-419.

31. Yeh, J. I., Charrier, V., Paulo, J., Hou, L. H., Darbon, E., Claiborne, A., Hol, W. G. J., and Deutscher, J. Structures of enterococcal glycerol kinase in the absence and presence of glycerol: Correlation of conformation to substrate binding and a mechanism of activation by phosphorylation. *Biochemistry* **2004**, 43, 362-373.

32. Gibbs, A. C., Abad, M. C., Zhang, X., Tounge, B. A., Lewandowski, F. A., Struble, G. T., Sun, W., Sui, Z., and Kuo, L. C. Electron density guided fragment-based lead discovery of ketohexokinase inhibitors. *J. Med. Chem.* **2010**, 53, 7979-7991.

33. Bunker, R. D., Bulloch, E. M., Dickson, J. M., Loomes, K. M., and Baker, E. N. Structure and function of human xylulokinase, an enzyme with important roles in carbohydrate metabolism. *J. Biol. Chem.* **2013**, 288, 1643-1652.

34. Kori, L. D., Hofmann, A., and Patel, B. K. Expression, purification, crystallization and preliminary X-ray diffraction analysis of a ribokinase from the thermohalophile *Halothermothrix orenii*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2012**, 68, 240-243.

35. Wang, L., Wang, H., Ruan, J., Tian, C., Sun, B., and Zang, J. Cloning, purification, crystallization and preliminary crystallographic analysis of a ribokinase from *Staphylococcus aureus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2009**, 65, 574-576.

36. Park, J., van Koeverden, P., Singh, B., and Gupta, R. S. Identification and characterization of human ribokinase and comparison of its properties with *E. coli* ribokinase and human adenosine kinase. *FEBS Lett.* **2007**, 581, 3211-3216.

37. Ogbunude, P. O. J., Lamour, N., and Barrett, M. P. Molecular cloning, expression and characterization of ribokinase of *Leishmania major*. *Acta Biochim. Biophys. Sin. (Shanghai)* **2007**, 39, 462-466.

38. Tourneux, L., Bucurenci, N., Saveanu, C., Kaminski, P. A., Bouzon, M., Pistotnik, E., Namane, A., Marliere, P., Barzu, O., de la Sierra, I. L., Neuhard, J., and Gilles, A. M. Genetic and biochemical characterization of *Salmonella enterica* Serovar Typhi deoxyribokinase. *J. Bacteriol.* **2000**, 182, 869-873.

39. Koga, Y., Haruki, M., Morikawa, M., and Kanaya, S. Stabilities of chimeras of hyperthermophilic and mesophilic glycerol kinases constructed by DNA shuffling. *J. Biosci. Bioeng.* **2001**, 91, 551-556.

40. Schnick, C., Polley, S. D., Fivelman, Q. L., Ranford-Cartwright, L. C., Wilkinson, S. R., Brannigan, J. A., Wilkinson, A. J., and Baker, D. A. Structure and non-essential function of glycerol kinase in *Plasmodium falciparum* blood stages. *Mol. Microbiol.* **2009**, 71, 533-545.

41. Pawlyk, A. C., and Pettigrew, D. W. Subcloning, expression, purification, and characterization of *Haemophilus influenzae* glycerol kinase. *Protein Expr. Purif.* **2001**, 22, 52-59.

42. Kralova, I., Rigden, D. J., Opperdoes, F. R., and Michels, P. A. M. Glycerol kinase of *Trypanosoma brucei* - Cloning, molecular characterization and mutagenesis. *Eur. J. Biochem.* **2000**, 267, 2323-2333.

43. Huang, H. S., Kabashima, T., Ito, K., Yin, C. H., Nishiya, Y., Kawamura, Y., and Yoshimoto, T. Thermostable glycerol kinase from *Thermus flavus*: Cloning, sequencing, and expression of the enzyme gene. *Biochim. Biophys. Acta* **1998**, 1382, 186-190.

44. Huang, H. S., Ito, K., Yin, C. H., Kabashima, T., and Yoshimoto, T. Cloning, sequencing, high expression, and crystallization of the thermophile *Thermus aquaticus* glycerol kinase. *Biosci. Biotechnol. Biochem.* **1998**, 62, 2375-2381.

45. Chua, T. K., Seetharaman, J., Kasprzak, J. M., Ng, C., Patel, B. K. C., Love, C., Bujnicki, J. M., and Sivaraman, J. Crystal structure of a fructokinase homolog from *Halothermothrix orenii*. *J. Struct. Biol.* **2010**, 171, 397-401.

46. Helanto, M., Aarnikunnas, J., von Weymarn, N., Airaksinen, U., Palva, A., and Leisola, M. Improved mannitol production by a random mutant of *Leuconostoc pseudomesenteroides*. *J. Biotechnol.* **2005**, 116, 283-294.

47. Qu, Q., Lee, S. J., and Boos, W. Molecular and biochemical characterization of a fructose-6-phosphate-forming and ATP-dependent fructokinase of the hyperthermophilic archaeon *Thermococcus litoralis*. *Extremophiles* **2004**, 8, 301-308.

48. Martinez-Barajas, E., Krohn, B. M., Stark, D. M., and Randall, D. D. Purification and characterization of recombinant tomato fruit (*Lycopersicon esculentum* Mill.) fructokinase expressed in *Escherichia coli*. *Protein Expr. Purif.* **1997**, 11, 41-46.

49. Asipu, A., Hayward, B. E., O'Reilly, J., and Bonthron, D. T. Properties of normal and mutant recombinant human ketohexokinases and implications for the pathogenesis of essential fructosuria. *Diabetes* **2003**, 52, 2426-2432.

50. Trinh, C. H., Asipu, A., Bonthron, D. T., and Phillips, S. E. Structures of alternatively spliced isoforms of human ketohexokinase. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, 65, 201-211.

51. Wang, R., Zhang, L., Wang, D., Gao, X., and Hong, J. Identification of a xylulokinase catalyzing xylulose phosphorylation in the xylose metabolic pathway of *Kluyveromyces marxianus* NBRC1777. *J. Ind. Microbiol. Biotechnol.* **2011**, 38, 1739-1746.

52. Bunker, R. D., Dickson, J. M., Caradoc-Davies, T. T., Loomes, K. M., and Baker, E. N. Use of a repetitive seeding protocol to obtain diffraction-quality crystals of a putative human D-xylulokinase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2012**, 68, 1259-1262.

53. Jin, Y. S., Jones, S., Shi, N. Q., and Jeffries, T. W. Molecular cloning of XYL3 (D-xylulokinase) from *Pichia stipitis* and characterization of its physiological function. *Appl. Environ. Microbiol.* **2002**, 68, 1232-1239.

54. Lee, J. Y., Cheong, D. E., and Kim, G. J. A novel assay system for the measurement of transketolase activity using xylulokinase from *Saccharomyces cerevisiae*. *Biotechnol. Lett.* **2008**, 30, 899-904.

55. Hemmerlin, A., Tritsch, D., Hartmann, M., Pacaud, K., Hoeffler, J. F., van Dorsselaer, A., Rohmer, M., and Bach, T. J. A cytosolic Arabidopsis D-xylulose kinase catalyzes the phosphorylation of 1-deoxy-D-xylulose into a precursor of the plastidial isoprenoid pathway. *Plant Physiol.* **2006**, 142, 441-457.

56. Dmytruk, O. V., Dmytruk, K. V., Abbas, C. A., Voronovsky, A. Y., and Sibirny, A. A. Engineering of xylose reductase and overexpression of xylitol dehydrogenase and xylulokinase improves xylose alcoholic fermentation in the thermotolerant yeast *Hansenula polymorpha*. *Microb. Cell Fact.* **2008**, 7, 21.

57. Hope, J. N., Bell, A. W., Hermodson, M. A., and Groarke, J. M. Ribokinase from *Escherichia coli* K12: Nucleotide sequence and overexpression of the *rbsK* gene and purification of ribokinase. *J. Biol. Chem.* **1986**, 261, 7663-7668.

58. Chuvikovsky, D. V., Esipov, R. S., Skoblov, Y. S., Chupova, L. A., Muravyova, T. I., Miroshnikov, A. I., Lapinjoki, S., and Mikhailopulo, I. A. Ribokinase from *E. coli*: Expression, purification, and substrate specificity. *Bioorg. Med. Chem.* **2006**, 14, 6327-6332.

59. Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to

1.8 angstrom resolution: Insights into a new family of kinase structures. *Structure* **1998**, 6, 183-193.

60. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.

61. Hansen, T., and Schonheit, P. ATP-dependent glucokinase from the hyperthermophilic bacterium *Thermotoga maritima* represents an extremely thermophilic ROK glucokinase with high substrate specificity. *FEMS Microbiol. Lett.* **2003**, 226, 405-411.

62. Mukai, T., Kawai, S., Mori, S., Mikami, B., and Murata, K. Crystal structure of bacterial inorganic polyphosphate/ATP-glucomannokinase - Insights into kinase evolution. *J. Biol. Chem.* **2004**, 279, 50591-50600.

63. Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., and Huisman, G. W. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, 25, 338-344.

Chapter V

BIORETROSYNTHESIS AS A PATHWAY CONCEPTUALIZATION AND CONSTRUCTION METHOD

Introduction

Many societally important synthetic and natural molecules are currently generated using biocatalytic processes. In most cases, individual enzyme-catalyzed steps are optimized for incorporation into multi-step synthetic pathways, such as in the syntheses of the blockbuster drugs sitagliptin[(1)], montelukast[(2)], and simvastatin[(3)], among others[(4)]. With less frequency, secondary metabolites, their intermediates, and/or analogs are synthesized via recapitulating and improving existing biosynthetic pathways (for example, artimesinic acid[(5)], taxadiene[(6)]) or by modifying native pathways (such as, pactamycin[(7)], macbecin[(8)]). Given the apparent multiplicative benefits of combining biotransformations into pathways, it is notable that de novo multistep engineered biosynthetic pathways for the synthesis of unnatural compounds are quite uncommon (1,2,4-butanetriol[(9)], 1,4-butanediol[(10)]), and the small subset of such pathways that employ cascades of more than one engineered/evolved enzyme, as in the production of the Atorvastatin side chain[(11)], are truly exceptional.

One possible reason for this deficiency of engineered non-natural pathways may be the lack of a systematic pathway conceptualization strategy to make the challenge of *de novo* pathway design more approachable. Borrowing a technique from synthetic chemistry, the process used in the practices of chemical and biochemical retrosynthetic analyses[(12, 13)] can be adapted to the context of a biosynthetic pathway in 'bioretrosynthesis'[(14)], where an analogous series of biotransformations, rather than chemical transformations, can be proposed to convert a target non-natural product into

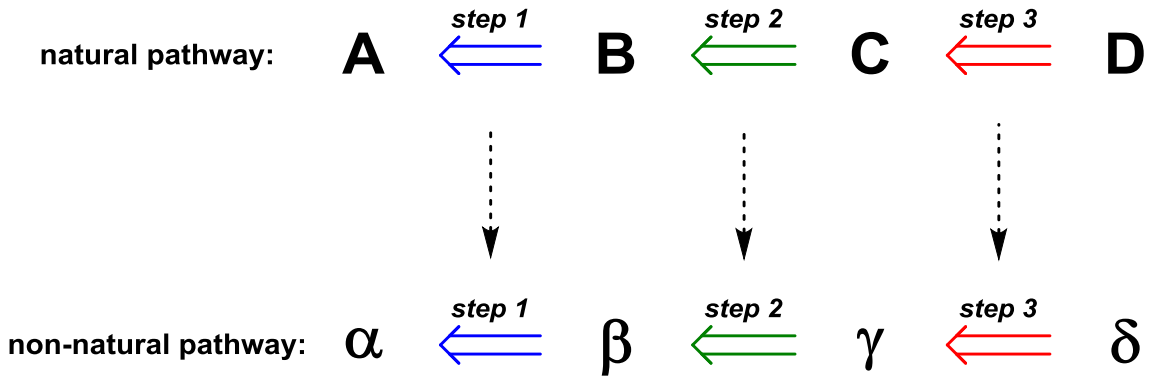

Figure 5-1. Bioretrosynthesis applied as pathway planning tool. Known enzymatic transformations of natural substrates in the biosynthesis of **D** from precursor **A** (top, Steps 1-3) are suggested to serve as a biocatalytic route to form non-natural product **δ** from precursor **α**, catalyzing an analogous series of reactions on non-natural substrates (Greek letters) (bottom, Steps 1-3).

increasingly simpler precursors (and precursors of precursors) using the body of known enzymatic reactions (Figure 5-1). This approach offers a widely applicable paradigm for planning a non-natural pathway through providing a sequence of biotransformations toward the target compound and simultaneously delineating a putative and tractable biosynthetic pathway, much like how retrosynthesis is used in planning a chemical synthesis scheme.

Within this chapter we apply bioretrosynthesis as an approach to construct an *in vitro* pathway for the synthetic nucleoside analog didanosine (dideoxyinosine, ddI), an archetypal off-patent pharmacological inhibitor of HIV-1 reverse transcriptase. The selection of this target was based on the following criteria: (1) dideoxyinosine is representative of a widely prescribed class of drugs used in antiviral and anticancer indications with new members currently in clinical trials, (2) manufacturing costs contribute over 75% of the final therapy costs for dideoxyinosine and other nucleoside analogs[(15)], and (3) metabolic progenitor enzymes can be readily identified for each step in a conceptual bioretrosynthesis.

We developed our retrosynthetic pathway using logical parallels to the natural synthesis of inosine. While there are several biosynthetic routes to inosine from primary metabolic precursors, the shortest sequence consists of only three enzymes to produce inosine from the natural sugar ribose: ribokinase (RK), phosphopentomutase (PPM) and purine nucleoside phosphorylase (PNP) (Figure 5-2a). In the proposed non-natural pathway (Fig. 5-2b, blue box), dideoxyinosine is generated by a purine nucleoside phosphorylase catalyzed addition of hypoxanthine to 2,3-dideoxyribose 1-phosphate. Retroconsecutively, 2,3-dideoxyribose 1-phosphate may be accessed by phosphopentomutase catalyzed isomerization of 2,3-dideoxyribose 5-phosphate, which in turn can be formed by phosphorylation of 2,3-dideoxyribose by ribokinase.

Beyond the multifold challenges entailed in designing new biochemical pathways[16] and engineering multiple enzymes for unnatural substrates[17], a critical requirement for advancing unnatural pathway engineering is the development of generalizable methods for the assembly and optimization of unnatural biosynthetic pathways. Construction of an engineered multistep pathway can progress in one of two directions. In the forward direction, pathway evolution proceeds from the beginning of a proposed pathway, recruiting and evolving enzymes for each step, and assembling steps in the direction of biosynthesis (Fig. 5-2b, top). Engineering a pathway in this fashion potentially requires evolving each enzyme and developing a unique screening/selection strategy for each step in the designed pathway (Fig 5-2b, gray boxes). Assay design can be challenging and, in the event of an intransigent intermediary step, new pathway strategies may be required.

An alternative approach to pathway construction may be inspired by the hypothesis of retrograde evolution[18]. This hypothesis asserts that biosynthetic pathways may be evolved in a stepwise fashion, optimizing each enzyme for its required function, but that the order of enzyme recruitment is in the reverse direction of synthesis,

beginning with the terminal step (Fig 5-2b, bottom). In this approach, pathway product formation is the single selection criterion for evolution at each retro-consecutive step (Fig 5-2b, orange boxes), thereby reducing the number of selections/screening strategies required to a single assay. This model can be experimentally applied to non-natural pathway creation as the second aspect of bioretrosynthesis, where non-natural pathway construction and evolution is inspired by this model of retrograde evolution, and the strategy is to screen for pathway product formation via increasingly tandem enzyme

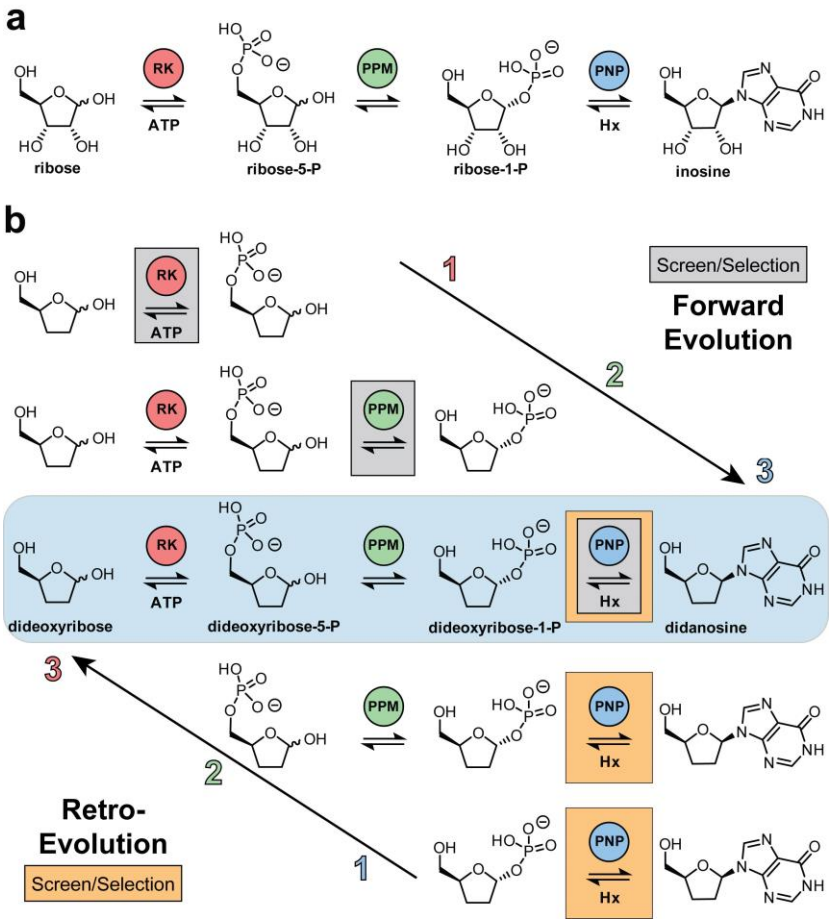

Figure 5-2. Model inosine biosynthetic pathway and proposed bioretrosynthesis of dideoxyinosine. (**a**) The three enzyme metabolic pathway for inosine used as a model to construct a dideoxyinosine pathway. (**b**) Comparison of the forward and retro-evolution strategies of pathway construction. Enzymes evolved in the forward direction proceed in the order of biosynthesis (RK, PPM then PNP), requiring individual screening assays for each enzymatic step (gray boxes). Retro-evolution requires one screening assay for terminal enzyme activity (orange boxes) for evolution of each enzyme in the reverse order of biosynthesis (PNP, PPM then RK) in tandem assays.

assays in a serially retro-consecutively extended pathway. Once optimized through enzyme engineering, this pathway consisting of ribokinase, phosphopentomutase and purine nucleoside phosphorylase is anticipated to form the biocatalytic portion of a semisynthetic route to dideoxyinosine from dideoxyribose, which can be synthesized in a short synthetic sequence from glutamic acid (Figure 5-3).

