DIRECTED BIOSYNTHESIS OF THE NUCLEOSIDE ANALOG DRUG DIDANOSINE

By

David Patrick Nannemann

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

Chemistry

May, 2011

Nashville, Tennessee

Approved:

Brian O. Bachmann

Jens Meiler

Richard N. Armstrong

F. Peter Guengerich

To my wife, Allison, and family,

for their never-ending love and support.

ACKNOWLEDGEMENTS

Any expressions of thanks I write here will certainly be insufficient. Many people have helped me in my path to becoming “Dr. Dave”, a term coined long ago when I first knew I would get an upper level science degree. That I would attain a degree in Chemistry was not foreseen. My initial plans took me through medical school and on to forensic pathology which was a plan hatched early and worked at for many years. Deciding to abandon that career path and look for another was only accomplished through much prayer and Godly inspiration. Which brings me to my first major “thank you”: I thank God the Father for His infinite love and for sending his Son to reconcile my sins to Him.

Secondly, I would like to thank my wife for supporting me through graduate school and her leap of faith in me to come to Nashville so I could attend Vanderbilt University. Diamond earrings are an insufficient token of my love and affection for her, and sincere gratitude for caring for us while I concentrate on school. I first fell in love with Allison while watching her care for and give unconditionally to numerous people at Camp Eagle in her roles on the activities staff, in the Camp Store, as Emergency Room Expert, and all-around mischief maker. That she is a great mother to my first child, Katelyn, is not surprising and I know Allison will never fail in caring for Kate and any other children we have.

Katelyn is always a joy. Her smile and excitement for my arrival at home each evening makes the time away bearable and the day’s stresses disappear. I pray that she takes her enthusiasm and energy into all aspects of her life.

A major component of my graduate work utilizes directed evolution. I would not have had the confidence to take on such a project without the guidance of two undergraduate professors, Dr. Richard Garner and Dr. Diana Flanagan. Dr. Garner

pointed out to me that God created the earth, and it doesn't matter how. Whether the universe came into existence through spoken word, a big bang, evolution or some combination is immaterial to my salvation in Jesus Christ. Since then it has become my belief that God created an earth resembling this one. And with the original sin and fall of man, the created world became imperfect and began to function so. With mutations accrued through imperfect DNA polymerases the created world has drifted to its current state, with speciation through natural selection a part of that drift. I thank Dr. Flanagan for challenging me in her molecular biology classes (I think I took them all!) teaching me the techniques and theory I needed to survive in a multi-disciplinary science, and for holding me to a higher standard, never afraid to tell me I could do better despite my work being sufficient for the best grade. I thank both for teaching me early that scientific research does not stop and there is always more to learn about the way things work.

Much gratitude goes to my graduate advisor, Dr. Brian O. Bachmann. The enthusiasm he has for science gave me the energy to push through difficult experiments and find the answer, no matter how painful. His direction and ideas were kept me moving. I appreciate the multi-disciplinary nature of his background, in engineering, organic synthesis, molecular biology and biochemistry. His seamless combination of these fields is an inspiration. I thank Brian for teaching me how to analyze a problem, devise an experiment to answer the questions and for patiently teaching me the techniques to carry out experiments in the lab.

Additional thanks go to Dr. Jens Meiler, who became a co-advisor and mentor in computational structural biology, and suffered my infrequent attendance in his lab as I grappled with wet-lab work. I most sincerely appreciate his offer for post-doctoral study, where I can master the methods I dabbled in during graduate work. With this I can become a well-rounded, multi-disciplinary scientist with the ability to tackle a problem from multiple angles.

I am indebted to Dr. Kristian W. Kaufmann for his collaboration early in my graduate career. Kristian trained me in Rosetta and spent many hours discussing ligand binding energy and explaining the function of Rosetta, as impossible as that is sometimes. I could not have accomplished chapter 2 of this work without him. I also thank Joe Crivelli for his persistent work on peptide binding energies and for explaining the complex math involved in the normalized sequence differentials.

A significant portion of my thesis was accomplished in close collaboration with Dr. Tim Panosian. Tim crystallized phosphopentomutase and pushed forward the characterization with never-ending creative ideas and hypotheses, resulting in chapter 4 and many more papers to come. His enthusiasm and imagination frequently sent the project into un-expected areas, some fruitful and others not so much. I will always remember our futile attempts to isolate fully-activated PPM. From Tim I learned to think outside the box and ask imaginative questions rather than accept the obvious result as the only solution. Our differing approaches to problem solving allowed this project to succeed and I sincerely enjoyed the collaboration.

I would also like to thank Dr. Tina Iverson, an excellent collaborator and Tim's advisor, for her work in making the phosphopentomutase studies a success. A huge appetite for the next set of data frequently provided excitement and motivation for the next experiment. All scientists should strive to have this trait.

I have spent most of my waking hours in Nashville in the lab. I could not have survived without my colleagues in the Chemistry Department and the Laboratory for Biosynthetic Studies. I thank Dr. Rob Scism for his pioneering efforts in directed evolution in the Bachmann lab; I could not have succeeded without your guidance and expertise brought about through many frustrating experiences. I am also thankful to William Birmingham who is taking over the project and finishing all the experiments I was unable to complete. I hope his path is a little easier for the many efforts Rob and I put

forth in learning, and attempting, each technique. I would also like to thank Dr. Vanessa Phelan and Glenna Kramer, for ensuring the lab is a light-hearted and enjoyable place to work. I thank Vanessa specifically for showing me what scientific perseverance is and for frequently listening to my incomprehensible musings as I tried to solve a problem. To the other lab members, Sunny, Ioanna, Yun Feng, Ahmad, Kasia, Danielle, Andrea and Ruth, I thank you for listening to ideas, sharing your experiences, and most importantly, suffering through painful group meetings and presentations. Your sincere and truthful comments have helped me become a better speaker.

Also enabling my time in graduate school was the financial support provided by the department, advisors and training fellowships. I thank the Chemistry Department for three years of stipend and tuition support in the form of a teaching assistantship. I also thank the Interdisciplinary Training program in Therapeutic Drug Discovery. My project never had dedicated funding so I thank my advisors for their investment to get this study off the ground.

Two people who kept me sane outside of the lab are Michael and Mary Schreuder. Michael became my best friend in Nashville and I cherished our time together. In particular, I enjoyed his friendly competition in everything, from pick-up soccer games, to marathon training runs, to board games, to one-on-one conversation. To Michael there was always a winner; and I'd never admit it was him. I thank Mary for her friendship and support of Allison as we moved here not knowing anyone. In recognition of excellent playing skills (and another dig at Michael): Mary, "Hail, Lord of Cataan."

Finally, I want to thank my family for their support and un-ending love. My parents have supported me and provided guidance for my entire life. I give credit to them for the person I am today, a hard working, family-oriented man of God. My mom was an inspiration to get the project done and done with my very best, no matter what. My dad is

an inspiration for hard work and has modeled self-sacrifice for the sake of his family. I thank my sisters for listening, however patronizingly, when I start to share science with them, and for training me well in living with girls. God bless them for that if I have more daughters.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

BIOCATALYTIC PATHWAY DESIGN AND ENZYME OPTIMIZATION

Introduction

The field of synthetic biology aims to harness and control biological function to accomplish specific tasks[1] and utilizes techniques and concepts from molecular biology and engineering[2]. The ultimate goal of synthetic biologists is to predict and design biological components and systems which respond in a particular and defined fashion[3]. To do this synthetic biologists combine gene 'parts' into 'devices' which perform a defined function[3]. Devices are then combined to form 'systems'. By providing the system with a 'chassis'[3] – or biological host – synthetic biologists have reported tasks such as synthesis of fine chemicals and commodities[4, 5], reporting of and response to local environmental conditions by bacteria[6, 7], and logic calculations which control function based on a particular input or combination of inputs[8-10].

One particularly fruitful area of synthetic biology is the synthesis of fine chemicals, commodities, fuels and drugs in microbial factories. Synthetic biologists and metabolic engineers generate microbial factories by expressing biosynthetic pathways in microbial hosts. Metabolic flux, the rate at which precursors are moved through the pathway to form product, is optimized to increase product yield and the concentration of bioproducts by controlling pathway expression through gene repressors and promoters, modulating relative protein concentrations by varying the strength of ribosomal binding sites (*rbs*) or adjusting mRNA stability, and co-localizing pathway components using protein scaffolds.

A microbial factory provides a number of advantages over large scale extractions or *de novo* chemical synthesis. The aforementioned classical production methods are

often associated with low yields or high waste streams and therefore higher costs. Often, producing organisms are not available at the scale required for large-scale extraction, as is the case for the anticancer drug paclitaxel and the oseltamavir precursor shikimic acid[11]. These limitations have been circumscribed through generation of culturable plant growths and construction of a microbial factory, respectively. In *de novo* synthesis great effort is often required to establish a chemical synthetic pathway yielding enantiopure product. Even still, chemical routes may suffer from incomplete conversion or the generation of side-products making purification of the final compound a necessity. Many, if not most, chemically synthesized molecules are not natural products, though, meaning they cannot be directly generated in a microbial factory and a biosynthetic pathway must be created for bacterial synthesis.

In contrast, microbial factories take advantage of enzymes to synthesize complex entities with excellent stereo- and regiocontrol in an environmentally friendly process. Processes utilize ambient temperatures and pressure, water as solvent and renewable catalysts in the form of enzymes which are generated through catabolism of useable and renewable feed-stocks. Even outside of cells enzyme biocatalysts are becoming readily available for many critical reactions. The synthesis of key chiral precursors of the cholesterol drug atorvastatin and anti-diabetic sitagliptin have been greatly improved through incorporation of enzymatic steps[12, 13].

In this work, the generation of biosynthetic pathways for construction of microbial factories is introduced and methods for the pathway optimization are examined. These principles are applied to the generation of a biocatalytic pathway for directed biosynthesis of the nucleoside analog 2´,3´-dideoxyinosine. 2´,3´-Dideoxyinosine (ddI, didanosine, Videx®) is critical to the treatment of individuals infected with Human Immunodeficiency Virus (HIV), which as of 2010 affected 33.3 million people worldwide. At this same time, the World Health Organization estimated that government subsidized

treatment costs were approximately US $209/patient/year making the drug unaffordable for most of the 33.3 million people affected by HIV[14]. Furthermore, with unsubsidized costs in the United States currently at US $4 478.55/patient/year, a significant pressure is placed on the government to treat citizens[15].

It has been shown that ~66% of the manufacturing cost for ddI is derived from synthesis of the active pharmaceutical ingredient[16]. A large part of this cost can be attributed to stereocontrol of the 1´- and 4´-carbons during synthesis. Natural nucleosides are often utilized as starting materials to take advantage of the previously set stereocenters. However, bulk adenosine and inosine is quite expensive, costing several hundred dollars per kg, driving up the price of the final product. Methods incorporating *de novo* sugar synthesis into formation of ddI utilize glutamate as an inexpensive starting material available in bulk and therefore do not suffer from high starting material costs. Nevertheless, these syntheses suffer from low yields which are compounded by the generation of a mixture of products (e.g. α- and β-anomers or regioisomers). Furthermore, the use of exotic catalysts and column chromatography in chemical synthesis results in large waste streams containing hazardous materials. Methods have been developed that utilize enzymes to form the final product, thereby controlling the stereochemistry and simplifying purification in a process resembling kinetic resolution. Although a stereochemically pure product is formed, chemical synthesis of the precursor molecule is still problematic. Chemical routes for the synthesis of ddI are discussed in detail later in Chapter 1.

The proposed pathway, outlined in Figure 1, utilizes chemical synthesis to generate 2,3-dideoxyribose from glutamic acid and enzymes to accomplish transformation into ddI. Early incorporation of enzymes allows for use of the inexpensive starting material glutamate but is likely to present an enhanced overall yield and a significantly reduced waste stream over previous syntheses.

While the advantages of directed biosynthesis are clear, transformations catalyzed by naturally-occurring enzymes are then utilized to form the final product. Because these enzymes are native to other substrates, namely ribose derivatives, function is typically sub-optimal for immediate process application. Enzyme engineering methods such as rational design and directed evolution are applied to improve turnover for the targeted substrate. Ultimately, lower net manufacturing costs will broaden the availability of effective anti-HIV drugs to peoples currently provided with less expensive, albeit less effective, regimens[17].

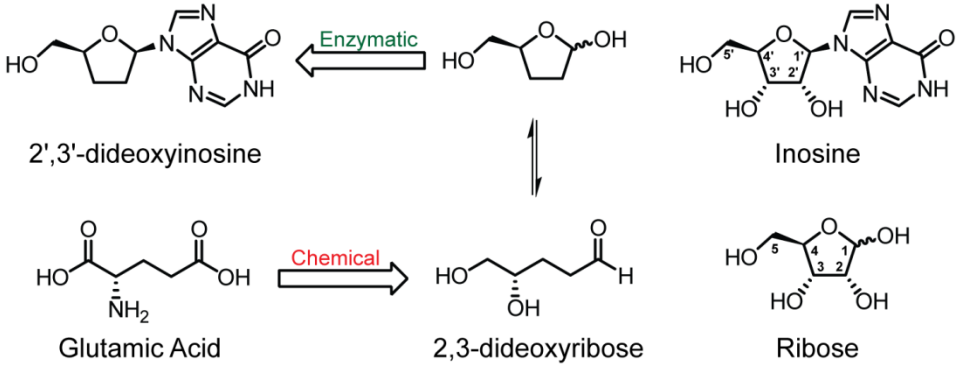

Figure 1-1. Proposed directed biosynthesis of 2´,3´-dideoxyinosine from glutamic acid proceeds through 2,3-dideoxyribose. 2´,3´-Dideoxyinosine is a nucleoside analog of inosine. 2,3-Dideoxyribose is an unnatural sugar, similar to ribose. The numbering scheme for ribose and inosine carbon atoms is shown.

Formation of microbial factories and pathway construction

Efficient, large-scale production of complex molecules such as β-lactams, lovastatin and vancomycin has long been accomplished through fermentation of metabolically engineered organisms. These early microbial factories represent cases of innate biosynthesis, as the compound is synthesized in the naturally producing organism[5]. While numerous compounds have been generated in this fashion[18], for a number of reasons many valuable products are not available through innate

biosynthesis. For example, producers may not be amenable to culture, fermentation or genetic manipulation or cannot be persuaded to up-regulate the production of a desired metabolite.

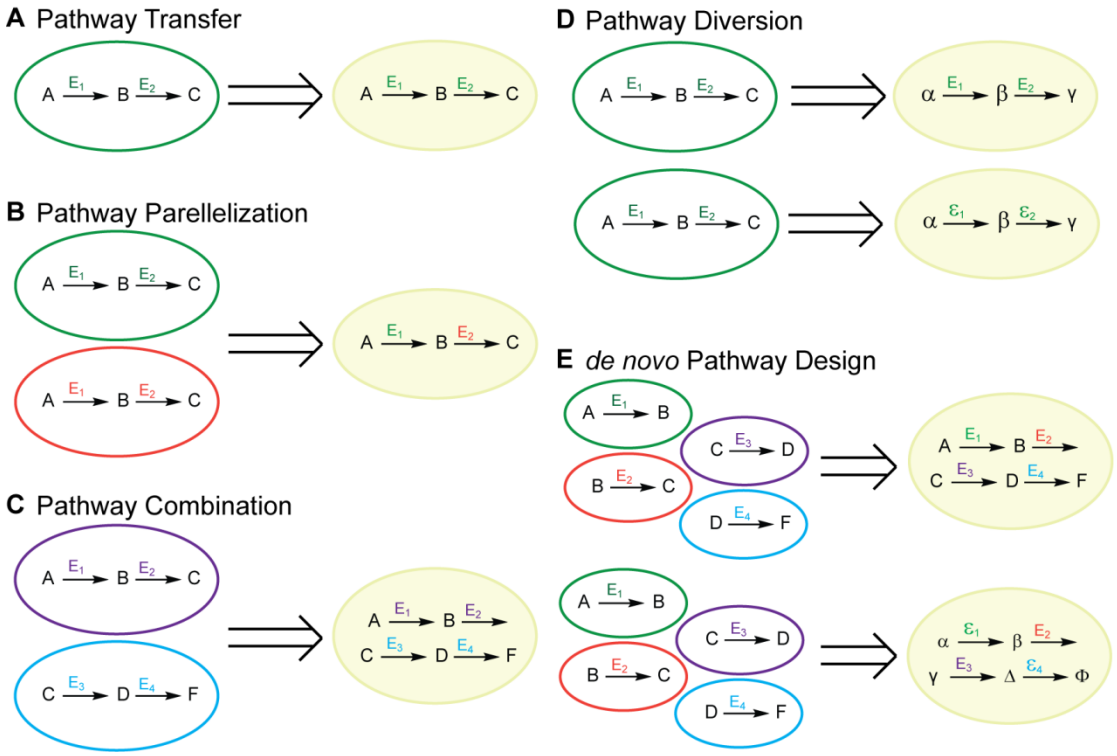

Figure 1-2. Heterologous pathway construction for biocatalytic synthesis of noninnate bioproducts. Enzymes (E_1, E_2, E_n) from biosynthetic pathways are transferred from the native producer (colored ovals) to form heterologous pathways (beige cells). Substrates, intermediates and products are represented by letters (A, B, C, D, F). Non-natural substrates or products are represented by greek letters (α, β, γ, Δ, ϕ). Engineered enzymes are denoted by the greek letter ε. Enzymes can be recombined in any order and any combination, as shown for *de novo* pathway design (E).

To overcome these difficulties, pathway genes can be transplanted into a heterologous host for expression and metabolic optimization[19]. This process is depicted in Figure 1-2A. Such a compound is known as a noninnate bioproduct[5]. This form of heterologous pathway transfer is the most straightforward. Recapitulation of the pathway in a secondary host allows process engineers to decouple production from regulatory

elements present in the native host, thereby increasing product titers[19]. An example of heterologous pathway transfer is the generation of 1-butanol from acetyl-CoA, originally biosynthesized in *Clostridium acetobutylicum* and transferred to *Escherichia coli*[20]. Engineering of the native host proved difficult as butanol formation is tied to pH, redox conditions and sporulation. Transfer to a secondary host enabled increased production levels of butanol outside the realm of host control.

While transfer of the pathway to a heterologous host can be advantageous it also presents its own challenges. Functional expression of an enzyme outside its native environment is often poor due to inclusion body formation or other environmental differences. Low enzyme levels thereby decrease product formation. Transfer of a related enzyme from a tertiary species, in a process called pathway parallelization, may enable more effective production of the target molecule (Figure 1-2B). By screening the activity of various homologs, Steen and coworkers were able to generate a noninnate pathway for *n*-butanol bioproduction in *Sacchoromyces cerevisiae*[21]. Similarly, concatenation of homologous enzymes from *C. acetobutylicum*, *E. coli* K-12, *Clostridium beijerinckii*, and *Thermoanaerobacter brockii* enabled the synthesis of isopropanol in concentrations as high as 45 mM in 10 hours[22].

A particular challenge in heterologous pathway transfer is the supply of precursor substrates into the pathway. In directed biosynthesis precursors are generated exogenously or obtained from commercial sources and product is formed through biocatalytic conversion upon addition of the precursor to the microbial culture. The directed biosynthesis process is also known as a 'hosted' biocatalytic pathway since the primary function of the microorganism is the expression and maintenance of pathway DNA and enzymes. Alternatively, 'integrated' biocatalytic pathways utilize precursors generated as part of the native metabolism of the host which enables fermentation and the transformation of inexpensive or complex feed-stocks into the desired product[23]. The

generation of *n*-butanol and isopropanol in *E. coli* by *C. acetobutylicum* pathway genes is an integrated pathway as acetoacetyl-CoA is formed by the host and fed into the heterologous pathway[20-22].

Frequently, a host organism does not generate the necessary precursors for biosynthesis of a target compound so pathways from two unrelated organisms are combined into a single heterologous pathway for optimized biosynthesis of a non-native product[19]. Pathway combination is depicted in Figure 1-2C. The generation of artemisinic acid, a direct precursor of the anti-malarial drug artemisinin is a well-heralded case of pathway combination. Genes for the biosynthesis of isopentenylpyrophosphate through the mevalonate pathway were transferred from *S. cerevisiae* into *E. coli*, thereby bypassing native control of isopentenylpyrophosphate synthesis through the deoxyxylulose 5-phosphate pathway and allowing for orthogonal control of enzyme function[24]. Subsequent turnover by amorphadiene synthase and a cytochrome P450 monooxygenase from the native producer *Artemisia annua* affords the noninnate bioproduct artemisinic acid[24, 25].

In addition to naturally occurring molecules, heterologous pathways can be re-appropriated to synthesize other molecules or non-natural molecules through off-target, promiscuous activities of enzymes in the pathway (Figure 1-2D). In further work by Atsumi and Hanai, isobutanol and other higher-order alcohols were generated by hijacking the 2-keto acid precursors of valine, threonine, norvaline, phenylalanine, isoleucine, and leucine[26]. Combination with the final steps of the Ehrlich pathway provides the alcohol products through a broad substrate-range 2-keto acid decarboxylase from *Lactococcus lactis* and an alcohol dehydrogenase from *S. cerevisiae*. Control of the predominant product is provided through upregulation of specific amino acid precursors. For instance, isobutanol production reached 23 mM

when the *ilv*IHCD genes for valine biosynthesis were over-expressed to enhance 2-ketoisovalerate synthesis.

Promiscuous activities can also be introduced through mutagenesis of key enzymes in the pathway. A pathway similar to that of Atsumi and Hanai for branched chain alcohol biosynthesis[26] was utilized to generate the C_6 branched alcohol 3-methyl-1-pentanol. In this pathway, enzymes in the leucine biosynthesis pathway were altered to accept 2-keto-3-methylvalerate in place of 2-ketoisovalerate. The 2-keto acid decarboxylase was then mutated to accept the larger substrate to increase product titers[27]. Runguphan and O'Connor were able to generate unnatural alkaloids in hairy-root cell cultures of *Catharansus roseus* upon introduction of a mutated strictosidine synthase[28, 29]. Entrance into the monoterpene indole alkaloid pathway is controlled by strictosidine synthase, which catalyzes the condensation of tryptamine and secologanin. Tryptamine analogs were fed to hairy-root cultures harboring the mutant gate keeper enzyme resulting in the biosynthesis of various unnatural alkaloids as the unnatural strictosidine product filtered through the downstream pathways. In another case, Tsuge *et al*[30] altered the make-up of biological polyhydroxyalkanoates (PHA), potential biodegradable thermoplastics, to include C_8–C_{12} subunits. Through structure-guided mutations the substrate binding pocket of the *R*-specific enoyl-coenzyme A hydratase portion of a PHA synthase was widened and deepened to accept the larger substrates[30].

Many enzymes already possess a desired promiscuous activity but the high-level native activity consumes significant amounts of precursor compounds or other off-target function decreases the yield of desired product. Mutagenesis alters the enzyme specificity and funnels the pathway to a particular product. This reduces the flow of precursors to unwanted side products and is a key strategy to maximize pathway flux. To produce propionate from α-ketobutyrate, Chang and Cronan sought a mutant pyruvate oxidase with increased activity[31]. Rational mutagenesis and site-directed

saturation mutagenesis were used to identify a mutant deficient in normal pyruvate function. The final mutant possessed wild-type levels of α-ketobutyrate activity but significantly reduced pyruvate activity. Pyruvate would act as an intracellular competitive inhibitor and an enzyme with increased discrimination should be a better producer of propionate. Overall, by increasing the enzymes specificity the yield of product relative to initial substrate is improved and, furthermore, the unbalancing effect of high enzyme activities on innate metabolites in the host organism is removed.

De novo pathway design is the generation of non-existing or non-natural pathways from unrelated enzymes (Figure 1-2E). It enables streamlined synthesis of compounds with natural pathways too complex for heterologous expression or the formation of unnatural compounds for which a natural pathway does not exist[19]. Few examples of *de novo* pathways, as they are defined here, have been demonstrated. Moon *et al* constructed a pathway for bacterial production of glucaric acid from glucose which utilizes 5 disparate enzymes from three independent sources[32]. This pathway provides a significant savings over mammalian pathways for glucaric acid production, which typically require 10 conversion steps. A second example of *de novo* pathway design for a natural compound exists in the synthesis of 1,3-propanediol from glucose, utilizing glycerol as an intermediate[33]. Flux through the pathway was optimized for both energy efficiency and overall product yield for each glucose molecule entering the pathway. Import and phosphorylation of glucose was improved by replacing the endogenous system with a designed system. Furthermore, an *E. coli* aldehyde dehydrogenase, previously uncharacterized but now known to be involved in oxidative stress reduction[34], was incorporated in the pathway which utilizes NADPH instead of NADH. This, along with other metabolic engineering elements, increased product titers significantly over the previously used fermentative 1,3-propanediol production methods. Finally, to my knowledge, the sole example of a *de novo* pathway for the biosynthesis of

an unnatural molecule is the generation of D- and L-1,2,4-butanetriol in *E. coli* from D-xylose and L-arabinose, respectively[35]. Enzymes from *Pseudomonas fragi* and *Pseudomonas putida*, along with endogenous enzymes, were brought together to form the final product, a critical precursor to nitro-containing explosives.

One of the most important steps in *de novo* pathway design is the generation of the pathway itself. Development of a *de novo* biocatalytic pathway is aided by a broad knowledge of enzyme functions from primary metabolic pathways and, additionally, biochemical reactions in the formation of secondary metabolites. Using this information, one can imagine a design strategy which utilizes similar planning processes to classical pathway construction in organic synthesis, namely, a retrosynthetic analysis to identify simple precursors which are commercially available or present in innate metabolism. The design of *de novo* biosynthetic pathways in a retrosynthetic manner has been termed bioretrosynthesis[36, 37].

The bioretrosynthetic approach is empowered by re-appropriation of enzyme functionalities which catalyze a desired chemical transformation, albeit on a different but similar substrate. This is analogous to general practices in organic chemistry, where reaction conditions shown to affect a particular chemical change, such as the reduction of a ketone to an alcohol, are recreated in the presence of a novel precursor[37]. Enzyme re-appropriation significantly increases the number of possible biocatalytic transformations. However, nature has only generated enzymes for a limited number of chemical rearrangements on a limited number of substrates. Enzyme engineering methodologies such as rational design and directed evolution enable tuning of enzymes to possess particular characteristics, e.g. increased activity, enantioselectivity or thermostability. Furthermore, significant progress has been made in the design of enzymes with totally novel chemistries that do not exist at all in nature. Enzyme

engineering methodologies and examples of their application are discussed in more detail later in Chapter 1.

Many enzyme databases have been generated to organize the immense amount of data that has been generated through genome sequencing and enzyme and pathway characterization; these include Swiss-Prot, BRENDA, KEGG, and MetaCyc. Most of these databases are organized by substrate, enzyme classification or species making them difficult to search for common functional group interchanges, a necessity for recycling enzyme functionality. Indeed, several groups have generated classification schemes based on chemical conversion, which have been reviewed[19], but a consensus has not yet been reached.

In the past decade several algorithms have been proposed which attempt to automate the pathway design process. In general, they categorize enzyme activity using carefully curated libraries of transformations based on enzyme classification number or by functional group interconversion. Suggested pathways, of which there are sometimes thousands, are prioritized by thermodynamic feasibility, maintenance of stoichiometric balance, effect upon native metabolism or heterologous expression of the pathway on host growth and/or predicted rate and flux. These algorithms have been applied to the prediction of pathways for amino acids[38] and polyketides[39] (BNICE), degradation of xenobiotics (University of Minnesota-Pathway Prediction System, UM-PPS) [40, 41], novel carbon fixation[42], branched-chain alcohols[43], and other value-added compounds (DESHARKY)[44]. A major criticism of some algorithms is a limitation of target compounds to previously observed biological molecules, as is the case for some applications of BNICE and also UM-PPS which is limited to generating pathways for degradation.

It is possible that thousands of potential pathway solutions can be found for the generation of a compound from metabolic precursors as a single precursor can potentially undergo multiple reactions to form multiple products, which can again

undergo multiple reactions. Many algorithms enumerate all possible pathways for a particular target resulting in computationally intense calculations and rigorous prioritization of potential pathways. For instance, in the prediction of pathways for isopropanol and 3-hydroxyproprionate, experimentally demonstrated novel pathways were ranked 17th out of 42,344 and second out of 4,524, respectively[26, 43]. Rodrigo *et al* overcame combinatorial and enumerative issues by implementing a Monte Carlo-based algorithm (DESHARKY) which allows for reverse steps during the design calculations which removes enzymes from the proposed pathway. The decision to go backwards in the design process uses a probability based on the number of introduced steps and unnatural compounds proposed. This keeps pathways short and ensures precursors are metabolically available, thereby improving convergence of the algorithm on a single answer. The step-removal heuristic in DESHARKY is enabled by the bioretrosynthetic approach: pathways which would require long or circuitous routes on the way to predetermined metabolic precursors are easily recognized and avoided. Amazingly, a biocatalytic pathway for glucaric acid was proposed by DESHARKY with remarkable similarity to the one engineered by Moon *et al*[32].