Here we assemble and evolve these individual enzymes in a retrograde fashion, according to the bioretrosynthetic process, into a five-step pathway including an ATP regeneration cycle. The resulting pathway displays a 9,500-fold change in pathway substrate selectivity and a 50-fold increase in dideoxyinosine production in comparison to the progenitor primary metabolic pathway. An unplanned result of the bioretrosynthetic strategy was the discovery of a pathway-shortening biochemical bypass and a previously unreported phosphorylation activity elicited by ribokinase. Inevitably, this activity was detected through testing for final product formation in the bioretrosynthesis evolution strategy rather than possibly having gone overlooked by screening for the anticipated dideoxyribose 5-phosphate intermediate.

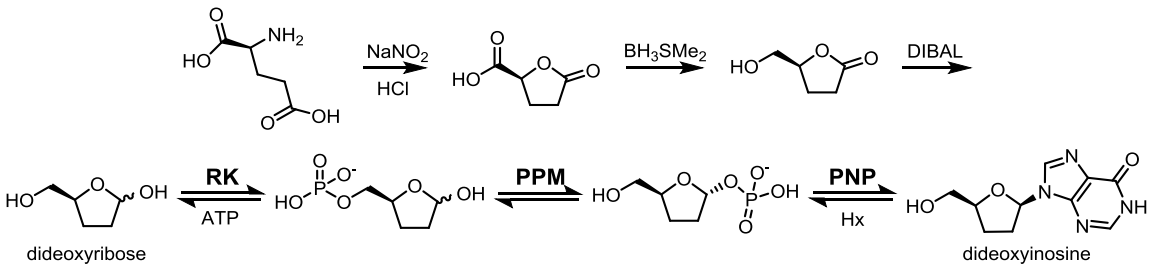

Figure 5-3. Semisynthetic pathway for production of dideoxyinosine. The non-natural sugar dideoxyribose is synthesized via cyclization of glutamic acid and two sequential reductions before entering the enzymatic pathway consisting of ribokinase (RK), phosphopentomutase (PPM) and purine nucleoside phosphorylase (PNP).

Methods

Ribokinase mutagenesis

Site-directed mutagenesis was performed using the QuikChange II mutagenesis kit (Stratagene) with wild-type ribokinase (Genbank Accession Number ACT45432.1) in pCDFDuet vector[19]. *DpnI* restriction endonuclease was used to digest the template plasmid DNA prior to purification of the mutant plasmid by QIAquick PCR Purification Kit (QIAgen, Inc.) and subsequent transformation. PCR sample preparation and thermal cycling for recombination of mutations were performed following the recommended protocol by the kit manufacture's manual. Primers used are listed in Table 5-1. Sample preparation for cloning protocols was performed as recommended in kit manuals.

Enzyme expression and purification

Plasmids containing PPM variants, PNP variants[20], RK variants, adenylate kinase or pyruvate kinase[19] were transformed into *E. coli* BL21(DE3) and grown at 37 °C in LB medium supplemented with 50 µg/mL kanamycin or 50 µg/mL streptomycin (RK variants) and induced with 1 mM IPTG for 3 – 6 h after OD_{600} had reached 0.5 – 0.6. Cell pellets were resuspended in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole,

Table 5-1. Primers used in site directed mutagenesis of RK. Mutations in each sequence are underlined.

Primer Name	Nucleotide Sequence
Asp16Leu for	AGCATTAATGCTCTGCACATTCTTAATCTTCAATC
Asp16Leu rev	GATTGAAGATTAAGAATGTGCAGAGCATTAATGCT
Asp16Asn for	AGTCTTAATGCTAACCACATTCTTAATCTTCAATC
Asp16Asn rev	GATTGAAGATTAAGAATGTGGTTAGCATTAATGCT
Asp16Ala for	AGCATTAATGCTGCCCACATTCTTAATCTTCAATC
Asp16Ala rev	GATTGAAGATTAAGAATGTGGGCAGCATTAATGCT

pH 7.4) and disrupted by passing through a French Pressure cell (1400 psi). The clarified lysate was applied to a HisTrapFFcrude Nickel affinity column (GE Healthcare, Inc.) and washed at 10% Buffer B (Buffer A with 500 mM imidazole). Protein was eluted by a linear gradient from 10% Buffer B to 60% Buffer B, before a step up to 100% Buffer B to fully elute the column. Purified enzymes were concentrated, desalted and exchanged into 25 mM Tris-HCl, pH 8 or 5 mM $MgCl_2$ 25 mM Tris-HCl, pH 8 (ribokinase variants) before storage at -80 °C. All enzyme concentrations were determined using the BCA Protein Assay Kit (Thermo Scientific, Inc.).

In vitro production of inosine and dideoxyinosine

Nucleoside production turnover through the one step (PNP only), two step (PPM and PNP) and full three step (RK, PPM and PNP) pathways was measured using purified enzymes. For the one step pathway, sugar 1-phosphate was generated *in situ* from (2,3-dideoxy)ribose 5-phosphate by preincubation with either wild-type PPM or PPM-4H11. The PPM variant (5.27 µM) was activated in the presence of 0.11 mM $MnCl_2$, 5.27 µM glucose 1,6-bisphosphate and 3.34 mM hypoxanthine in 25 mM Tris, pH 8 for 15 min in a volume of 900 µL before addition of 50 µL 50 mM sugar 5-phosphate (2.63 mM final) and incubated for 4 h to create the sugar 1-phosphates. Each reaction was then passed through a 10 kDa molecular weight cutoff filter at 4 °C to remove PPM and aliquoted to 95 µL portions on a 96-well plate before 5 µL of 0.4 µM PNP variant (for inosine production) or 10 µM PNP variant (for dideoxyinosine production) was added to initiate the reaction. The final reaction performed in duplicate contained 20 nM or 500 nM PNP variant, 0.1 mM $MnCl_2$ and 3 mM hypoxanthine in 25 mM Tris, pH 8 with equilibrium concentrations of (2,3-dideoxy)ribose 5-phosphate and (2,3-dideoxy)ribose 1-phosphate and trace amounts of glucose 1,6-bisphosphate. Assays were incubated for 5 min (for ribose 1-phosphate) and 20 min (for dideoxyribose 1-phosphate) before being

quenched by addition of 5 µL 2 M NaOH then neutralized after 30 min by addition of 5 µL 2 M HCl/1 M $CaCl_2$. Assay plates were centrifuged to pellet the precipitates and samples were prepared for LC/MS analysis as described in previous chapters.

Catalysis via the two step pathway was performed on a 96-well plate in 100 µL reactions in duplicate comparing nucleoside production through multiple pairs of PPM and PNP variants from sugar 5-phosphates. Samples were initiated by addition of 10 µL of 10 mM sugar 5-phosphate to 90 µL of assay mix containing all other reaction components. Final reaction conditions were 0.1 mM $MnCl_2$, 0.5 µM glucose 1,6-bisphosphate, 0.1 µM PPM variant, 2 µM PNP variant, 2 mM hypoxanthine and 1 mM sugar 5-phosphate. Reactions were incubated for 5 min (ribose 5-phosphate) and 20 min (dideoxyribose 5-phosphate) before quenching, neutralizing and preparing samples for LC/MS as previously described.

Inosine and dideoxyinsone production through the full three step pathway and ATP regeneration cycle was determined comparing production of all three wild-type enzymes to production after incorporating the engineered PNP, PPM and RK variants in the presence of an ATP regeneration cycle. Conditions for inosine production were 0.6 mM $MnCl_2$, 0.6 mM $MgCl_2$, 30 mM KCl, 1 µM PNP variant, 0.1 µM PPM variant, 1 µM RK variant, 5 µM adenylate kinase, 5 µM pyruvate kinase, 0.5 µM glucose 1,6-bisphosphate, 0.5 mM ATP and 1 mM ribose. Reactions for dideoxyinosine production were as listed above with 10 µM PNP variant, 10 µM PPM variant, 100 µM RK variant, 15 µM glucose 1,6-bisphosphate and 1mM dideoxyribose. Duplicate assays were initiated by addition of ribose or dideoxyribose and incubated at room temperature for 5 min (ribose) or 10 hr (2,3-dideoxyribose). Aliquots of 250 µL were quenched by loading onto Oasis® HLB (3 mL, 60 mg) solid-phase extraction cartridges (Waters) preconditioned with 3 mL methanol and 3 mL water on a vacuum manifold. Loaded cartridges were then washed with 1 mL water and nucleosides were eluted with 1.5 mL methanol[21]. The methanol

fraction was evaporated to dryness and the remaining residue was reconstituted in 250 µL water (ribose pathway) or 50 µL water (2,3-dideoxyribose pathway). Reactions for production of nucleosides without PPM were performed identically without the addition of the PPM variant and were incubated for 10 h before quenching. Samples were prepared for LC/MS as previously described using 10 µM 2-deoxyguanosine as the internal standard. Turnover in the reaction mix without the sugar substrate incubated for the same time were subtracted from the final turnover measurements.

Results

Bioretrosynthetic Step 1: Nucleoside Phosphorylase

We previously improved the terminal reaction of the proposed pathway by evolving human purine nucleoside phosphorylase (hPNP) to phosphorolyze dideoxyinosine[20], but had not evaluated improvements in the synthesis direction. We therefore began evaluating the application of bioretrosynthesis as method in non-natural pathway construction by assaying the product forming enzyme for competence in the synthesis of inosine and dideoxyinosine by the wild-type and hPNP-46D6 variants from the respective sugar 1-phosphates via HPLC/MS. The wild-type hPNP exhibited a 660-fold higher turnover of ribose 1-phosphate than dideoxyribose 1-phosphate (Figure 5-4). This selectivity was reduced to 4.7-fold preference for ribose 1-phosphate in the optimized hPNP-46D6 variant after 16-fold increase in the rate of dideoxyinosine production combined with a corresponding 8.7-fold decrease in inosine formation. With a 140-fold change in substrate selectivity and a turnover rate of 0.37 µM/min/µM PNP, this enzyme was considered to possess sufficient activity for the *in vitro* tandem assays of PPM or PPM-RK required for implementing a bioretrosynthetic strategy for engineering the targeted dideoxyinosine pathway.

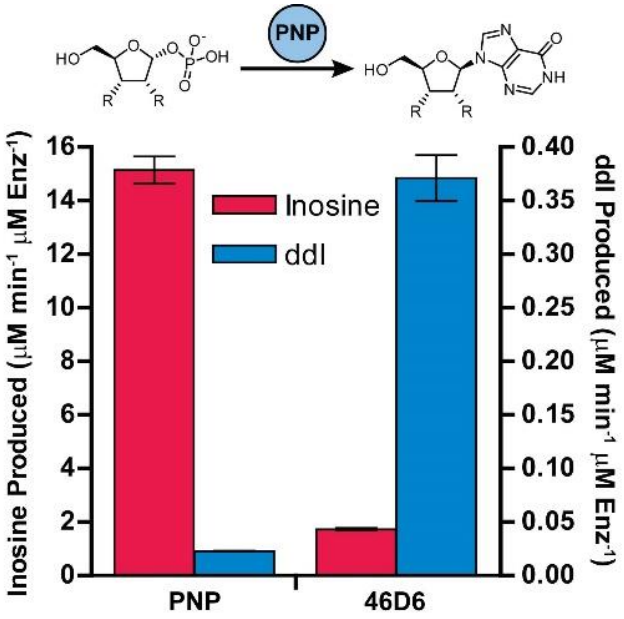

Figure 5-4. In vitro biosynthetic production of inosine and dideoxyinosone catalyzed by PNP. Nucleoside production was measured via HPLC/MS analysis from *in situ* generated sugar 1-phosphates in one step biocatalysis (PNP only). Enzyme variants used in each reaction are listed under bars. Turnover was normalized to PNP variant concentration and assay length. Data are mean ± s.d. (n=2). Ribosyl substrates, R=OH. Dideoxyribosyl substrates, R=H.

Bioretrosynthetic Step 2: Phosphopentomutase

The two step pathway showed further production selectivity improvements after pairing the two optimized enzymes (Fig. 5-5). The combination of wild-type PPM and wild-type PNP showed a 1420-fold bias for ribose 5-phosphate over the dideoxy substrate. Incorporating the evolved PNP variant into the pathway provided a 28.9-fold increase in dideoxyinosine production and a small loss in inosine formation. The effect on inosine biosynthesis was further compounded after pairing the two optimized enzymes, showing a 342-fold total change in substrate selectivity to create a tandem evolved biosynthetic pathway with 32.5-fold improved dideoxyinosine production and only a 4.2-fold preference for the natural substrates. Although the evolved pair of enzymes did not provide much increase in dideoxyinosine production over the wild-type

PPM with the evolved PNP under the tested conditions, the PPM-4H11 variant did provide a significant change in selectivity by reducing inosine productivity.

Bioretrosynthetic Step 3: RK

As previously mentioned in Chapter IV, at this point in the bioretrosynthetic construction it was also necessary to include a cofactor recycling system as ATP, ADP and AMP were found to inhibit PPM. To satisfy this requirement, the high energy phosphate donor phosphoenolpyruvate was added to the *in vitro* production reactions along with pyruvate kinase and adenylate kinase to regenerate ATP. This allowed the reaction to contain a lower concentration of ATP and prevent loss of cofactor equivalents due to hydrolysis over the extended incubation lengths. Including this cofactor

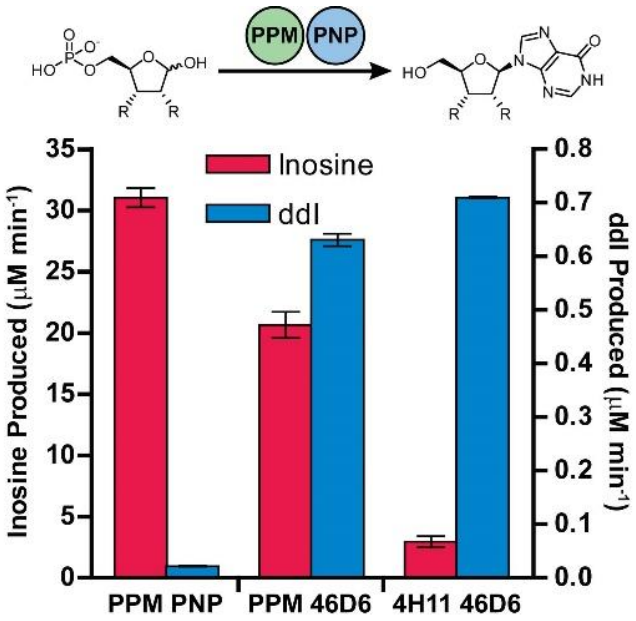

Figure 5-5. In vitro biosynthetic production of inosine and dideoxyinosine catalyzed by PPM and PNP in tandem. Nucleoside production was measured via HPLC/MS analysis from sugar 5-phosphates in the two-step tandem pathway (PPM and PNP). Enzyme variants used in each reaction are listed under bars. Production was normalized to incubation time. Data are mean ± s.d. (n=2). Ribosyl sybstrates, R=OH. Dideoxyribosyl substrates, R=H.

regeneration cycle extended the full system to a five enzyme biosynthetic pathway.

Unsurprisingly, the complete pathway consisting of all wild-type enzymes efficiently produced inosine, reaching 208 µM in five minutes, while dideoxyinosine production was only 0.6 µM after 10 h (Fig. 5-6). Retroconsecutively introducing the

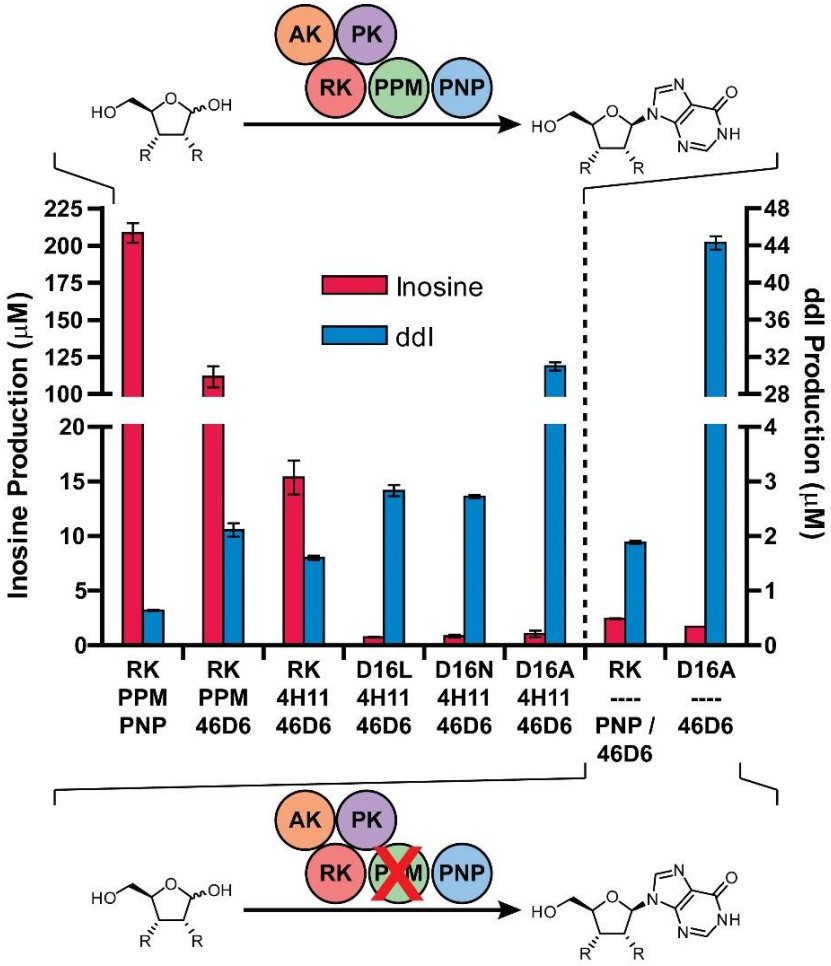

Figure 5-6. In vitro biosynthetic production of inosine and dideoxyinosine catalyzed by the full biosynthetic pathway. Nucleoside production was measured via HPLC/MS analysis from ribose and dideoxyribose in the full biosynthetic pathway (RK, PPM and PNP) or the pathway without PPM. Enzyme variants used in each reaction are listed under bars (dashed line means no variant included). Direct phosphorylation of the sugar C1 position by wild-type RK was tested in tandem with the appropriate PNP variant for best detection of activity. The full inosine pathway was incubated for 5 min, while dideoxyinosine production was for 10 h. Reactions without PPM were incubated for 10 h, however inosine production was normalized to 5 min do be directly comparable to production by the full pathway with PPM. Data are mean ± s.d. (n=2). Ribosyl sybstrates, R=OH. Dideoxyribosyl substrates, R=H.

evolved hPNP-46D6 and then the evolved PPM-4H11 variants reduced inosine production by 2-fold and 13.6-fold, respectively, while providing 3-fold and 2.5-fold gains in dideoxyinosine production, respectively (Fig. 5-6).