Numerous products have been generated in microbial factories through heterologous pathway expression. Several excellent reviews have been written highlighting examples of noninnate biosynthesis and heterologous pathway construction[5, 18, 19, 23, 45].

Bioretrosynthetic Pathway Construction

In addition to a pathway design method, the bioretrosynthetic approach has potential as an engineering paradigm[37]. To date, pathways have been assembled and optimized in the order of the reactions in the pathway (Figure 1-3). A bioretrosynthetic approach to pathway assembly would start with the product-forming enzyme, the

ultimate enzyme, and retro-consecutively incorporate enzymes to the front of the pathway in a reverse step-wise manner to the penultimate enzyme, and so on (Figure 1-3).

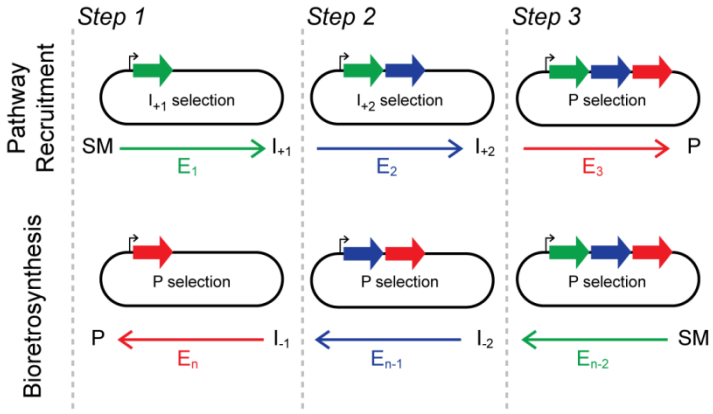

Figure 1-3. Paradigms for *de novo* heterologous pathway construction include pathway recruitment which optimizes enzymes in a front-to-back manner and bioretrosynthesis which optimizes the ultimate enzyme first and proceeds in a reverse stepwise manner thereby requiring a single screen. SM – starting material; P – product; I – intermediate; E – enzyme.

Bioretrosynthesis as an engineering paradigm is inspired by the retro-evolution hypothesis for pathway evolution in nature[37], first proposed by Norman Horowitz in 1945[46]. Here, early organisms have developed in a nutrient rich environment. As a critical nutrient, such as a nucleoside or amino acid, is consumed the organism develops a means to synthesize it from available precursors. As the precursor is consumed a second enzyme evolves to synthesize the precursor from other available molecules, and so on. The formation of pathways in this way uses product selection as the primary selective pressure on the organism. The retro-evolutionary model contrasts with the recruitment model, where genes are assembled into cassettes in a forward step-wise manner, starting with basic environmental components and progressively incorporating genes for enzymes which generate more complex molecules. This, however, implies that

intermediates generated during the early evolution of the pathway provide a selective advantage to the organism. Correspondingly, the retro-evolution hypothesis assumes that intermediates and precursors are available from the environment and limits the evolution of pathways to compounds available in the prebiotic earth, like those generated in Miller–Urey-type experiments[47].

The bioretrosynthetic engineering strategy provides a number of advantages over the traditional front-to-back protocol. First, it requires only a single assay for the product molecule or a method for measurement of consumption of the penultimate intermediate. Traditional forward pathway construction would require *n* assays for *n* enzymes in the pathway (Figure 1-2A). The bioretrosynthetic approach, in contrast, reduces the number of assays required for pathway optimization from *n* to one. Second, in a forward-constructed pathway it is generally necessary to re-optimize intermediate enzymes after completion of the pathway to maximize pathway flux by balancing the rate of formation and concentration of intermediates, expression levels of enzymes, reduce feedback inhibition or competition from structurally related molecules, decrease the metabolic burden of enzyme production, and so forth. In contrast, reverse step-wise construction optimizes pathway flux for high product yields throughout the building process. By beginning with optimization of the ultimate enzyme intermediate formation rates by early enzymes are improved relative to the final enzyme. Therefore, some intermediate enzymes must not be maximally improved but improved only to the point that they are no longer the rate-limiting step. Furthermore, a fully optimized pathway would utilize each enzyme in a stoichiometric ratio alleviating the need for engineering of expression and allowing the entire pathway to be placed behind a single promoter and each enzyme behind a similar *rbs*.

Bioretrosynthetic pathway engineering has certain prerequisites to its laboratory implementation. First, suitable enzymes to catalyze desired transformations, and their

genetic material, must be identified and collected. In addition to the databases and software previously discussed for pathway design, metagenomic analyses have increased the collection of gene sequences potentially encoding enzymes for useful chemical transformations[2, 19, 45]. Second, intermediate and starting materials must be available to feed into the pathway. The acquisition of intermediates and/or starting materials sometimes requires chemical synthesis of numerous molecules as the pathway is formed. Finally, a robust assay for detecting product formation or consumption of a penultimate intermediate must be in place. The assay must be sensitive enough to measure small levels of enzyme turnover but also possess a large functional range so as to enable comparisons of enzyme turnover during the late stages of optimization.

Engineering Strategies for Enzymes

The utility of biochemical reactions in the synthesis of organic molecules, providing regio- and enantiospecifity along with environmentally friendly conditions, has long been recognized[48] and synthetic biology efforts have enabled biocatalytic compound production in microbial factories from inexpensive starting materials, as illustrated above. However, broad application of enzymes in synthesis is severely limited by their inherent instability, inability to accommodate alternative substrates, limitation to a small range of reaction conditions, and frequent requirement for expensive cofactors. These limitations are characteristics of finely-tuned enzyme active site dynamics and global protein architecture and must be overcome for controlled application of biocatalytic transformations in organic processes, directed biosynthesis, and microbial factories.

To overcome these limitations, enzymes with new or improved functions are increasingly generated from existing enzymes through a 2-fold strategy consisting of

rational active site remodeling through mutagenesis, to modify binding specificity for a desired reaction or substrate, followed by more stochastic methods such as directed evolution. First, a suitable progenitor enzyme or scaffold with a measurable starting activity must be identified along with acquisition of some knowledge of the active site geometry from structural data or homology models. If an enzyme meeting the aforementioned criteria is identified, first or second shell active site residues are selected for mutation and functional assessment[49-51]. Saturation mutagenesis is often applied to generate the full complement of amino acids at targeted active site residues, as even the most prudently selected active site mutations may result in unexpected catalytic consequences[28].

To streamline efforts to redesign substrate specificity, it would be desirable to identify a subset of potentially beneficial mutations, particularly in cases in which permutations of several active site residues may be required. Computational protein design strategies allow for large numbers of enzymes to be generated and analyzed *in silico* for alternative ligand specificity and improved activity on non-cognate substrates[52]. Many such algorithms for the design of novel small molecule binding sites have been developed and will be described briefly below.

Stochastic methods identify mutations outside the binding site which have a pronounced affect on enzyme function and enable enzyme engineering even in the absence of structural data. Nature fine-tunes the function of enzymes and metabolic pathways through natural selection, which embodies a two-fold process: the occurrence of a mutation and the advantage that mutation confers on the organism. Directed evolution is the embodiment of natural selection in the laboratory. It has been utilized to improve biocatalytic utility, in terms of turnover rate and enantioselectivity, thermostability and stability in solvents other than water or in detergents, to name a few. Directed evolution requires the formation of diverse libraries of mutant variants and

screening to identify proteins or enzymes with a desired trait. A review of mutagenesis and screening methods, as applied to the improvement of enzymes, is provided below and in Appendix A.

Computational Enzyme Engineering

Successful enzyme turnover requires a complex balance of interdependent forces which bring about recognition and binding of substrate and transition state, release of product, placement of catalytic residues, control of local environmental characteristics (such as pK_a) and carefully coordinated entropy trade-offs (e.g. desolvation and side chain orientation), not to mention equilibrative forces such as conformational flexibility and macromolecular protein dynamics. It is often the case in natural enzymes that contributions to catalysis are made by six or more individual amino acids[53]; a thorough search of all combinations would require 20^6, or 6.4×10^7, variants making experimental analysis of all possible combinations impractical. Computational methods, though, enable the analysis of large numbers of protein variants. However, accurate modeling of active site amino acid positions and their local environment is critical for successful engineering of enzyme activity. Many strategies and algorithms have been developed which aim to properly predict and stabilize the interaction between small molecules and proteins, in both natural and *de novo* designed enzymes. In general, more research has been performed in *de novo* enzyme design, with notable successes. While fewer studies on computational redesign of enzymes exist, there are several notable examples and many protein receptors have been the target of substrate redesign campaigns. As the interdependent catalytic forces in natural enzymes are pre-optimized the engineering process is essentially reduced to a binding problem, making studies on receptor redesign applicable to enzyme redesign.

Initial *de novo* designed substrate-binding proteins and enzymes were generated in four-helix bundles. DeGrado and coworkers introduced O_2-dependent phenol oxidase activity into a heterotetrameric four-helix bundle hosting a di-iron site utilizing a design algorithm based primarily on *ab initio* calculations[54]. Later studies constructed a homotetrameric bundle capable of binding a nonbiological cofactor[55]. Control of the orientation and placement of histidine residues for coordination of the cofactor was of critical importance and accomplished by outlining specific criterion to constrain the design. More recently, Koder *et al* extended this model paradigm resulting in a four-helix bundle capable of O_2 transport[56]. Although Koder *et al* did not utilize computational design methodologies, their control of cofactor redox potential and binding orientation in such a simple fold-type is nonetheless impressive.

Enzymatic activity has also been generated by building active sites into existing protein scaffolds. Substrates and catalytic residues are first modeled in ideal positions followed by placement of secondary residues which appropriately orient the reaction center. Bolon and Mayo first applied this strategy to build histidine-mediated nucleophilic hydrolysis of *p*-nitrophenyl acetate into thioredoxin[57]. They performed their search around a covalent histidine/*p*-nitrophenyl acetate high-energy intermediate. The resulting protein catalyst possessed a K_m of 170 μM and k_{cat} of 4.6×10^{-4} sec^{-1}, but only a 180-fold increase over the uncatalyzed reaction.

More recently, the Rosetta protein structure prediction and design algorithms have been extended to identify potential catalytic sites in novel scaffolds. First, high level quantum mechanical calculations are performed to optimally position enzyme functional groups which stabilize a modeled reaction transition state. Second, using an inverse rotamer search strategy, the RosettaMatch hashing algorithm identifies scaffolds capable of hosting functional groups in the idealized positions[58]. Finally, residues adjacent to catalytic sites and transition state are then designed combinatorially to hold

catalytic residues in position and form a substrate-specific pocket. Designs are ranked based on maintenance of the idealized reaction coordinates and transition state binding energies calculated using a knowledge-based potential. Three successful examples highlight this approach.

An enzyme catalyzing the Kemp elimination was generated using this approach[59]. RosettaMatch identified 100 000 scaffolds with potential to host the reaction. Subsequent design of the active sites in each scaffold resulted in 59 variant designs in 17 individual scaffolds. Eight demonstrated detectable activity with k_{cat}/K_m values of 6–160 $M^{-1}s^{-1}$. One of these candidates, KE07, crystallized readily and became the template for a directed evolution campaign in which 7 rounds of random mutagenesis and gene shuffling (with cassette mutagenesis of specific sites) produced variants >200-fold increased catalytic efficiency (12 to 2 600 $M^{-1}s^{-1}$). Further characterization indicates that the directed evolution process corrected a mistake in the initial design and optimized the pK_a and electrostatic environment of catalytic residues in the binding pocket[60].

In a comparable approach, a retro-aldol catalyzing enzyme using a Schiff base mechanism was developed[61]. From 181 555 RosettaMatch solutions, 72 designs in 10 different scaffolds were selected for synthesis and testing. Active designs were produced with TIM-barrel and jelly roll folds with the best design catalyzing a 0.74 $M^{-1}s^{-1}$ turnover.

In addition to the Kemp eliminase and retro-aldolase, the Rosetta methods have been applied to the design of a Diels–Alderase which catalyzes the stereo- and regiospecific cycloaddition of 4-carboxybenzyl *trans*-1,3-butadiene-1-carbamate and *N,N*-dimethylacrylamide[62]. Greater than 100 scaffolds identified through RosettaMatch were winnowed to 84 designs, of which 50 were stable upon expression. Two active enzymes were generated in six bladed β-propellor and keto-steroid isomerase scaffolds. Iterative saturation mutagenesis of residues adjacent to activating residues (‘catalytic residues’) optimized 100- and 20-fold each design.

De novo enzyme design provides biocatalytic engineers with additional tools to develop synthetic pathways, particularly in cases where the chemistry is not naturally catalyzed. Additionally, it tests the basic understanding of enzyme function and catalytic principles. However, it is clear that significant improvement in design methodologies is required to match the turnover levels of natural enzymes and, furthermore, functionally new enzymes are not always necessary to complete a biocatalytic pathway.

The improvement of enzyme turnover through incorporation of computationally predicted mutations has been successfully demonstrated, as have changes in substrate preference. To test the ORBIT energy function and HERO design algorithm for successful recognition of enzyme-like active sites, Lassila *et al* simultaneously mutated 18 residues in the binding site of *E. coli* chorismate mutase and allowed the algorithm to identify the optimal residue at each position. Five variants were constructed, three of which contained catalytic efficiency on par to or greater than the wild-type enzyme indicating the algorithms ability to recognize and design enzyme active sites[63]. Further probing of the most active mutant, Ala32Ser, indicates this mutation as most optimal for this position[64]. The computationally-identified residue is not in direct contact with the substrate but instead forms a hydrogen bond with a critical catalytic residue; this is a rare example of successful secondary residue design by computational techniques.

More intense QM/MM calculations were performed to predict mutations necessary to improve the catalytic efficiency of a cocaine hydrolase[65]. An impressive 2 000-fold improvement was found with a variant containing five mutations. Prescreening of all possible mutations with a very simple hydrogen bond-based function enabled timely execution of the protocol despite its reliance on high-level calculations.

Alteration of the substrate preference of GrsA-PheA, the phenylalanine activating domain of the NRPS gramicidin synthase, was accomplished with an ensemble-based scoring algorithm, K*, which relies on a Boltzmann-based scoring function to efficiently

prune mutations and conformations from the available search-space. Initial application found a double mutant that switched the binding specificity to leucine[66]. The algorithm was extended to predict mutations outside the binding pocket resulting in a further 2-fold improvement in catalytic efficiency[67]. To my knowledge, this is the first case of successful *in silico* prediction of distal residues which affect activity.

Algorithms within the Rosetta framework were utilized to predict the length and amino acid make-up of an active site loop in a guanine deaminase with the end-goal of generating cytosine deaminase activity[68]. The removal of two residues and a complete change in local sequence identity were necessary to control the precise placement of a catalytic asparagine residue. The new variant was able to catalyze deamination of ammelide, a structural intermediate of guanine and cytosine, with turnover improved over the wild-type by 2 orders of magnitude to 0.15 $M^{-1}s^{-1}$. While the new activity has been introduced, turnover is still seven orders of magnitude lower than guanine deaminase with guanine.

Directed Evolution of Biocatalytic Function

Directed evolution is the embodiment of natural selection in the laboratory, where variation in the enzyme population is first introduced through mutagenesis and improved variants are identified and selected based on specific functional criteria. By reproducing natural search strategies directed evolution can, in theory, identify solutions to the catalytic problem in the absence of complete biochemical or structural characterization or other knowledge-based biases. In practice though, structural and functional data are frequently garnered to guide directed evolution experimental designs and increase the likelihood of success.

Proteins and enzymes have been evolved to possess a number of traits not present in the wild type enzyme. Variants have been generated which catalyze new

reactions; or the same reaction with increased rate of turnover under specific conditions; or with a new substrate or broader substrate scope; to generate conformational stability under specific conditions; and more. Many successes have been published and recently reviewed[48, 49, 69, 70]; however, a few selected cases are outlined here to illustrate the potential of directed evolution.

P450 monooxygenases have been developed by Arnold and coworkers for the biochemical fermentation of alcohols from straight chain alkanes. Using a medium chain fatty acid oxidase as a progenitor enzyme, directed evolution methodologies introduced catalytic competence for the oxidation of successively shorter and less oxidized alkyl chain precursors, first octane[71] and later to propane[72]. Step-wise improvements via mutagenesis throughout the individual domains of the monooxygenase variants ultimately resulted in an efficient P450, propane monooxygenase[73]. Furthermore, one of the variants resulting from these campaigns was evolved to convert ethane to ethanol, currently an important biofuel for use in automobiles[74]. Zhao and coworkers have demonstrated how the cofactor regeneration problem may be addressed using directed evolution of phosphite dehydrogenase. Through a combination of multiple mutagenesis methods over several generations the $t_{1/2}$ of phosphite dehydrogenase at 45°C was improved greater than 23 000-fold from the parent enzyme without sacrificing catalytic efficiency, providing a useful biochemical method for NADH cofactor regeneration[75, 76]. Impressive results have also been reported in the area of improving enzyme enantioselectivity for new substrates through the process of iterative saturation mutagenesis and Combined Active site Saturation Testing (CASTing).[50] To demonstrate the utility of these methods, Reetz applied iterative saturation mutagenesis to *Pseudomonas aeuriginosa* lipase, a well studied enzyme, to increase the enantioselectivity for a selected chiral ester, succeeding in generating a mutant with an enantiomeric ratio (E-value) 594-fold improved over the original enzyme. In comparison,

previous efforts using conventional mutagenesis protocols (epPCR, saturation mutagenesis of hot spots, gene shuffling and recombination) identified a lipase possessing a comparatively lower E-value of 51[77].

Every directed evolution campaign has a varied and specific goal but each one shares two common elements: the introduction of sequence variety and the identification of enzyme variants possessing the desired trait through screening and selection. The first step in directed evolution is the generation of a mutant library from a template protein. Most mutagenesis protocols fit into three categories: random mutagenesis, genetic recombination and site-directed, or targeted, mutagenesis (Figure 1-4). Random mutagenesis, as the name implies, introduces variation throughout the target gene without any methodological prejudice which controls the outcome. Common random mutagenesis methods include error-prone PCR, irradiation with ultraviolet light or plasmid replication in mutator strains. Error-prone PCR can be accomplished using mutation-prone polymerases, by the addition of noncanonical nucleic acids capable of 'wobble' base pairing, or chemical modification of the parent DNA such that the hydrogen bonding pattern is altered. In each case, polymerases are used which do not possess an error-checking function. Gene shuffling aims to capture and combine desirable traits from multiple templates. Recombination can occur at random points, as in the protocols utilizing DNAse and developed by Stemmer or in StEP (STaggered Extension Process) which was initially introduced by Arnold and coworkers. ITCHY (Incremental Truncation for the Creation of Hybrid Enzymes) and SHIPREC (Sequence Homology Independent Protein Recombination) protocols allow for controlled crossover at domain boundaries and between secondary structure elements.

Targeted mutagenesis protocols use knowledge-based criteria to thoroughly and systematically alter the residue identity at particular locations. In site-directed saturation mutagenesis proteins are mutated at a single site to all possible amino acids with

primers containing randomized codons. As a first step, site-directed saturation mutagenesis is most effective when residues contacting the substrate, or in a second shell, are targeted. As a second round of evolution, saturation mutagenesis is often targeted to sites identified through random mutagenesis, enabling the refinement of mutational hotspots. Occasionally, every residue in a protein is individually saturated to find the single best mutation at any position. In many more cases, saturation mutagenesis has been performed iteratively, at one site followed by a second and then a third. While step-wise improvement is frequently found through iterative saturation mutagenesis, the evolutionary landscape can become stuck in local minima identified in early rounds; furthermore, the order of mutagenesis is often arbitrary and it is possible that a different result could be identified if sites were mutated in another fashion. To bypass such concerns, Reetz and coworkers introduced the combinatorial active site saturation test, or CASTing, which applies primer cassettes with multiple randomized codons to mutate several proximal residues simultaneously[50]. CASTing requires that larger libraries are screened to ensure complete coverage of the mutant library but also ensures that the evolutionary process is not biased by local minima or order of mutation.

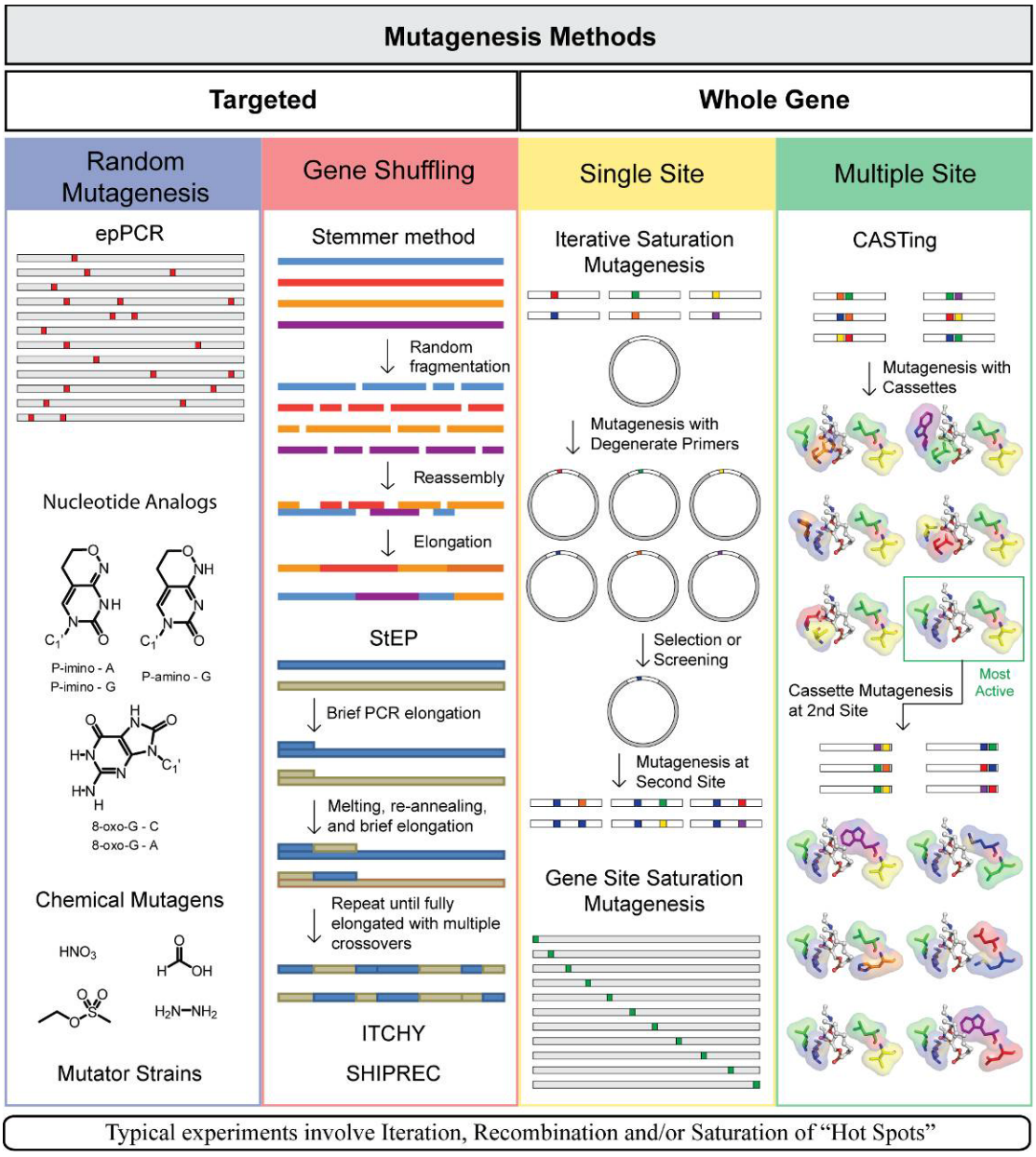

Figure 1-4. Mutagenesis methods for the generation of variant enzyme libraries. Methods target single sites or the whole gene. Iterative application of a method allows for accumulation of improved mutations. Recombination of the most improved mutants from a single round is frequently accomplished through gene shuffling and "hot spots", positions identified through random mutagenesis, are further probed by saturation mutagenesis.

Identification of improved variants is a second element shared by all directed evolution studies. Selection-based methods and biochemical screening-based

approaches represent the two primary techniques for identification of improved mutants. Aside from differences in the mechanism of improved variant identification, screening and selection methods can be differentiated by the number of variants the procedure can screen in a realistic time period (Figure 1-5). Larger library sizes allow for the activity of more variants to be probed and benefit from increased sequence coverage. A short description of screening and selection methods is given here to provide context for the following sections. Screening and selection methods in directed evolution have been rigorously reviewed[78-82] and a more detailed analysis is found in Appendix A.

In selection-based methods, a desired enzymatic activity, such as amino acid synthesis, is linked to host cell survival through growth under selective conditions, such as in minimal medium lacking the particular amino acid. Typical examples include the synthesis of an essential amino acid[83-85] or nucleotide[81, 82] or degradation of an environmental toxin[86-94]. Positive selections can be performed on agar plates or rely on enrichment of cells harboring enzymes of the desired activity in a liquid culture[95]. Negative selection methods, in which the desired activity results in cell death, often entail that cells must be sorted and replicated into both library reference and assay microtiter plates to facilitate backtracking and resurrection of the desired library variants, thereby making them lower throughput[96, 97]. In general, selection methods provide the capability to screen 10^6 or more variants; however, it is often difficult to link cellular response to molecule production, correlate the response in concentration-dependent manner, and ensure products are impermeable to the cellular membrane.

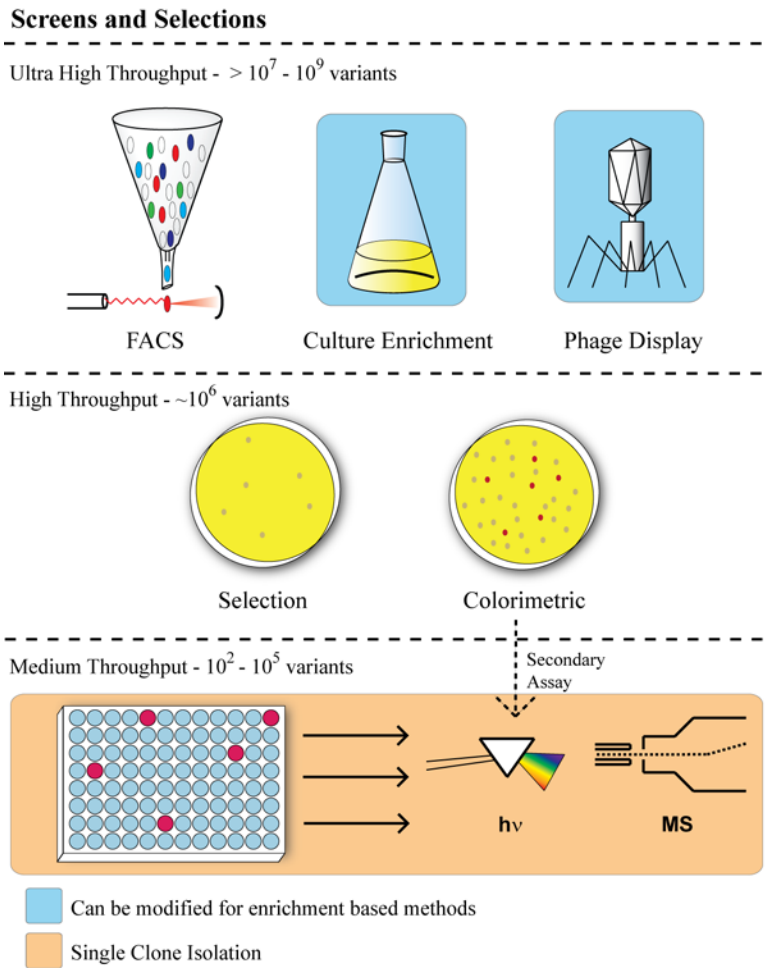

Figure 1-5. Screening and selection methods for directed evolution. Representative methods for identification of improved enzyme variants are classified by throughput capacity. Examples of methodologies capable of selecting or screening ultra large libraries include FACS, enrichment, and phage display. High throughput methods include selection or screening on agar plates. Many biochemical assay methods can be modified to 96-well format, but require that colonies are picked into individual wells prior to screening. Some plate-based colorimetric screens require secondary analysis in this format.

In screening-based methods, the rate of substrate turnover is determined via biochemical assay, often via colorimetric or fluorometric measurements. *In vitro* (cell-free) activity assays generally require the isolation of individual colonies into 96- or 384-well plates, followed by lysis and multiple reagent transfers. Thus, screening-based methods are frequently limited by cost and time to less than 10^6 mutants (10^3-10^4 is more common). Just as it is challenging to link enzymatic reactions to cell viability, it is

also difficult to contrive reporter systems that induce a change in color or fluorescence *in vivo*, or in compartments. Such systems often require an enzymatic assay for activity in cell lysates from colonies isolated in microtiter plates. Although these assays require the growth of individual cultures for assay, there is no limit to the varieties of instrumental analysis that can be performed, and most assays can be implemented in a 96-well format and auto-sampling technologies. Examples of these methods include analytical HPLC[98], mass spectrometry[87], TLC[99, 100], radioactivity[101], and UV/Vis spectroscopy (e.g. NADH reduction)[102-104].

Some colorimetric screens can be modified and applied to agar plate colony screening. Here, colonies displaying a desired color within a certain time window are selected as active. Frequently, the colorimetric agar plate assay is used as a primary screen to identify enzyme variants with functional turnover. Subsequently, a secondary and sometimes tertiary activity assay is performed in a 96-well microtiter plate to confirm and identify the most active clones[105-107]. Colorimetric agar plate colony screening methods benefit from the larger library sizes assayable via the initial agar plate screen before encountering the time and cost limitations of single colony isolation techniques.

Flow cytometry and phage display methods also permit the assay of larger mutant libraries, without requiring that product formation is linked to cell growth. Fluorescence activated cell sorting (FACS) is a type of flow cytometry in which a mixture of cells is sorted, one cell at a time, based upon the specific light scattering and fluorescent characteristics of each cell. In phage display screening, the DNA of a gene library is fused to a viral coat protein gene, and combined in a phagemid (a vector containing both plasmid and phage origins of replication). The phagemid DNA then becomes packaged in the phage capsid that displays the protein. The phage are subjected to chromatographic separation (affinity binding) to select for desired activity[108]. Very rarely is phage display connected directly to enzyme activity although an increase

in a biosynthetic enzyme activity can be identified through tighter binding to a tethered transition state analog[109]. Other display methods, such as cell surface display, provide the advantage of a large library size in addition to facile subsequent screening capability[110].

Statistical Analysis of Directed Evolution Methods

With the large number of available methods for stochastic improvement of enzyme function it is difficult to know which methods offer the best chance for success. While there are many impressive success stories, a cursory survey of the literature also reveals a substantial number of studies in which the enhancements are more modest. Also, complicating the comparison of the success of many studies is that many articles report improvement in properties that do not refer directly to kinetic parameters (e.g. total yield of reaction). Presumably, this is not an intentional obfuscation. Rather, many directed evolution studies are simply goal oriented, and the goal is a desired phenotype. Desired phenotypes may be “an organism survives”, “a colony changes color”, or that “a percent conversion is achieved”. Indeed, almost all reported studies to date appear to succeed in generating their desired phenotype or chemo-phenotype and by these criteria, can be viewed as bona fide successes in directed evolution. Therefore, it is difficult to identify and emulate a particular strategy during the enzyme evolution process.

Given the substantial body of literature describing successes in directed evolution, the diversity of the methods employed and the variety of enzyme classes improved, it would be interesting to perform an objective assessment of success of the methods in directed evolution for the creation or optimization of discreet biosynthetic enzymes; in particular, those studies which aimed to improve activity for a non-native substrate. To accomplish this, articles from the last ten years were selected using

objective selection criteria, and classified each study, primarily, by the mutagenesis method and protocol for selection or screening of variant function. Mutagenesis protocols were categorized into four basic subsets: random mutagenesis, gene shuffling, targeted mutagenesis, and studies which used a combination of methods. Screening and selection methods were generally classified by the number of mutants which could be screened in a reasonable time frame (~1–2 weeks).

Examples of improvements in biocatalytic enzyme activity for non-native substrates were collected from ~3,500 total articles discussing directed evolution. As part of this objective comparison, it was required that improved turnovers were not due to protein stabilization or improved production levels. Only articles reporting velocities normalized to enzyme concentration were included in the analysis; normalization to total protein content of cell free extract was also accepted. Furthermore, only articles that report improvements in biotransformation reactions were collected. Articles with the goal of improving protein stability under non-native reaction conditions (e.g. thermostablization) were not included. Studies reporting on the optimization of restriction enzymes, DNA/RNAzymes, catalytic antibodies or protein:protein interactions were excluded as they did not generally report on the improvement of a small molecule biosynthetic enzyme. Reported improvements in activity for the natural substrate (or other common substrate, as is the case for guaiacol and cytochrome c peroxidases) were considered improvements in native reaction conditions parameters and therefore were not included. The directed evolution process is defined by randomized mutation of a gene and screening for a desired new activity; as such, it was required that more than one site be mutated, whether it be simultaneously (e.g. epPCR or CASTing) or through an iterative process (iterative saturation mutagenesis). This study also required that a selection scheme or activity screen be employed as the primary method to identify improved mutants and that the compound used during screening was also the

compound used for catalytic characterization (determination of V_{max}, k_{cat}, k_{cat}/K_m); in some cases substrate derivatives which enabled high throughput screening and/or selection (i.e. biotinylation[111] or fluorescent tagging[112]) were included.

Studies reportedly aimed at broadening substrate specificity were excluded by the requirement that the substrate in the employed selection is the most faithful reporter of improvements in activity. Many studies report mutants functional on multiple substrates and it was often unclear how a given mutant was identified through screening with a particular substrate. Additionally, studies that reported on the generation of enantioselective enzymes were typically not included on the grounds that the desired function was not strictly increased turnover of a desired substrate but a ratio of turnovers for two enantiomers. Where studies aiming to generate enantioselectivity were able to screen or select for improved activity on a particular enantiomer in a single step (one assay per mutant, as opposed to an assay with each enantiomer), and additionally meet the remaining conditions, they were included[112, 113].

Using the GoPubMed.com semantic server[114], the top 20 research journals publishing articles with the search phrases “directed evolution” or “directed molecular evolution” were identified. Additionally, all citations comprised studies that 1) reported improvements in a small molecule biotransformation enzyme, 2) reported kinetic parameters, including k_{cat}/K_m and/or k_{cat} (or V_{max}) and/or K_m, for both the progenitor enzyme and the evolved enzyme, 3) are target-based studies; in other words, the reported improved turnover was for the substrate used in the screen and 4) more than one site was mutated, reflecting a departure from single site saturation mutagenesis.

Database searches for the terms “directed evolution” and “directed molecular evolution” were performed in the PubMed and ISI Web of Knowledge databases resulting in the accumulation of ~3,500 unique articles. Patents, conference proceedings, and non-English language listings were removed. To focus these results, a

search was performed with the GoPubMed.com semantic analysis web-server[114] to identify the top 20 journals publishing articles on “directed evolution” or “directed molecular evolution” during the last 10 years (2000 to present). Journals that primarily published review articles (Current Opinions in Biotechnology, Current Opinions in Chemical Biology, and Trends in Biotechnology) or methods (Methods in Molecular Biology) were excluded from the top 20 list. Including commentaries and review articles, these 20 research journals have published ~1,275 articles since the year 2000. The journals included in this study were: Protein Engineering, Design and Selection, Journal of Molecular Biology, Proceedings of the National Academy of the Science, ChembioChem, Nucleic Acids Research, Angewandte Chemie International Edition in English, Applied Environmental Microbiology, Applied Microbiology and Biotechnology, Biochemistry, Journal of Biotechnology, Nature Biotechnology, Journal of Biological Chemistry, Journal of the American Chemical Society, Biotechnology and Bioengineering, Chemistry & Biology, Protein Science, FEBS Journal, Nature, Science, and Biochemical and Biophysical Research Communications.

Using these criteria the ~1,275 articles in the GoPubMed database were scanned, first by title, then abstract, and finally the full text, and extracted information on enzyme class, screening method, mutagenesis protocol and enzyme turnover parameters. Initially, the database was limited to articles reporting k_{cat} and K_m; finding only 32 articles the criteria was expanded to include articles reporting V_{max}, k_{cat}, K_m or k_{cat}/K_m. The final database contains 81 improved enzymes from 77 papers published in 18 journals, and is listed in Table 1 of Appendix B, with references contained therein. The pruning results are depicted in Figure 1-6.

Our initial goal was to identify articles that reported a fold improvement in kinetic parameter for wild-type enzyme and evolved enzyme. As the preceding discussion demonstrates, this ultimately was not always a clear cut process. Thus an important

caveat to be mentioned here is that while every effort was made to identify articles within the 1,275 member pool that met the stringent selection criteria, it is doubtless the case that some qualifying articles were overlooked.

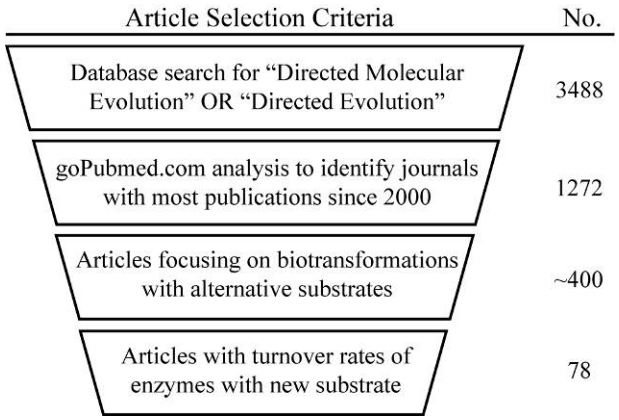

Figure 1-6. Article selection criteria for meta-analysis of directed evolution methodologies. The number of articles remaining after each round of pruning is labeled.

While all enzyme classes (EC 1–7) are reported to have been optimized by directed evolution methods, the majority of the 77 articles meeting the selection criteria fell into three E.C. classes, (1) 23 oxidoreductases, (2) 22 transferases and (3) 28 hydrolases (Figure 1-7A). Within the oxidoreductase category six enzymes were P450 or other heme based oxidases and only one was a cofactor-dependent reductase. The prevalence of hydroxylation enzymes in recent directed evolution campaigns (7 cases) may be reflective of the importance of hydroxylation in the alternative energy industry and the difficult nature of C–H bond activation reactions in the context of preparative organic synthesis. Additionally, formal oxidation of an amino or hydroxyl group to a keto group was well represented with 6 instances of class 1. The transferase category is the most functionally diverse and included members with such assorted activities as aminotransferase, kinase and glycosyltransferase. Within the dominant class, the

hydrolase category, the leading representatives were lipases, glycosyl hydrolases and esterases representing 3, 6, and 9 experiments respectively. The poorest represented enzyme class overall was that of ligase enzymes (e.g. synthetases). The low population of this class may be a result of the inherent difficulty in evolving multi-substrate enzymes, the lack of identification of interesting targets for evolution within this enzyme class, or absence of reported kinetic parameters meeting the criteria of this analysis.

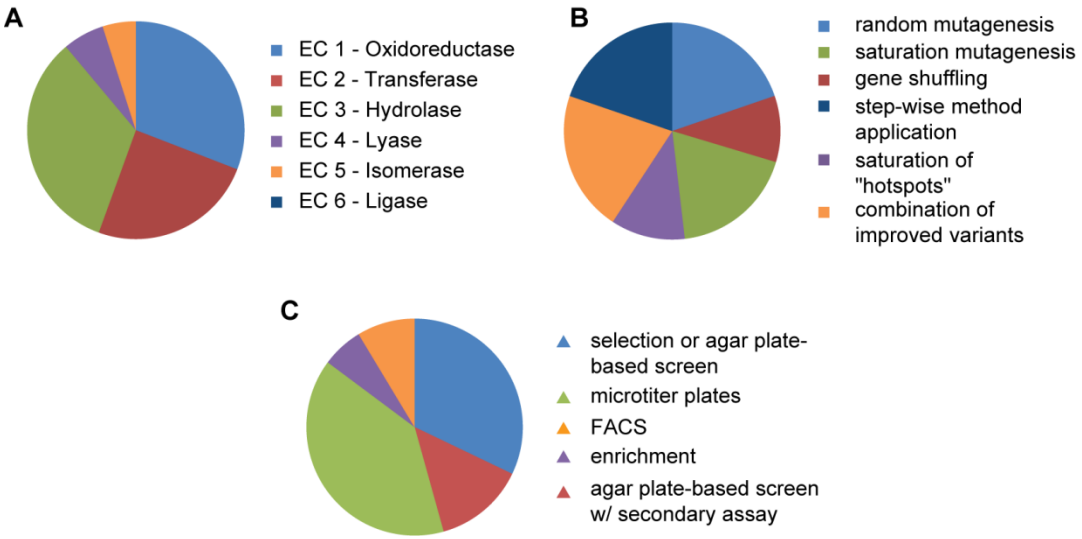

Figure 1-7. Statistical analysis of selected directed evolution experiments by enzyme class (A), mutagenesis method (B) or screen (C). Methods are described in detail in the text and in Appendix A.

As previously discussed, there are a great variety of methods for introducing mutations into progenitor genes. In practice, most studies employed a combination of methodologies over several generations to achieve an optimized enzyme (Figure 1-7B). The workhorse of directed evolution appears to be error prone PCR, and it was this methodology that was responsible for the origination of many mutations and/or the determination of hot spots within a progenitor gene. Of the cases in this study, 51 used epPCR at some point in the evolutionary process. Saturation mutagenesis comprised

the second largest category, being used 36 times, revealing the importance of using structural and/or experimental prescience of active site residues to guide the directed evolution experiment. Gene shuffling was the least employed as a stand-alone method for introducing diversity. However, it was most frequently applied as the last step to combine beneficial mutations, and to remove deleterious or neutral changes.

A popular aphorism of directed evolution is “you get what you screen for”[80]. To those practicing in the field, this may be prefixed with the qualification, “if you get anything…”. Indeed, it is impossible to determine the statistics for failed directed evolution experiments, although many campaigns reported null or minimal improvements in some rounds of mutagenesis and screening. In these cases, alternative mutagenesis strategies were devised to generate mutants with the desired improvement[115, 116]. Of successful reported experiments meeting the selection criteria, microtiter well-based assays were identified as the most common method of activity screening, although agar plate based selection and screening was also frequently employed (Figure 1-7). Liquid culture enrichment and FACS were each used in only a small fraction of the selected articles. This distribution of methods may be a reflection of the limited range of enzymatic activities that can be tied to cell survival or a fluorescent signal in whole cell imaging. Microtiter well-based assays in clarified cell lysate allow analysis via tracking changes in absorbance in the UV or visible light range and permit the use of tandem enzyme assays that may be required to link the desired reaction to a measurable parameter. Additionally, assays designed for microtiter plates can be subsequently analyzed by using autosamplers and HPLC, MS or capillary electrophoresis for direct product quantitation, making these approaches the most broadly applicable methods the for analysis of a wide variety of enzymatic reactions, albeit with substantially lower throughput.

The success of directed evolution studies can be measured in a variety of fashions, most typically by the attainment of the desired phenotype or the desired reaction yield. While providing valid criteria for success, these results form a difficult basis for evaluating these success in the context of the degree of improvement (i.e. almost all publications report success in the attainment of a desired chemo-phenotype). Correspondingly, selected papers were evaluated in the context of kinetic parameters, specifically k_{cat}, K_m and/or k_{cat}/K_m, the last parameter being the most frequently cited parameter. Interestingly, while the average fold improvement for k_{cat}, K_m, and k_{cat}/K_m were 366-, 12-, and 2 498-fold respectively, the median fold improvements were significantly lower at 5.4-, 3-, and 15.6-fold respectively. This disparity is a result of the small number of extremely large fold rate enhancements (see Table 1 in Appendix B, and discussion below). In both statistics, the improvements in K_m were relatively low; this may be reflective of the lower dynamic range for K_m values (often 10 μM to 10 mM) and the difficulty in selecting for improved binding during catalysis using common activity screening.

The prevalence of a given classification or method may not predict its ultimate success. Figure 1-8 illustrates the relationship of the parameter k_{cat}/K_m to the classification schemes presented herein. Significant trends between the enzyme class, mutagenesis method or screening paradigm and the fold improvement were not identified. Enzymes with the greatest fold improvement (>100) were represented in every enzyme class, excepting lyases, using each mutagenesis method, and from every screening method.

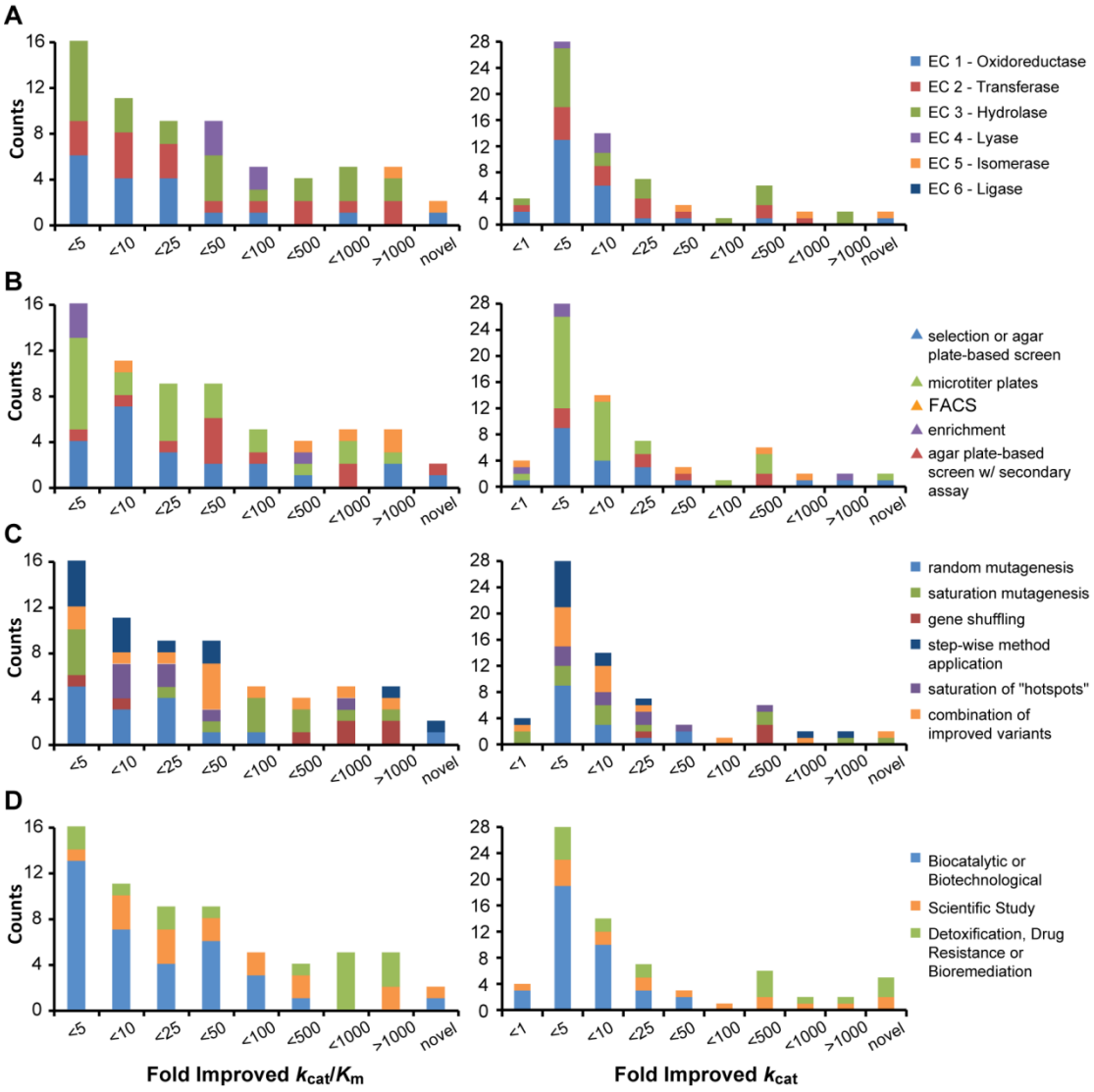

Figure 1-8. Breakdown of enzymes improved by directed evolution studies by catalytic efficiency and turnover rate. A. Enzyme classification; B. Screening method; C. Mutagenesis method; D. Application.

Of note, 7 out of the 14 cases that were identified reporting >100-fold improvement for a new substrate employed a form of saturation mutagenesis in their directed evolution strategy (Figure 1-8B). This correlation may be expected as it has been demonstrated that mutations closer to the binding pocket have a greater impact on enzymatic activity.[117] There are numerous examples of studies comparing saturation

mutagenesis of residues in the binding site to whole-gene random mutagenesis. In each case, targeted mutagenesis strategies allow for smaller library sizes and equal, or better, levels of improvement. While this has been thoroughly documented[77, 118, 119], this data suggests that saturation mutagenesis is not the only method resulting in large fold improvements, and it also does not guarantee a high level of success.

An assumption of directed evolution is that a more complete search of the mutational space may allow for identification of more improved variants. While large library screens have been reported (e.g. $>10^7$ by FACS), the fold improvements did not scale with the size of the library screened. For instance, screening of only 300 clones in each of two rounds identified a carboxylesterase with 840-fold improved catalytic efficiency[99] whereas a similar FACS based study resulted in a modest 5.5-fold improvement.[120]

Finally, in the context of the generation of new biocatalytic enzymes the intent of each directed evolution study was examined and sorted each case into three types: 1) biocatalytic or biotechnological application studies; 2) studies which aimed to answer a scientific question, such as the evolutionary process; and 3) enzymes which were evolved to remove toxins for survival, for bioremediation, or to study drug resistance mechanisms (Figure 1-8D). Despite making up >50% of the cases, only one study out of 14 with >100-fold improvement in enzyme function has biocatalytic application as the ultimate goal. Overall 41 studies of 81 were enzymes with potential biocatalytic application.

The magnitude of rate enhancements may be perceived as modest from a biophysical perspective, with a median value of 5.4-fold for k_{cat}. However, this improvement, and the improvements frequently observed as a result of improved protein production levels, may be deemed highly significant from a biotechnological perspective as the time and/or cost savings scale with increased turnover.

Biocatalytic Synthesis of Nucleoside Analogs

To demonstrate the feasibility and utility of *de novo* biosynthetic pathways in reducing the cost of complex unnatural compounds, the chemo-enzymatic synthesis of 2´,3´-dideoxyinosine has been studied. In terms of chemical structure, didanosine is the simplest dideoxynucleoside drug and representative of the broader class of dideoxynucleoside analogs (Figure 1-9)[121]. As a close analog of inosine, extensive biochemical and structural studies exist for both *de novo* biosynthesis and purine salvage pathways of inosine[122, 123]. Since a large fraction of nucleoside analogs are variants of 2´,3´-dideoxynucleosides, methods developed for ddI can be applied to the generation of biocatalytic pathways and biosynthetic production of many other analogs.

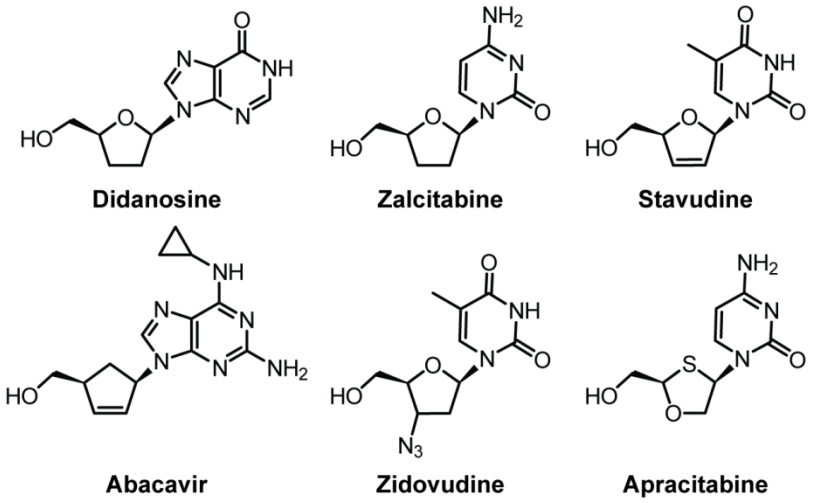

Figure 1-9. 2´,3´-Dideoxynucleosides for treatment of HIV include didanosine, zalcitabine, stavudine, abacavir, zidovudine, and the experimental compound apracitabine.

Nucleoside analogs have long been recognized for their utility in the treatment of viral infections and cancer. Since the synthesis of the first analogs in the 1950s, nucleoside analog drugs have become indispensable as therapeutic molecules in the treatment of the herpesviruses, HIV, papillomavirus, polyomaviruses, adenoviruses,

poxviruses, hepatitis B, and hepatitis C viruses and in the treatment of carcinomas such as leukemias, colorectal, pancreatic, bladder, breast, and non-small cell lung cancers. Furthermore, nucleoside analogs are utilized in DNA sequencing methodologies and other biotechnological techniques. As such a critical class of compounds the synthesis and biological activity of nucleoside analogs is a major area of study.

Despite the broad efficacy of nucleoside analogs, high prices prohibit broad therapeutic application in some geographical areas with significant need, such as treatment of the AIDS epidemic in Sub-Saharan Africa. The cost of the manufacturing of the active ingredients of some drugs in this class comprise up to 55–99% of the final therapeutic price.[16] These costs can be attributed to low yield chemical steps, many requiring chromatographic separation, and expensive starting materials. To illustrate this, a discussion of chemical synthetic methods for ddI is shared below. Because of the difficulty in synthesis even modest improvements in process yields or a switch to a less expensive starting material will facilitate broad distribution of critical drugs worldwide.

Chemical Synthetic Methods for Production of Dideoxyinosine

Webb *et al* first reported the synthesis of ddI in the literature in 1988[124]. Efficacy towards treatment of HIV was first reported in 1986 by Mitsuya and Broder[125], who used commercially available sources for their studies[17]. It was approved for clinical use in 1991 and remains a component in highly active anti-retroviral therapies and drug cocktails. Since then, a number of syntheses have been developed for ddI, falling into three general strategies, each utilizing a different starting material. Initial syntheses of ddI, and many of the other dideoxynucleosides, were accomplished from inosine and deoxyinosine[126]. Webb *et al* treated 5´-benzoyl protected 2´-deoxyinosine with 1,1´-thiocarbonyldiimidazole to form a 3´-O-imidazolothionocarbonate derivative. Subsequent reduction with tributyltin oxide and AIBN/polymethylhydrosiloxane, followed by

ammonolysis for deprotection, affords the final product in 33% yield (Figure 1-10).[124] This same process was previously reported for the synthesis of 2´,3´-dideoxyadenosine, 2´,3´-dideoxyguanosine, 2´,3´-dideoxythymidine, and 2´,3´-dideoxycytosine.[127]

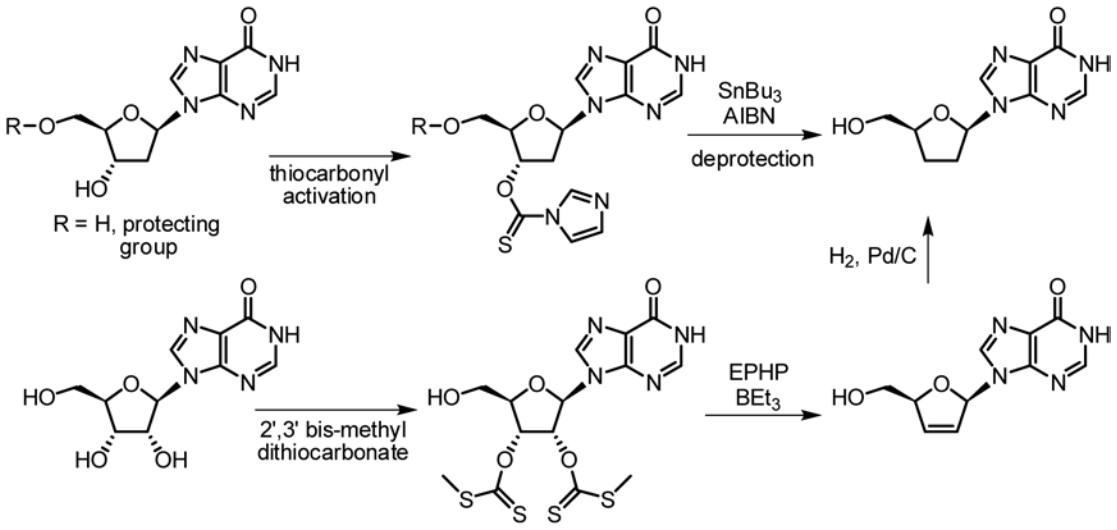

Figure 1-10. Chemical synthesis of ddI by Barton deoxygenation of 2´-deoxyinosine and inosine. Reaction intermediates are shown from the syntheses of Webb *et al*[124] and Torii *et al*[133]. A variety of thiocarbonyl derivatives have been demonstrated in this process.

The Barton deoxygenation process has been improved upon[128, 129] and also utilized with phenoxythionocarbonate[130, 131] derivatives of 2´-deoxyinosine and β-cyanoethyl-dithiocarbonate derivatives of inosine[132]. The application of 1-ethylpiperidinium hypophosphite (EPHP) as a reducing agent and triethylborane as radical initiator by Torii *et al* avoids the necessary removal of tin oxide by silica chromatography and avails the Barton deoxygenation process to large-scale processes where remaining tin is toxic[133]. Torii and coworkers report a 60% yield from inosine to ddI. Where inosine is the starting material, a frequent intermediate is 2´,3´-didehydro-2´,3´-dideoxyinosine; this intermediate is easily reduced by hydrogenation with palladium on carbon[126]. The Corey–Winters reduction of cyclic thiocarbonates, utilizing triethyl

phosphite, has been successfully applied, generating the unsaturated 2´,3´-didehydro-2´,3´-dideoxyinosine intermediate and, subsequently, ddI[132, 134].