The functional three enzyme system facilitated preliminary mutagenesis studies of RK to observe the effects on pathway product formation. Analysis of the X-ray costructure of *E. coli* ribokinase with ribose (1GQT[22]) shows that the active site residue Asp16 interacts with the C2 and C3 hydroxyls of ribose[23] (Fig. 5-7). As these hydroxyl functionalities are not present in dideoxyribose, we replaced the charged aspartate with non-polar and space filling leucine, the polar isostere asparagine, or alanine (Asp16Leu/Asn/Ala mutants) and measured nucleoside production through the engineered pathway. All three variants reduced inosine formation to ≤1 μM at five minutes, diminishing productivity 200 - 275-fold from the wild-type pathway (Fig. 5-6). The Asp16Leu/Asn mutations provided slight increases in dideoxyinosine productivity, however the Asp16Ala RK (RK-Asp16Ala) resulted in a 20-fold increase in

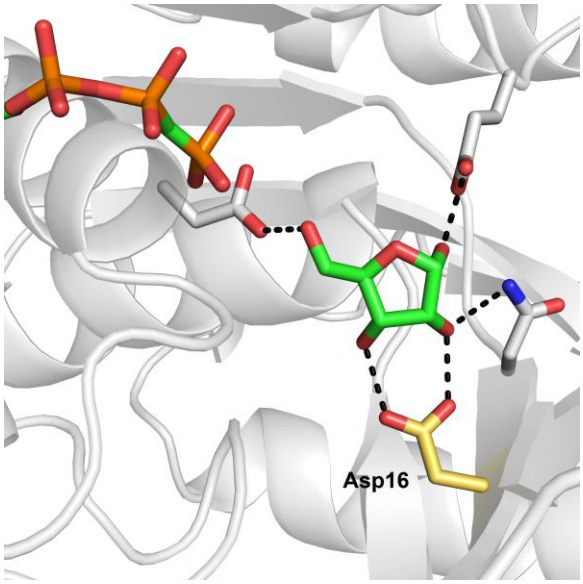

Figure 5-7. Orientation of ribose by Asp16 in RK. The interaction between the active site residue Asp16 and bound ribose is highlighted in wild-type *E. coli* RK (PDB entry 1GQT[22]). This interaction between Asp16 (gold) and ribose (green) was targeted for removal in preliminary mutagenesis studies by mutation to leucine, asparagine and alanine (Asp16Leu/Asn/Ala).

dideoxyinosine production compared to wild-type RK, forming 31 μM after 10 h (Figure 5-6). Combined with the reduced production of inosine, the pathway of RK-Asp16Ala, PPM-4H11 and hPNP-46D6 resulted in a 9,544-fold total change in nucleoside production selectivity through the three evolved enzymes and a total of 50-fold increased dideoxyinosine production.

The unexpectedly large increase in productivity in the pathway containing RK-Asp16Ala prompted investigation into the functional role of the improved ribokinase. In particular, we considered the possibility that previously unobserved anomeric (C1) phosphorylation of dideoxyribose could more directly provide the substrate for PNP and increase pathway productivity. Indeed, under optimized complete pathway conditions but excluding PPM-4H11, the RK-Asp16Ala in tandem with hPNP-46D6 provided the highest dideoxyinosine production (verified by HPLC/MS comparison to synthetic dideoxyinosine standard) at 44 μM, or 4.4% yield from dideoxyribose and 70-fold greater than the wild-type three enzyme pathway. Of note, the turnover of the RK-Asp16Als/hPNP-46D6 two-enzyme system was 40% higher than the three enzyme pathway including PPM-4H11 (Figure 5-6). This suggests that the majority of the RK-Asp16Ala effect on activity in the full pathway was in fact due to direct phosphorylation of the anomeric hydroxyl group rather than expected O-phosphorylation at the C5 position, which diverts ribose into the more indirect (PPM-dependent) pathway. A low level of this activity was also observed in the evolved pair of enzymes when ribose was the substrate, and additionally in wild-type RK on both substrates (Figure 5-6). Consequently, the decreased production in the presence of PPM-4H11 is likely the result of partitioning the generated dideoxyribose 1-phosphate into dideoxyribose 5-phosphate by the engineered PPM.

Discussion

Through using bioretrosynthesis as a pathway construction method and assaying for final product formation, we are able to examine the effect on production that each enzyme has at all three biosynthetic steps by simply extending the pathway. In this way, we can compare production from natural to non-natural substrates by wild-type and evolved enzymes. Nucleoside production from sugar 1-phosphates by wild-type PNP and the evolved hPNP-46D6 displayed a 16-fold increase in dideoxyinosine production along with a 140-fold change in substrate selectivity by the variant enzyme (Fig. 5-4). In addition to creating a greater driving force for total production, improving the turnover rate of PNP was necessary to allow the evolution of PPM for this non-natural pathway. The product of PPM catalysis is an isomer of the substrate and therefore provides no mass change for detection my mass spectrometry, and also lacks a detectable chromophore for direct measurement by UV/Vis absorbance. Here, PNP in tandem acts as the reporter enzyme enabling PPM activity measurements by either hypoxanthine consumption or inosine/dideoxyinosine formation in typical assays to determine PPM activity[(24)], and additionally constitutes the next biosynthetic step within our designed non-natural pathway. The improved turnover rate of hPNP-46D6 increased the dynamic range of the assay by providing a much more efficient and sensitive method for detection of PPM activity on dideoxyribose 5-phosphate. Use of the tandem assay also begins to connect the enzymes as a biosynthetic pathway through enabling the evolution of PPM in the presence of the evolved PNP.

As the product-forming enzyme, the efficiency of PNP defines the total yield through the pathway, as indicated by the increase in dideoxyinosine production after introducing the engineered hPNP-46D6 into biosynthetic pathways (Figures 5-4, 5-5, 5-6). Selectivity, however, can be affected at each biosynthetic step by integrating evolved enzymes into the pathway and is evident in the change in nucleoside production

observed after pairing the evolved hPNP-46D6 and PPM-4H11 variants (Figure 5-5). This engineered system highlights two main aspects of retro-consecutive pathway construction: 1) the substrate selectivity advantage that can be gained through combining two evolved enzymes in tandem (342-fold in this pathway by combining the evolved PNP and PPM), and 2) demonstrating the use of fundamentally the same assay for measuring improvements at two enzymatic steps. As stated previously, this can have great implications to simplify the assay design requirement when evolving multiple enzymes, as the same assay can be modified to fit the conditions of the tandem system rather than needing to design and validate a completely new assay for each enzymatic step.

Another key determinant of total productivity is substrate flux into the pathway. Manipulation of these so called 'gatekeeping enzymes' can determine the production scope of the pathway based on the turnover and substrate selectivity in this first committed step[(25)]. For high titer production of one desired compound, this enzyme would ideally be substrate specific to reduce wasteful consumption of starting materials. Ultimately, this initial biosynthetic step sets the pace for the remaining pathway by producing and maintaining the dedicated substrate pool for the subsequent biosynthetic enzymes. In terms of this dideoxyinosine biosynthetic pathway, screening multiple kinases for production of dideoxyinosine from dideoxyribose in tandem with the evolved PPM and PNP enzymes identified *E. coli* ribokinase as the highest producer and most comprehensively characterized and therefore was the best candidate for the pathway (Chapter IV).

The full pathway with *E. coli* RK demonstrates the *in vitro* biosynthesis of both inosine and dideoxyinosine from simple sugars, indicating the viability of this pathway for production of the non-natural nucleoside analog. As anticipated, wild-type RK is fully capable of maintaining ribosyl substrate flux into the pathway, quickly generating high

levels of inosine (Figure 5-6). However, production of dideoxyinosine through the full pathway from dideoxyribose is very inefficient, providing significantly lower yields even at long incubations and much higher enzyme concentrations. This result wasn't completely unexpected since despite the noted dideoxyinosine production increase by PPM-4H11 and hPNP-46D6 in tandem (Figure 5-5), wild-type RK was predicted to be quite inefficient in phosphorylating dideoxyribose due, in part, to the reduced number of functional groups of the non-natural substrate and the lack of distinct substrate modifiers (such as phosphate or Coenzyme-A) that help lower K_M by providing more interactions with the enzyme[26]. This particularly low activity by the first biosynthetic enzyme severely limits production through the non-natural pathway, and emphasizes the high level of improvement required to become an efficient biocatalyst.

Targeting the highly conserved active site Asp16 residue for mutagenesis was hypothesized to provide a less favorable binding environment for ribose, but possibly also improve dideoxyribose binding by reducing the active site polarity to match the non-natural substrate. The leucine and asparagine mutations did deliver substantial losses in ribose activity, however a considerable improvement in dideoxyinosine formation was only found after removing side chain functionality in the Asp16Ala mutation, increasing production 20-fold over wild-type RK and a total of 50-fold in the three step engineered pathway compared to the all wild-type pathway. Surprisingly, more detailed characterization of the ribokinase revealed that the dideoxyinosine produced in this system is not due to an antepenultimate retro-extension, but is instead primarily the result of a previously unobserved C1 phosphorylation regiochemistry for the engineered ribokinase. This newly identified activity enables a pathway shortening bypass of the typically required PPM catalyzed isomerization by directly forming a substrate for the ultimate, PNP-catalyzed step. Having PPM present in the pathway with RK-Asp16Ala actually lowered total yield, presumably by sequestering a portion of

substrate as the unproductive dideoxyribose 5-phosphate intermediate due to the enzymatic equilibrium. As an underlying level this bypass activity was also detected in the wild-type RK variant, the Asp16Ala mutation did not introduce the capacity for direct C1 phosphorylation but rather afforded a substantial enhancement, seemingly through promoting the unanticipated dideoxyribose binding orientation. Additionally, this C1 phosphorylation appears to be the predominant activity observed on dideoxyribose since production through the full pathway with dideoxyribose is comparable to production in the absence of PPM (Figure 5-6). This can also be observed in the pathway with wild-type RK through the slight decrease in production after incorporating the more efficient PPM-4H11 variant, capable of diverting a larger amount of substrate to the non-productive dideoxyribose 5-phosphate pool (Fig. 5-6). Detecting this unexpected activity is the direct consequence of the requirement of screening for product formation in the bioretrosynthetic model, as this bypass activity may have likely gone unobserved if assays were designed to measure production of the anticipated dideoxyribose 5-phosphate product. Also of note, in a survey of members of the PfkB subfamily of the ribokinase-like superfamily to which *E. coli* RK belongs, the critical Asp16 is conserved throughout adenosine kinases and ribokinases[27], aminoimidazole riboside kinases and ADP-dependent glucokinases[28], ADP-dependent 6-phosphofructokinases[29] and ketohexokinases[30] and has also been observed in phosphofructokinase-2[31] and tagatose 6-phosphate kinase[32]. To our knowledge, no enzyme has been reported to phosphorylate ribose at the 1-position in this or any other superfamily. Therefore, this previously unreported activity of the RK-Asp16Ala variant potentially provides a much more unique foundation for further engineering dideoxyribose turnover and selectivity.

Conceptualization of the pathway using bioretrosynthetic analysis and assembly from primary metabolic progenitor enzymes resulted in a pathway that produced low levels of dideoxyinosine with selectivity vastly favoring the native substrates (Figure

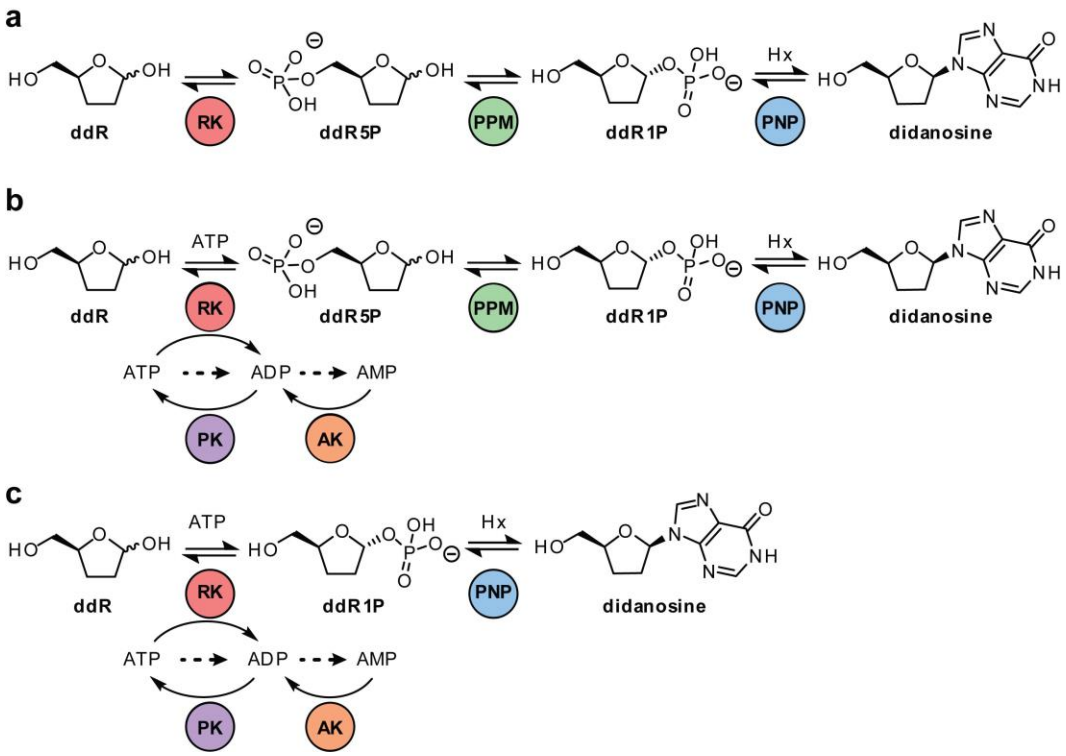

Figure 5-8. Progression of the dideoxyinosine biosynthetic pathway components through stages of bioretrosynthetic optimization. (**a**) The original proposed three enzyme biosynthetic pathway. (**b**) The full pathway including the ATP regeneration cycle after identifying the final kinase progenitor for phosphorylation of dideoxyribose. (**c**) The two step dideoxyinosine biosynthetic pathway as a result of the discovered bypass activity of the engineered RK. Dashed arrows indicate expected degradation products.

5-8a). Subsequent to the directed evolution of the ultimate pathway enzyme hPNP, we improved pathway turnover and selectivity and found it beneficial to add an ATP-recycling system to optimize precursor concentrations and yield. Evolution of the antepenultimate PPM resulted in a new five-enzyme prototype pathway retaining turnover with greatly increased selectivity (Figure 5-8b). Finally, engineering of the antepenultimate RK provided an unexpected biochemical bypass within the pathway and a substantial improvement in turnover (Figure 5-8c). The final improvements in substrate selectivity (9,500-fold) and improvements in turnover for target substrates (50-fold) of the full pathway illustrate the successful demonstration and practically useful application of the retro-evolution hypothesis *in vitro*. These levels of improvement were also observed

in the shortened pathway without PPM, which gave 8,300-fold improved substrate selectivity and 70-fold increased production of dideoxyinosine.

Conclusions

In this study, we provide a proof-of-concept for modeling non-natural pathway optimization and construction on the hypothesis of retrograde evolution in a practically useful laboratory process termed bioretrosynthesis. Although dideoxyinosine production overall is modest in this *in vitro* demonstration, these results nonetheless demonstrate the potential of this pathway as a biosynthetic route to dideoxyinosine. Further optimization of this pathway for yield and production conditions may provide a scalable biosynthetic route to dideoxyinosine and could aid in increasing accessibility to treatment by reducing therapeutic costs. This biocatalytic system may also provide access to a larger dideoxynucleoside analog pool after a forward extension of the pathway to an engineered nucleoside 2'-deoxyribosyltrasnferase capable of exchanging the nucleobase component of dideoxynucleosides[(33)]. An additional retro-extension can also be envisioned, recruiting an alcohol dehydrogenase to generate dideoxyribose from (*S*)-1,2,5-pentanetriol (i.e. dideoxyribitol), further simplifying the chemical synthesis aspect.

While the RK-Asp16Ala/hPNP-46D6 system may be the most efficient route for the generation of dideoxynucleosides through the enhanced C1 phosphorylation activity of the RK variant, direct anomeric phosphorylation of 2- or 3-substituted ribosides via RK-Asp16Ala may not be possible due to differences in substrate binding. Biosynthesis of these types of substituted nucleosides may therefore require the development and use of PPM-based strategies for sugar analog activation described here, potentially beginning engineering from the generalist PPM-4H11 variant evolved in this work. Given the breadth of diversity found in the sugar moieties of nucleoside analog drugs (Figure

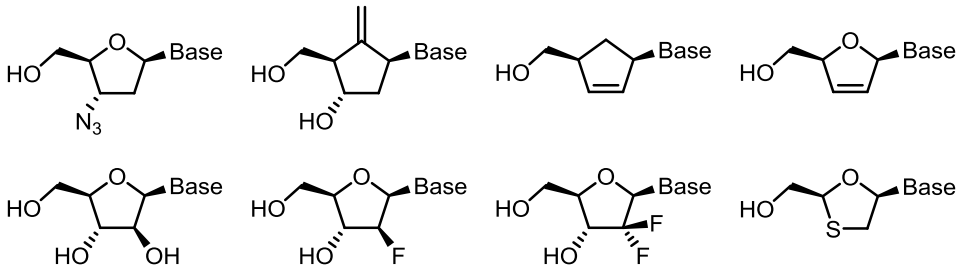

Figure 5-9. Examples of sugar moieties found in non-natural nucleoside analogs.

5-9), PPM variant enzymes capable of activating the anomeric (C1) position of these sugar analogs may find complementary application in biocatalytic generation of a variety of compounds within this class of therapeutics. Each enzyme may also be individually useful in the ever-growing collection of engineered enzymes for application in other biocatalysis studies, and likewise in 'biocatalytic retrosynthesis'[13] evaluations of new or existing chemical synthesis routes to determine if enzymes may be able to replace traditional chemical synthesis reactions.

Retro-evolution can not only be considered a theoretical model of pathway evolution but can also be employed as a laboratory method in pathway design and optimization for a target non-natural compound. In addition to simplifying the assay design requirement for evolution of multi-enzyme pathways, by screening for final product formation rather than putative pathway intermediates, bioretrosynthesis can inherently allow for the detection of unanticipated pathway bypass reactions due to unpredictable activity gains. However, these bypass activities may likely not be a commonly observed outcome. This bioretrosynthetic method is a biosynthetic equivalent to and a true bench-top application of the established retrosynthetic analysis concept used in planning chemical synthesis[12] and has the potential to become a widely applicable strategy for engineering multistep non-natural biosynthetic pathways for both natural and non-natural products.

Acknowledgements

This work was contributed to by William R. Birmingham, David P. Nannemann and Brian O. Bachmann. Bioretrosynthetic evolution of a didanosine biosynthetic pathway. W.R.B. purified and tested enzymes in the biosynthetic pathway studies. D.P.N. established initial synthesis routes of dideoxyribose 5-phosphate. B.O.B. supervised biochemical experiments. We thank V. Phelan for assistance with establishing the mass spectrometry based analysis methods.

References

1. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.

2. Liang, J., Lalonde, J., Borup, B., Mitchell, V., Mundorff, E., Trinh, N., Kochrekar, D. A., Cherat, R. N., and Pai, G. G. Development of a biocatalytic process as an alternative to the (-)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org. Process. Res. Dev.* **2010**, 14, 193-198.

3. Gao, X., Xie, X., Pashkov, I., Sawaya, M. R., Laidman, J., Zhang, W., Cacho, R., Yeates, T. O., and Tang, Y. Directed evolution and structural characterization of a simvastatin synthase. *Chem. Biol.* **2009**, 16, 1064-1074.

4. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, 485, 185-194.

5. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013**, 496, 528-532.

6. Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., and Stephanopoulos, G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **2010**, 330, 70-74.

7. Ito, T., Roongsawang, N., Shirasaka, N., Lu, W., Flatt, P. M., Kasanah, N., Miranda, C., and Mahmud, T. Deciphering pactamycin biosynthesis and engineered production of new pactamycin analogues. *Chembiochem* **2009**, 10, 2253-2265.

8. Zhang, M.-Q., Gaisser, S., Nur-E-Alam, M., Sheehan, L. S., Vousden, W. A., Gaitatzis, N., Peck, G., Coates, N. J., Moso, S. J., Radzom, M., Foster, T. A., Sheridan, R. M., Gregory, M. A., Roe, S. M., Prodromou, C., Pearl, L., Boyd, S. M., Wilkinson, B., and Martin, C. J. Optimizing natural products by biosynthetic engineering: Discovery of nonquinone Hsp90 inhibitors. *J. Med. Chem.* **2008**, 51, 5494-5497.

9. Niu, W., Molefe, M. N., and Frost, J. W. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **2003**, 125, 12998-12999.

10. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **2011**, 7, 445-452.

11. Ma, S. K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G. W., and Sheldon, R. A. A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem.* **2010**, 12, 81.

12. Corey, E. J. The logic of chemical synthesis - Multistep synthesis of complex natural carbogenic molecules. *Angew. Chem. Int. Ed. Engl.* **1991**, 30, 455-465.

13. Turner, N. J., and O'Reilly, E. Biocatalytic retrosynthesis. *Nat. Chem. Biol.* **2013**, 9, 285-288.

14. Bachmann, B. O. Biosynthesis: Is it time to go retro? *Nat. Chem. Biol.* **2010**, 6, 390-393.

15. Pinheiro, E., Vasan, A., Kim, J. Y., Lee, E., Guimier, J. M., and Perriens, J. Examining the production costs of antiretroviral drugs. *AIDS* **2006**, 20, 1745-1752.

16. Medema, M. H., van Raaphorst, R., Takano, E., and Breitling, R. Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* **2012**, 10, 191-202.

17. Eriksen, D. T., Lian, J., and Zhao, H. Protein design for pathway engineering. *J. Struct. Biol.* **2013**.

18. Horowitz, N. H. On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U.S.A.* **1945**, 31, 153-157.

19. Scism, R. A., and Bachmann, B. O. Five-component cascade synthesis of nucleotide analogues in an engineered self-immobilized enzyme aggregate. *Chembiochem* **2010**, 11, 67-70.

20. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

21. Bezy, V., Morin, P., Couerbe, P., Leleu, G., and Agrofoglio, L. Simultaneous analysis of several antiretroviral nucleosides in rat-plasma by high-performance liquid chromatography with UV using acetic acid/hydroxylamine buffer - Test of this new volatile medium-pH for HPLC-ESI-MS/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2005**, 821, 132-143.

22. Andersson, C. E., and Mowbray, S. L. Activation of ribokinase by monovalent cations. *J. Mol. Biol.* **2002**, 315, 409-419.

23. Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 angstrom resolution: Insights into a new family of kinase structures. *Structure* **1998**, 6, 183-193.

24. De Groot, H., De Groot, H., and Noll, T. Enzymic determination of inorganic phosphates, organic phosphates and phosphate-liberating enzymes by use of nucleoside phosphorylase-xanthine oxidase (dehydrogenase)-coupled reactions. *Biochem. J.* **1985**, 230, 255-260.

25. Umeno, D., Tobias, A. V., and Arnold, F. H. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol. Mol. Biol. Rev.* **2005**, 69, 51-78.

26. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **2011**, 50, 4402-4410.

27. Datta, R., Das, I., Sen, B., Chakraborty, A., Adak, S., Mandal, C., and Datta, A. K. Homology-model-guided site-specific mutagenesis reveals the mechanisms of substrate binding and product-regulation of adenosine kinase from *Leishmania donovani*. *Biochem. J.* **2006**, 394, 35-42.

28. Zhang, Y., Dougherty, M., Downs, D. M., and Ealick, S. E. Crystal structure of an aminoimidazole riboside kinase from *Salmonella enterica*: Implications for the evolution of the ribokinase superfamily. *Structure* **2004**, 12, 1809-1821.

29. Currie, M. A., Merino, F., Skarina, T., Wong, A. H. Y., Singer, A., Brown, G., Savchenko, A., Caniuguir, A. S., Guixe, V., Yakunin, A. F., and Jia, Z. ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3 -Structure determination and biochemical characterization of PH1645. *J. Biol. Chem.* **2009**, 284, 22664-22671.

30. Trinh, C. H., Asipu, A., Bonthron, D. T., and Phillips, S. E. Structures of alternatively spliced isoforms of human ketohexokinase. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, 65, 201-211.

31. Cabrera, R., Baez, M., Pereira, H. M., Caniuguir, A., Garratt, R. C., and Babul, J. The crystal complex of phosphofructokinase-2 of *Escherichia coli* with fructose-6-phosphate - Kinetic and structural analysis of the allosteric ATP inhibition. *J. Biol. Chem.* **2011**, 286, 5774-5783.

32. Miallau, L., Hunter, W. N., McSweeney, S. M., and Leonard, G. A. Structures of *Staphylococcus aureus* D-tagatose-6-phosphate kinase implicate domain motions in specificity and mechanism. *J. Biol. Chem.* **2007**, 282, 19948-19957.

33. Kaminski, P. A., Dacher, P., Dugue, L., and Pochet, S. In vivo reshaping the catalytic site of nucleoside 2'-deoxyribosyltransferase for dideoxy- and didehydronucleosides via a single amino acid substitution. *J. Biol. Chem.* **2008**, 283, 20053-20059.

Chapter VI

DISSERTATION SUMMARY AND FUTURE DIRECTIONS

Synopsis

Engineered biocatalysts are increasingly being developed to replace traditional chemical synthesis steps for materials ranging from drug intermediates[1-5] and active ingredients[6, 7] to value-added industrial chemicals[8-11], however most are used in single step enzymatic transformations. While multi-enzyme systems are common in catalyzing cofactor regeneration[12] and dynamic kinetic resolution[13] reactions, there are very few examples that incorporate enzyme evolution in multiple biosynthetic steps, such as in the production of the atorvastatin side chain intermediate[14]. Instead, engineering of multi-step biosynthetic pathways has primarily focused on manipulating wild-type enzyme expression and precursor production to guide metabolic flux[15-18] and the heterologous recapitulation of full pathways[19, 20], all toward the goal of increasing yields of natural products.

A major difficulty in delineating a non-natural pathway comes in the knowledge base and experimental strategy necessary for *de novo* construction and evolution. We have described a useful theoretical approach to propose a biosynthetic route for a non-natural product that is borrowed from the retrosynthesis model in organic synthesis[21]. By performing a similar regressive analysis to increasingly simpler substrates using the body of known enzymatic transformations—a process we have termed 'bioretrosynthesis'[22]—a putative biosynthetic pathway can be established for a target compound. In this work, the target product was the non-natural nucleoside analog reverse transcriptase inhibitor dideoxyinosine (ddI, didanosine). Bioretrosynthesis was also validated as an experimental demonstration of the model of retrograde evolution[23]

as the core concepts of this theory are merged into a practical laboratory method of pathway optimization and construction by evolving the proposed enzymes in the reverse order of biosynthesis and testing for final product formation.

As dideoxyinosine is not known to be a naturally produced nucleoside analog, a hypothetical biosynthetic pathway was proposed by applying a bioretrosynthetic analysis through suggesting a sequence of enzymatic transformations that would reduce substrate complexity to simpler compounds in a reverse step-wise manner. Modeling a non-natural biosynthetic pathway for dideoxyinosine after biosynthetic routes of natural nucleoside production, the primary metabolic enzymes ribokinase (RK), phosphopentomutase (PPM) and purine nucleoside phosphorylase (PNP) were chosen for biosynthesis of dideoxyinosine from the non-natural sugar dideoxyribose. Enzyme variants from *E. coli*, *Bacillus cereus* and human, respectively, were identified as progenitor enzymes for evolving activity on the corresponding non-natural dideoxy substrates.

Applying the bioretrosynthesis approach as a practical application of the retrograde evolution concept and using dideoxyinosine as a model non-natural product, we began evolution of the product forming enzyme, PNP and developed a variant of human PNP with enhanced phosphorolysis of dideoxyinosine[(24)]. Here we have demonstrated that this enzyme also shows improved synthesis of dideoxyinosine from dideoxyribose 1-phosphate and hypoxanthine. Next, we provide a retro-extension to the biosynthetic pathway through engineering PPM from *Bacillus cereus* for conversion of dideoxyribose 5-phosphate to dideoxyinosine in tandem assays with the evolved PNP. A third retro-extension to an *E. coli* RK variant provided further improvements in dideoxyinosine production and ~9,500-fold change in nucleoside production selectivity in a multi-step engineered biosynthetic pathway.

The PPM engineering aspect of this work resulted in new understanding of the enzyme structure and the relationship of the active site architecture to effective turnover. Crystallographic evidence with natural or non-natural substrates bound indicated the potential of multiple active site residues to affect substrate selectivity and turnover. Targeted saturation mutagenesis at multiple positions validated Ser154 and Val158 as directly contributing to this selectivity and overall catalytic efficiency, and screening identified Ser154Gly and Val158Leu as individual beneficial mutations in the active site.

Removing the residue side chain in the Ser154Gly mutation provided an increased active site volume that held the potential to allow later mutations accumulated through directed evolution to slightly adjust the substrate binding pocket in favor of dideoxyribose-5-phosphate. In addition, the mutation eliminated a polar element of the active site that was hypothesized to be contributing to the observed disordered dideoxyribose-5-phosphate binding. Although cocrystal structures of this variant with dideoxyribose-5-phosphate did not indicate more defined ligand density for the bound substrate, kinetic characterization did indeed show a significant loss in ribose-5-phosphate selectivity.

In a separate lineage, the side chain methylene extension in the Val158Leu mutation decreased active site volume resulting in a generalist enzyme with no substrate preference. This variant was not able to be cocrystallized with any substrate (glucose 1,6-bisphosphate, ribose-5-phosphate or dideoxyribose-5-phosphate), which supported the theory that the leucine side chain creates a stearic hindrance toward substrate binding. Kinetic analysis of this variant indicates that the more bulky residue has a greater effect on the binding of ribose-5-phosphate than dideoxyribose-5-phosphate, possibly due to the stearic and polarity conflicts with the C2 and C3 hydroxyls of the ribosyl ring.

Random mutagenesis and subsequent recombination of mutations created variants 2G8 and 4H11 with each showing approximately 3-fold improved production of dideoxyinosine in cell lysate over the wild-type enzyme. The 4H11 variant also maintained the activity of a generalist enzyme, providing a total 710-fold change in substrate selectivity over the wild-type PPM. Several of these newly acquired mutations appeared to destabilize known hydrogen bond interactions within the core and catalytic domains of PPM, possibly contributing to an increased interdomain flexibility that improved catalysis. Unfortunately, no improvements were observed after additional mutagenesis using 2G8 and 4H11 as templates.

Turning to the proposed antepenultimate retro-extension, we surveyed a panel of five potential progenitor kinases and identified *E. coli* RK as the candidate most functionally capable of generating dideoxyinosine in tandem with the evolved PPM and PNP variants. Mutational analysis of the active site Asp16 residue to a series of amino acids aimed to analyze the influence on substrate binding after changing the hydrophobicity, polarity and stearic bulk at the position revealed that disrupting ribose activity was fairly simple, as all mutations drastically reduce activity. However an increase in dideoxyinosine production was observed only once the side chain functionality was removed via mutation of aspartate to alanine, providing 20-fold higher production than wild-type RK and a total of 50-fold greater in the three step engineered pathway compared to the all wild-type pathway. Further characterization of the activity of this Asp16Ala RK indicated that production of dideoxyinosine is predominantly the result of a previously unreported activity where the C1 position is phosphorylated, rather than the typical C5 position, forming a pathway shortening bypass around PPM instead of an antepenultimate retro-extension from PPM. The ability to detect this unexpected efficiency is the direct consequence of applying the bioretrosynthetic model for pathway construction, where the requirement is screening for final pathway product formation,

rather than possibly having gone overlooked by screening for the anticipated dideoxyribose 5-phosphate intermediate.

Significance

The work described here in the directed evolution of multiple enzymes and production of dideoxyinosine through an engineered biosynthetic pathway demonstrates the potential for providing a biocatalytic route to this drug, and possibly an extension to include other nucleoside analogs. As a high percentage of the final price per year of treatment with many nucleoside analog drugs is tied to chemical manufacturing of the active ingredient[25], eventually replacing or, at a minimum, supplementing a portion of the production volume with a more environmentally friendly and economical biocatalytic route could directly lead to a reduction in production costs that would be passed on in the form of lower prices to patients in need of treatment. A productive and industrially scalable biosynthetic pathway of this kind could considerably affect the total cost and availability of this drug given the features of a short chemical synthesis from inexpensive starting material, stereoselective production, use of renewable resources and reduction in waste generation and would result in an increased availability of treatment.

From a pathway evolution standpoint, the bioretrosynthesis paradigm demonstrated here provides the first evidence supporting the viability of the hypothesis of retrograde evolution as a method of biosynthetic pathway formation first proposed in 1945[23]. It additionally serves as a model strategy for the planning and construction of *de novo* biosynthetic pathways toward either natural or non-natural products. The terminal (product forming) enzyme activity based screening strategy can greatly simplify the methods and assay development required for engineering multiple enzymes. In general, the same detection scheme can be adapted and used in increasingly tandem assays to

screen for activity of multiple enzymes in the non-natural pathway, rather than requiring design of individual assays for each biotransformation. Cofactor regeneration cycles can be added as necessary through retro-extensions to catalyze reactions in sequence to reform required pathway components. As indicated here, this product forming assay guided aspect of bioretrosynthesis also has the added benefit of discovering an unexpected pathway shortening bypass due to unexpected gains in activity acquired through directed evolution.

Although engineered biocatalysts have become quite commonly used for individual biotransformations, there are a number of issues as to why biocatalysts have not been coupled together into non-natural biosynthetic pathways. First and foremost, the process of enzyme engineering is not trivial, and it is likely that many if not all enzymes in the proposed biosynthetic pathway would require some degree of evolution for activity on the non-natural substrate. Second, the chance of success in reaching an appropriate level of activity for practical application cannot be determined prior to investing substantial effort into the research, and there may be just as many unreported failures as there are published successes meeting the desired goal. A pathway requiring the engineering of multiple enzymes increases this risk of not meeting the desired production or activity goal as each enzyme must be evolved for optimal flux through the pathway. A final factor in contributing to the low number of non-natural biosynthetic pathways is a lack of a generalizable approach to conceptualizing, constructing and evolving the pathway for desired production. An individual case by case optimization strategy does not allow for readily transferable methods to apply in other scenarios.

Our demonstration of non-natural pathway construction and optimization through bioretrosynthesis begins to address a number of these limiting factors. The premise that all enzymes in the pathway are evolved and selected for activity based on turnover of the product forming enzyme theoretically requires that only one assay need be

developed for screening, and in some instances it may be easily adapted to detect formation of the final product or consumption of the final cosubstrate. This can significantly reduce the burden of developing individual screening assays for evolution of each enzyme in the pathway and allow a more streamlined process to engineer each biosynthetic enzyme.

Although the degree of enzyme improvement through directed evolution can never be known *a priori*, an advantage of bioretrosynthesis is that the activity of each enzyme is engineered in the presence of all enzymes required for subsequent biosynthetic steps. In this way not only is the target enzyme activity evolved, but also the production of the pathway is engineered as a cooperative unit, providing initial measures for developing a system with optimal metabolic flux. Pathway construction by this method may prevent the obstacle in some cases where absolute maximal turnover rate by an enzyme actually decreases pathway product yield due to inhibition by high levels of substrates or products and can give indications for future optimization by regulated gene expression.