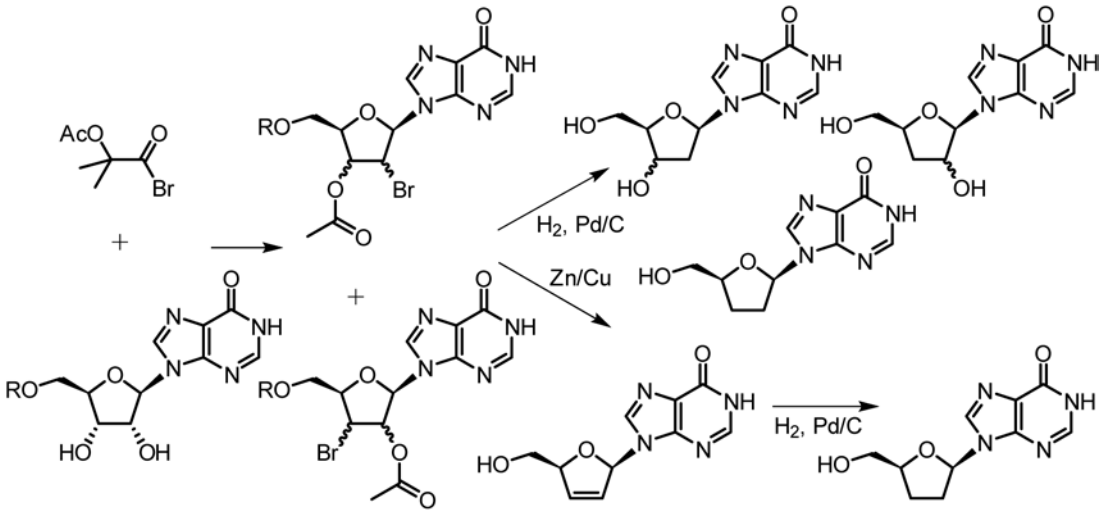

Figure 1-11. Chemical synthesis of ddI through application of the Mattocks reaction performs direct reduction of the bromoacetate to generate a mixture of deoxy- and dideoxy- products. Reductive elimination of the 2´ and 3´ functional groups to form the unsaturated nucleoside prior to hydrogenation forms the desired product, alleviating purification steps.

A second method utilizing inosine as the starting material takes advantage of the specific reaction of 1,2-diols with α-acetoxyisobutyryl bromide, known as the Mattocks reaction, and depicted in Figure 1-11. The resulting acetoxy bromide can be directly reduced by catalytic hydrogenation although this frequently results in a mixed product composition of 2´- and 3´-deoxyinosine and ddI and a lower yield of pure ddI[128]. Reductive elimination processes with zinc or copper catalysts have generated the 2´,3´-didehydro-2´,3´-dideoxyinosine species and hydrogenation with Pd/C generates ddI[135]. Modifications of the Eastwood olefination process, whereby a five-membered cyclic orthoformate fragments in the presence of an acid catalyst, has been used to generate 2´,3´-dideoxyuridine, 2´,3´-didehydro-2´,3´-dideoxyadenosine, and 2´,3´-didehydro-2´,3´-dideoxythymidine. This process is not widely applicable to generation of

dideoxynucleosides though, as the compounds are prone to elimination of the heterocyclic base and cleavage of the glycosidic bond under acidic conditions[126].

De novo synthesis of the unnatural sugar followed by addition of the nucleobase offers an alternative to inosine and deoxyinosine as the starting material. In 1989, Farina *et al*[136] published the synthesis of ddI from glutamic acid and hypoxanthine (Figure 1-12A). Deamination and cyclization of glutamic acid can be accomplished stereospecifically and reduction of the carboxylate results in a five carbon cyclic lactone. Upon protection of the 5-hydroxyl, further reduction of the lactone and acetylation affords the acetyl lactol as a mixture of anomers. Treatment with trimethylsilylbromide and addition of bis-silyladenine forms a racemic mix of ddA anomers upon deprotection. Dideoxyinosine is generated by enzymatic reaction with adenosine deaminase, which is selective for the β-anomer. The product can be purified through recrystallization in methanol and ddI is collected in <10% overall yield from glutamic acid[136]. Precedence for the use of adenosine deaminase to generate ddI from ddA comes from studies on the application of ddA as a prodrug[137, 138]. Acid cleavage of ddA in the stomach, however, generates high levels of adenine which causes nephrotoxicity upon crystallization in the kidneys[17]. Dideoxyinosine has been synthesized directly, without the use of adenosine deaminase, in a method similar to Farina *et al*[136]. The generation of a 2-selenophenyl acetyl lactol allows for stereocontrolled addition of the nucleobase (Figure 1-12B), providing the β-anomer as the major product. ddI is formed through reduction of the 2´-selenophenyl-ddI through reduction with tributyltin hydride and triethylborane in 44% total yield[139]. This strategy, whereby the unnatural sugar is first synthesized, often results in the generation of enantiomeric products resulting in the loss of significant material or requiring chromatographic purification.

Semi-synthetic methods for production of ddI have also been reported. In addition to the use of adenosine deaminase[136-138], *E. coli* BL21 was studied as a

biocatalyst for production of nucleosides and nucleoside analogs by feeding uridine derivatives as the sugar-donor and either adenine or hypoxanthine as the base acceptor but found *E. coli* BL21 to be an inefficient catalyst for formation of ddI[140]. A screening program identified *E. coli* AJ 2595 cells that were capable of synthesizing 32 mM ddI from 100 mM hypoxanthine and 100 mM dideoxyuridine, which had been chemically synthesized, over 24 hours at optimal conditions[141]. Additionally, several patents address methods of ddI synthesis that use immobilized adenosine deaminase or whole cell cultures[142, 143].

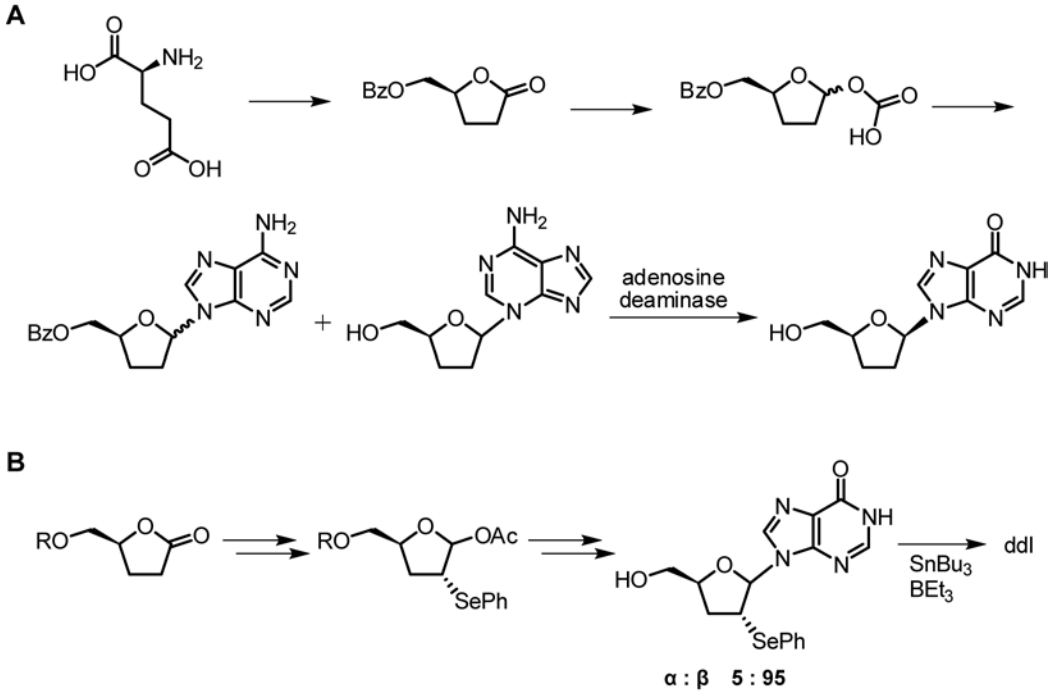

Figure 1-12. *De novo* sugar synthesis for the chemical production of ddI often results in an anomeric mixture of product. Methods reported by Farina *et al*[136] (A) and Beach *et al*[139] (B) are highlighted.

Bioretrosynthetic Design of a Chemo-enzymatic Pathway for ddI

A biocatalytic pathway for ddI synthesis was designed using retrosynthetic and bioretrosynthetic principles. Potential transformations were chosen based on the

structure and reactivity of each molecule while matching those characteristics to other studies involving molecules of some similarity. This process, as applied to the design of a chemo-enzymatic pathway for ddI, will be outlined here, using the KEGG database to identify enzymes which catalyze the desired reaction on substrates similar to the target enzyme. Numerous pathways are possible, with some having steps with high risk levels or unlikelihood and others having the advantage of fewer steps. Several possible pathways are illustrated in Figure 1-13 and their benefits and challenges, in terms of available literature and engineering probability, will be discussed.

Nucleosides are generated intracellularly through two pathways: *de novo* construction of nucleotides from phosphoribosyl pyrophosphate (PRPP), amino acids and folate and through the nucleoside and nucleotide salvage pathways. The *de novo* pathway is made up of 11–13 enzymes for purines and 6–14 for pyrimidines, making it an energy-expensive mechanism for nucleotide generation. Because of its length, modification of the *de novo* pathways to optimally synthesize ddI, from a simple sugar precursor, would be a tremendous undertaking, likely requiring the improvement of >10 enzymes.

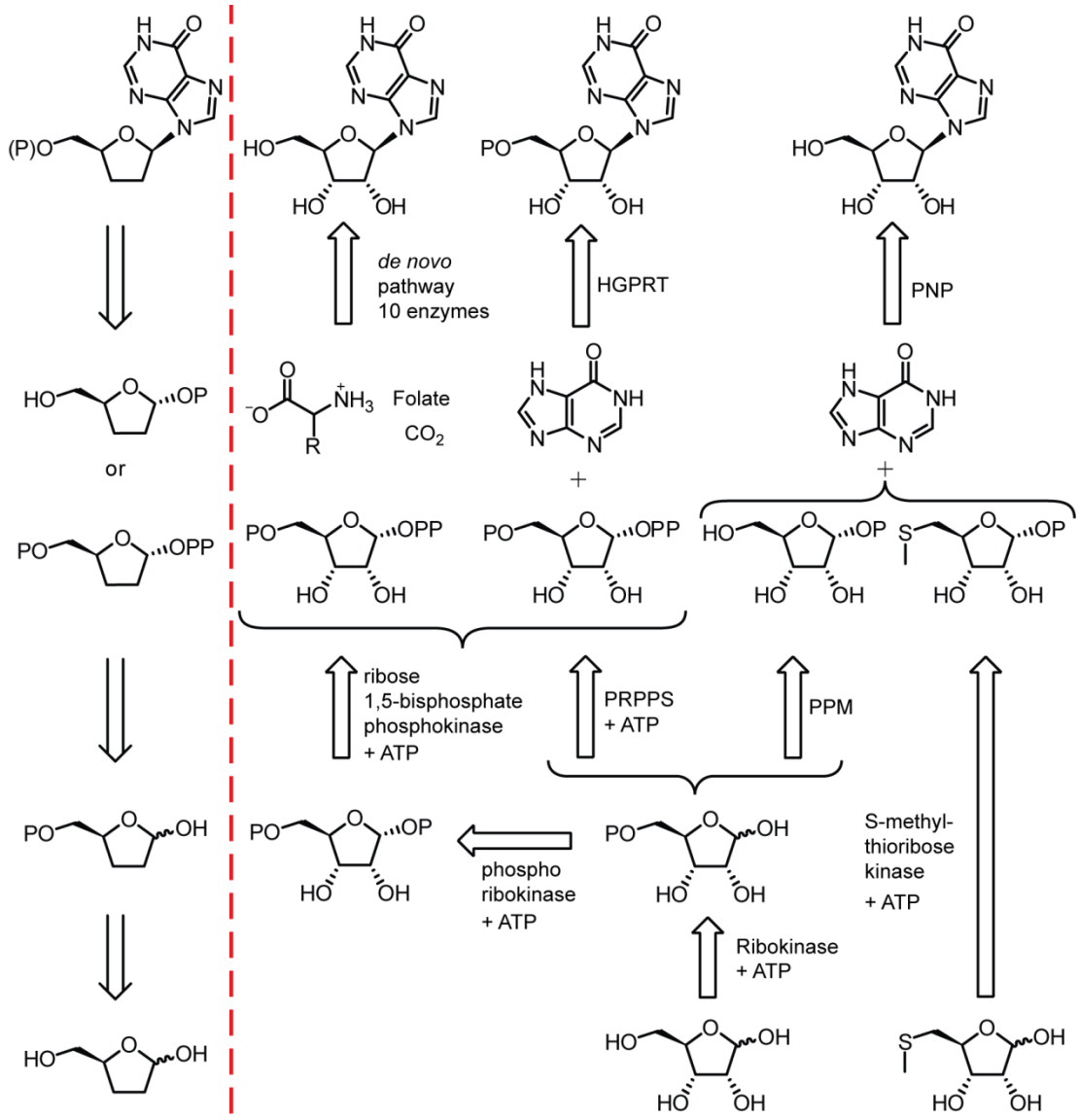

Figure 1-13. Bioretrosynthetic analysis of ddI for identification of potential biocatalytic pathways. Enzymatic transformations which generate molecules similar to ddI (e.g. inosine or inosine monophosphate) are collected and compared. Each backward step brings about a simpler molecule. Analogous molecules for ddI synthesis are shown to the left of the red line. HGPRT – hypoxanthine-guanine phosphoribosyl transferase; PNP – purine nucleoside phosphorylase; PRPPS – phosphoribosyl pyrophosphate synthase; PPM – phosphopentomutase.

In nature, salvage enzymes exist to scavenge exogenous sources of sugar and nucleobase to bypass the *de novo* synthesis process. Hypoxanthine-guanine phosphoribosyl transferase (HGPRT, EC 2.4.2.8) generates inosine monophosphate

from hypoxanthine and PRPP, an early precursor in nucleotide synthesis. Alternatively, inosine is formed by purine nucleoside phosphorylase (PNP, 2.4.2.1) from ribose 1-phosphate and hypoxanthine, which is then phosphorylated to form the metabolically useful monophosphate. HGPRT and PNP are well-studied enzymes, each having been crystallized and with significant literature background available. To generate ddI utilizing a PRPP analog, however, will require dephosphorylation of the final product; if necessary, this could be accomplished by certain nucleotidases[144]. Potential substrates for these enzymes include 5-phospho-2,3-dideoxyribose α-1-pyrophosphate and α-2,3-dideoxyribose 1-phosphate.

Along this line, PRPP is formed intracellularly from other ribose phosphates, namely ribose 1,5-bisphosphate and ribose 1-phosphate. Ribose 1,5-bisphosphate phosphokinase (EC 2.7.4.23) and phosphoribosyl pyrophosphate synthetase (PRPPS, EC 2.7.6.1), utilize ATP to form PRPP and ADP or AMP, respectively. PRPPS from multiple species have been studied extensively, crystallized numerous times, and characterized with numerous substrates and through mutagenesis. An abundance of data makes PRPPS a promising target for enzyme engineering. Ribose 1,5-bisphosphate phosphokinase, on the other hand, has only recently been described as a gene product in an *E. coli* NAD salvage pathway[145], making a difficult and risky enzyme to work with. Ribose 1,5-bisphosphate, in turn, is synthesized from ribose 5-phosphate by phosphoribokinase (EC 2.7.1.18), another enzyme which has been minimally studied.

Ribose 1-phosphate can be formed directly from ribose 5-phosphate by phosphopentomutase (EC 5.4.2.7), a second enzyme in the nucleoside salvage pathway. Phosphopentomutases (PPMs) have been identified in mammalian and bacterial species and activity for dideoxyribose phosphates shown to exist in certain bacterial isolates[146]. However, minimal structural data is available on PPMs and there are certain inconsistencies in the required cofactors and likely mechanism. (These

inconsistencies are described and explained in Chapter 4.) The formation of ribose 5-phosphate from the general precursor ribose can be afforded by application of ribokinase (EC 2.7.1.15) and ATP. Ribokinase has been extensively studied and its mechanism of action described for many species. 2,3-dideoxyribose is, therefore, a potential precursor for ribokinase on the way to ddI. Thus, the pathway going through PRPP requires 3–4 enzymes to reach inosine monophosphate and the equivalent of 3 molecules of ATP. Incorporation of PPM into the pathway affords the use of a single molecule of ATP and three total enzymes.

Other enzymes are available to phosphorylate cyclic sugar derivatives outside of those for ribose described above. S-Methyl-5-thioribose kinase (EC 2.7.1.100) acts in the bacterial methionine salvage pathway enabling the consumption of *S*-methyl-5-thioribose and *S*-methyl-5´-thioadenosine[147]. *S*-Methyl-5-thioribose kinase has been crystallized with S-methyl-5-thioribose and nucleotide substrates enabling the determinants of substrate specificity to be elucidated. Incorporation of S-methyl-5-thioribose kinase would allow for direct biosynthesis of dideoxyribose 1-phosphate from dideoxyribose and thus, yield a short, two enzyme pathway to ddI. However, in addition to mutations required for changes of the substrate at the 2- and 3-position, mutations which allow the kinase to bind 5-hydroxy sugars will be necessary, making this a higher-risk pathway.

Dideoxyribose is a simple sugar derivative which can be chemically synthesized from glutamic acid using methods similar to the synthesis of ddI by Farina *et al*[136]. As stated previously, this strategy maintains the stereochemistry of the starting material. Glutamic acid, a commodity material with application in flavorings, cyclizes under acidic conditions in the presence of $NaNO_2$ to form (*S*)-γ-butyrolactone-γ-carboxylic acid. Step-wise reduction of the carboxylic acid and the lactone will form the desired (*S*)-γ-hydroxymethyl-γ-butyrolactol. Modification of the Farina *et al* synthesis pathway to utilize

enzymes earlier in the process alleviates the generation of α-nucleosides, immediately increasing the yield by reducing the amount of side products formed.

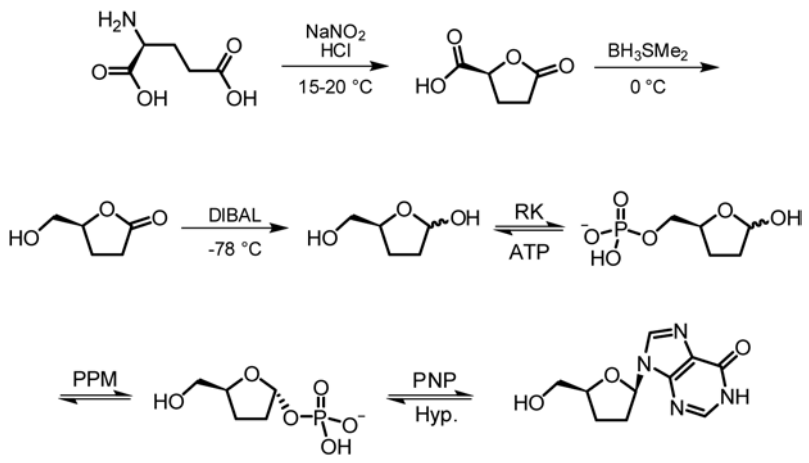

Figure 1-14. Biocatalytic pathway for directed biosynthesis of 2´,3´-dideoxyinosine starting from glutamic acid and utilizing ribokinase (RK), phosphopentomutase (PPM) and purine nucleoside phosphorylase (PNP)

Each of the previously described enzymes naturally functions on ribose or a ribose-like molecule and activity is likely to be low for the dideoxy derivative. Because of this, enzyme engineering will be a necessary part of the pathway optimization. Despite great effort and years of study, some reported here, a limited understanding of enzyme mutagenesis on structure and function makes the outcome of mutagenesis studies difficult to predict, even under ideal conditions. For this reason, the shorter pathway utilizing *S*-methyl-5-thioribose kinase carries a certain element of risk. A pathway involving PNP, PPM, and ribokinase was chosen as a probable pathway for synthesis of ddI. The resulting pathway is depicted in Figure 1-14.

Dissertation Statement

The main topic of this dissertation is the generation of a *de novo* biocatalytic pathway for ddI. A three enzyme pathway for biosynthesis of ddI from glutamic acid and

hypoxanthine has been designed using a bioretrosynthetic approach. Stereo- and regioselective properties of enzymes will greatly simplify the synthesis of ddI. Production of this critical anti-HIV compound from commodity materials could provide a significantly less expensive method for ddI synthesis as compared to the natural nucleoside precursors likely used in chemical production. Since a significant portion of nucleoside analog drug cost is accrued during production of the active ingredient, even a small decrease in price will alleviate the financial burden placed on HIV-infected individuals in poor countries. An *in vitro* demonstration of the pathway is discussed in chapter 5.

For biocatalytic production to be feasible at process levels, the turnover of each enzyme in the pathway must be improved. Rational active site redesign involving site-directed mutagenesis based on structural, functional and binding data and directed evolution are methods which enable tuning of enzymatic properties to the required task. The application of computational design tools and directed evolution to improve the final enzyme in the proposed pathway, human purine nucleoside phosphorylase, is demonstrated in chapters 2 and 3, respectively.

Chapter 4 focuses on the biochemical and structural characterization of phosphopentomutase, the penultimate enzyme in the proposed pathway. Rational design requires knowledge of the substrate binding mode within the enzyme. Prior to this work, no crystal structures of a bacterial phosphopentomutase had been described in the literature. Furthermore, the catalytic cycle had not been fully described. This work identifies critical enzyme-substrate interactions and clarifies the turnover cycle of PPM.

Microbial factories make possible efficient production of complex molecules utilizing environmentally friendly reaction conditions and offer many advantages over large-scale chemical syntheses. The work presented in this dissertation provides a base for subsequent *in vivo* pathway demonstration, improvement of each enzyme in the proposed pathway and metabolic engineering of organisms for microbial ddI production.

References

1. Voigt, C. A., Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **2006,** 17, (5), 548-57.

2. Marner, W. D., 2nd, Practical application of synthetic biology principles. *Biotechnol. J.* **2009,** 4, (10), 1406-19.

3. Endy, D., Foundations for engineering biology. *Nature* **2005,** 438, (7067), 449-53.

4. Keasling, J. D., Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* **2008,** 3, (1), 64-76.

5. Na, D.; Kim, T. Y.; Lee, S. Y., Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.* **2010,** 13, (3), 363-370.

6. Levskaya, A.; Chevalier, A. A.; Tabor, J. J.; Simpson, Z. B.; Lavery, L. A.; Levy, M.; Davidson, E. A.; Scouras, A.; Ellington, A. D.; Marcotte, E. M.; Voigt, C. A., Synthetic biology: engineering *Escherichia coli* to see light. *Nature* **2005,** 438, (7067), 441-442.

7. Basu, S.; Gerchman, Y.; Collins, C. H.; Arnold, F. H.; Weiss, R., A synthetic multicellular system for programmed pattern formation. *Nature* **2005,** 434, (7037), 1130-4.

8. Anderson, J. C.; Voigt, C. A.; Arkin, A. P., Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **2007,** 3, 133.

9. Tamsir, A.; Tabor, J. J.; Voigt, C. A., Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. In *Nature*, 2010/12/15 ed.

10. Regot, S.; Macia, J.; Conde, N.; Furukawa, K.; Kjellen, J.; Peeters, T.; Hohmann, S.; de Nadal, E.; Posas, F.; Sole, R., Distributed biological computation with multicellular engineered networks. *Nature* **2011,** 469, (7329), 207-211.

11. Draths, K. M.; Knop, D. R.; Frost, J. W., Shikimic acid and quinic acid: replacing isolation from plant sources with recombinant microbial biocatalysis. *J. Am. Chem. Soc.* **1999,** 121, (7), 1603-1604.

12. 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.; Huisman, G. W., Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007,** 25, (3), 338-344.

13. 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.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010,** 329, (5989), 305-309.

14. Organization, W. H., Transaction prices for antiretroviral medicines and HIV diagnostics from 2008 to March 2010. *GPRM - Global Price Reporting Mechanism* **2010,** May 2010, 1-44.

15. Bartlett, J. G.; Auwaerter, P. G.; Phan, P. A., *The Johns Hopkins POC-IT Center ABX Guide: diagnosis & treatment of infectious diseases, second edition*. Jones & Bartlett Learning: Baltimore Maryland, 2010.

16. Pinheiro, E.; Vasan, A.; Kim, J. Y.; Lee, E.; Guimier, J. M.; Perriens, J., Examining the production costs of antiretroviral drugs. *Aids* **2006,** 20, (13), 1745-1752.

17. Martin, J. C.; Hitchcock, M. J. M.; De Clercq, E.; Prusoff, W. H., Early nucleoside reverse transcriptase inhibitors for the treatment of HIV: A brief history of stavudine (D4T) and its comparison with other dideoxynucleosides. *Antiviral Res.* **2010,** 85, (1), 34-38.

18. Lee, S. Y.; Kim, H. U.; Park, J. H.; Park, J. M.; Kim, T. Y., Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov. Today* **2009,** 14, (1-2), 78-88.

19. Martin, C. H.; Nielsen, D. R.; Solomon, K. V.; Prather, K. L., Synthetic metabolism: engineering biology at the protein and pathway scales. *Chem. Biol.* **2009,** 16, (3), 277-286.

20. Atsumi, S.; Cann, A. F.; Connor, M. R.; Shen, C. R.; Smith, K. M.; Brynildsen, M. P.; Chou, K. J.; Hanai, T.; Liao, J. C., Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **2008,** 10, (6), 305-311.

21. Steen, E. J.; Chan, R.; Prasad, N.; Myers, S.; Petzold, C. J.; Redding, A.; Ouellet, M.; Keasling, J. D., Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* **2008,** 7, 36.

22. Hanai, T.; Atsumi, S.; Liao, J. C., Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* **2007,** 73, (24), 7814-8.

23. Dalby, P. A.; Baganz, F.; Lye, G. J.; Ward, J. M., Protein and pathway engineering in biocatalysis. *Chim. Oggi* **2009,** 27, (4), 18-22.

24. Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D., Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **2003,** 21, (7), 796-802.

25. 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.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** 440, (7086), 940-3.

26. Atsumi, S.; Hanai, T.; Liao, J. C., Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008,** 451, (7174), 86-89.

27. Zhang, K.; Sawaya, M. R.; Eisenberg, D. S.; Liao, J. C., Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (52), 20653-8.

28. Bernhardt, P.; McCoy, E.; O'Connor, S. E., Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **2007,** 14, (8), 888-897.

29. Runguphan, W.; O'Connor, S. E., Metabolic reprogramming of periwinkle plant culture. *Nat. Chem. Biol.* **2009,** 5, (3), 151-153.

30. Tsuge, T.; Hisano, T.; Taguchi, S.; Doi, Y., Alteration of chain length substrate specificity of *Aeromonas caviae* R-enantiomer-specific enoyl-coenzyme A hydratase through site-directed mutagenesis. *Appl. Environ. Microbiol.* **2003,** 69, (8), 4830-6.

31. Chang, Y. Y.; Cronan, J. E., Jr., Conversion of *Escherichia coli* pyruvate oxidase to an 'α-ketobutyrate oxidase'. *Biochem. J.* **2000,** 352 Pt 3, 717-24.

32. Moon, T. S.; Yoon, S. H.; Lanza, A. M.; Roy-Mayhew, J. D.; Prather, K. L., Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* **2009,** 75, (3), 589-95.

33. Nakamura, C. E.; Whited, G. M., Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **2003,** 14, (5), 454-459.

34. Perez, J. M.; Arenas, F. A.; Pradenas, G. A.; Sandoval, J. M.; Vasquez, C. C., *Escherichia coli* YqhD exhibits aldehyde reductase activity and protects from the harmful effect of lipid peroxidation-derived aldehydes. *J. Biol. Chem.* **2008,** 283, (12), 7346-53.

35. Niu, W.; Molefe, M. N.; Frost, J. W., Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **2003,** 125, (43), 12998-12999.

36. Prather, K. L.; Martin, C. H., *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **2008,** 19, (5), 468-74.

37. Bachmann, B. O., Biosynthesis: is it time to go retro? *Nat. Chem. Biol.* **2010,** 6, (6), 390-393.

38. Hatzimanikatis, V.; Li, C.; Ionita, J. A.; Henry, C. S.; Jankowski, M. D.; Broadbelt, L. J., Exploring the diversity of complex metabolic networks. *Bioinformatics* **2005,** 21, (8), 1603-9.

39. Gonzalez-Lergier, J.; Broadbelt, L. J.; Hatzimanikatis, V., Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J. Am. Chem. Soc.* **2005,** 127, (27), 9930-8.

40. Ellis, L. B.; Gao, J.; Fenner, K.; Wackett, L. P., The University of Minnesota pathway prediction system: predicting metabolic logic. *Nucleic Acids Res.* **2008,** 36, (Web Server issue), W427-32.

41. Fenner, K.; Gao, J.; Kramer, S.; Ellis, L.; Wackett, L., Data-driven extraction of relative reasoning rules to limit combinatorial explosion in biodegradation pathway prediction. *Bioinformatics* **2008,** 24, (18), 2079-85.

42. Bar-Even, A.; Noor, E.; Lewis, N. E.; Milo, R., Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** 107, (19), 8889-94.

43. Cho, A.; Yun, H.; Park, J. H.; Lee, S. Y.; Park, S., Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.* **2010,** 4, 35.

44. Rodrigo, G.; Carrera, J.; Prather, K. J.; Jaramillo, A., DESHARKY: automatic design of metabolic pathways for optimal cell growth. *Bioinformatics* **2008,** 24, (21), 2554-6.

45. Carothers, J. M.; Goler, J. A.; Keasling, J. D., Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.* **2009,** 20, (4), 498-503.

46. Horowitz, N. H., On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U. S. A.* **1945,** 31, (6), 153-157.

47. Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L., The Miller volcanic spark discharge experiment. *Science* **2008,** 322, (5900), 404.

48. Tao, J. H.; Xu, J. H., Biocatalysis in development of green pharmaceutical processes. *Curr. Opin. Chem. Biol.* **2009,** 13, (1), 43-50.

49. Jackel, C.; Kast, P.; Hilvert, D., Protein design by directed evolution. *Annu. Rev. Biophys.* **2008,** 37, 153-173.

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

51. Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z. G., Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, (7), 3778-83.

52. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **2009,** 13, (1), 26-34.

53. Baker, D., An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **2010,** 19, (10), 1817-1819.

54. Kaplan, J.; DeGrado, W. F., *De novo* design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, (32), 11566-11570.

55. Cochran, F. V.; Wu, S. P.; Wang, W.; Nanda, V.; Saven, J. G.; Therien, M. J.; DeGrado, W. F., Computational *de novo* design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J. Am. Chem. Soc.* **2005,** 127, (5), 1346-1347.

56. Koder, R. L.; Anderson, J. L.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L., Design and engineering of an O(2) transport protein. *Nature* **2009,** 458, (7236), 305-309.

57. Bolon, D. N.; Mayo, S. L., Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, (25), 14274-14279.

58. Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Rothlisberger, D.; Baker, D., New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **2006,** 15, (12), 2785-2794.

59. 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.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** 453, (7192), 190-U4.

60. Khersonsky, O.; Rothlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S., Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **2010,** 396, (4), 1025-42.

61. 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.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., *De novo* computational design of retro-aldol enzymes. *Science* **2008,** 319, (5868), 1387-1391.

62. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010,** 329, (5989), 309-313.

63. Lassila, J. K.; Keeffe, J. R.; Oelschlaeger, P.; Mayo, S. L., Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity. *Protein Eng. Des. Sel.* **2005,** 18, (4), 161-163.

64. Lassila, J. K.; Keeffe, J. R.; Kast, P.; Mayo, S. L., Exhaustive mutagenesis of six secondary active-site residues in *Escherichia coli* chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalyst. *Biochemistry* **2007,** 46, (23), 6883-6891.

65. Zheng, F.; Yang, W.; Ko, M. C.; Liu, J.; Cho, H.; Gao, D.; Tong, M.; Tai, H. H.; Woods, J. H.; Zhan, C. G., Most efficient cocaine hydrolase designed by virtual screening of transition states. *J. Am. Chem. Soc.* **2008,** 130, (36), 12148-55.

66. Lilien, R. H.; Stevens, B. W.; Anderson, A. C.; Donald, B. R., A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase a phenylalanine adenylation enzyme. *J. Comput. Biol.* **2005,** 12, (6), 740-761.

67. Chen, C. Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R., Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (10), 3764-9.

68. Murphy, P. M.; Bolduc, J. M.; Gallaher, J. L.; Stoddard, B. L.; Baker, D., Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (23), 9215-20.

69. Dougherty, M. J.; Arnold, F. H., Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **2009,** 20, (4), 486-91.

70. Johannes, T. W.; Zhao, H. M., Directed evolution of enzymes and biosynthetic pathways. *Curr. Opin. Microbiol.* **2006,** 9, (3), 261-267.

71. Glieder, A.; Farinas, E. T.; Arnold, F. H., Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* **2002,** 20, (11), 1135-1139.

72. Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H., Regio- and enantioselective alkane hydroxylation with engineered cytochromes P450 BM-3. *J. Am. Chem. Soc.* **2003,** 125, (44), 13442-50.

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

74. Meinhold, P.; Peters, M. W.; Chen, M. M. Y.; Takahashi, K.; Arnold, F. H., Direct conversion of ethane to ethanol by engineered cytochrome P450BM3. *ChemBioChem* **2005,** 6, (10), 1765-1768.

75. Johannes, T. W.; Woodyer, R. D.; Zhao, H., Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. *Appl. Environ. Microbiol.* **2005,** 71, (10), 5728-34.

76. McLachlan, M. J.; Johannes, T. W.; Zhao, H., Further improvement of phosphite dehydrogenase thermostability by saturation mutagenesis. *Biotechnol. Bioeng.* **2008,** 99, (2), 268-74.

77. Reetz, M. T.; Prasad, S.; Carballeira, J. D.; Gumulya, Y.; Bocola, M., Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **2010,** 132, (26), 9144-9152.

78. Dietrich, J. A.; McKee, A. E.; Keasling, J. D., High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* **2010,** 79, 563-90.

79. Boersma, Y. L.; Droge, M. J.; Quax, W. J., Selection strategies for improved biocatalysts. *FEBS J.* **2007,** 274, (9), 2181-95.

80. Arnold, F. H.; Georgiou, G., *Directed enzyme evolution : screening and selection methods*. Humana Press: Totowa, N.J., 2003; p xv, 383 p.

81. Olsen, M.; Iverson, B.; Georgiou, G., High-throughput screening of enzyme libraries. *Curr. Opin. Biotechnol.* **2000,** 11, (4), 331-337.

82. Wahler, D.; Reymond, J. L., Novel methods for biocatalyst screening. *Curr. Opin. Chem. Biol.* **2001,** 5, (2), 152-158.

83. Akanuma, S.; Yamagishi, A.; Tanaka, N.; Oshima, T., Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. *Protein Sci.* **1998,** 7, (3), 698-705.

84. Rothman, S. C.; Kirsch, J. F., How does an enzyme evolved *in vitro* compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J. Mol. Biol.* **2003,** 327, (3), 593-608.

85. Yano, T.; Oue, S.; Kagamiyama, H., Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** 95, (10), 5511-5515.

86. Barak, Y.; Ackerley, D. F.; Dodge, C. J.; Banwari, L.; Alex, C.; Francis, A. J.; Matin, A., Analysis of novel soluble chromate and uranyl reductases and generation of an improved enzyme by directed evolution. *Appl. Environ. Microbiol.* **2006,** 72, (11), 7074-7082.

87. 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.; Lassner, M. W., Discovery and directed evolution of a glyphosate tolerance gene. *Science* **2004,** 304, (5674), 1151-1154.

88. Cho, C. M. H.; Mulchandani, A.; Chen, W., Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Appl. Environ. Microbiol.* **2004,** 70, (8), 4681-4685.

89. Claren, J.; Malisi, C.; Hocker, B.; Sterner, R., Establishing wild-type levels of catalytic activity on natural and artificial (β/α)(8)-barrel protein scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (10), 3704-3709.

90. Gulick, A. M.; Fahl, W. E., Forced evolution of glutathione-S-transferase to create a more efficient drug detoxication enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** 92, (18), 8140-8144.

91. Hoseki, J.; Yano, T.; Koyama, Y.; Kuramitsu, S.; Kagamiyama, H., Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus*. *J. Biochem.* **1999,** 126, (5), 951-956.

92. Landis, D. M.; Loeb, L. A., Random sequence mutagenesis and resistance to 5-fluorouridine in human thymidylate synthases. *J. Biol. Chem.* **1998,** 273, (40), 25809-25817.

93. Stemmer, W. P. C., DNA shuffling by random fragmentation and reassembly - *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** 91, (22), 10747-10751.

94. Walter, K. U.; Vamvaca, K.; Hilvert, D., An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* **2005,** 280, (45), 37742-37746.

95. Zhang, K. C.; Li, H.; Cho, K. M.; Liao, J. C., Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** 107, (14), 6234-6239.

96. Black, M. E.; Newcomb, T. G.; Wilson, H. M.; Loeb, L. A., Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, (8), 3525-9.

97. Christians, F. C.; Scapozza, L.; Crameri, A.; Folkers, G.; Stemmer, W. P., Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* **1999,** 17, (3), 259-264.

98. Hibbert, E. G.; Senussi, T.; Costelloe, S. J.; Lei, W. L.; Smith, M. E. B.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase activity on non-phosphorylated substrates. *J. Biotechnol.* **2007,** 131, (4), 425-432.

99. Ohki, T.; Shibata, N.; Higuchi, Y.; Kawashima, Y.; Takeo, M.; Kato, D.; Negoro, S., Two alternative modes for optimizing nylon-6 byproduct hydrolytic activity from a carboxylesterase with a beta-lactamase fold: X-ray

crystallographic analysis of directly evolved 6-aminohexanoate-dimer hydrolase. *Protein Sci.* **2009,** 18, (8), 1662-1673.

100. Wei, C. L.; Yang, Y. B.; Deng, C. H.; Liu, W. C.; Hsu, J. S.; Lin, Y. C.; Liaw, S. H.; Tsai, Y. C., Directed evolution of *Streptomyces clavuligerus* deacetoxycephalosporin C synthase for enhancement of penicillin G expansion. *Appl. Environ. Microbiol.* **2005,** 71, (12), 8873-8880.

101. Bhuiya, M.-W.; Liu, C.-J., Engineering monolignol 4-O-methyltransferases to modulate lignin biosynthesis. *J. Biol. Chem.* **2010,** 285, (1), 277-285.

102. Hsu, C. C.; Hong, Z.; Wada, M.; Franke, D.; Wong, C. H., Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. *Proc Natl Acad Sci U S A* **2005,** 102, (26), 9122-6.

103. Williams, G. J.; Domann, S.; Nelson, A.; Berry, A., Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** 100, (6), 3143-8.

104. Woodhall, T.; Williams, G.; Berry, A.; Nelson, A., Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics. *Angew. Chem., Int. Ed.* **2005,** 44, (14), 2109-2112.

105. Matsumura, I.; Wallingford, J. B.; Surana, N. K.; Vize, P. D.; Ellington, A. D., Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase. *Nat. Biotechnol.* **1999,** 17, (7), 696-701.

106. Schmidt-Dannert, C., Engineering novel carotenoids in microorganisms. *Curr. Opin. Biotechnol.* **2000,** 11, (3), 255-261.

107. Schmidt-Dannert, C.; Umeno, D.; Arnold, F. H., Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.* **2000,** 18, (7), 750-3.

108. Hida, K.; Hanes, J.; Ostermeier, M., Directed evolution for drug and nucleic acid delivery. *Adv. Drug Delivery Rev.* **2007,** 59, (15), 1562-78.

109. Fernandez-Gacio, A.; Uguen, M.; Fastrez, J., Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **2003,** 21, (9), 408-414.

110. Jose, J., Autodisplay: efficient bacterial surface display of recombinant proteins. *Appl. Microbiol. Biotechnol.* **2006,** 69, (6), 607-14.

111. Sunbul, M.; Marshall, N. J.; Zou, Y. K.; Zhang, K. Y.; Yin, J., Catalytic Turnover-Based Phage Selection for Engineering the Substrate Specificity of Sfp Phosphopantetheinyl Transferase. *J. Mol. Biol.* **2009,** 387, (4), 883-898.

112. Antipov, E.; Cho, A. E.; Wittrup, K. D.; Klibanov, A. M., Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (46), 17694-17699.

113. Droge, M. J.; Boersma, Y. L.; van Pouderoyen, G.; Vrenken, T. E.; Ruggeberg, C. J.; Reetz, M. T.; Dijkstra, B. W.; Quax, W. J., Directed evolution of *Bacillus subtilis* lipase A by use of enantiomeric phosphonate inhibitors: crystal structures and phage display selection. *ChemBioChem* **2006,** 7, (1), 149-157.

114. Doms, A.; Schroeder, M., GoPubMed: Exploring PubMed with the gene ontology. *Nucleic Acids Res.* **2005,** 33, W783-W786.

115. Fujii, R.; Nakagawa, Y.; Hiratake, J.; Sogabe, A.; Sakata, K., Directed evolution of *Pseudomonas aeruginosa* lipase for improved amide-hydrolyzing activity. *Protein Eng. Des. Sel.* **2005,** 18, (2), 93-101.

116. Hawwa, R.; Larsen, S. D.; Ratia, K.; Mesecar, A. D., Structure-based and random mutagenesis approaches increase the organophosphate-degrading activity of a phosphotriesterase homologue from *Deinococcus radiodurans*. *J. Mol. Biol.* **2009,** 393, (1), 36-57.

117. Morley, K. L.; Kazlauskas, R. J., Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2005,** 23, (5), 231-7.

118. Paramesvaran, J.; Hibbert, E. G.; Russell, A. J.; Dalby, P. A., Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. *Protein Eng. Des. Sel.* **2009,** 22, (7), 401-411.

119. Parikh, M. R.; Matsumura, I., Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of β-fucosidase from β-galactosidase. *J. Mol. Biol.* **2005,** 352, (3), 621-628.

120. Liu, L. F.; Li, Y. F.; Liotta, D.; Lutz, S., Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res.* **2009,** 37, (13), 4472-4481.

121. Clerq, E. D., Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents* **2009,** 33, (4), 307-320.

122. Pugmire, M. J.; Ealick, S. E., Structural analyses reveal two distinct families of nucleoside phosphorylases. *Biochem. J.* **2002,** 361, 1-25.

123. Schramm, V. L., Enzymatic transition states: thermodynamics, dynamics and analogue design. *Arch. Biochem. Biophys.* **2005,** 433, (1), 13-26.

124. Webb, R. R.; Wos, J. A.; Martin, J. C.; Brodfuehrer, P. R., Synthesis of 2',3'-Dideoxyinosine. *Nucleosides Nucleotides* **1988,** 7, (2), 147-153.

125. Mitsuya, H.; Broder, S., Inhibition of the *in vitro* infectivity and cytopathic effect of human lymphotrophic-T virus type III/lymphadenopathy-associated virus (HTLV-III/LAV) by 2',3'-dideoxynucleosides. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** 83, (6), 1911-1915.

126. Huryn, D. M.; Okabe, M., AIDS-driven nucleoside chemistry. *Chem. Rev. (Washington, DC, U. S.)* **1992,** 92, (8), 1745-1768.

127. Prisbe, E. J.; Martin, J. C., A novel and efficient preparation of 2',3'-dideoxynucleosides. *Synth. Commun.* **1985,** 15, (5), 401-409.

128. Shiragami, H.; Amino, Y.; Honda, Y.; Arai, M.; Tanaka, Y.; Iwagami, H.; Yukawa, T.; Izawa, K., Synthesis of 2',3'-dideoxypurinenucleosides via the palladium catalyzed reduction of 9-(2,5-di-O-acetyl-3-bromo-3-deoxy-β-D-xylofuranosyl)purine derivatives. *Nucleosides Nucleotides* **1996,** 15, (1-3), 31-45.

129. Zeidler, J.; Golankiewicz, B., Synthesis and H-1 and C-13 NMR spectral characteristics of 8-bromo-2',3'-dideoxyguanosine and 8-bromo-2',3'-dideoxyinosine. *Nucleosides Nucleotides* **1996,** 15, (5), 1077-1095.

130. Ciuffreda, P.; Casati, S.; Santaniello, E., Lipase-catalyzed protection of the hydroxy groups of the nucleosides inosine and 2'-deoxyinosine: a new

chemoenzymatic synthesis of the antiviral drug 2',3'-dideoxyinosine. *Bioorg. Med. Chem. Lett.* **1999,** 9, (11), 1577-1582.

131. Bussolari, J. C.; Panzica, R. P., Synthesis and anti-HIV evaluation of 2',3'-dideoxy imidazo- and nu-triazolo[4,5-d]pyridazine nucleosides. *Bioorg. Med. Chem.* **1999,** 7, (11), 2373-2379.

132. Chu, C. K.; Bhadti, V. S.; Doboszewski, B.; Gu, Z. P.; Kosugi, Y.; Pullaiah, K. C.; Vanroey, P., General syntheses of 2',3'-dideoxynucleosides and 2',3'-didehydro-2',3'-dideoxynucleosides. *J. Org. Chem.* **1989,** 54, (9), 2217-2225.

133. Torii, T.; Izawa, K.; Cho, D. H.; Jang, D. O., Synthesis of 2',3'-dideoxyinosine via radical deoxygenation. *Nucleosides, Nucleotides Nucleic Acids* **2007,** 26, (8-9), 985-988.

134. Saito, Y.; Zevaco, T. A.; Agrofoglio, L. A., Chemical synthesis of C-13 labeled anti-HIV nucleosides as mass-internal standards. *Tetrahedron* **2002,** 58, (47), 9593-9603.

135. Bhat, V.; Stocker, E.; Ugarkar, B. G., A new synthesis of 2',3'-dideoxyinosine. *Synth. Commun.* **1992,** 22, (10), 1481-1486.

136. Farina, V.; Benigni, D. A., A new synthesis of 2',3'-dideoxynucleosides for AIDS chemotherapy. *Tetrahedron Lett.* **1988,** 29, (11), 1239-1242.

137. Ford, H.; Siddiqui, M. A.; Driscoll, J. S.; Marquez, V. E.; Kelley, J. A.; Mitsuya, H.; Shirasaka, T., Lipophilic, acid-stable, adenosine deaminase-activated anti-HIV prodrugs for central-nervous-system delivery. 2. 6-halo and 6-alkoxy prodrugs of 2'-β-fluoro-2',3'-dideoxyinosine. *J. Med. Chem.* **1995,** 38, (7), 1189-1195.

138. Meier, C.; Knispel, T.; DeClercq, E.; Balzarini, J., ADA-bypass by lipophilic cycloSal-ddAMP pro-nucleotides a second example of the efficiency of the cycloSal-concept. *Bioorg. Med. Chem. Lett.* **1997,** 7, (12), 1577-1582.

139. Beach, J. W.; Kim, H. O.; Jeong, L. S.; Nampalli, S.; Islam, Q.; Ahn, S. K.; Babu, J. R.; Chu, C. K., A highly stereoselective synthesis of anti-HIV 2',3'-dideoxynucleosides and 2',3'-didehydro-2',3'-dideoxynucleosides. *J. Org. Chem.* **1992,** 57, (14), 3887-3894.

140. Rogert, M. C.; Trelles, J. A.; Porro, S.; Lewkowicz, E. S.; Iribarren, A. M., Microbial synthesis of antiviral nucleosides using *Escherichia coli* BL21 as biocatalyst. *Biocatal. Biotransform.* **2002,** 20, (5), 347-351.

141. Shirae, H.; Kobayashi, K.; Shiragami, H.; Irie, Y.; Yasuda, N.; 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, (2), 419-424.

142. Skonezny, P. M. Process for preparing dideoxyinosine using adenosine deaminase enzyme. US 2004/0175804 A1, Sept. 9, 2004, 2004.

143. Yokozeki, K. Method of producing 2',3'-dideoxyinosine. 4970148, Nov. 13, 1990, 1990.

144. Cihlar, T.; Ray, A. S., Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **2010,** 85, (1), 39-58.

145. Hove-Jensen, B.; Rosenkrantz, T. J.; Haldimann, A.; 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, (9), 2793-2801.

146. Hamamoto, T.; Noguchi, T.; Midorikawa, Y., Phosphopentomutase of *Bacillus stearothermophilus* TH6-2: the enzyme and its gene ppm. *Biosci., Biotechnol., Biochem.* **1998,** 62, (6), 1103-1108.

147. Ku, S. Y.; Yip, P.; Cornell, K. A.; Riscoe, M. K.; Behr, J. B.; Guillerm, G.; Howell, P. L., Structures of 5-methylthioribose kinase reveal substrate specificity and unusual mode of nucleotide binding. *J. Biol. Chem.* **2007,** 282, (30), 22195-22206.

Chapter II

RATIONAL AND COMPUTATIONAL DESIGN OF THE SUBSTRATE BINDING POCKET OF HUMAN PURINE NUCLEOSIDE PHOSPHORYLASE[1]

Introduction

Nucleoside analog drugs are primary therapeutics in the treatment of viral infections including HIV[2] and hepatitis[3]. Indeed, it has recently been reported that of the 25 drugs currently approved in the USA for the treatment of HIV, eight are nucleoside analogs[2]. Nucleoside analogs are also currently in advanced stage clinical trials for the treatment of various cancers including leukemia[4]. The broad efficacy of these compounds contrasts with their price. The cost of the manufacturing of the active ingredients of some of these drugs comprise up to 55–99% of the final therapeutic price[5], a fact which has spurred the continuing development of new methods for the synthesis of nucleoside analogs.

Currently, there is great interest in developing biocatalytic and chemo-enzymatic alternatives for the synthesis of nucleoside analogs, which will enable regio- and enantioselective synthesis of these critical compounds. Toward this aim, 2′,3′-dideoxyinosine was targeted for directed biosynthesis as a representative of the broader class of dideoxynucleoside drugs. Dideoxyinosine (ddI) is a nucleoside reverse transcriptase inhibitor and is a close analog of the primary metabolite inosine. Both *de novo* biosynthesis and purine salvage pathways for inosine have been extensively characterized in biochemical and structural studies[6, 7]. As a large fraction of nucleoside analogs are variants of 2′,3′-dideoxynucleosides, methods developed for ddI may have broader application[2].

Enzymes with new or improved functions are increasingly generated from existing enzymes or scaffolds by a 2-fold strategy consisting of (i) rational mutational active site remodeling, to modify binding specificity for a desired reaction or substrate, followed by (ii) optimization of global protein function by more stochastic methods such as directed evolution. Prerequisites to the first stage are the identification of a progenitor enzyme or scaffold with a suitable starting activity and acquisition of some knowledge of the active site geometry from structural data or homology models. If a suitable enzyme is identified, first or second shell interacting active site residues are selected for mutation and functional assessment[8-10]. As even the most prudently selected active site mutations may result in unexpected catalytic consequences, the complete complement of amino acids at targeted active site residues is often generated by saturation mutagenesis[11].

To focus these efforts, it would be desirable to identify a subset of potentially beneficial mutations, particularly in cases in which permutations of several active site residues may be required simultaneously. Many computational protein design strategies have been developed to engineer alternative ligand specificity into proteins and generate enzyme variants with improved activity on non-cognate substrates[12]. Recently, methods incorporated in the Rosetta program have been developed to estimate relative ligand–protein interaction energies with conformational flexibility. RosettaLigand employs a Monte Carlo-based search algorithm with protein side chains replaced by residues from a rotamer library[13, 14]. It is easily combined with RosettaDesign[15] allowing for redesign of binding pockets[16]. Indeed, the Rosetta framework has been extended to the design of enzymes catalyzing both natural (retro-aldol[17]) and nonnatural (Kemp elimination[18]) reactions in the context of novel protein scaffolds[19]. The protocol identifies critical interactions between transition state models and catalytic residues. Subsequently, adjacent amino acids are placed in the binding site to optimize stability of the catalytic

residues and specificity for the ligand. The initial computational designs displayed low catalytic activity and were further improved through directed evolution and screening.

Herein, the identification of human purine nucleoside phosphorylase (hPNP) as an engineering candidate for nucleoside analog biocatalysis is described. Human PNP catalyzes the reversible synthesis or phosphorolysis of 6-oxopurine (deoxy)nucleosides. As PNPs play a critical role in human and pathogen physiology, numerous studies have been performed to biochemically and structurally characterize this class of enzymes. This knowledge permitted the identification of a single first shell active site residue, Tyr-88 (Y88), as a potential modulator of ribose substrate selectivity[20, 21]. Additionally, computational design of the hPNP binding site in the presence of inosine and ddI suggested Y88 as critical for differentiation of the two substrates. This result demonstrates that the computational method is generally applicable to focus experimental studies on specific sites and selected mutants thereby reducing experimental effort and accelerating research. RosettaLigand was used to predict changes in transition state binding free energy, or activation energy, upon mutation of this position to all genetically encoded amino acids except glycine and proline. Mutation to phenylalanine was consistently predicted to increase catalytic activity with respect to ddI. Experimental testing finds Y88F is 23.6-fold more efficient for ddI phosphorolysis than wild type hPNP and 2.9-fold less efficient for inosine. Systematic experimental analysis of all Y88X single mutants finds good agreement of Rosetta predictions with experimental data for all but negatively charged amino acids.

The present study differs from previous enzyme designs in that Rosetta is tested in a limiting setup where scaffold, binding mode and even a single site of mutation are predetermined. Thereby success hinges on accurate prediction of binding free energy changes for mutations at the Y88 site. The results demonstrate the general ability of Rosetta to identify favorable mutations in enzyme catalytic sites and enabled a

customization of the Rosetta energy function to improve correlation between predicted and experimentally determined transition state binding affinities to R = 0.65. However, these results point also to inaccuracies in handling electrostatics, in particular for charged amino acids, where further improvement of the Rosetta energy function is needed.

The Y88F variant then became the starting point for a directed evolution study. A high-throughput assay was developed and combined with error-prone PCR (epPCR) to generate and test libraries of PNP mutants. Three rounds of mutation and selection resulted in an enzyme with a modest 3-fold improvement in turnover compared with Y88F in *in vitro* assays. The directed evolution of PNP is described in Chapter 3.

To the best of our knowledge, this study comprises the first engineering of a nucleoside phosphorylase for the biosynthesis of an unnatural dideoxynucleoside. This work demonstrates that a combination approach of targeted active site mutation and directed evolution may find future application in nucleoside analog biosynthetic pathway engineering.

Methods

Transition State Model

Transition state models for docking and calculation of binding energies incorporate critical characteristics of the mechanism and transition state. The binding modes of substrates (PDB ID 1m73, 1a9s, 1rct, 1v2h, 1pwy, 1rfg), substrate analogs (PDB ID 1a9t, 1v41, 1v3q, 1v45, 1rt9) and transition state analog inhibitors (PDB ID 1pf7, 1rsz, 1rr6, 1b8o) crystallized with human and bovine PNP were studied for their resemblance to the experimentally-derived transition state (Figure 2-1A)[22]. Coordinates of ligands with characteristics matching the known transition state were used as internal coordinates of the transition state model. Phosphate coordinates are derived from ribose

1-phosphate crystallized in bovine PNP (PDB ID 1a9t) as O_P is oriented appropriately for nucleophilic attack (Figure 2-1B). The P–O_P bond in this structure is ~0.22 Å longer than the other P–O bonds recapitulating the known bond lengthening of ~0.23 Å as the reaction coordinate progresses[23]. The purine of DADME-Immucilin-H (PDB ID 1rsz) is the aromatic base for both inosine and ddI transition state models (Figure 2-1C)[24]. Separate coordinates for the sugar residues of inosine and ddI transition state models (Figure 2-1D,E) were taken from Immucilin-H (PDB ID 1pf7) and ddI (PDB ID 1v3q), respectively. The ring shape in ddI is flatter than that of nucleosides with hydroxyls at the 2′ or 3′ positions. Internal coordinates of the Immucilin-H sugar moiety were used for the inosine transition state model to best replicate the position of the 5′-OH over the 4′-O. The O5′-C5′-C4′-C3′ dihedral of Immucilin-H in 1pf7 is 62.72°. For ddI, the 5′-OH of the dideoxyribose moiety was rotated manually to match that of Immucilin-H. The ddI transition state model has an O5′-C5′-C4′-C3′ dihedral angle of 65.41°. The sugar moiety was aligned with the average C3′, C4′, and C5′ position from the aligned structures and a point directly between N9 of the purine and O_P of the phosphate which put the anomeric carbon 2.56 Å from N9 and 2.66 Å from O_P, both of which are shorter than the known value of 3 Å calculated from kinetic isotope effect[22].

Computational Mutation and Docking

Trimeric hPNP coordinates were obtained from the Protein Data Bank (PDB ID 1rct, 1pf7, 1rr6, 1v3q) and relaxed using a combination of Monte Carlo rotamer replacement and gradient based minimization[25] to generate an ensemble of ten energy minimized models for each parent structure. Inosine and ddI transition state models were placed in the binding site of the minimized backbone ensemble for the ensemble derived from 1pf7 for docking and design.