Finally, and most significantly, applied bioretrosynthesis introduces what has the potential to become a widely applicable strategy for practical conceptualization, construction and optimization of *de novo* non-natural pathways. The focus on final product formation minimizes the chance of pathway deviations through creating side products, and reduces the risk of assembling production into dead end pathways. Aside from the unexpected discovery of activity creating a pathway bypass, which is still indeed an added advantage although likely a more rare finding, the results described within this work present a variety of likely general benefits of this approach and indicate the potential of bioretrosynthesis as a widely applicable and practical strategy for non-natural pathway construction to produce both natural and non-natural products.

Future Directions

Directed evolution was successfully used to increase activity of PPM and RK for the non-natural substrates dideoxyribose-5-phosphate and dideoxyribose, respectively. Additionally, the full pathway consisting of the engineered RK-PPM-PNP and the partial pathway of evolved RK-PNP were shown to produce dideoxyinosine at significantly higher levels than the wild-type enzymes. However, production through this system is still much too low and inefficient to be useful on a large scale. Further mutagenesis and activity engineering for all three enzymes RK, PPM and PNP would likely be necessary to increase production to titers more appropriate for scaled up biosynthetic studies.

PPM was evolved through saturation mutagenesis, epPCR and recombination of mutations, however, improvements failed after the third round of mutagenesis in a second and third attempt at epPCR even under altered conditions to slightly increase the rate of mutagenesis. Continued evolution of this enzyme may require application of different mutagenesis methods, such as gene shuffling[26] or a much higher mutagenesis rate in epPCR, to increase catalysis. Additionally, targeting active site residues for mutation by saturation mutagenesis or CASTing[27] in the 4H11 variant my yield new beneficial results now that the active site has been slightly adjusted by the Val158Leu mutation. The flexible Arg193 residue might be an interesting target for this strategy due to the induced change in conformation upon ribose-5-phosphate binding in wild-type PPM[28] and the proposed stearic clash with the Val158Leu mutation based on observations from the 4H11 crystal structure.

Preliminary engineering of *E. coli* RK resulted in a surprising gain in dideoxyribokinase activity in the Asp16Ala variant via a previously undetected phosphorylation regiochemistry at the C1 position. This activity is a very significant benefit for pathway production and provides a unique foundation to build upon through further mutagenesis, which would be absolutely required for increased production

volumes. Initial analysis of RK crystal structures suggest that Asn14 and Gly41 might be interesting targets for saturation mutagenesis to gain activity or substrate selectivity, and could be performed either individually or simultaneously to determine if there may be a cooperative effect due to residue identity at these positions. Considerations for whether activity would solely be analyzed in the shortened pathway (RK and PNP), the full pathway (RK, PPM and PNP) or both would need to be taken into account in the assay design since additional active site mutations may improve only one or the other of C1 or C5 phosphorylation activity on dideoxyribose and some beneficial activity gains may be missed without full characterization.

In addition to further evolution for greater improvement in turnover and substrate selectivity of the enzymes currently in the dideoxyinosine biosynthetic pathway, more enzymes could be added to extend the biosynthetic component in either direction. Optimization of precursor supply and product efflux may each have beneficial influences on production titers. An enzymatic retro-extension could entail recruitment of an alcohol dehydrogenase to oxidize (*S*)-1,2,5-propanetriol (dideoxyribitol) to dideoxyribose. Production of the polyol is easily accessible in two steps from L-glutamic acid with relatively high yield[(29)], and would reduce the chemical synthesis requirement through replacement with an additional biocatalytic step. On the other end of the pathway, a forward extension to an already engineered nucleoside 2'-deoxyribosyltransferase[(30)] could diversify the pool of available dideoxynucleosides through exchanging hypoxanthine for a different natural or non-natural nucleobase.

Although it would likely require the engineering of several more enzymes, it may be possible to extend biosynthesis back to simple and abundant natural precursors. For example, pyruvate and glycolaldehyde could undergo an aldol condensation to form a 4,5-dihydroxy-2-keto acid that is subsequently extended by one carbon unit through the activity of an engineered 2-hydroxy-3-oxoadipate synthase. This six carbon sugar acid

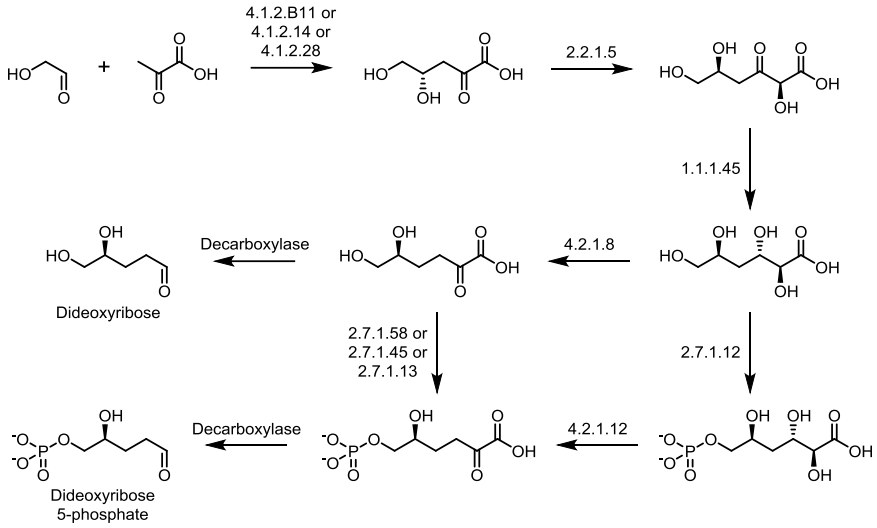

Figure 6-1. Possible biosynthetic routes from pyruvate and glycolaldehyde to dideoxyribose and dideoxyribose 5-phosphate. Enzyme names corresponding to each EC number and the naturally catalyzed reactions are provided in Figure 6-2.

analog could then proceed through one of several conceivable pathways consisting of sugar acid modifying enzymes and a decarboxylase to yield either dideoxyribose or possibly the advanced intermediate dideoxyribose 5-phosphate (Figure 6-1). The specific route to either substrate, however, would ultimately be determined by which enzymes are active or evolvable for activity on the non-natural substrates. Again, this may require the evolution of several enzymes since many of the proposed intermediates are non-natural substrates for the suggested enzymes (Figure 6-2), nonetheless, production through this or a similar pathway could provide additional advantages to the biosynthetic process through linking the pathway to central metabolism.

The series of RK, PPM and PNP presented here, likely after further directed evolution for activity changes, could also possibly find use in biosynthetic production of other 2'- and 3'-substituted nucleoside analogs from the corresponding substituted sugar analogs or the enzymes may even be found individually useful in other biocatalytic processes. In summary, the enzyme evolution and pathway optimization work described in this dissertation provides a foundation for biosynthetic production of nucleoside analogs that can be further optimized, evolved, extended and engineered to increase

yield and broaden application to produce a diverse group of molecules. The bioretrosynthesis model of pathway design described and demonstrated here also has the potential to be a widely applicable pathway construction and optimization strategy, hopefully enabling greater success in accessing non-natural products through *de novo* biosynthetic pathways.

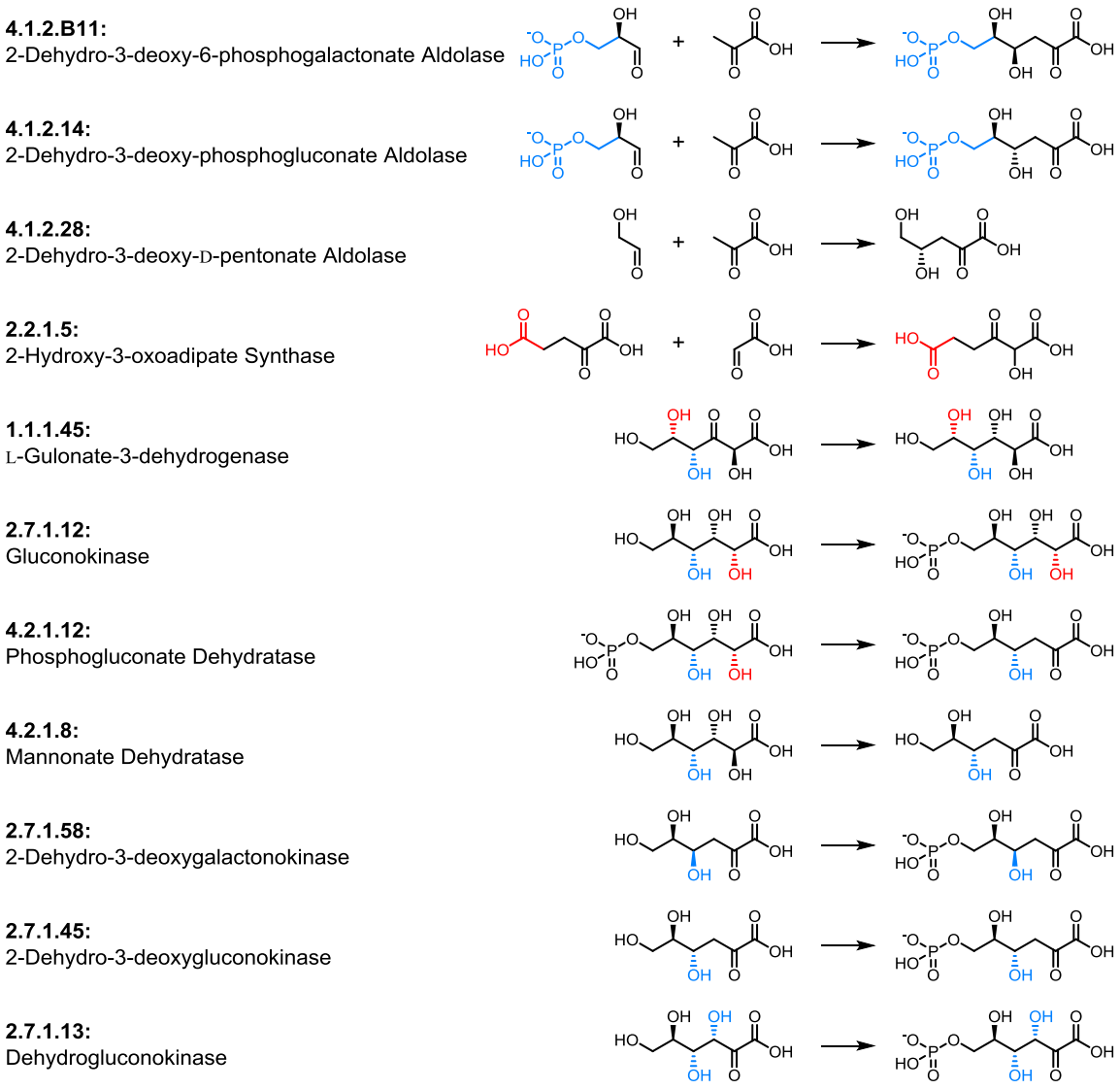

Figure 6-2. Natural reactions catalyzed by enzymes proposed for dideoxyribose and dideoxyribose 5-phosphate biosynthesis. Portions highlighted in blue indicate aspects not present in the non-natural substrates proposed for the pathway. Red indicates functional groups that are altered (i.e. different stereochemistry) in the non-natural substrates.

References

1. Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., and Huisman, G. W. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007**, 25, 338-344.

2. Greenberg, W. A., Varvak, A., Hanson, S. R., Wong, K., Huang, H., Chen, P., and Burk, M. J. Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 5788-5793.

3. Liang, J., Lalonde, J., Borup, B., Mitchell, V., Mundorff, E., Trinh, N., Kochrekar, D. A., Cherat, R. N., and Pai, G. G. Development of a biocatalytic process as an alternative to the (-)-DIP-Cl-mediated asymmetric reduction of a key intermediate of montelukast. *Org. Process. Res. Dev.* **2010**, 14, 193-198.

4. Patel, R. N. Biocatalysis: Synthesis of key intermediates for development of pharmaceuticals. *ACS Catal.* **2011**, 1, 1056-1074.

5. Ran, N., and Frost, J. W. 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.* **2007**, 129, 6130-6139.

6. Gao, X., Xie, X., Pashkov, I., Sawaya, M. R., Laidman, J., Zhang, W., Cacho, R., Yeates, T. O., and Tang, Y. Directed evolution and structural characterization of a simvastatin synthase. *Chem. Biol.* **2009**, 16, 1064-1074.

7. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, 329, 305-309.

8. Nakamura, C. E., and Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **2003**, 14, 454-459.

9. Niu, W., Molefe, M. N., and Frost, J. W. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **2003**, 125, 12998-12999.

10. Xie, D. M., Shao, Z. Y., Achkar, J. H., Zha, W. J., Frost, J. W., and Zhao, H. M. Microbial synthesis of triacetic acid lactone. *Biotechnol. Bioeng.* **2006**, 93, 727-736.

11. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **2011**, 7, 445-452.

12. Liu, W., and Wang, P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol. Adv.* **2007**, 25, 369-384.

13. Oroz-Guinea, I., and Garcia-Junceda, E. Enzyme catalysed tandem reactions. *Curr. Opin. Chem. Biol.* **2013**, 236-249.

14. Ma, S. K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G. W., and Sheldon, R. A. A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem.* **2010**, 12, 81.

15. Atsumi, S., Hanai, T., and Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, 451, 86-U13.

16. Dellomonaco, C., Clomburg, J. M., Miller, E. N., and Gonzalez, R. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **2011**, 476, 355-U131.

17. Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **2010**, 463, 559-U182.

18. Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N. S., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D., and Yoshikuni, Y. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **2012**, 335, 308-313.

19. Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., and Stephanopoulos, G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **2010**, 330, 70-74.

20. Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Diola, D., Dorin, G., Ellens, K. W., Fickes, S., Galazzo, J., Gaucher, S. P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L. F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P. S., Keasling, J. D., Reiling, K. K., Renninger, N. S., and Newman, J. D. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013**, 496, 528-532.

21. Prather, K. L., and Martin, C. H. *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **2008**, 19, 468-474.

22. Bachmann, B. O. Biosynthesis: Is it time to go retro? *Nat. Chem. Biol.* **2010**, 6, 390-393.

23. Horowitz, N. H. On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U.S.A.* **1945**, 31, 153-157.

24. Nannemann, D. P., Kaufmann, K. W., Meiler, J., and Bachmann, B. O. Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010**, 23, 607-616.

25. Pinheiro, E., Vasan, A., Kim, J. Y., Lee, E., Guimier, J. M., and Perriens, J. Examining the production costs of antiretroviral drugs. *AIDS* **2006**, 20, 1745-1752.

26. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 10747-10751.

27. Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. X., and Vogel, A. Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* **2005**, 44, 4192-4196.

28. Panosian, T. D., Nannemann, D. P., Watkins, G. R., Phelan, V. V., McDonald, W. H., Wadzinski, B. E., Bachmann, B. O., and Iverson, T. M. *Bacillus cereus* phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle. *J. Biol. Chem.* **2011**, 286, 8043-8054.

29. Hercouet, A., Bessieres, B., LeCorre, M., and Toupet, L. Synthesis of optically active 2,3-methanopipecolic acid. *Tetrahedron Lett.* **1996**, 37, 4529-4532.

30. Kaminski, P. A., Dacher, P., Dugue, L., and Pochet, S. In vivo reshaping the catalytic site of nucleoside 2'-deoxyribosyltransferase for dideoxy- and didehydronucleosides via a single amino acid substitution. *J. Biol. Chem.* **2008**, 283, 20053-20059.

APPENDIX A

NMR Spectra

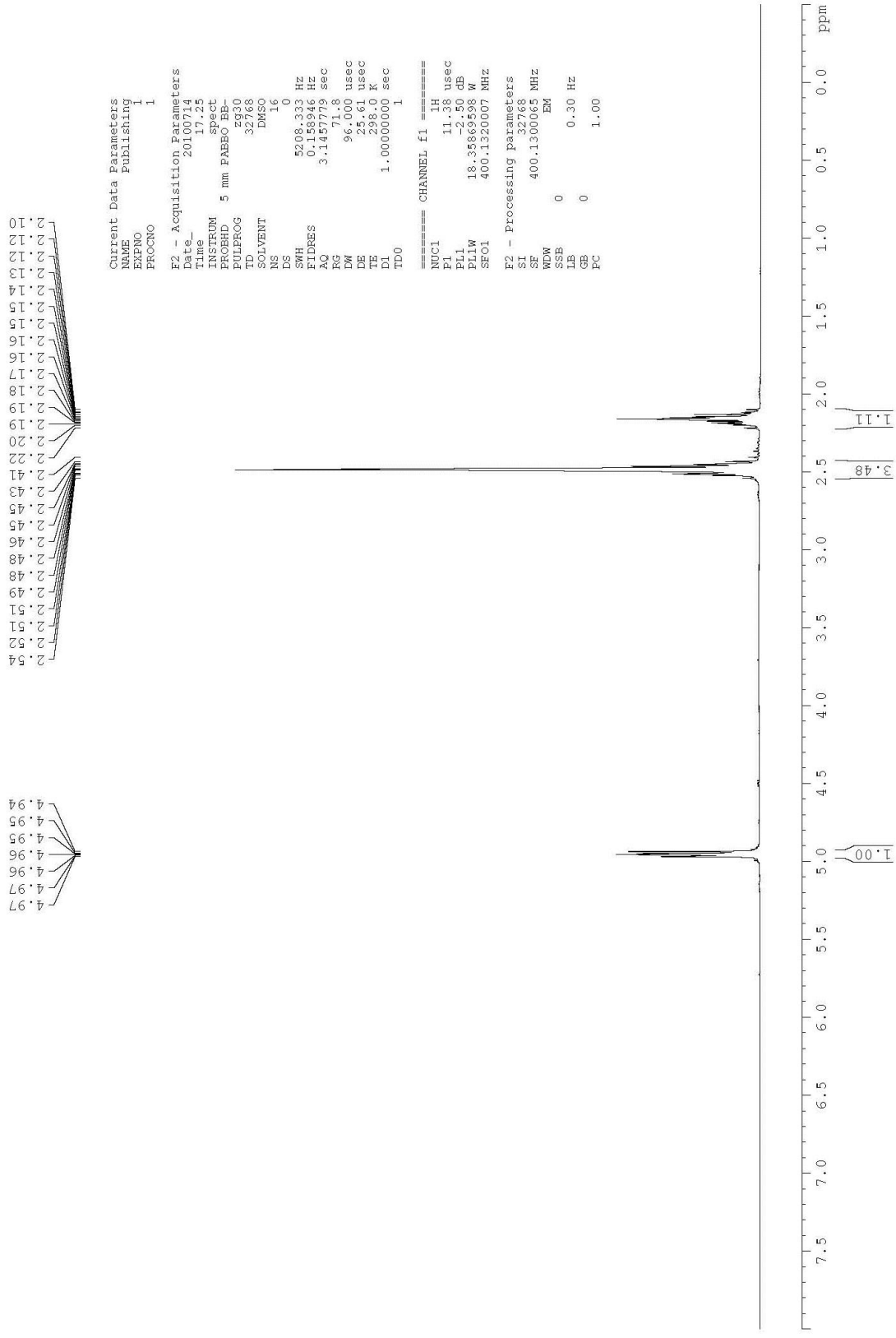

Figure A-1. ^{1}H NMR of (S)-γ-butyrolactone-γ-carboxylic acid in d_6-DMSO.