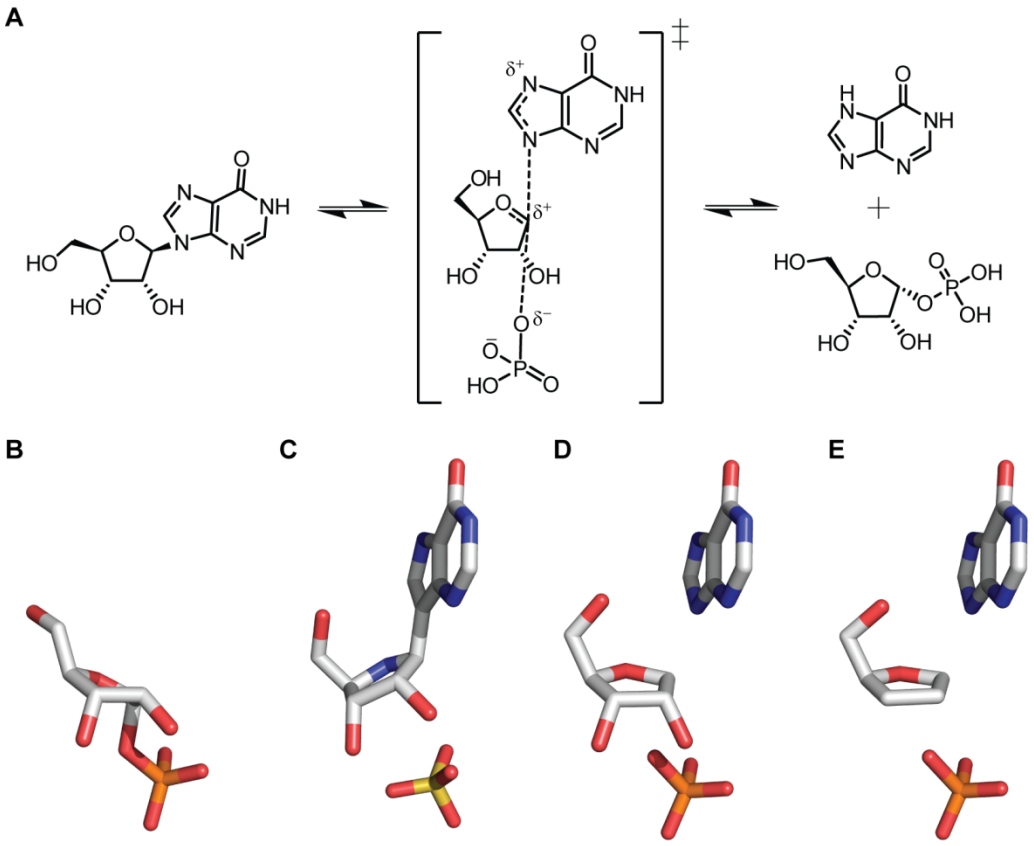

Figure 2-1. Construction of inosine and dideoxyinosine transition state models. A. The human PNP transition state structure consists of a highly dissociative ribooxocarbenium ion transition state. Transition state models for inosine (D) and ddI (E) were generated through combination of the purine coordinates from Immucilin-H (C, PDB ID 1pf7) and the phosphate coordinates of ribose 1-phosphate (B, PDB ID 1a9t).

Residues in the hPNP binding site involved in substrate selectivity of inosine over ddI were identified by redesigning the binding pocket around each transition state model. As shown in Figure 2-2, 23 residues with Cα or Cβ atoms within 8 angstroms of the ligand and Cα/Cβ vectors oriented toward the binding site were allowed to mutate. To facilitate an unbiased design process the amino acid identity for all 23 residues was converted to alanine. The position of inosine and ddI transition state models was optimized during design of these 23 residues along with concurrent repacking of additional side chains in the vicinity of the bindsite. Three iterative rounds of design were

performed to pare down the list of mutable residues. Residues that Rosetta filled with the wild-type amino acid were converted back to the original identity after each round and fixed in this identity for subsequent rounds along with residues that lack interaction with substrate transition state models.

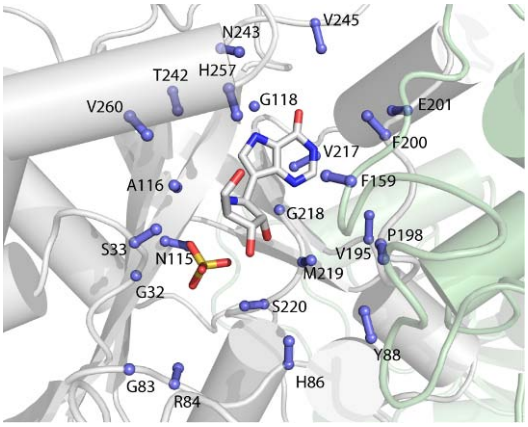

Figure 2-2. Identification of selectivity modulating amino acids in the hPNP binding site by computational design. The Cα and Cβ atoms of residues undergoing design are represented as blue spheres and sticks. The transition state analog Immucilin-H and sulfate are depicted as sticks (PDB ID 1pf7). This image copyright of Oxford University Press. Reproduced with permission. Adapted from Nannemann, *et al*.[1]

To identify sites for which enable differentiation of inosine and ddI the Rosetta-suggested mutations for the remaining six sites were compared for each substrate. These sites were computationally analyzed for predicted selectivity and ddI transition state binding energy in single mutations by docking and by a comparison of the information-weighted Euclidian distance for the design results of the final round. For each position the Euclidean distance d_i between substrate A and substrate B (in this case inosine and ddI, respectively) at design position i is calculated as $d_i = \sqrt{\sum_{k=1}^{n}\left(b_i^A v_{i_k}^A - b_i^B v_{i_k}^B\right)^2}$. The vector of amino acid frequencies at position i is represented by v_i and b_i is the total number of bits of information at that position,

calculated as $b_i = \log_2 n + \sum_{k=1}^{n} v_{i_k} \log_2 v_{i_k}$. The position with the greatest information-weighted Euclidian distance is implicated as a primary selectivity controlling position.

Based on this computational analysis and biochemical data Y88 was identified as a likely modulator of substrate selectivity; therefore, the identity of Y88 was altered to each possible amino acid (excluding Pro and Gly). All side-chains in the active site were optimized using a backbone-dependent rotamer library[26, 27] while all other amino acids were held in their minimized conformation. High-resolution docking of transition state models and optimization of side-chain interactions was achieved through Monte Carlo minimization of side-chain rotamers and intense sampling of the ligand orientation.

Energies for the unbound and bound forms of the enzyme were calculated using the Rosetta energy function. The function is a linear combination of weighted scores including a Lennard–Jones attractive and repulsive potential, an orientation dependent hydrogen-bonding potential, Coulomb electrostatics and an implicit solvation model. Transition state binding energy is calculated as $\Delta\Delta G = \Delta \mathrm{G}_{\mathrm{bound}} - \Delta \mathrm{G}_{\mathrm{unbound}}$[28-30]. Weights for individual parameters of the energy function were established using multiple linear regression. The correlation to the experimental activation energy ($\Delta \mathrm{G}_{TS}^{\ddagger} = -RT \ln(k_{\mathrm{cat}}/K_{\mathrm{m}})$)[31] of each mutant with each substrate was optimized in a leave-one-out (LOO) cross-validation scheme. For each substrate/mutant/backbone combination the scores of the top ten models were averaged to minimize noise in the predicted binding free energy[32].

Cloning, production and purification of hPNP and site directed mutants

Expression vector pCRT7/NT-TOPO-PNP, containing wild-type hPNP, was generously provided by Prof. Vern Schramm[22] and NdeI/HindIII restriction sites were added to facilitate cloning into pET28a (forward primer: 5´-CATATGGAGAACGG-

ATACACCTATGAAGATTATAAG-3′, reverse primer: 5′-AAGCTTCAAGTGGCTTTGTCAG-GGAGTG-3′). Site directed mutations at position 88 were generated using the Quikchange-II system (Stratagene, La Jolla, CA) (forward primer: 5′-GCAGGTTCCACATGXXXGAAGGGTACCCACTCTGG-3′, reverse primer: 5′-CCAGAGTGGGTACCCTTCXXXCATGTGGAACCTGC-3′, where XXX was the variable codon). Mutations were verified via DNA sequencing.

Individual PNP mutants were over-expressed as N-terminal hexahistidine tagged constructs in *E. coli* BL21(DE3). Strains harboring mutant PNP constructs were grown with shaking in 500 mL of LB broth with 50 μg/mL kanamycin at 37 °C. At $OD_{600} \approx 0.6$ cultures were induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside and allowed to incubate for an additional 6–10 h. Cells were harvested by centrifugation and cell pellets were frozen at -80 °C until immediately before purification. Cells were resuspended in Binding Buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8), disrupted by passage through a French pressure cell and centrifuged to remove cellular debris. The soluble proteins were purified in a single step via Ni-affinity chromatography using a HisTrap FF column on an AKTA FPLC (GE Healthcare Life Sciences). Proteins were eluted using a linear gradient from 100% Binding Buffer to 100% Elution Buffer (50 mM Na_2HPO_4, 300 mM NaCl, 500 mM Imidazole, pH 8). The sample was desalted and stored in Exchange Buffer (100 mM Tris HCl, 0.1 mM EDTA, 0.1 mM DTT, pH 7.5) at -80 °C[23]. All enzyme concentrations were determined via λ_{280} measurements and extinction coefficients were estimated using Accelrys DSGene 1.5.

Biochemical Assays of PNP

PNP assays were performed in the phosphorolysis direction by continuously monitoring the formation of hypoxanthine[33]. A catalytic excess of xanthine oxidase was used in a tandem assay converting hypoxanthine to uric acid with concomitant reduction

of iodonitrotetrazolium chloride (INT) to form a formazan chromophore (λ_{max}: 546 nm). Assay Mix buffer contained 50 mM potassium phosphate saturated with O_2, 50 mM HEPES, 0.075% Triton X-100, 1 mM INT, and xanthine oxidase from buttermilk (Sigma). Substrates were dissolved in Assay Mix at concentration ranges of 20–200 μM or 100–1000 μM inosine and 250–2500 μM or 750–7500 μM ddI (Fluka and 3B Medical Systems, Inc.) depending on preliminary substrate concentration response curves. For each mutant, the concentration of enzymes diluted in Assay Mix were adjusted to produce rates within the dynamic range of the assay and < 15% consumption of substrate was observed over time course measurements. Correspondingly, PNP concentration ranges of 0.006–2 μM for measuring inosine rates and 0.02–10 μM for measuring ddI rates were used with substrate ranges spanning the K_m region, wherever possible. The assays were performed as follows: 100 μL of substrate dissolved in Assay Mix was transferred into wells of a flat bottomed 96-well plate, equilibrated at 25 °C, followed by addition of 100 μL of assay mix containing PNP. The rate of hypoxanthine formation was observed by monitoring tandem formazan formation by its unique absorbance at 546 nm over 5 minutes. A hypoxanthine standard curve was performed in parallel with each concentration series to convert absorbance numbers into molar turnover values.

Results

Progenitor Identification and Design Rationale

An understanding of the effects of enzyme structure on activity and substrate preference is critical to enzyme selection for biocatalytic application and subsequent rational redesign. Nucleoside phosphorylases have been studied extensively because of their role in human disease and potential for drug development. Deficiencies in PNP result in a marked effect on T cell and cellular immune function, while B cell and humoral

immunity remain unaffected[21]. Thus, inhibition of the enzyme shows potential in treatment of T-cell and auto immune diseases[34]. Since nucleoside phosphorylases play a critical role in metabolism of nucleoside and nucleotide analog drugs the enzyme is a target for treatment of pathogens, such as *Plasmodium falciparum*[22] and *Mycobacterium tuberculosis*[35, 36]. For these reasons, the mechanism has been studied in detail.

Nucleoside phosphorylases have one of two folds which are broken down into the NP-I and NP-II families. NP-I family members possess an α/β-subunit fold while the NP-II family forms a dimeric quaternary structure consisting of an α-domain and a mixed α/β domain. In this work, the purine nucleoside phosphorylases (PNPs), which are NP-I family members, are of primary interest. PNPs of the NP-I family exist in either trimeric or hexameric quaternary structures but maintain a strikingly similar tertiary fold throughout the family (Figure 2-3A–D). PNPs from higher organisms and prokaryotic PNPs have the trimeric quaternary fold while lower organisms primarily possess the hexameric form. As such, trimeric and hexameric PNPs are typically represented by human (hPNP) and *E. coli* (ecPNP) variants.

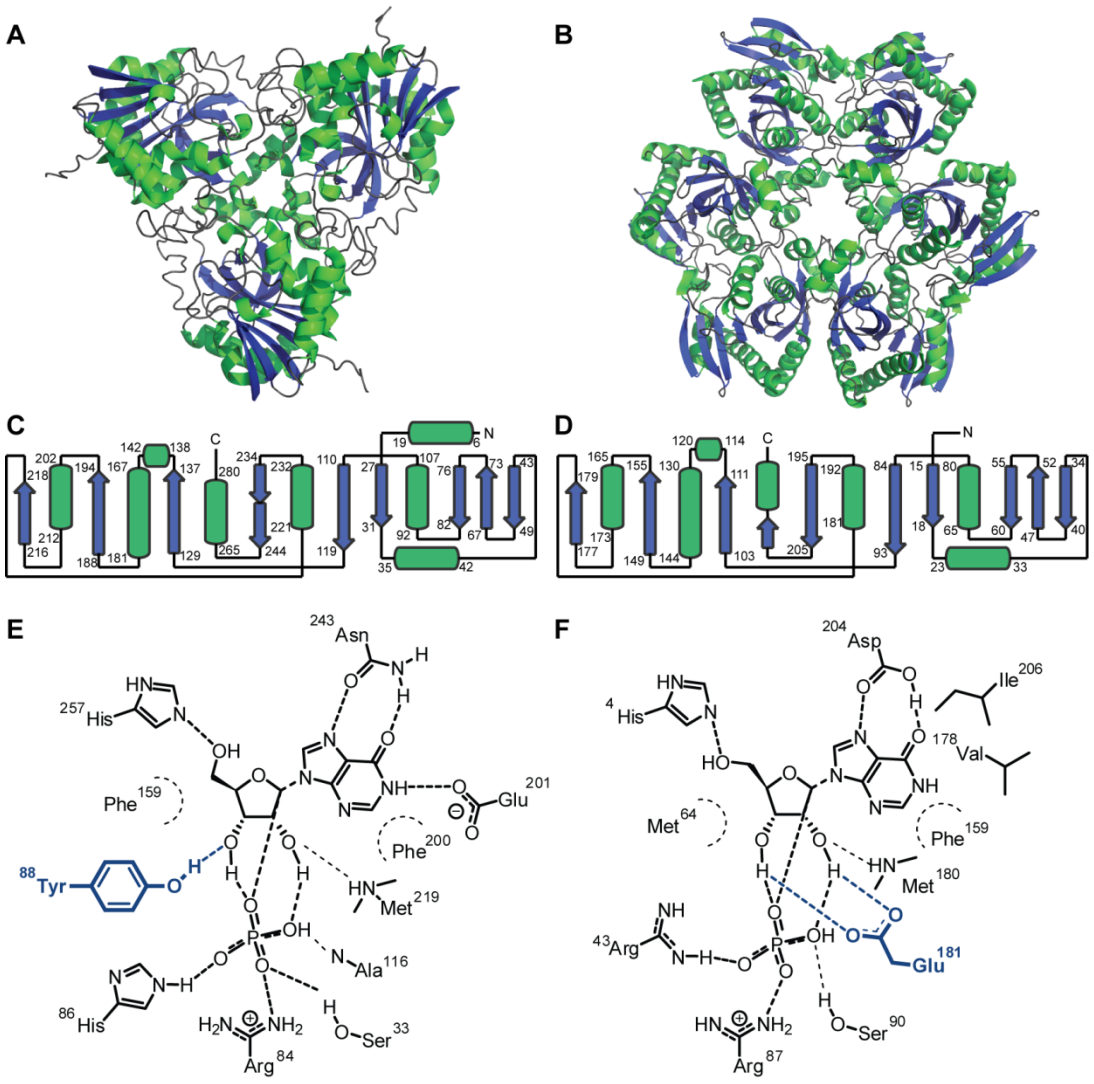

Figure 2-3. Structural overview of purine nucleoside phosphorylases in the NP-1 fold family. Human and *E. coli* PNP represent trimeric (A) and hexameric (B) quaternary structures of the NP-1 family (PDB ID 1rct and 1pr0, respectively). Fold diagrams for trimeric (C) and hexameric (D) PNP demonstrate a similar tertiary fold within each monomer. Quaternary structure differences are reflected in the binding site of trimeric (E) and hexameric (F) PNPs. Critical interactions between protein and sugar are shown in blue.

Differences in the quaternary fold are reflected in the binding site and significantly affect the substrate selectivity (Figure 2-3E,F). Trimeric PNPs are specific to the 6-oxo purines hypoxanthine and guanine whereas hexameric PNPs are indiscriminate at the nucleobase. Each fold orients phosphate in a similar manner with

the trimeric PNPs having an additional Glu-His diad which serves to activate phosphate. The target compound, ddI, and the natural substrate, inosine, differ only in the atomic make-up of the sugar, so this region is of particular importance to a rational redesign effort. Both trimeric and hexameric PNPs utilize backbone and side chain contacts to orient substrate. Each PNP forms a hydrogen bond between the 2´-hydroxyl and the backbone nitrogen of a methionine. The fact that both quaternary forms utilize 2´-deoxynucleosides as substrate is an indication that this backbone – side chain interaction is not critical for function. Because of this, and that it is currently difficult to engineer changes in backbone–substrate contacts, interactions between protein side chains and the substrate have been targeted. In hPNP, a hydrogen bond is present between Y88 and the 3´-hydroxyl of inosine while ecPNP forms a bidentate interaction between Glu-181 and ribosyl nucleosides (Figure 2-3E,F). Human PNP has been chosen as a starting point for development of dideoxynucleoside phosphorylase catalytic activity for two reasons: first, side chain interactions are limited to one hydroxyl on the substrate; and second, mutations at Y88 of hPNP are tolerated to a greater degree than mutations at Glu-181 of ecPNP (Erion *et al*[20] and unpublished results).

Upon identification of hPNP as an ideal starting point for engineering, a full understanding of the hPNP mechanism and knowledge of the contributions of each residue in the binding pocket is necessary for application of computational bindsite optimization. Overall, the transition state for nucleoside phosphorolysis is highly dissociative[37] and there is strong ribooxacarbenium ion character[24]. Throughout the reaction coordinate of hPNP the purine ring is positioned via hydrogen bonding interactions with Asn-243 and Glu-201. The nucleophilic phosphate is positioned by a complex network of hydrogen bonding interactions (Figure 2-3E) and activated for addition to C1´ by a His-86/Glu-89 diad. As noted earlier, there appear to be relatively few significant ribose binding interactions, namely His-257 contacting O5´ and Y88,

which forms a hydrogen bond with O3´. The O3´ hydroxyl is also implicated in coordinating to the phosphate ligand and has been suggested to play a role in optimizing the geometry of the transition state.[20] As side chain contacts between the ligand and substituted sugar atoms are limited to Y88, this residue is a primary contributor to substrate selectivity in the ribose binding region and, thus, a main target for rational redesign.

Binding Energy Calculations

An initial gradient-based energy minimization[25] of the experimental protein structure was carried out in the absence of substrate to obtain an unbiased starting conformation of the protein that occupies a minimum in the energy landscape. The lowest energy structure had a Cα-rmsd of 0.64 Å from the experimental coordinates (PDB ID 1pf7). The average Cα-rmsd of the ten lowest energy minimized structures was ~0.86 Å.

Transition state models for inosine and ddI were constructed based on experimental and crystallographic data (see methods section and Figure 2-1). Docking of the transition state models allows for calculation of transition state binding energies and prediction of enzymatic activity. For this purpose the transition state models were placed into the minimized protein structures and the position and orientation of the transition state model was optimized together with the conformations of residues near the binding site using the RosettaLigand energy function[13, 14].

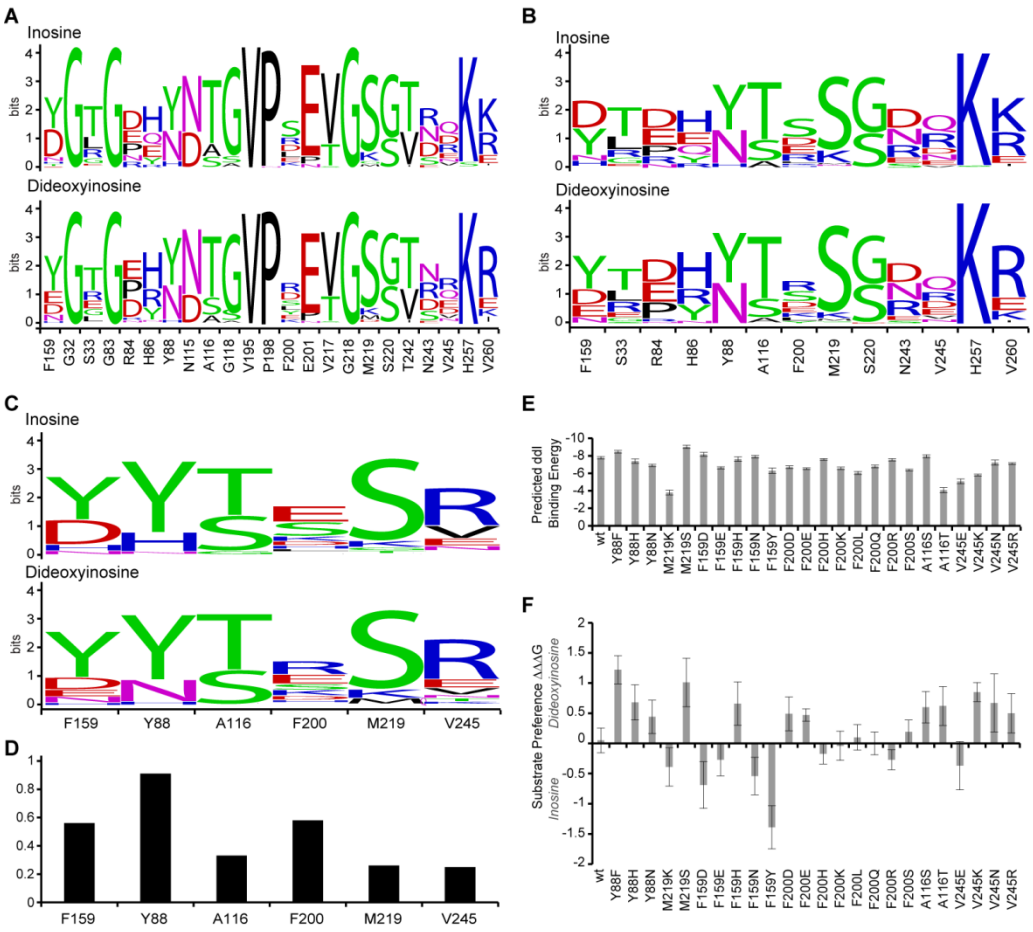

Figure 2-4. Identification of hPNP substrate selectivity residues through differential analysis of designed residues and binding energies. Results of active-site redesign of residues in Figure 2-2 for inosine and ddI are shown as the frequency of a given mutation at each position by plotting information content using the WebLogo server[38]. 23 residues were allowed to design during the first round (a), 13 during the second round (b) and 6 during the third round (c). Mutations altering substrate selectivity were identified through calculation of the information-normalized Euclidean distance (d), the predicted binding energy ($\Delta\Delta G$) for ddI (e) and substrate binding preferences ($\Delta\Delta\Delta G$) (f). This image copyright of Oxford University Press. Reproduced with permission. Adapted from Nannemann, *et al*.[1]

To demonstrate the applicability of the docking protocol to the identification of specificity encoding sites the RosettaDesign algorithm was used to identify amino acids in the binding site that maintain tight binding upon mutation but display a differential mutation profile in the presence of inosine and ddI. There are 23 amino acids that have Cα or Cβ atoms within 8 Å of the transition state analog ImmucilinH (PDB ID 1pf7) and

point toward the binding site (Figure 2-2). Each of these amino acids was allowed to mutate to any other amino acid in the presence of ddI or inosine transition state analogs. The frequency of each amino acid in each position was evaluated using the publicly available WebLogo server[38]. Residues populated by Rosetta with the wild-type identity are not expected to affect a change in substrate binding characteristics. Similarly, residues designed by Rosetta to interact with neighboring amino acids or turned toward solvent would not be expected to alter the substrate activity profile. With such an analysis 17 residues had been reverted to wild-type after two rounds of iterative refinement (Figure 2-4A-B). Five of the six remaining residues (F159, Y88, A116, F200, M219 and V245) interact with the sugar moiety of the transition state models. The obtained mutation profiles were rather similar suggesting that minimal differentiation between the two ligands will be achieved (Figure 2-4C). Suggested mutations for Y88 indicate that tyrosine is ideal for this location; secondary predictions, however, differ significantly for inosine and ddI (His and Asn, respectively) indicating the potential for modulating the substrate selectivity and specificity by mutating this position. The information-normalized Euclidean distance for each design position was calculated to quantify the impact of each residue on selectivity. This analysis further points to Y88 as a selectivity-modulating residue.

Next, step-wise analysis of all proposed single mutants was performed. All six positions were mutated sequentially to all amino acids proposed by Rosetta and affinity for the ddI transition state model as well as preference for the ddI transition state model above the inosine transition state model was computed (Figure 2-4D-E). Y88 remains as the only site that fulfills criteria for both tight binding to transition state models and displays a differential mutation profile. Consequentially, Y88 was mutated *in silico* through replacement of the wild-type side chain to all possible amino acids (except Pro and Gly) and the transition state models were re-docked.

Using this protocol a first round of seven single mutants was initially characterized: Y88F/A/C/I/L/W/K. These mutants were selected for predicted high ddI activity, a range in predicted activity and variety in the side-chain character. In a second round of experiments all Y88X mutants were generated, purified, and characterized to comprehensively analyze the ability of RosettaLigand to rank individual mutations (Table 2-1). Mutants predicted to form tight complexes with the ddI transition state model include Y88F, Y88L, and Y88M as well as the negatively charged amino acid mutations Y88D and Y88E (Figure 2-5).

Table 2-1. Kinetic Characteristics for hPNP-Y88X[a]

Variant	k_{cat} (s^{-1})		K_m (μM)		k_{cat}/K_m (s^{-1} M^{-1})	
	Inosine	ddI	Inosine	ddI	Inosine (x 10^2)	ddI
Wild-type	43.9 ± 0.6	0.9 ± 0.01	48 ± 2	1030 ± 20	9210 ± 140	875 ± 8
Y88F	28.3 ± 0.7	10.5 ± 0.1	73 ± 4	450 ± 10	3890 ± 110	23140 ± 230
Y88H	8.6 ± 0.2	2.5 ± 0.1	74 ± 3	675 ± 20	1160 ± 23	3670 ± 60
Y88W	5.6 ± 0.1	0.09 ± 0.01	500 ± 25	980 ± 40	113 ± 3	91 ± 2
Y88A	17.6 ± 0.4	2.8 ± 0.05	80 ± 3	1670 ± 50	2190 ± 51	1670 ± 22
Y88V	2.4 ± 0.1	0.6 ± 0.01	1200 ± 60	4410 ± 160	19.6 ± 0.7	144 ± 3
Y88L	15.9 ± 0.3	6.1 ± 0.05	94 ± 4	391 ± 8	1690 ± 34	15670 ± 130
Y88I	1.3 ± 0.1	0.5 ± 0.02	1420 ± 55	4670 ± 270	9.4 ± 0.3	115 ± 4
Y88M	12.8 ± 0.1	5.2 ± 0.1	345 ± 7	645 ± 20	371 ± 3	8090 ± 90
Y88C	5.5 ± 0.2	2.5 ± 0.1	170 ± 10	810 ± 30	321 ± 13	3070 ± 60
Y88S	8.8 ± 0.1	2.7 ± 0.1	430 ± 10	4060 ± 110	203 ± 3	652 ± 10
Y88T	2.3 ± 0.1	0.4 ± 0.01	1200 ± 70	8160 ± 280	19 ± 1	44 ± 1
Y88N	9.7 ± 0.2	1.5 ± 0.1	845 ± 25	9530 ± 540	114 ± 2	153 ± 6
Y88Q	1.4 ± 0.1	0.15 ± 0.01	855 ± 50	860 ± 25	17 ± 0.6	177 ± 2
Y88D	0.8 ± 0.1	0.02 ± 0.001	2600 ± 120	12700 ± 500	2.9 ± 0.1	1.8 ± 0.5
Y88E	0.7 ± 0.1	0.01 ± 0.001	4860 ± 410	14900 ± 1200	1.4 ± 0.1	0.9 ± 0.1
Y88K	ND	ND	ND	ND	ND	ND
Y88R	0.01 ± 0.001	ND	370 ± 20	ND	0.0184 ± 0.0005	ND

a. Assay conditions: 50 mM phosphate buffer, pH = 7, 50 mM HEPES, pH 7, 0.075% Triton X-100, 1 mM INT, xanthine oxidase. Inosine: 20–200 or 100–1000 μM. ddI: 250–2500 or 750–7500 μM. Enzymes were assayed at concentrations which ensure steady state conditions.

In wild-type hPNP, the phosphorolysis of inosine is three orders of magnitude more efficient than that of ddI. The values of K_m and k_{cat} reported here are similar to those previously reported for inosine[20, 21]. Each single mutant had decreased k_{cat} and

increased K_m values for inosine. Positively charged Y88K and Y88R mutants were not sufficiently active for kinetic characterization at enzyme concentrations more than five hundred times those used for the wild-type enzyme; accordingly, they were not included in the *in silico* analyses. In line with the *in silico* prediction, the Y88F mutant displayed the second highest catalytic activity.

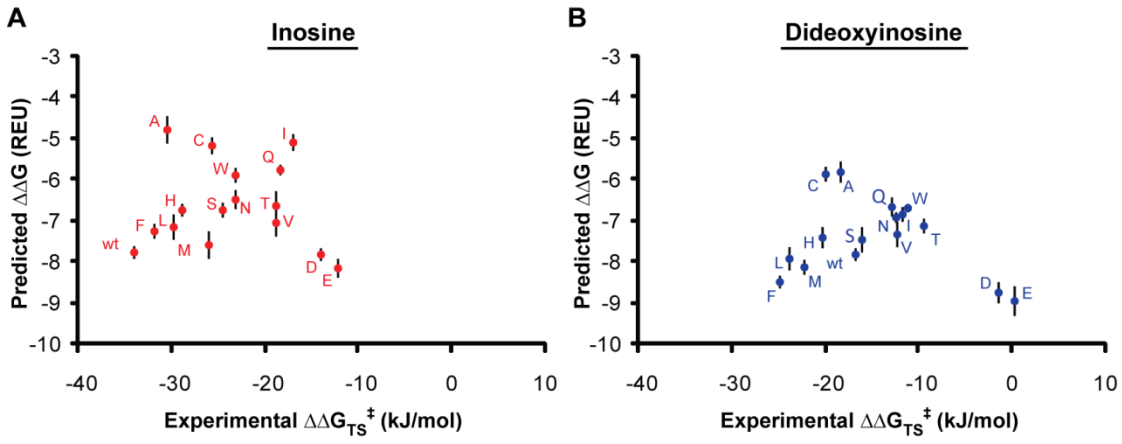

Figure 2-5. Correlation of experimental binding energy to binding energy predicted with the standard RosettaLigand weight-set for (A) inosine and (B) ddI. ●, inosine; ●, ddI. REU are Rosetta Energy Units. Error bars indicate standard deviation of top ten transition state bound models. This image copyright of Oxford University Press. Reproduced with permission. Adapted from Nannemann, *et al*.[1]

Six mutants demonstrated a higher catalytic efficiency for ddI than the wild-type enzyme. A similar profile for toleration of mutations was seen with ddI as with inosine however it appears that almost any mutation allows for some improvement in the catalytic efficiency ratio relative to wild-type hPNP. As predicted *in silico* Y88F, Y88L, and Y88M displayed the highest catalytic activities. Of these, hPNP-Y88F displayed the highest overall catalytic efficiency and the highest turnover rate. Y88D/E mutants did not show the predicted high activity. It is expected that addition of a negative charge to the binding site disrupts the sensitive arrangements of partial charges in the catalytic site, an

effect that is not considered by RosettaLigand as it solely optimizes binding affinity and has no means to directly assess catalytic competence of an active site from a mechanistic perspective.

Table 2-2. Predicted and experimental binding energies for PNP-Y88 mutants.

Variant	Experimental Binding Energy (kJ/mol)[a]		Predicted Binding Energy (REU)[b]		Predicted Binding Energy (kJ/mol)[c]	
	Inosine	ddI	Inosine	ddI	Inosine	ddI
Wild-type	-34.0	-16.8	-7.73	-7.78	-19.82	-23.39
Y88F	-31.9	-24.9	-7.24	-8.46	-29.70	-21.75
Y88H	-28.9	-20.3	-6.71	-7.39	-25.67	-14.08
Y88W	-23.1	-11.2	-5.86	-6.66	-20.40	-18.48
Y88A	-30.5	-18.4	-4.76	-5.78	-22.44	-13.28
Y88V	-18.8	-12.3	-7.01	-7.31	-25.84	-16.30
Y88L	-29.8	-23.9	-7.12	-7.91	-27.03	-15.61
Y88I	-17.0	-11.8	-5.06	-6.83	-21.27	-13.03
Y88M	-26.1	-22.3	-7.57	-8.09	-30.77	-22.41
Y88C	-25.7	-19.9	-5.15	-5.85	-21.81	-9.66
Y88S	-24.6	-16.1	-6.72	-7.44	-27.00	-12.24
Y88T	-18.8	-9.4	-6.62	-7.1	-16.60	-4.15
Y88N	-23.2	-12.5	-6.46	-6.9	-19.99	-13.52
Y88Q	-18.4	-12.8	-5.75	-6.64	-25.41	-16.50
Y88D	-14.1	-1.5	-7.78	-8.73	-20.92	-11.37
Y88E	-12.3	0.3	-8.13	-8.93	-19.12	-12.13
Y88K	n/d[d]	n/d[d]	-9.01	-9.14	n/d[d]	n/d[d]
Y88R	-7.21	n/d[d]	-7.66	-7.26	n/d[d]	n/d[d]

a. Experimental binding energy, $\Delta G_{TS}^{\ddagger} = -RT\ \ln(k_{cat}/K_m)$
b. Binding energies established using the Protein:Ligand weightset.
c. Binding energies established using the PNP:Nucleoside customized weightset.
d. Values for lysine and arginine mutants were not determined with the reweighted scoring function as kinetic data was incomplete for these mutants.

Training of Rosetta Scoring Function for PNP:Nucleoside Interactions

With a comprehensive dataset in hand, an attempt was made to test which fraction of the differences in experimental activation energy and predicted transition state binding free energies can be attributed to inaccurate weighting of the energy terms in RosettaLigand and which part must be attributed to inaccuracies in the RosettaLigand energy function and structural models used. Weights were developed for individual components of the Rosetta energy function using multi-linear regression to optimize the

correlation of predicted binding energies to experimental activation energies. Activation energies were calculated from experimental data as $\Delta G^{\ddagger}_{TS} = -RT \ln(k_{cat}/K_m)$ while predicted binding energies were calculated as $\Delta\Delta G = \sum w_i \cdot (s_{i_{bound}} - s_{i_{unbound}})$, where w_i is the weight applied to a particular score s_i[28-31]. Using a leave-one-out cross validation analysis weights are generated without the use of one data point and then binding energy predictions made for the data point left out (Table 2-2).

Table 2-3. Comparison of established weights for Rosetta applications.

Score	PNP : Nucleoside	Protein : Ligand[14]	Protein : Protein[29]	Standard[39]
Attractive	1 ± 0.05	1	1	1
Repulsive	0.33 ± 0.02	0.75	0.16	0.91
Solvation	0.56 ± 0.04	0.63	0.73	0.65
GBSol	0.37 ± 0.04	-	-	-
Hydrogen Bonding	-	1.5	1.11	1.74
Pair Elec.	-	0.63	-	0.34
Rotamer Probability	-	0.4	0.64	0.4
Phi Psi Probability	-	0.4	-	0.51

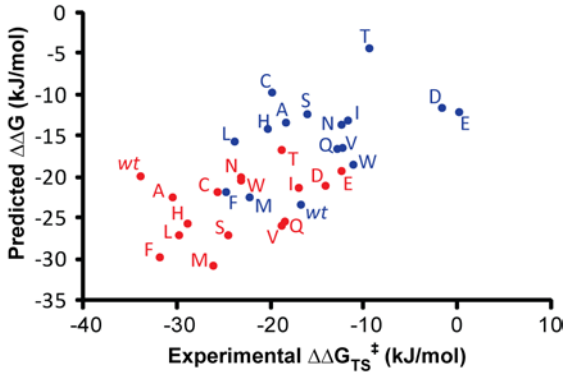

Figure 2-6. Correlation of experimental activation energy to activation energy calculated with a customized PNP : Nucleoside weight-set. ●, inosine; ●, ddI. REU are Rosetta Energy Units. Error bars indicate standard deviation of top ten transition state bound models. This image copyright of Oxford University Press. Reproduced with permission. Originally published by Nannemann, *et al*.[1]

Discussion

We have described the identification of human purine nucleoside phosphorylase for biocatalytic synthesis of dideoxyinosine and rational design of the binding site based on computational prediction of transition state binding energies. Based on analysis of previously reported structures and activity profiles, in addition to computational design studies, a single active site residue, Y88, was selected as a likely hotspot for improvements in sugar analog binding and created a comprehensive library of substitutions at this position. RosettaLigand was used to estimate the binding energies of inosine and ddI transition state models to *in silico* models of these mutations.

The present experiment evaluates Rosetta in a scenario likely to occur in enzyme design – assessing the change of enzyme specificity for a different but related ligand. Simultaneously this study evaluates a scenario in which options for computational design were limited to a single active site residue, Y88. The accuracy of the Rosetta algorithm needs to be sufficiently high to discriminate improved substrate/transition state binding properties for a single active site ensemble (hPNP). By comparing Rosetta calculated binding energies to energies derived from experimentally measured kinetic parameters for every possible mutation the ability of Rosetta to rank substitutions at Y88 could be assessed. RosettaLigand ranked the Y88F variant in the top three for transition state stabilization. However, RosettaLigand generally underestimated energies of variants with improved catalytic function and overestimated energies of mutants with decreased function; this trend was particularly evident in the case of wild-type hPNP which was predicted to process ddI with greater efficiency than inosine.

Phenomena outside the scope of the method may contribute to errors in the prediction of protein variant binding energies. For example, wild-type PNP crystallized

with ddI has an ordered water molecule in the binding pocket which forms a hydrogen bond with Y88 and is situated proximal to the position of the 3´-hydroxyl in inosine-bound structures. The presence of this water molecule was not evaluated in the Rosetta calculations and might be one source of error. It is possible that removal of the hydrogen bond acceptor by mutation to phenylalanine may be a significant source of improvement. Additionally, mutation to β-branched amino acids Y88V/I/T reduced the function of the enzyme relative to structurally similar counterparts, such as Y88L/S. Rosetta models predicted these mutants to have a destabilizing effect in the enzyme, evident by an increase in the total energy score caused by clashing between the Cβ-methyl substituent and Pro-198, which resides on an adjacent loop involved in purine binding. This destabilization is not revealed by binding energy calculations as the residue-residue clash is present in both bound and unbound structures. This has been previously noted in the design of protein:protein interfaces[40] and is confirmed experimentally in this work.

The empirical and computational datasets resulting from this experiment permitted the optimization of Rosetta parameters resulting in appreciable correlation with R = 0.65 (Figure 2-6). Reweighting of the scoring function corrects the gross misidentification of charged amino acids as catalytically advantageous and aligns inosine and ddI with respect to one another. This reweighting, while advantageous in terms of correlating global substrate preference, decreases the quality of ranking mutants with a particular substrate. Conversely, the unchanged RosettaLigand score function properly identifies the wild type and Y88F variants as most catalytically active for inosine and ddI (Figure 2-5A,B).

In this study, predicted transition state binding energies were correlated to experimental data by calculation of the activation energy as $\Delta G_{TS}^{\ddagger} = -RT\ \ln(k_{\mathrm{cat}}/K_{\mathrm{m}})$[31, 41]. This term can be rationalized through analysis of a simple, one substrate reaction

following Michaelis-Menten kinetics (Figure 2-7A). The free energies involved in the unimolecular reaction can be found through the application of transition state theory and equilibrium thermodynamics[31, 42, 43]. Suppose that the difference in energy between the ground state, S, and the transition state of the reaction, $\mathrm{S}^{\ddagger}$, is $\Delta\mathrm{G}^{\ddagger}$. From equilibrium thermodynamics the following relationship can be established: $[\mathrm{S}^{\ddagger}] = [S]e^{\frac{-\Delta\mathrm{G}^{\ddagger}}{RT}}$. The transition state $\mathrm{X}^{\ddagger}$ decomposes to X at the same rate as the vibration frequency v of the bond that is breaking. Therefore, the rate of decomposition of S is given by $\frac{-d[S]}{dt} = v[\mathrm{S}^{\ddagger}] = [S]\left(\frac{kT}{h}\right)e^{\frac{-\Delta\mathrm{G}^{\ddagger}}{RT}}$, where $v = \frac{kT}{h}$. For the Michaelis-Menten reaction, the overall rate equation is given by $k_{cat} = \frac{kT}{h}e^{\frac{-\Delta\mathrm{G}^{\ddagger}}{RT}}$ and the activation energy of the chemical steps of bond breaking and bond making is given by $\Delta\mathrm{G}^{\ddagger} = -RT\ \ln(k_{\mathrm{cat}})$. Using equilibrium thermodynamics, where $K_{\mathrm{m}} = K_s$ and K_s is the association of free enzyme and free substrate the substrate binding energy is given by $\Delta G_s = -RT\ \ln(K_{\mathrm{m}})$.

Following transition state theory the Michaelis-Menten reaction in Figure 2-7A can be expressed as the equilibrium between free enzyme and free substrate and the transition state complex (Figure 2-7B). The free energies for this process are diagrammed in Figure 2-7C and the substrate binding energy and the activation energy of the chemical steps are related to the total activation energy $\Delta\mathrm{G}^{\ddagger}_{TS}$, by $\Delta\mathrm{G}^{\ddagger}_{TS} = \Delta\mathrm{G}^{\ddagger} + \Delta\mathrm{G}_{\mathrm{s}}$, where $\Delta\mathrm{G}^{\ddagger}_{TS}$ and $\Delta\mathrm{G}^{\ddagger}$ are algebraically positive and $\Delta\mathrm{G}_{\mathrm{s}}$ is negative[42]. Substituting the above relationship into the overall rate equation gives $RT\ln(k_{\mathrm{cat}}/K_{\mathrm{m}}) = RT\ln\frac{kT}{h} - \Delta\mathrm{G}^{\ddagger} - \Delta\mathrm{G}_{\mathrm{s}}$ and relates $k_{\mathrm{cat}}/K_{\mathrm{m}}$ to the activation energy, $\Delta\mathrm{G}^{\ddagger}_{TS}$, in terms of transition state theory.

Enzymes function by decreasing the activation energy of the chemical steps through tighter binding of the transition state than the substrate, effectively lowering the

overall activation energy of the transformation[31, 42, 44]. The current method separates the quantum mechanical portion of the free energy of the enzyme-transition state complex from the geometrical component. In other words, the high energy of partially formed bonds is ignored and becomes implicit in the structure of the transition state. Because the enzyme has tighter binding to the transition state the free energy of the enzyme-transition state complex is effectively lower than the enzyme-substrate complex (Figure 2-7D).

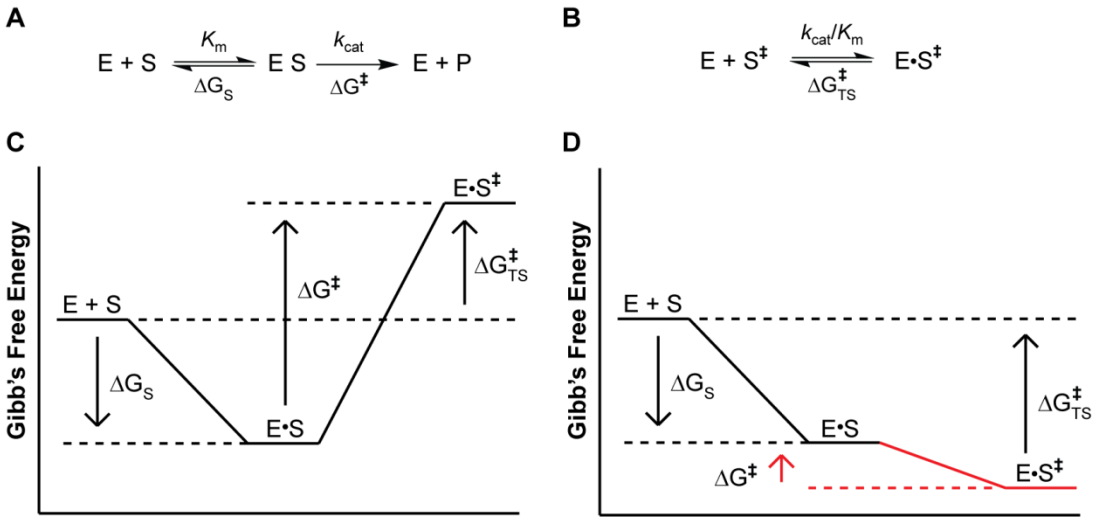

Figure 2-7. Gibbs free energy diagrams of enzyme catalysis in experimental and computational studies. A. Interaction of free enzyme and free substrate to form the Michaelis complex and then product. B. Equilibrium between free enzyme and an unbound transition state and the enzyme-transition state complex. C. The change in energy is shown for the binding of substrate from free enzyme and free substrate and the formation of the enzyme-transition state complex from the enzyme-substrate complex. The enzyme–transition state complex is high in free energy due to partially formed bonds. D. Internal transition state energies are not accounted for in the RosettaLigand docking method. Enzymes decrease the overall activation energy ($\Delta G^{\ddagger}_{TS}$) through tighter binding of the transition state over the substrate. Ignoring the high energy of partially formed bonds causes a decrease in free energy upon formation of the enzyme-transition state complex.

An ideal design method would be dependent only on the change in activation energy ($\Delta G^{\ddagger}$) realized upon a change in the binding site. However, by separating the

components of the chemical activation energy into quantum mechanical and structural portions the value of $\Delta G^{\ddagger}$ is effectively reduced to the difference in transition state binding energy and substrate binding energy (Figure 2-7D). This small value of $\Delta G^{\ddagger}$ is difficult to differentiate from the noise inherent in the framework of the current method. By correlating experimental data to the transition state binding energy, $\Delta G_{TS}^{\ddagger}$, instead of the activation energy, $\Delta G^{\ddagger}$, the substrate binding energy was incorporated into the score serving to increase the signal above the noise in the docking calculations.

By performing the design and docking studies with the transition state models mutants with greater predicted binding energy should possess a higher turnover rate. Because we have separated the quantum mechanical and structural components of the transition state and because RosettaLigand does not directly evaluate the mechanistic aspects of turnover one might argue that predicted transition state binding energies correlate more accurately with K_m, where $\Delta G_{TS}^{\ddagger} = -RT \ln(K_m)$. Nonetheless, at least in the present studies, this analysis yields somewhat reduced correlation coefficients for inosine and ddI, where R = 0.236 and 0.284, respectively, using binding energies established with the RosettaLigand weightset and excluding charged residues. These values are 0.362 and 0.438 for inosine and ddI, respectively, when correlating predicted binding energies to k_{cat}/K_m.

All terms in the RosettaLigand energy function were applied in the generation of enzyme:substrate ensembles but the attractive and repulsive, solvation and generalized Borne solvation terms were most critical for the final evaluation of binding energy. There are several notable differences in the weights developed here with those reported for use in protein:protein binding, the RosettaLigand weights used during docking, and the standard weight set. Relative to the Lennard–Jones attractive weight the weight for the Lennard–Jones repulsive score is twice as large as that used in protein docking but

much lower than the original ligand docking and standard weights. The weight for solvation is also slightly decreased compared to typical Rosetta weights. Of note is the absence of a hydrogen bonding term from the evaluation; while this term was used during the docking portion of the experiment its inclusion in the final weight set does not improve the correlation to experiment. It is not surprising that the weights of electrostatic terms were increased considering the number and role of charged and polar residues in binding and catalysis.

Conclusions

The process outlined herein was applied to improve the catalytic properties of hPNP for an alternate substrate, dideoxyinosine. The Rosetta method proved capable of identifying mutations which are likely to improve catalysis in a test case made particularly difficult by constraint to a single amino acid substitution and high substrate similarity. Furthermore, the methods potential for application to systems which, unlike hPNP, are not thoroughly characterized has been demonstrated.

In this case study, charged residues had a major impact on computational and experimental results. Addition of an electrostatic term allowed for moderate improvement in the results. The makeup of the scoring function highlights the characteristics of this particular test and points to the broad scope of applications available for the Rosetta algorithm. Indeed, many other nucleoside analogs contain modifications on the sugar moiety providing a handle for successful design.

Acknowledgements

Portions of this research were originally published in Protein Engineering, Design and Selection. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein*

Engineering, Design and Selection. 2010; 23:607-616. DPN designed and conducted all experiments and wrote the manuscript. KWK contributed to the calculation of transition state binding energies and analysis of *in silico* data. JM advised in the design of *in silico* experiments for calculation of transition state binding energies and in analysis of *in silico* data. BOB advised in the design of experiments for assay of hPNP point mutants and directed evolution mutagenesis and screening methods and in analysis of the directed evolution study.

References

1. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010,** 23, (8), 607-616.

2. Clerq, E. D., Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents* **2009,** 33, (4), 307-320.

3. Soriano, V.; Peters, M. G.; Zeuzem, S., New therapies for hepatitis C virus infection. *Clin. Infect. Dis.* **2009,** 48, (3), 313-320.

4. Li, F. J.; Maag, H.; Alfredson, T., Prodrugs of nucleoside analogues for improved oral absorption and tissue targeting. *J. Pharm. Sci.* **2008,** 97, (3), 1109-1134.

5. Pinheiro, E.; Vasan, A.; Kim, J. Y.; Lee, E.; Guimier, J. M.; Perriens, J., Examining the production costs of antiretroviral drugs. *Aids* **2006,** 20, (13), 1745-1752.

6. Pugmire, M. J.; Ealick, S. E., Structural analyses reveal two distinct families of nucleoside phosphorylases. *Biochem. J.* **2002,** 361, 1-25.

7. Schramm, V. L., Enzymatic transition states: thermodynamics, dynamics and analogue design. *Arch. Biochem. Biophys.* **2005,** 433, (1), 13-26.

8. Jackel, C.; Kast, P.; Hilvert, D., Protein design by directed evolution. *Annu. Rev. Biophys.* **2008,** 37, 153-173.

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

10. Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z. G., Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, (7), 3778-83.

11. Bernhardt, P.; McCoy, E.; O'Connor, S. E., Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **2007,** 14, (8), 888-897.

12. Damborsky, J.; Brezovsky, J., Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.* **2009,** 13, (1), 26-34.

13. Davis, I. W.; Baker, D., RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009,** 385, (2), 381-392.

14. Meiler, J.; Baker, D., ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **2006,** 65, (3), 538-48.

15. Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D., Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** 302, (5649), 1364-8.

16. Murphy, P. M.; Bolduc, J. M.; Gallaher, J. L.; Stoddard, B. L.; Baker, D., Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (23), 9215-20.

17. 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.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., *De novo* computational design of retro-aldol enzymes. *Science* **2008,** 319, (5868), 1387-1391.

18. 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.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** 453, (7192), 190-U4.

19. Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Rothlisberger, D.; Baker, D., New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **2006,** 15, (12), 2785-2794.

20. Erion, M. D.; Takabayashi, K.; Smith, H. B.; Kessi, J.; Wagner, S.; Honger, S.; Shames, S. L.; Ealick, S. E., Purine nucleoside phosphorylase. 1. Structure-function studies. *Biochemistry* **1997,** 36, (39), 11725-34.

21. Stoeckler, J. D.; Cambor, C.; Parks, R. E., Jr., Human erythrocytic purine nucleoside phosphorylase: reaction with sugar-modified nucleoside substrates. *Biochemistry* **1980,** 19, (1), 102-107.

22. Lewandowicz, A.; Schramm, V. L., Transition state analysis for human and *Plasmodium falciparum* purine nucleoside phosphorylases. *Biochemistry* **2004,** 43, (6), 1458-68.

23. Deng, H.; Lewandowicz, A.; Schramm, V. L.; Callender, R., Activating the phosphate nucleophile at the catalytic site of purine nucleoside phosphorylase: a vibrational spectroscopic study. *J. Am. Chem. Soc.* **2004,** 126, (31), 9516-7.

24. Ringia, E. A. T.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Murkin, A. S.; Schramm, V. L., Transition state analogue discrimination by related purine nucleoside phosphorylases. *J. Am. Chem. Soc.* **2006,** 128, (22), 7126-7127.

25. Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D., High-resolution structure prediction and the crystallographic phase problem. *Nature* **2007,** 450, (7167), 259-264.

26. Bower, M. J.; Cohen, F. E.; Dunbrack, R. L., Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* **1997,** 267, (5), 1268-1282.

27. Dunbrack, R. L.; Karplus, M., Backbone-dependent rotamer library for proteins - application to side-chain prediction. *J. Mol. Biol.* **1993,** 230, (2), 543-574.

28. Kaufmann, K. W.; Dawson, E. S.; Henry, L. K.; Field, J. R.; Blakely, R. D.; Meiler, J., Structural determinants of species-selective substrate recognition in human and *Drosophila* serotonin transporters revealed through computational docking studies. *Proteins* **2009,** 74, (3), 630-42.

29. Kortemme, T.; Baker, D., A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** 99, (22), 14116-21.

30. Morozov, A. V.; Havranek, J. J.; Baker, D.; Siggia, E. D., Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res.* **2005,** 33, (18), 5781-98.

31. Fersht, A. R., Catalysis, binding and enzyme-substrate complementarity. *Proc. R. Soc. London, Ser. B* **1974,** 187, (1089), 397-407.

32. Popov, V. M.; Yee, W. A.; Anderson, A. C., Towards *in silico* lead optimization: scores from ensembles of protein/ligand conformations reliably correlate with biological activity. *Proteins: Struct., Funct., Bioinf.* **2007,** 66, (2), 375-387.

33. Degroot, H.; Degroot, H.; 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, (1), 255-260.

34. Lewandowicz, A.; Ringia, E. A.; Ting, L. M.; Kim, K.; Tyler, P. C.; Evans, G. B.; Zubkova, O. V.; Mee, S.; Painter, G. F.; Lenz, D. H.; Furneaux, R. H.; Schramm, V. L., Energetic mapping of transition state analogue interactions with human and *Plasmodium falciparum* purine nucleoside phosphorylases. *J. Biol. Chem.* **2005,** 280, (34), 30320-8.

35. Basso, L. A.; Santos, D. S.; Shi, W. X.; Furneaux, R. H.; Tyler, P. C.; Schramm, V. L.; Blanchard, J. S., Purine nucleoside phosphorylase from *Mycobacterium tuberculosis*. Analysis of inhibition by a transition-state analogue and dissection by parts. *Biochemistry* **2001,** 40, (28), 8196-8203.

36. Shi, W. X.; Basso, L. A.; Santos, D. S.; Tyler, P. C.; Furneaux, R. H.; Blanchard, J. S.; Almo, S. C.; Schramm, V. L., Structures of purine nucleoside phosphorylase from *Mycobacterium tuberculosis* in complexes with immucillin-H and its pieces. *Biochemistry* **2001,** 40, (28), 8204-8215.

37. Nunez, S.; Antoniou, D.; Schramm, V. L.; Schwartz, S. D., Promoting vibrations in human purine nucleoside phosphorylase. A molecular dynamics and hybrid quantum mechanical/molecular mechanical study. *J. Am. Chem. Soc.* **2004,** 126, (48), 15720-9.

38. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* **2004,** 14, (6), 1188-90.

39. Kuhlman, B.; Baker, D., Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** 97, (19), 10383-10388.

40. Sammond, D. W.; Eletr, Z. M.; Purbeck, C.; Kimple, R. J.; Siderovski, D. P.; Kuhlman, B., Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J. Mol. Biol.* **2007,** 371, (5), 1392-404.

41. Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Winter, G., Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. *Biochemistry* **1983,** 22, (15), 3581-6.

42. Fersht, A., *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*. W.H. Freeman: New York, 1999.

43. Lienhard, G. E., Enzymatic catalysis and transition-state theory. *Science* **1973,** 180, (82), 149-154.

44. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G., How enzymes work: analysis by modern rate theory and computer simulations. *Science* **2004,** 303, (5655), 186-195.

Chapter III

DIRECTED EVOLUTION OF HUMAN PURINE NUCLEOSIDE PHOSPHORYLASE[1]

Introduction

Structure-based rational design strategies aim to improve protein properties such as rate of turnover, substrate acceptance, enantioselecity or protein stability by targeting mutations in the enzyme active site. There are numerous accounts proving the efficacy of rational active site mutation; however, there are surely just as many unreported failures in which the targeted function is not achieved. This is not to say that structure-based design strategies do not work but, instead, that current models of the complex interplay among residues in the active site and with the protein structural framework as a whole cannot account for the full function of the enzyme. Additionally, systems which are not fully characterized or for which structural data is unavailable are frequently the subject of engineering campaigns making rational design a shot in the dark at best. Directed evolution methodologies endeavor to fill this knowledge gap through the generation of random mutant libraries and subsequent screening for desired traits and/or activity.

Mutagenesis protocols in directed evolution have been developed which target the binding site or, alternatively, generate random sequence changes throughout the protein. Sufficient structural or sequence data must be available to target critical active-site residues. If this data is available site-directed saturation mutagenesis enables identification of active-site mutations which are difficult to predict or, by multi-site saturation mutagenesis, identification of sets of mutually beneficial mutations. On the other hand, random mutagenesis protocols can be applied in the absence of sequence

or structural data to identify beneficial mutations outside the binding site which are often difficult or impossible to predict.