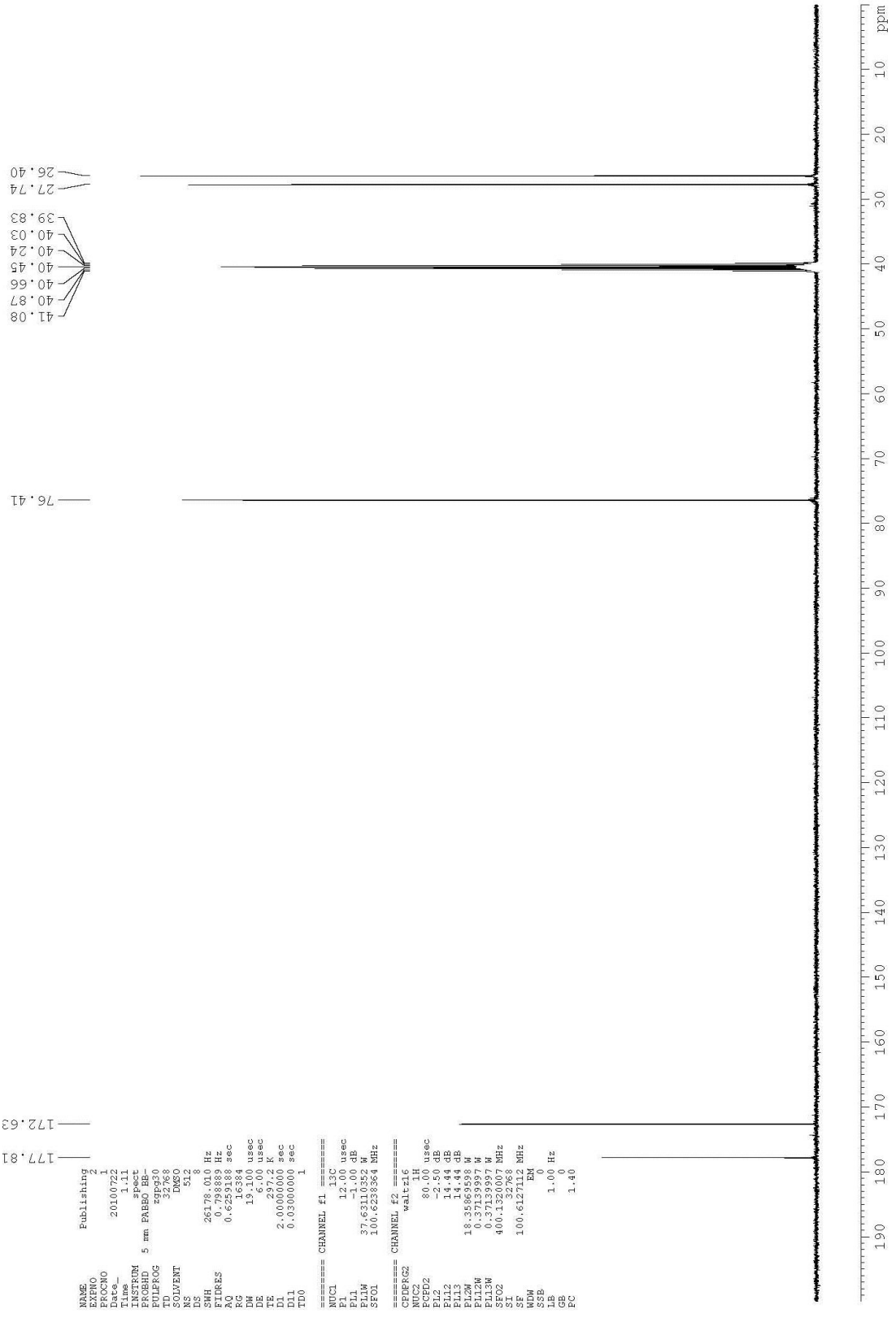

Figure A-2. ^{13}C NMR of (S)-γ-butyrolactone-γ-carboxylic acid in d_6-DMSO.

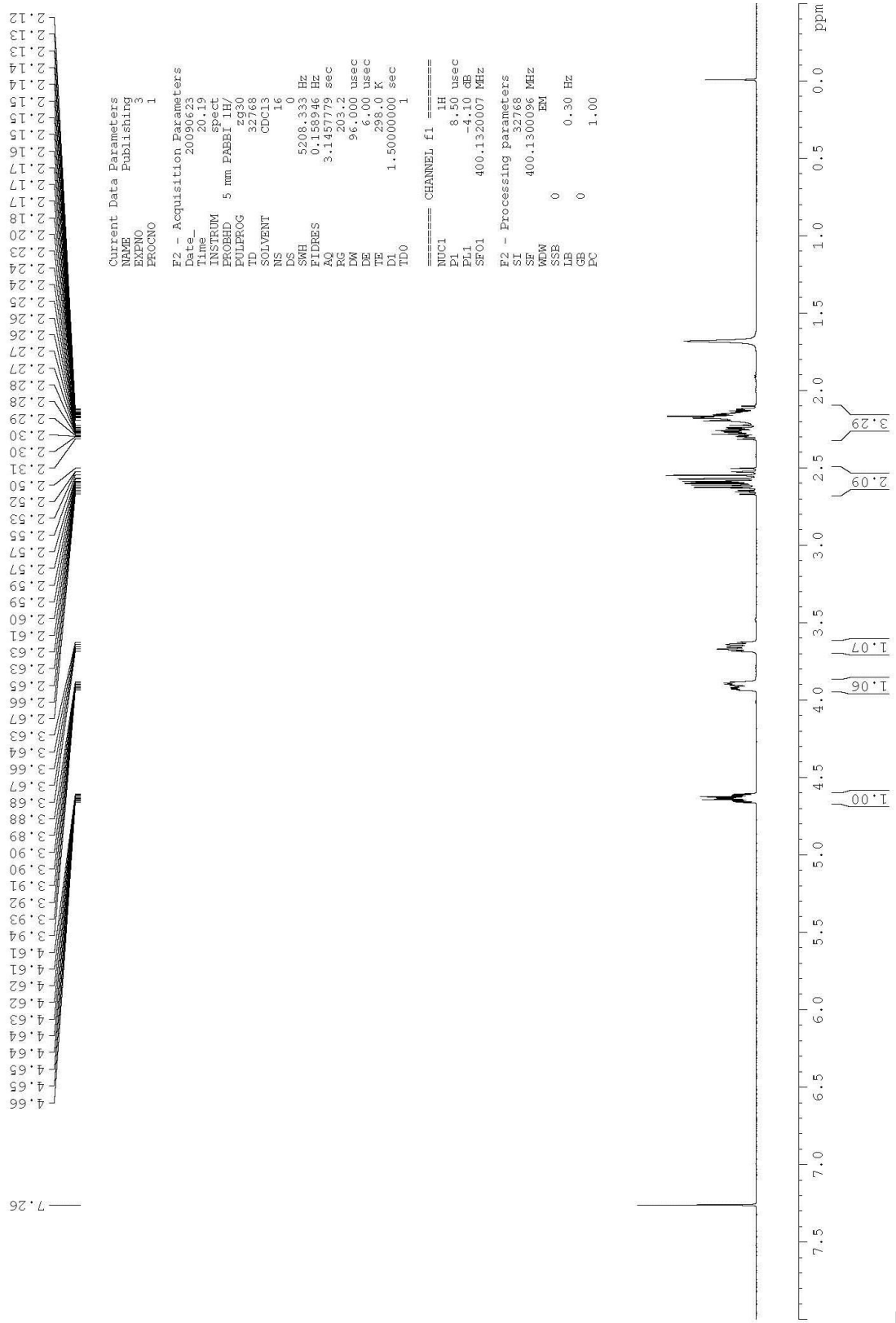

Figure A-3. ^{1}H NMR of (S)-γ-(hydroxymethyl)-γ-butyrolactone in $CDCl_3$.

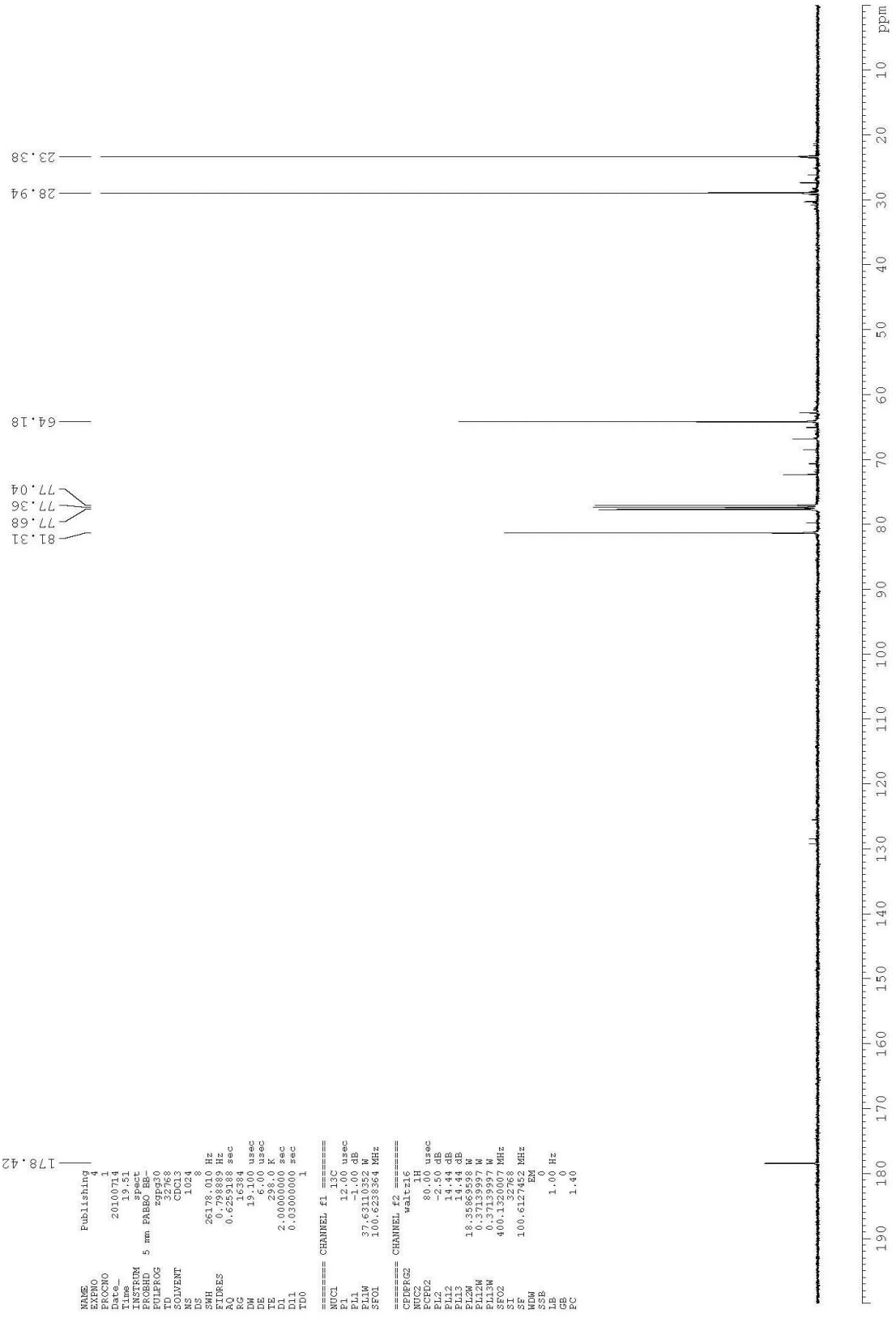

Figure A-4. ^{13}C NMR of (S)-γ-(hydroxymethyl)-γ-butyrolactone in $CDCl_3$.

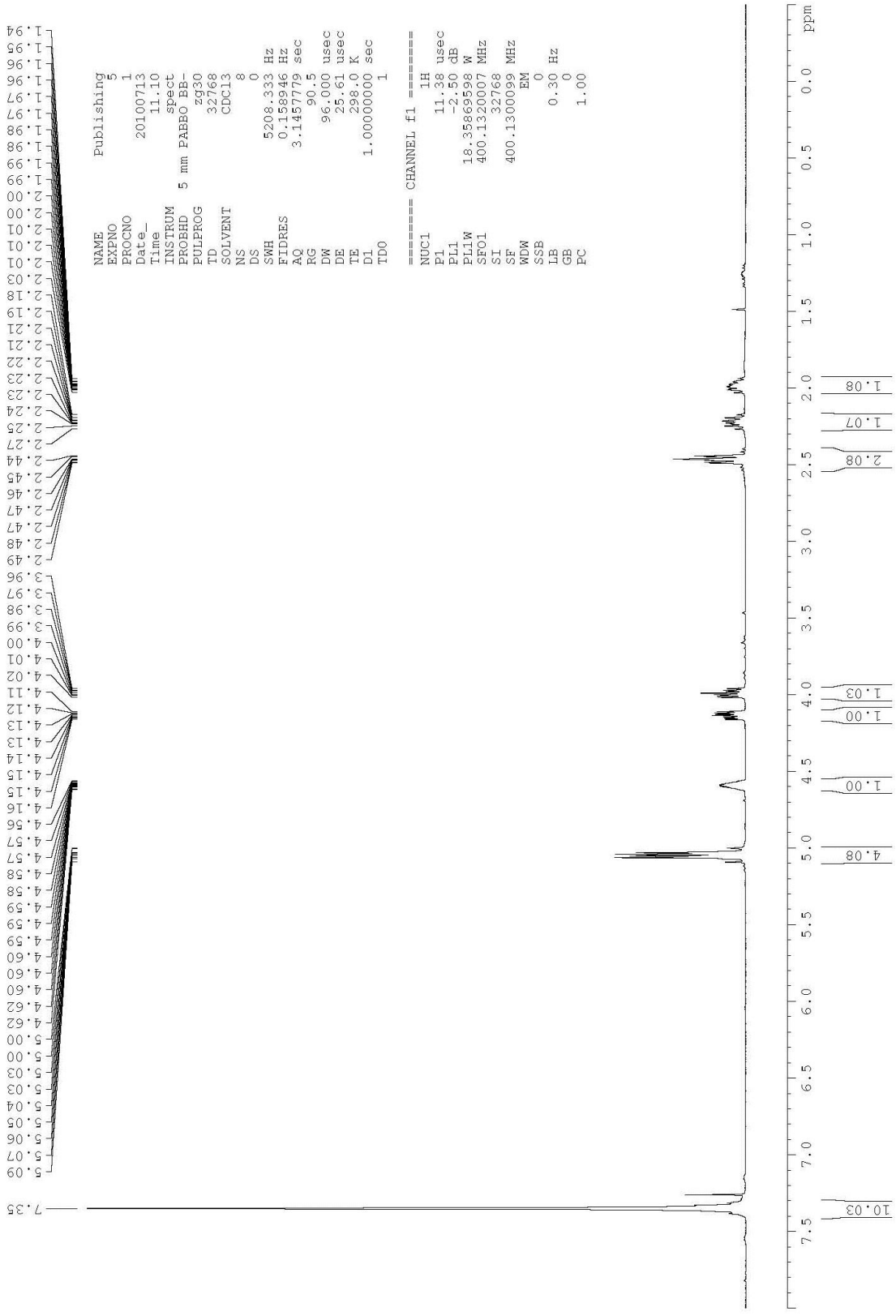

Figure A-5. ^{1}H NMR of (S)-γ-(dibenzylphosphomethyl)-γ-butyrolactone in $CDCl_3$.

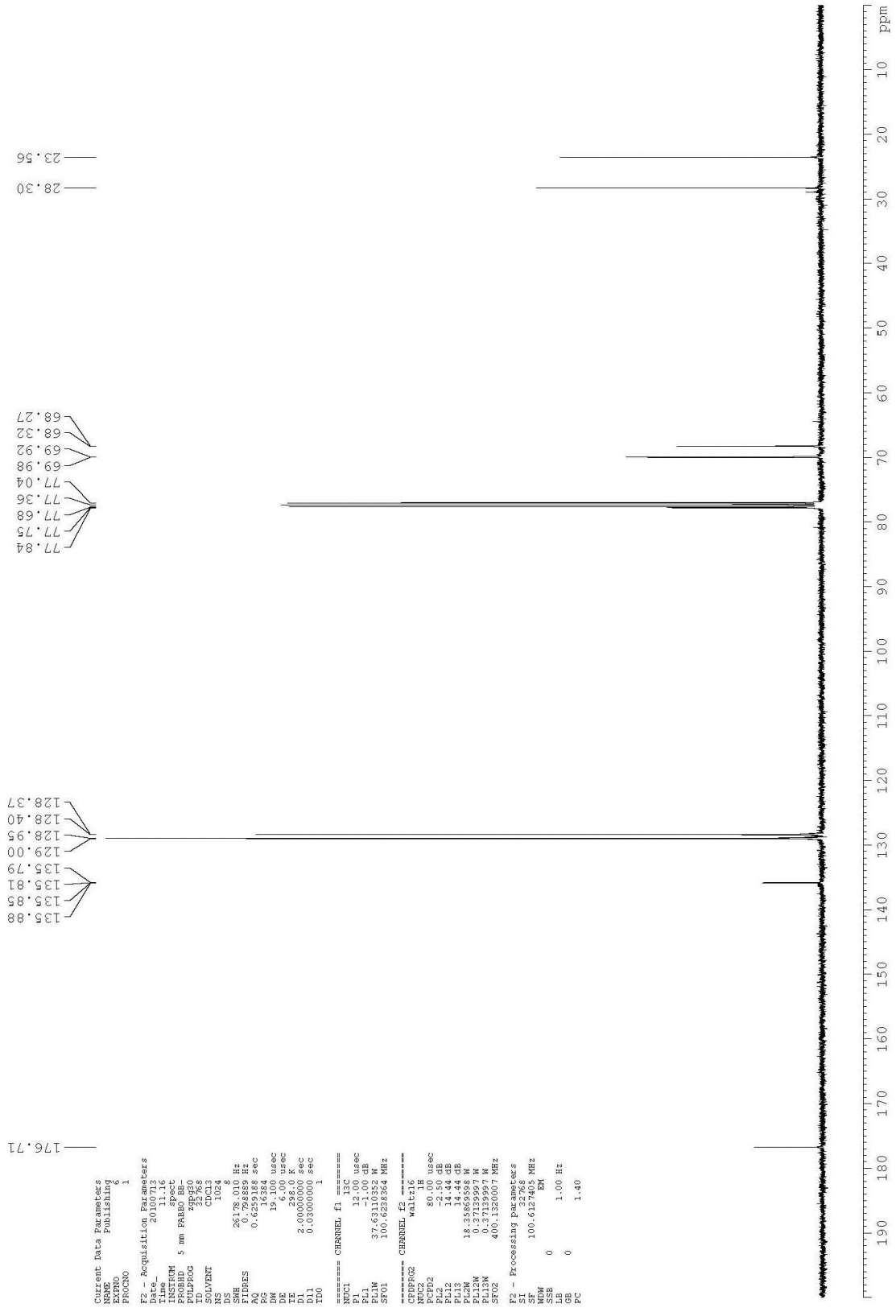

Figure A-6. ^{13}C NMR of (S)-γ-(dibenzylphosphomethyl)-γ-butyrolactone in $CDCl_3$.

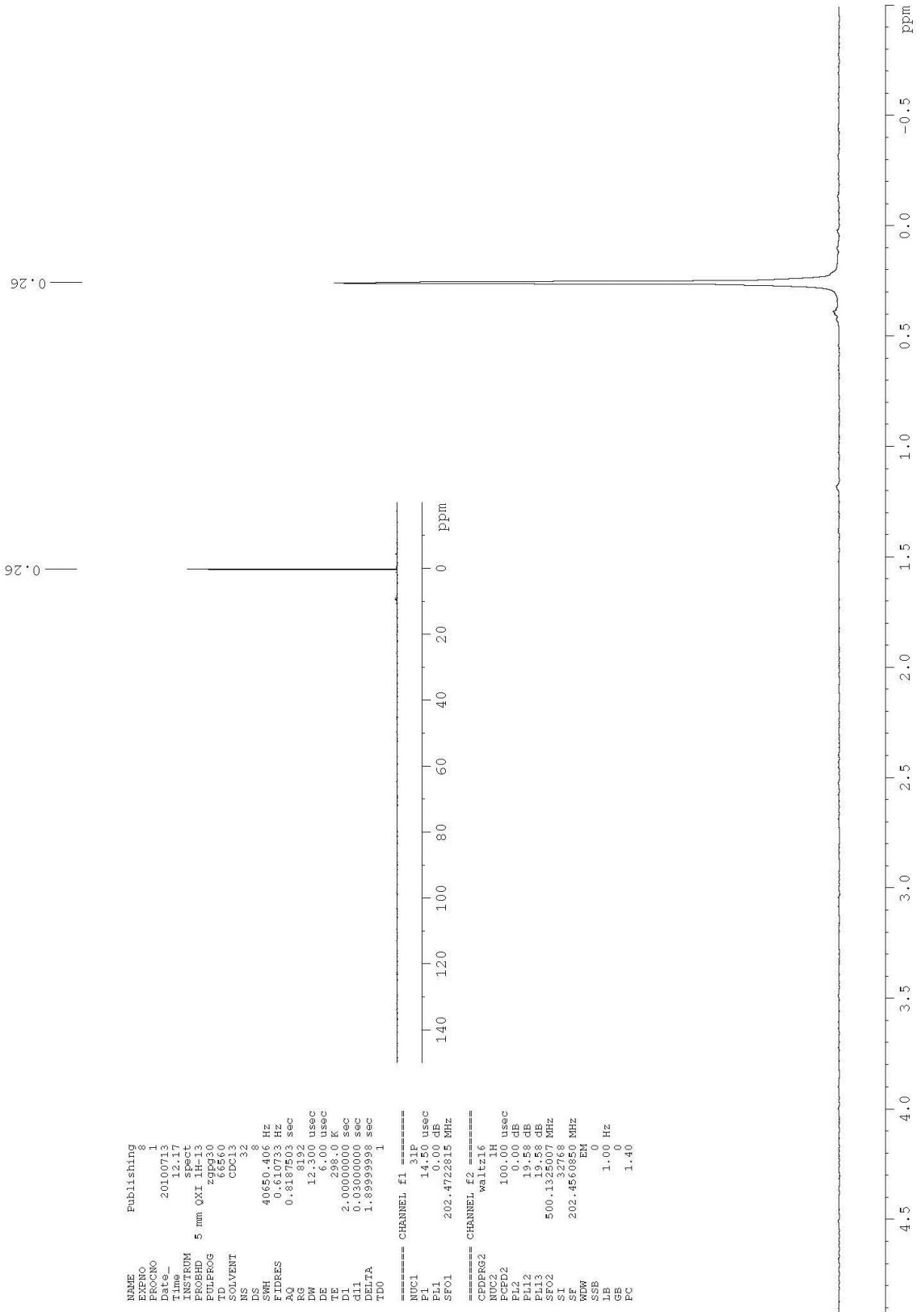

Figure A-7. ^{31}P $\{^{1}H\}$ NMR of (S)-γ-(dibenzylphosphomethyl)-γ-butyrolactone in $CDCl_3$.

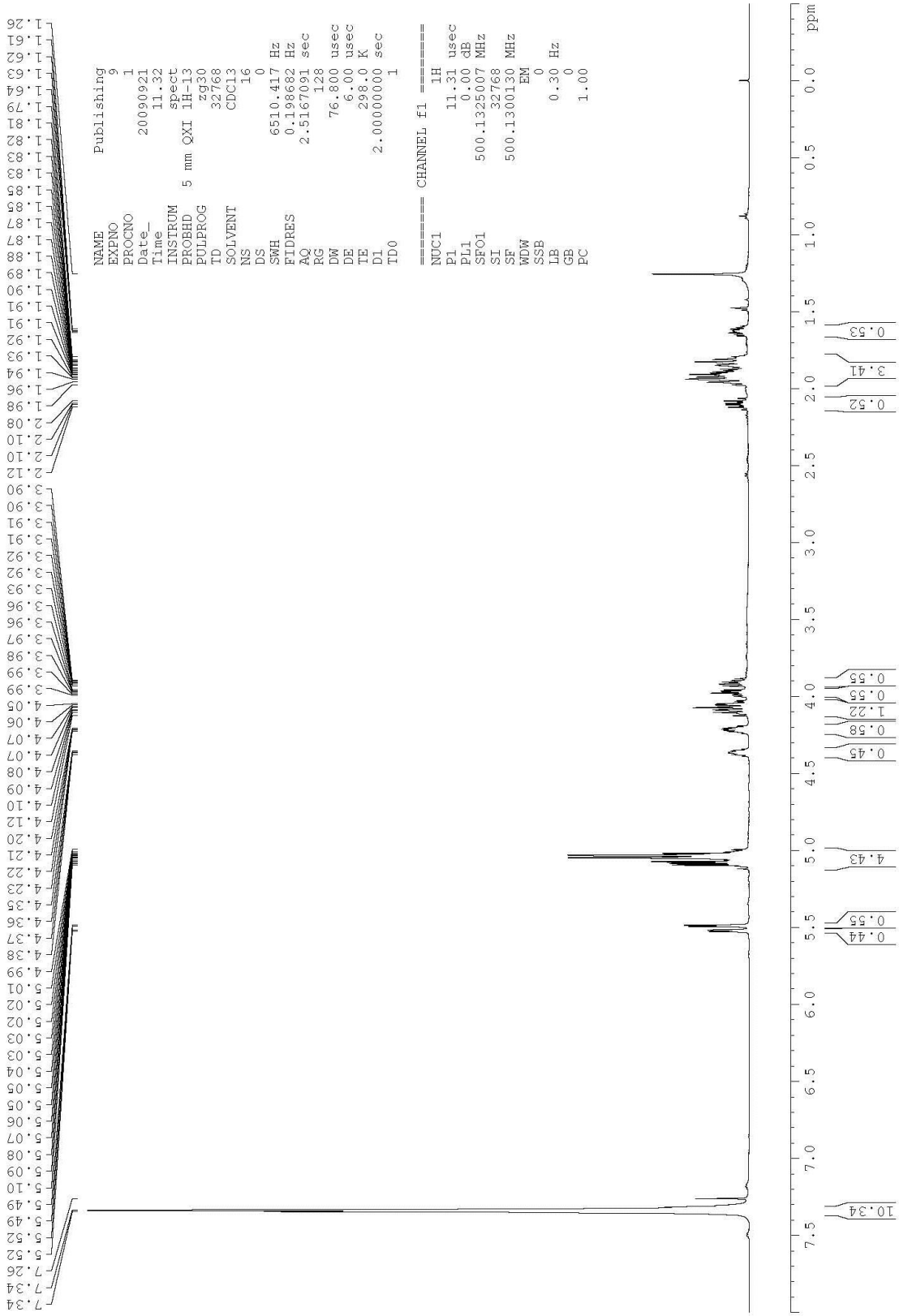

Figure A-8. ^{1}H NMR of 2,3-dideoxyribose-5-(di-O-benzyl)phosphate in $CDCl_3$.

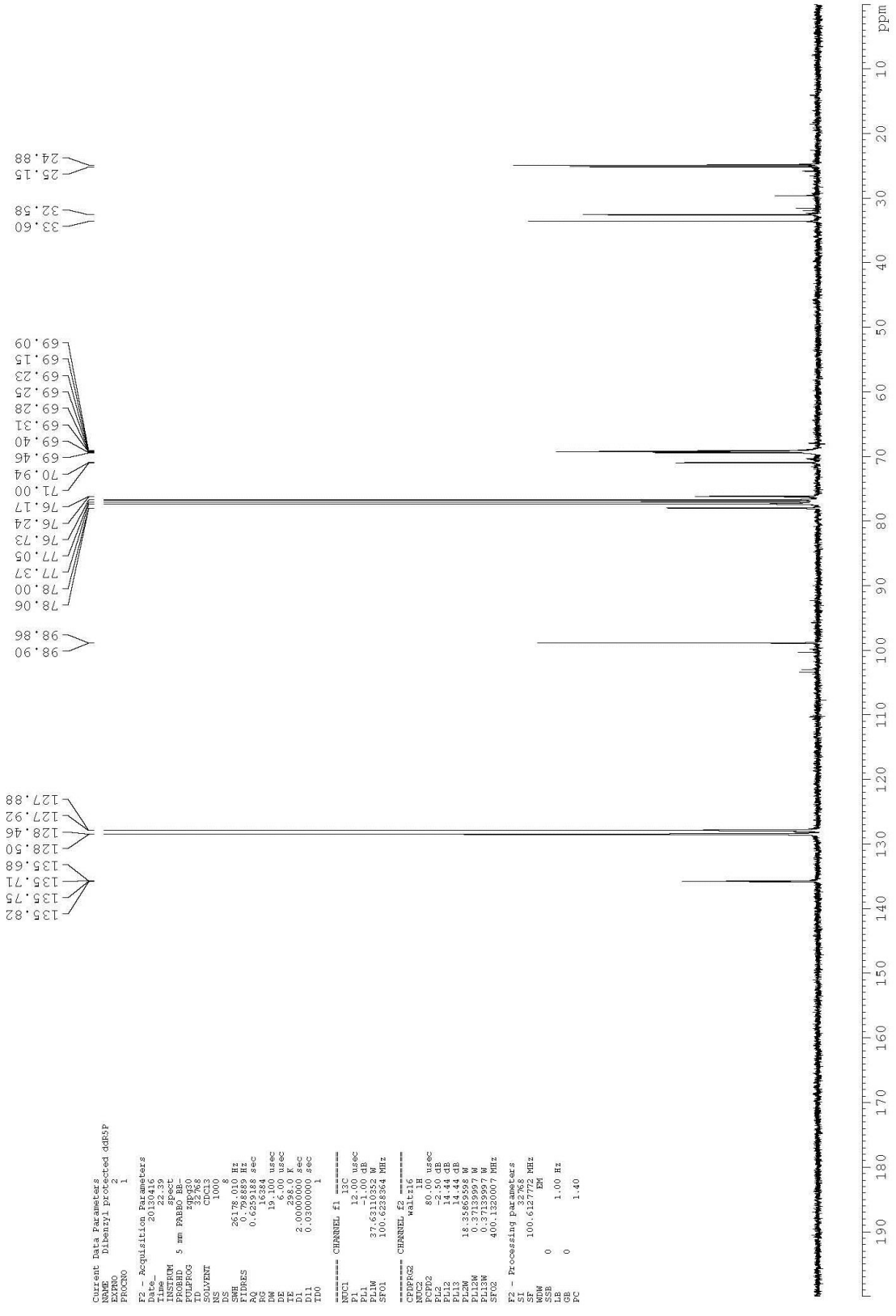

Figure A-9. ^{13}C NMR of 2,3-dideoxyribose-5-(di-O-benzyl)phosphate in $CDCl_3$.

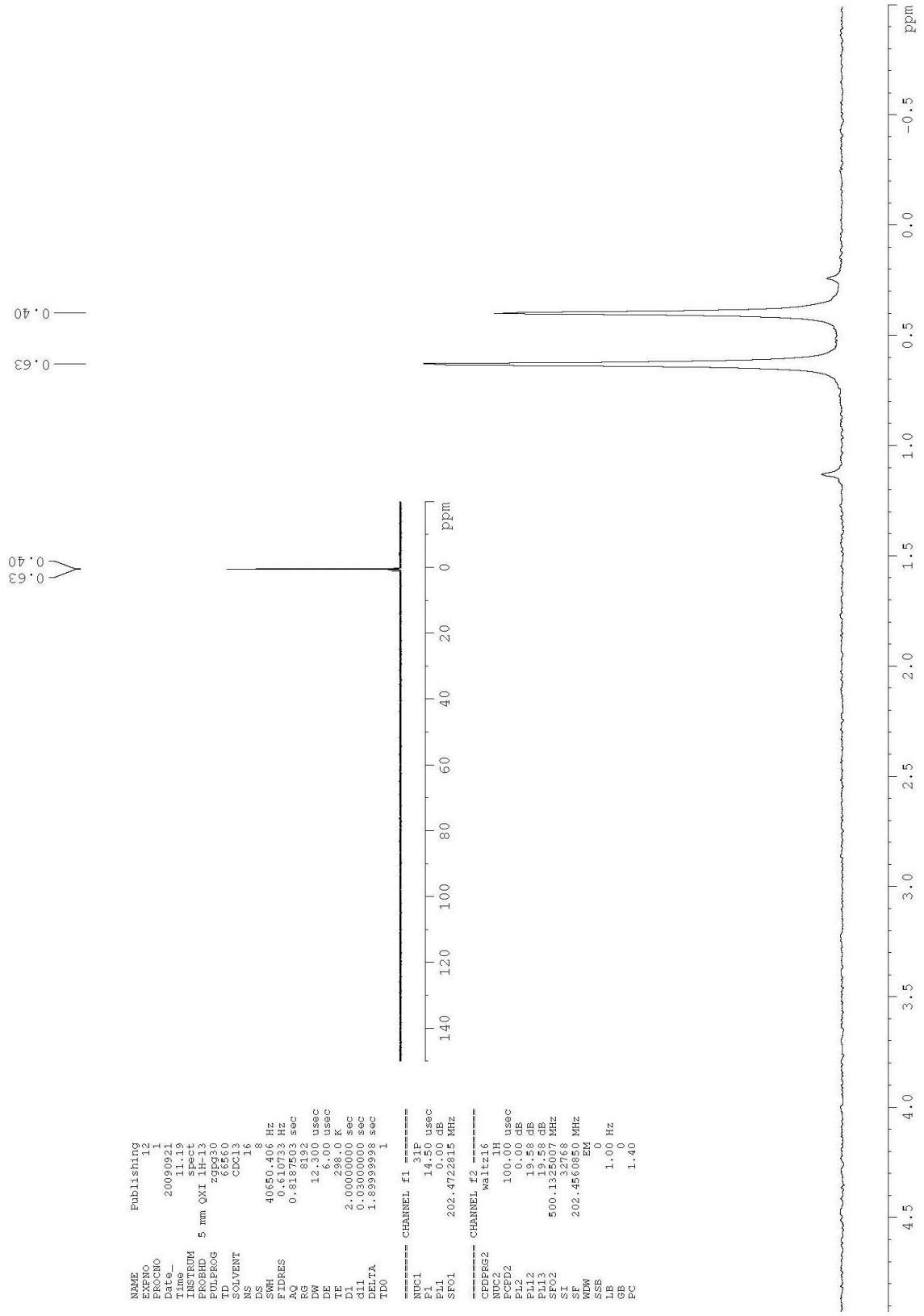

Figure A-10. ^{31}P $\{^{1}H\}$ NMR of 2,3-dideoxyribose-5-(di-O-benzyl)phosphate in $CDCl_3$.

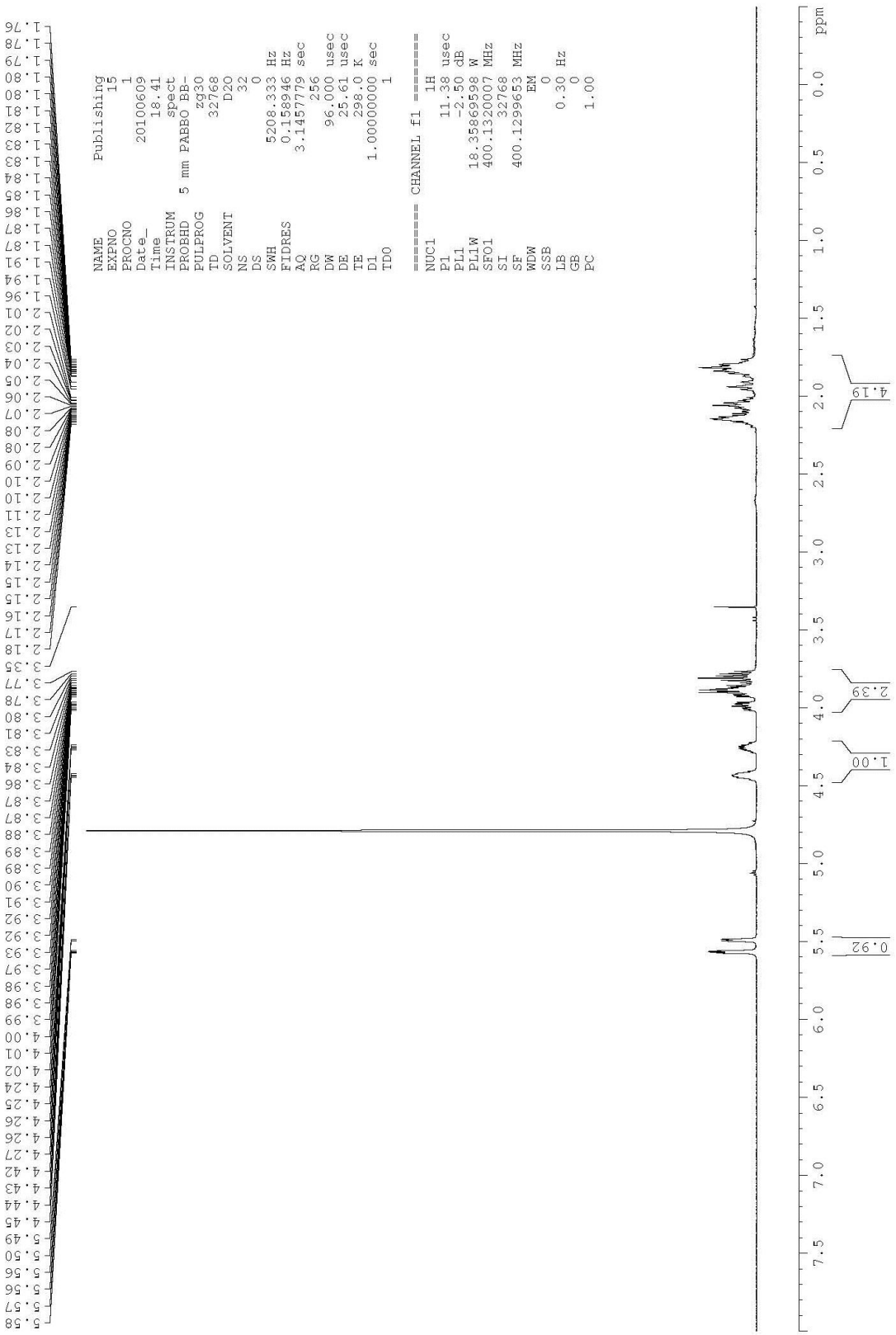

Figure A-11. ^{1}H NMR of 2,3-dideoxyribse 5-phosphate, disodium salt in D_2O.

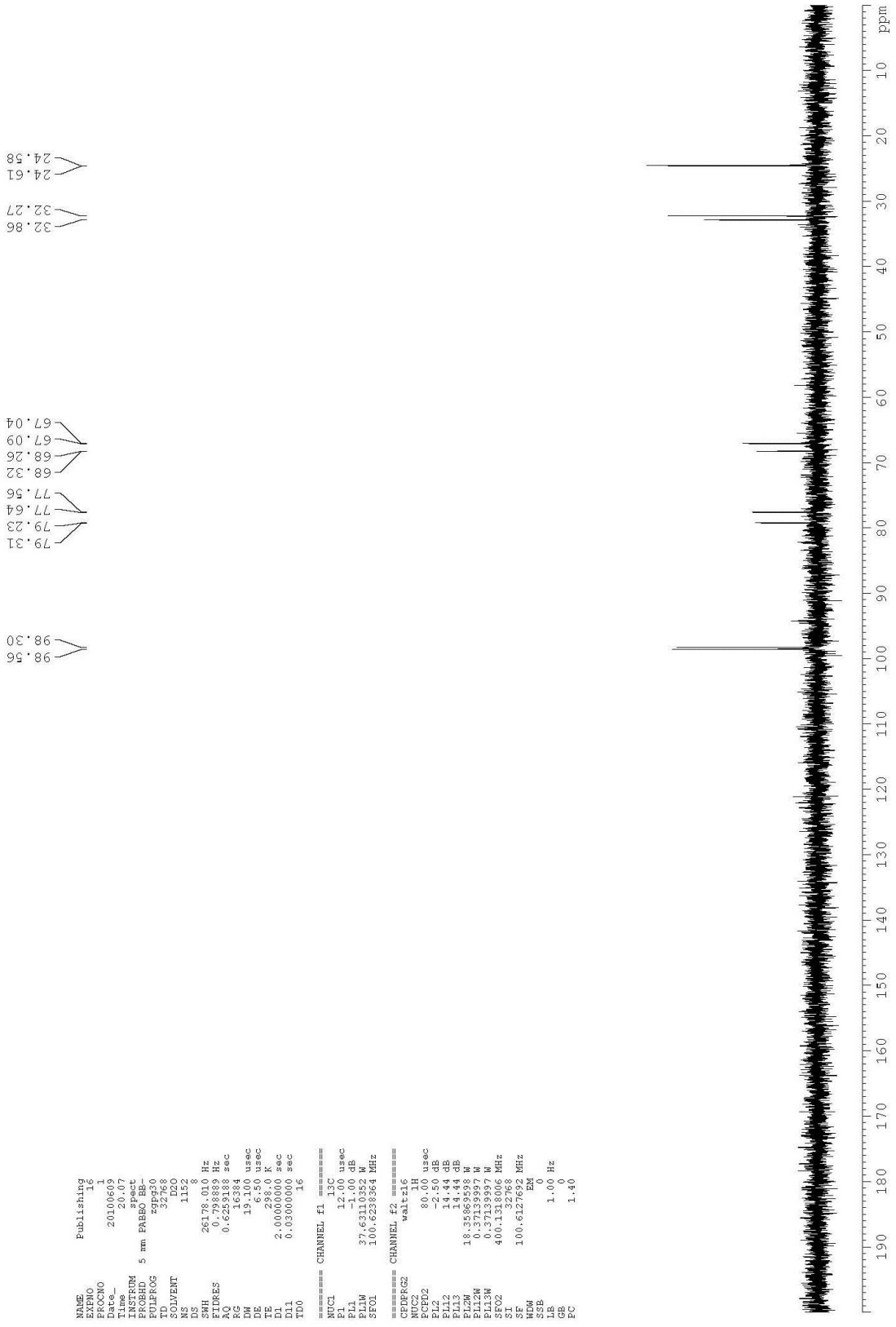

Figure A-12. ^{13}C NMR of 2,3-dideoxyribse 5-phosphate, disodium salt in D_2O.

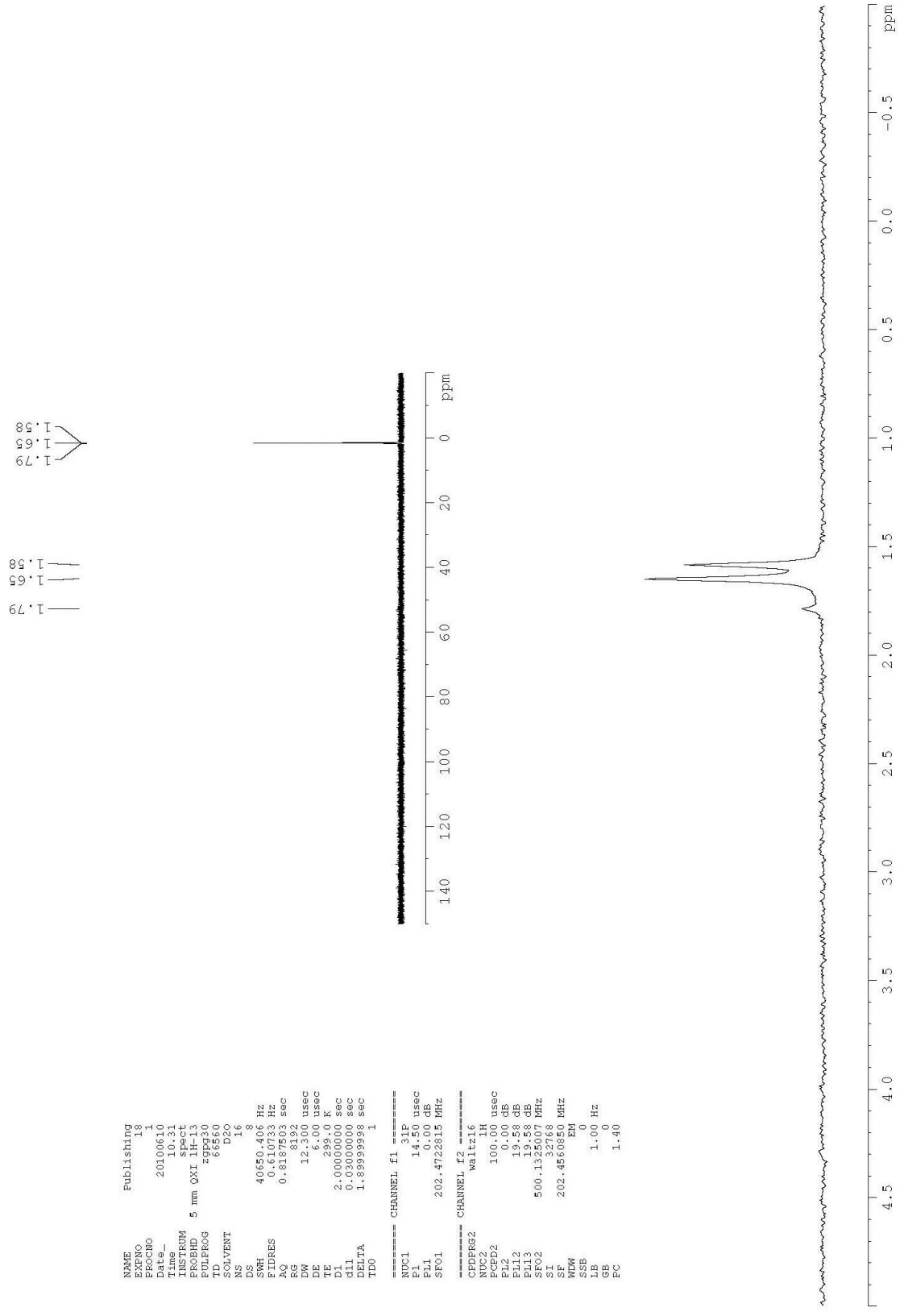

Figure A-13. ^{31}P {^{1}H} NMR of 2,3-dideoxyribse 5-phosphate, disodium salt in D_2O.

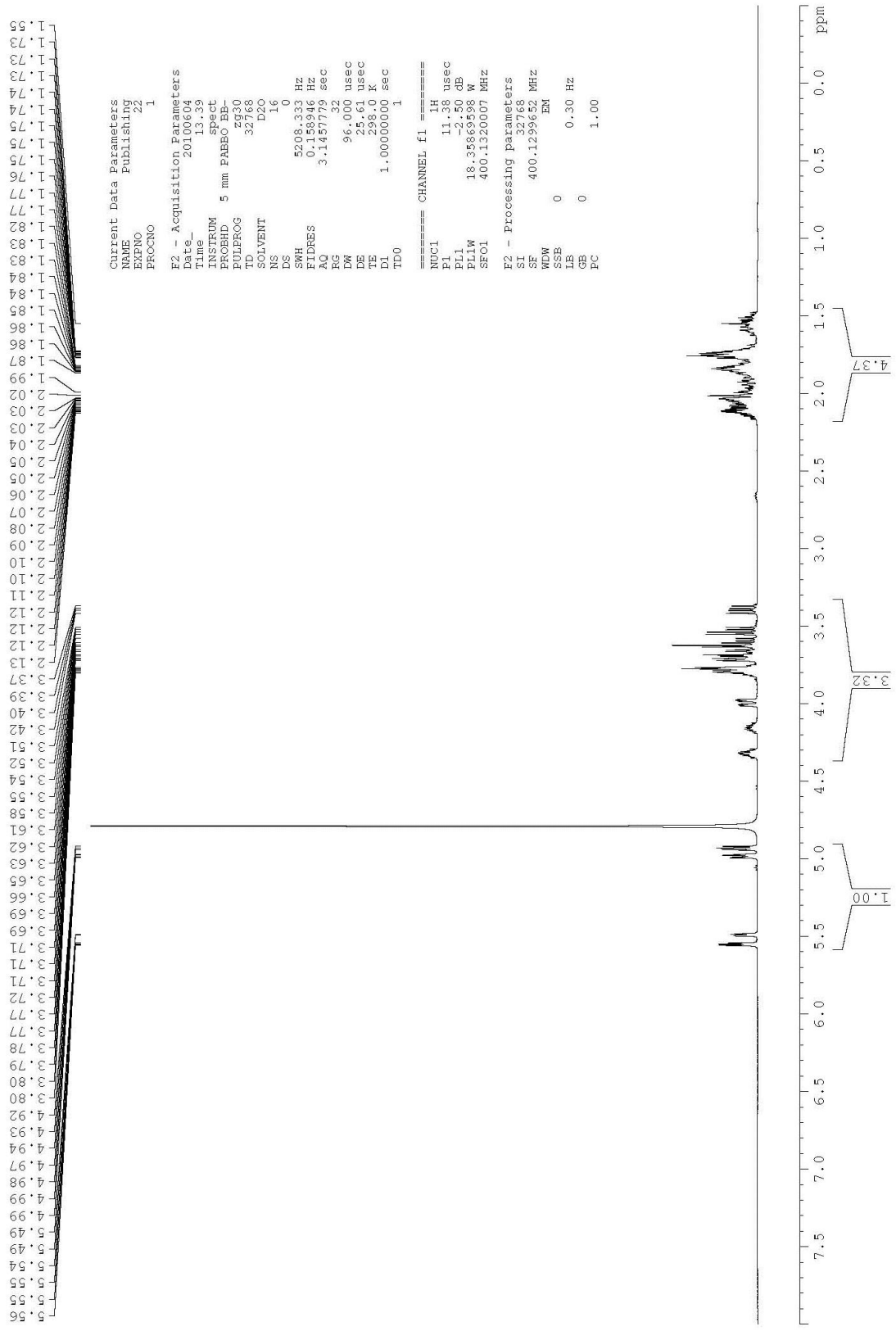

Figure A-14. ^{1}H NMR of 2,3-dideoxyribose in D_2O.

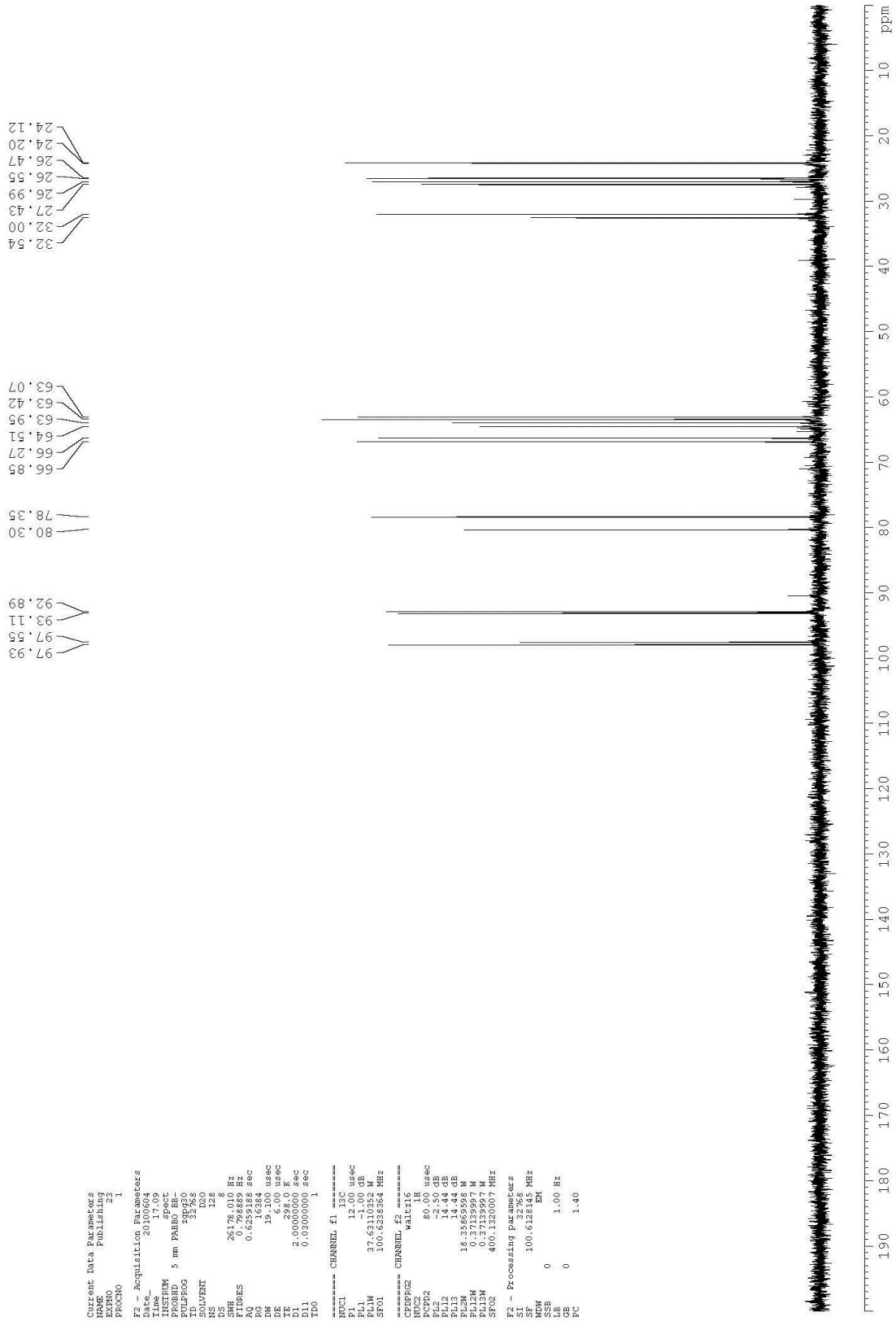

Figure A-15. ^{13}C NMR of 2,3-dideoxyribose in D_2O