In this chapter, the directed evolution of hPNP-Y88F, the best point mutant identified through computational redesign, is described with the aim of identifying mutations outside the binding site which affect ddI turnover. A high-throughput screen was developed allowing for the analysis of mutant libraries consisting of several thousand variants. Error-prone PCR was used to generate mutations throughout hPNP-Y88F, with a targeted dead rate of 30–40% resulting in a mutagenesis rate of 1.5–2 amino acid changes per variant. *In vitro* selection in cell free extracts allowed for identification of variants with improved ddI phosphorolysis. Overall, three rounds of directed evolution resulted in a mutant with six mutations and a ddI turnover rate in cell free extracts 36 times that of wild-type.

Methods

Library Generation and Screening

Mutant libraries were generated via epPCR using MutazymeII (Stratagene, Inc.) with forward primer, 5´-GCAGCAGCCATCATCATCATC-3´ and reverse primer, 5´-GGATCTCAGTGGTGGTGGTGG-3´ flanking the PNP coding region of pET28a-hPNP template constructs. To ensure efficient restriction digestion prior to ligation into pET28a, primers were designed to generate PCR product with overhangs 45 bp upstream and downstream of NdeI and HindIII restriction sites. pET28a-hPNP plasmid preparations were used as the template for directed evolution and subsequent rounds of directed evolution used plasmid preparations from the previous round as template. The rate of mutation was adjusted by a) varying template concentration in PCR reactions and b) the number of rounds of PCR. The mutation rate was titrated so that ca. 30% of PNP mutant subcloned into pET28a demonstrated <5% activity (assay described below).

Correspondingly, template concentrations were varied from 0.03–20 ng/μL with 20–30 cycles. The desired mutation rate was obtained with 20 ng/μL and 20 cycles. DNA sequencing of 10 random mutants at the 30% dead rate indicated a mutation rate of ~1.5–2 base pairs per kb.

PCR products were gel purified to remove template, digested with NdeI/HindIII and gel purified again prior to ligation into correspondingly restricted pET28a. Ligation reactions were performed using Promega T4 DNA ligase with 17 ng insert and 100 ng vector per 10 μL for 180 minutes. Transformation of the ligation reactions into *E. coli* BL21(DE3) cells was accomplished by electroporation. A random sampling of clones indicated all contained PNP insert.

Screening of hPNP variant activities

hPNP variants were expressed in *E. coli* BL21(DE3) cells harboring the pET28a-hPNP plasmid which was introduced by electroporation. The plasmid contains a kanamycin resistance marker allowing for selection of positive transformants on kanamycin-containing LB-agarose plates. Individual transformants were picked into 300 μL round bottomed 96-well plates containing 75 μL LB medium with 50 μg/mL kanamycin and grown to confluence (24 hours at 37°C with shaking at 220 rpm). Glycerol stocks of the library were generated by plate replication in LB medium prior to cells being collected by centrifugation at 3000 rpm. Supernatant was removed by inversion and the resulting pellets were frozen and stored at -80 °C until ready for assay. Directly prior to assay, frozen cell pellets were thawed and resuspended in 200 μL of a lysis mixture containing 50 mM phosphate buffer, 0.5 mg/mL egg white lysozyme (Sigma), 20 μg/mL deoxyribonuclease I (Sigma) and 0.1–0.125 mg/mL xanthine oxidase (to consume endogenous hypoxanthine) followed by a single freeze/thaw cycle from -80 °C to 37 °C[2]. Following centrifugation to remove cellular debris, 25 μL of cell free extract

from each well was transferred into 384-well flat bottom plates and assayed by addition of 50 µL of buffer containing 50 mM phosphate, pH 7, 50 mM HEPES, pH 7, 125 µM ddI, 0.0375% Triton X-100, 2 mM INT and 0.4–0.6 mg/mL xanthine oxidase. The reactions were followed continuously at 546 nm for 60 seconds. Hits from the primary screen were replicated from glycerol stocks (1–5 per plate) and assayed alongside cells expressing the parent PNP as a benchmark to eliminate false positives.

Plasmid preps from the best candidates were retransformed by electroporation and 1 mL of an overnight culture from the fresh transformants was used to inoculate 50 mL of LB broth and grown to OD_{600} ≈ 0.6–0.8 before addition of 1mM isopropyl-β-D-1-thiogalactopyranoside. After three hours of incubation at 37 °C, 2 mL aliquots were pelleted by centrifugation at 13,000 rpm and frozen at -80 °C. Pellets were lysed using BugBuster protein extraction reagent (Novagen Inc.) and the resulting cell free extract was diluted in Assay Mix to titrate the specific activity within the dynamic range of the kinetic assay as described above. Upon addition of substrate (125 µM ddI or 20 µM inosine) hypoxanthine release was measured continuously over 60 seconds. Turnover rates were calculated and normalized to the cell density (OD_{600}) at the time of harvest to yield a per-cell turnover rate for each PNP mutant. The kinetic parameters for selected mutant PNPs with high per cell turnover rates (2–3 per round) were established as described above. Of the mutants characterized in each round those with the most favorable kinetic parameters for ddI were selected as the source of template DNA for the next round of epPCR and screening.

Results

Screen optimization and epPCR mutagenesis

To further optimize the hPNP-Y88F mutant for turnover of ddI, a 96-well plate assay method for screening epPCR generated mutant libraries was designed and

implemented (Figure 3-1A). Variants were first cloned into pET28a and subsequent to transformation into electrocompetent *E. coli* BL21(DE3) single colonies were transferred into 96-well round bottom plates. The plates were incubated with shaking for 24 h and grown to confluence. Cells were harvested by centrifugation and lysed by addition of lysozyme and a single freeze/thaw cycle[2]. Upon removal of insoluble debris by centrifugation, cell-free extracts were transferred to 384-well flat-bottom plates to assay phosphorylase activity via a modified hypoxanthine formation assay. The assay is based on a continuous colorimetric assay for the conversion of hypoxanthine to uric acid via xanthine oxidase which is coupled to the reduction of iodonitrotetrazolium chloride to a purple formazan dye (λ_{max}: 546 nm)[3]. Reactions were initiated by addition of 125 μM ddI (~1/4 K_m) and the amount of enzyme was titrated to assure linearity of the reaction during the measurement window (≥60 s). To test the reproducibili ty of the screening methodology a single sequence library containing hPNP-Y88F was assayed. As shown in Figure 3-1B, under final assay conditions the single sequence library had a coefficient of variance of 11.6% within each plate.

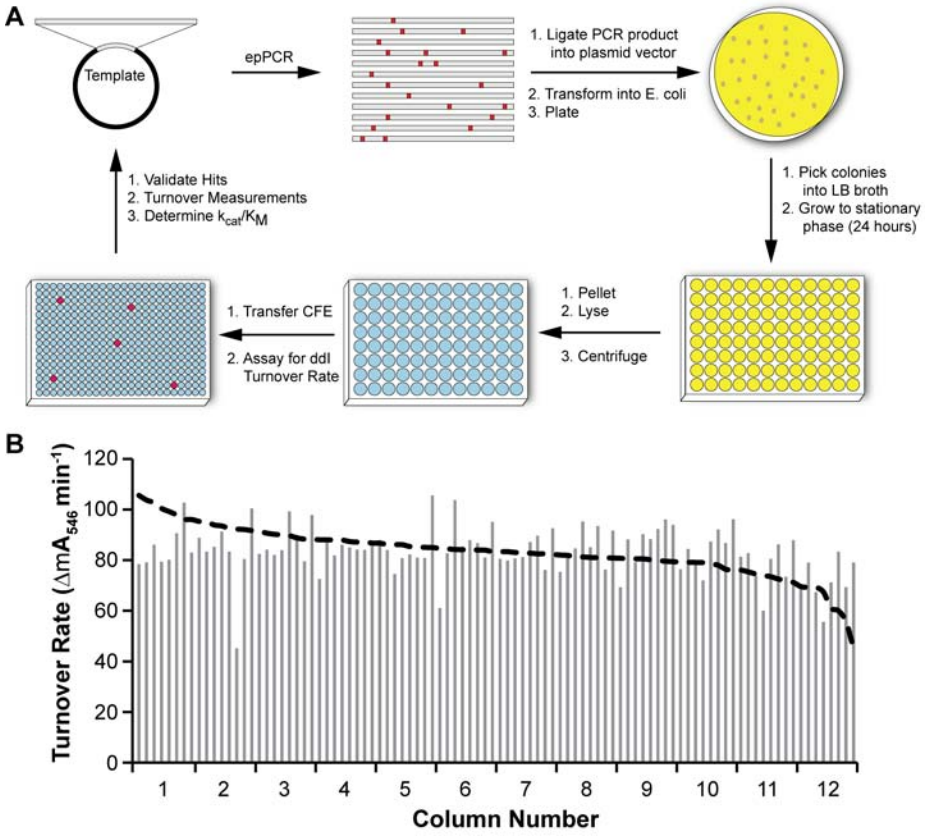

Figure 3-1. High-throughput screen for increased ddI phosphorolysis activity by PNP (A). Evaluation of screen reproducibility was accomplished by screening of a hPNP-Y88F single-sequence library (B). A coefficient of variance of 11.6% was calculated for each plate.

Diverse libraries were generated using an error-prone polymerase (Mutazyme, Stratagene, Inc) to amplify hPNP with Y88F hPNP cloned into pET28a serving as the template DNA. Primers were designed to be >45 bp upstream and downstream of restriction sites (NdeI and HindIII) in order to ensure efficient restriction of PCR products prior to ligation reactions. The rate of mutation was adjusted by varying the amount of template and number of PCR cycles such that the dead rate (colonies with <5% activity) was approximately 30–40% (Figure 3-2). Library sizes of 1 000–2 000 colonies had an average mutation rate of 1.5–2 mutations per kb, as established by sequencing 10

randomly selected clones. After screening, top hits from each plate were collected and re-screened from glycerol stocks of the error-prone library to eliminate false positives.

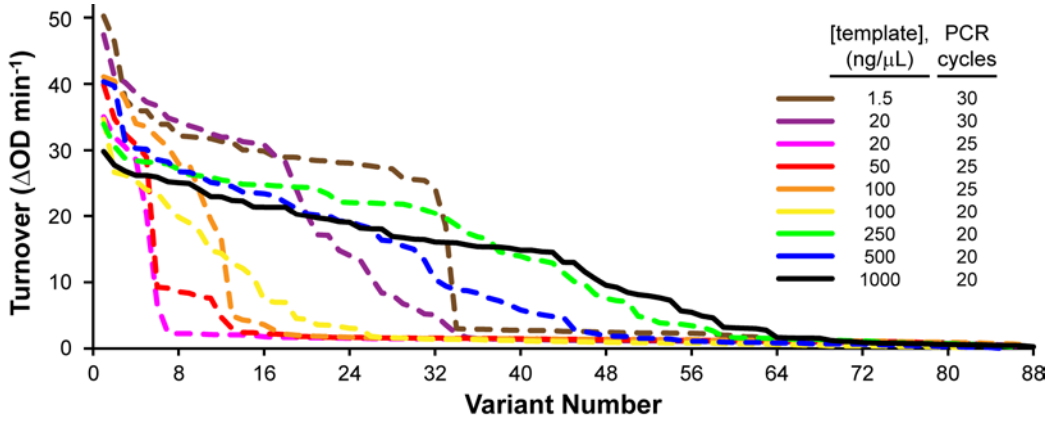

Figure 3-2. Mutagenesis rates established through variation of the template concentration and rounds of PCR. The solid black line indicates the selected epPCR conditions.

Library Screening and Mutant Characterization

Over three rounds of directed evolution the turnover rate of enzyme in *E. coli* extracts steadily increased, albeit modestly (Figure 3-3). The first round mutant, hPNP-19E2, has an increased normalized turnover rate for ddI with a concomitant decrease in turnover of inosine. Subsequent rounds provided a further increase in turnover of ddI but also an increase in inosine turnover. Catalytic constants, however, do not consistently improve for ddI throughout the directed evolution process (Table 3-1). The first round mutant hit (hPNP-19E2) has a k_{cat} and K_m of 5.8 ± 0.1 s^{-1} and 230 ± 15 μM which is a decrease in turnover rate but a significant improvement in K_m over the template. The kinetic parameters of hits from ensuing rounds of selection (hPNP-30F2 and hPNP-46D6) do not improve despite a regular increase in turnover in cell-free extracts. The final clone, hPNP-46D6, contained five amino acid mutations (G4E, Y88F, M170T,

Q172L, T177A) with the gene containing a sixth silent mutation (Table 3-1 and Figure 3-4).

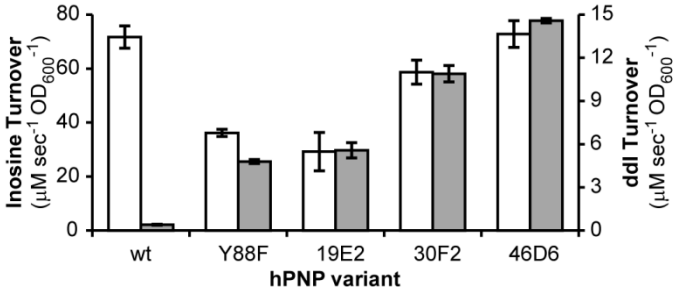

Figure 3-3. *In vitro* turnover of ddI increases throughout the directed evolution process. Turnover rates are normalized to cell density. Error bars are the standard deviation of three assays. Inosine, white; ddI, gray. This image copyright of Oxford University Press. Reproduced with permission. Originally published by Nannemann, *et al*.[1]

Table 3-1. Kinetic characteristics for mutants selected by directed evolution

Round	Variant	Acquired Mutation	Codon change	k_{cat} (s^{-1})		K_m (μM)		k_{cat}/K_m (M^{-1} s^{-1})	
				Inosine	ddI	Inosine	ddI	Inosine(x10^2)	ddI
	wild-type	-	-	43.9 ± 0.6	0.90 ± 0.01	48 ± 2	1030 ± 20	9210 ± 140	875 ± 8
rational	Y88F	Y88F	TAT→TTT	28.3 ± 0.8	10.5 ± 0.1	73 ± 4	450 ± 10	3890 ± 110	23140 ± 230
DE-1	19E2	M170T	ATG→ACG	9.6 ± 0.2	5.8 ± 0.1	26 ± 1	230 ± 15	3710 ± 70	25160 ± 620
DE-2	30F2	G4E Q172L C206C	GGA→GAA CAG→CTG TGT→TGC	14.5 ± 0.3	6.1 ± 0.1	37 ± 2	247 ± 10	3970 ± 90	24650 ± 330
DE-3	46D6	T177A	ACC→GCC	14.7 ± 0.7	4.5 ± 0.1	33 ± 4	235 ± 11	4390 ± 280	19240 ± 320

In addition to improving the turnover and kinetic characteristics of the mutant enzymes a marked improvement in substrate selectivity for ddI was observed. Each mutant has an improved specificity ratio (wild-type efficiency / mutant efficiency), though none of the variants are selective for ddI over inosine. Wild-type hPNP possesses a specificity ratio for inosine of >1000:1 whereas selected hPNP variants attained a specificity ratio of 15:1 subsequent the directed evolution process.

Discussion

The turnover rate of hPNP-Y88F was improved by directed evolution resulting in a 3-fold improvement in *in vitro* assays, and an overall 22-fold improvement in catalytic efficiency. By applying whole-gene random mutagenesis, mutations distal to the binding site that affect ddI turnover were identified. The acquired M170T mutation in hPNP-19E2 provides significant improvement in binding affinity with a modest decrease in turnover rate despite its location ~25 Å from the reaction center. In fact, it has been observed that distal, conservative mutations affect the dynamics and catalysis of hPNP and kinetic isotope effect studies indicate a significant change in the transition state structure[4-7]. These studies underpin the importance of dynamic motion coupling throughout the protein architecture to catalysis. Although the exact effect of the M170T mutation can only be speculated upon it favorably alters the PNP reaction coordinate for ddI phosphorolysis.

Successive rounds of directed evolution identified Q172L and T177A mutations which occur on the same helix as M170T (Figure 3-4). These may have accumulated to counteract any destabilizing effects of the first mutation or this gene region is particularly prone to mutation. Based on its location, the G4E mutant may re-orient and stabilize the recombinant hexahistidine tag which is not removed prior to screening or characterization. For instance, as judged by gel-filtration chromatography of wild type PNP, removal of the N-terminal hexahistidine tag used for purification significantly decreases the formation of higher-order oligomers (data not shown). Acquired mutations may improve expression efficiency or RNA stability as evidenced by the considerable increase in turnover of enzyme in *E. coli* extracts despite nominal improvements in catalytic efficiency.

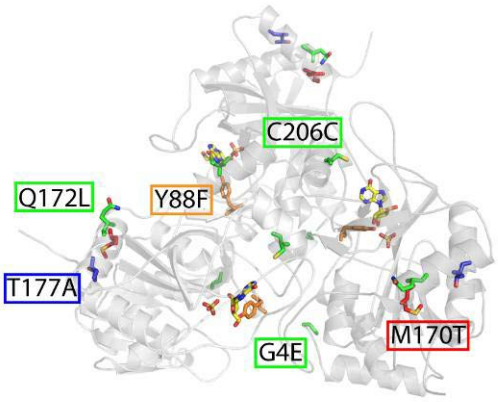

Figure 3-4. Mutations selected by directed evolution mapped onto the hPNP structure (PDB ID 1rct). Mutations are colored by lineage: orange, rational; red, DE-1; green, DE-2; blue, DE-3. This image copyright of Oxford University Press. Reproduced with permission. Adapted from Nannemann, *et al*.[1]

It is not surprising, however, that mutations in the first or second shell of the binding site were not identified in this work considering the moderate library sizes employed. However, directed evolution strategies that target residues in the first and second shell of the binding site (*i.e.* CASTing[8]) have found success in modifying or broadening the substrate specificity of enzymes. It is possible that additional mutations in the vicinity of Y88F may further improve the turnover of ddI as mutations closer to the active site are frequently observed to have greater effect on substrate selectivity[9]. Furthermore, formation of human:bovine PNP chimeras alters enzyme dynamics and catalysis, with a significant effect on the reaction profile[4-7]. A search of this conformational space through gene shuffling may identify mutants which utilize a reaction coordinate better fit to phosphorolysis of ddI.

Conclusions

Directed evolution using whole–gene epPCR and an *in vitro* selection scheme allowed for identification of residues distant to the binding site which improve the binding

affinity of the phosphorylase for ddI. Individual rounds resulted in modest improvements in kinetic parameters and/or turnover. The robustness of the *in vitro* tandem enzymatic assay is demonstrated by its ability to discriminate incremental improvements in activities. The final mutant (hPNP-46D6) has a catalytic efficiency 22-times greater than the wild-type and the specificity ratio was shifted to 15:1 from a starting point of 1000:1. On a per cell basis the turnover rate was 36-fold improved; it is possible that the additional improvement originates from increased protein production or a higher proportion of active enzyme. As these improvements were products of small libraries, it is likely that further rounds of directed evolution can improve the catalytic properties of the enzyme and identify residues in other regions of the protein that may enhance catalysis of ddI or further shift the selectivity.

In principle, catalysts for the phosphorolysis and formation of a variety of nucleoside analogs can be accessed using the methods described herein. Indeed, hPNP-46D6 has been tested for the biocatalytic generation of ddI as a component of tandem reactions and found it to be kinetically competent in the synthesis direction (to be presented in Chapter 5). Several nucleoside analogs with ribose substitution at 2´ and 3´ positions are of significant clinical importance as are nucleoside analogs with nucleobase substitutions[10-12]. While hPNP is specific for the 6-oxo nucleosides, mutations in the base binding region will likely improve turnover of compounds with nucleobase substitutions. For example, mutation of Asn-243 to aspartate has been shown to introduce 6-amino nucleoside phosphorolysis activity[13]. Of note, relatively few biotransformation-based approaches have been described for biocatalysis of dideoxynucleosides and other analogs[14-18]. Future experiments will endeavor to generate optimized enzymes and pathways capable of generating ribose precursors for PNP variants. Engineered pathways for nucleoside analogs have potential to provide economical alternatives to chemical synthesis of these valuable pharmaceuticals.

Acknowledgements

Portions of this research were originally published in Protein Engineering, Design and Selection. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Engineering, Design and Selection*. 2010; 23:607-616. DPN designed and conducted all experiments and wrote the manuscript. KWK contributed to the calculation of transition state binding energies and analysis of *in silico* data. JM advised in the design of *in silico* experiments for calculation of transition state binding energies and in analysis of *in silico* data. BOB advised in the design of experiments for assay of hPNP point mutants and directed evolution mutagenesis and screening methods and in analysis of the directed evolution study.

References

1. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010,** 23, (8), 607-616.

2. Hsu, C. C.; Hong, Z.; Wada, M.; Franke, D.; Wong, C. H., Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. *Proc Natl Acad Sci U S A* **2005,** 102, (26), 9122-6.

3. Degroot, H.; Degroot, H.; 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, (1), 255-260.

4. Ghanem, M.; Li, L.; Wing, C.; Schramm, V. L., Altered thermodynamics from remote mutations altering human toward bovine purine nucleoside phosphorylase. *Biochemistry* **2008,** 47, (8), 2559-64.

5. Li, L.; Luo, M.; Ghanem, M.; Taylor, E. A.; Schramm, V. L., Second-sphere amino acids contribute to transition-state structure in bovine purine nucleoside phosphorylase. *Biochemistry* **2008,** 47, (8), 2577-83.

6. Luo, M.; Li, L.; Schramm, V. L., Remote mutations alter transition-state structure of human purine nucleoside phosphorylase. *Biochemistry* **2008,** 47, (8), 2565-76.

7. Saen-Oon, S.; Ghanem, M.; Schramm, V. L.; Schwartz, S. D., Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase. *Biophys. J.* **2008,** 94, (10), 4078-88.

8. Fazelinia, H.; Cirino, P. C.; Maranas, C. D., Extending Iterative Protein Redesign and Optimization (IPRO) in protein library design for ligand specificity. *Biophys. J.* **2007,** 92, (6), 2120-2130.

9. Morley, K. L.; Kazlauskas, R. J., Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2005,** 23, (5), 231-237.

10. Li, F.; Maag, H.; Alfredson, T., Prodrugs of nucleoside analogues for improved oral absorption and tissue targeting. *J Pharm Sci* **2008,** 97, (3), 1109-34.

11. Soriano, V.; Peters, M. G.; Zeuzem, S., New therapies for hepatitis C virus infection. *Clin. Infect. Dis.* **2009,** 48, (3), 313-320.

12. Clerq, E. D., Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents* **2009,** 33, (4), 307-320.

13. Maynes, J. T.; Yam, W.; Jenuth, J. P.; Gang Yuan, R.; Litster, S. A.; Phipps, B. M.; Snyder, F. F., Design of an adenosine phosphorylase by active-site modification of murine purine nucleoside phosphorylase. Enzyme kinetics and molecular dynamics simulation of Asn-243 and Lys-244 substitutions of purine nucleoside phosphorylase. *Biochem. J.* **1999,** 344 Pt 2, 585-92.

14. Kaminski, P. A.; Dacher, P.; Dugue, L.; 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, (29), 20053-20059.

15. Komatsu, H.; Awano, H.; Ishibashi, H.; Oikawa, T.; Ikeda, I.; Araki, T., Chemo-enzymatic syntheses of natural and unnatural 2'-deoxynucleosides. *Nucleic Acids Res. Suppl.* **2003**, (3), 101-102.

16. Medici, R.; Porro, M. T.; Lewkowicz, E.; Montserrat, J.; Iribarren, A. M., Coupled biocatalysts applied to the synthesis of nucleosides. *Nucleic Acids Symp. Ser.* **2008**, (52), 541-2.

17. Rogert, M. C.; Trelles, J. A.; Porro, S.; Lewkowicz, E. S.; Iribarren, A. M., Microbial synthesis of antiviral nucleosides using *Escherichia coli* BL21 as biocatalyst. *Biocatal. Biotransform.* **2002,** 20, (5), 347-351.

18. Shirae, H.; Kobayashi, K.; Shiragami, H.; Irie, Y.; Yasuda, N.; 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, (2), 419-424.

Chapter IV

BIOCHEMICAL CHARACTERIZATION, STRUCTURAL ANALYSIS AND DETERMINATION OF CATALYTIC CYCLE FOR *BACILLUS CEREUS* PHOSPHOPENTOMUTASE[1]

Introduction

Enzyme-catalyzed phosphoryl transfer forms the basis for many biological, bioenergetic and regulatory processes, and is one of the most common cellular reactions[2]. Numerous enzyme families have evolved mechanistically distinct solutions for phosphoryl transfer[3]. Phosphomutases are phosphotransfer enzymes that rearrange the position of phosphate within a substrate molecule through either *intra*molecular (*i.e.* the phosphate is transferred to a different position on the same molecule) or *inter*molecular phosphoryl transfer (*i.e.* the phosphate is transferred from one substrate molecule to another).

Bacterial phosphopentomutases (PPMs, EC 5.4.2.7) interconvert α-D-ribose 1-phosphate (ribose 1-phosphate) and α-D-ribose 5-phosphate (ribose 5-phosphate), which bridges glucose metabolism and RNA biosynthesis[4]. The importance of this reaction has recently been underscored by the observation that targeted deletion of the gene encoding PPM in the pathogen *Francisella tularensis* (*deoB*) results in markedly decreased virulence[5]. PPMs appear to be biochemically and structurally distinct from their human congeners[6, 7] making them potential targets for antibiotic development.

Sequence clustering classifies prokaryotic PPMs within the alkaline phosphatase superfamily of metalloenzymes, which includes a range of functionally diverse enzymes such as cofactor-independent phosphoglycerate mutase, phosphodiesterase, and estrone and aryl sulfatases[8]. The majority of alkaline phosphatase superfamily enzymes catalyze a hydrolase reaction; however, both PPM[6] and the co-factor independent phosphoglycerate mutase catalyze phosphomutase reactions[9, 10].

All previously-characterized alkaline phosphatase superfamily members follow a unified general reaction mechanism[11], which is shown in Figure 4-1. In alkaline phosphatase itself[12, 13], the catalytically competent enzyme has an unphosphorylated catalytic nucleophile, Ser-102 (Figure 4-1, State 1). Turnover is initiated when the metallocenter activates a phosphoester donor substrate (R_D-OPO_3H^-) (Figure 4-1, State 2) to transfer the phosphoryl group to the hydroxyl of Ser-102 (Figure 4-1, State 3). This results in a covalent phosphoenzyme intermediate (E-OPO_3H^-) (Figure 4-1, State 4). A second phosphoryl transfer from the enzyme to the acceptor water molecule (Figure 4-1, States 5,6) completes the reaction cycle. This general reaction mechanism has also been verified for cofactor-independent phosphoglycerate mutase[9, 10], a phosphomutase in the alkaline phosphatase family. In this case, the incoming substrate is also the acceptor molecule, thus defining it as an *intra*molecular phosphomutase.

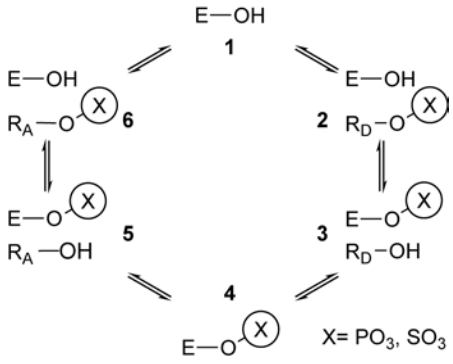

Figure 4-1. Generalized catalytic cycle for alkaline phosphatase superfamily proteins. For all previously characterized alkaline phosphatase family members, the catalytic cycle begins with a dephosphorylated (or desulforylated) enzyme, E-OH, State 1. Donor substrate (R_D-O-X, where X=PO_3H^-, SO_3H.) binds, and is attacked by the catalytic nucleophile, States 2, 3. The dephosphorylated donor substrate, R_D-OH is released, leaving a covalently modified enzyme intermediate, E-O-X, State 4. An acceptor substrate, R_A-OH, attacks the phosphoryl enzyme, forming a phosphorylated product, R_A-O-X, and returning the enzyme to the dephosphorylated state, E-O-H, States 5, 6, 1.

By comparison, characterization of bacterial PPM has been limited. Biochemical analysis of the *E. coli* enzyme revealed that *E. coli* PPM catalyzes Mn^{2+}- or Co^{2+}-dependent

interconversion of ribose 5-phosphate and ribose 1-phosphate[6] and that α-D-glucose 1,6-bisphosphate (glucose 1,6-bisphosphate) stimulates activity. Given the wealth of information available for alkaline phosphatase superfamily members, one might reasonably expect to infer the mechanism of bacterial PPM based on its classification within that family. This would predict that, like cofactor-independent phosphoglycerate mutase, PPM acts as an *intra*molecular phosphomutase. However, to our knowledge, no study has directly investigated the catalytic cycle of PPM.

This work investigates the biochemical mechanism of *Bacillus cereus* PPM in the context of the alkaline phosphatase superfamily fold. Surprisingly, the data demonstrate that PPM acts as an *inter*molecular transferase, which would require an adaptation of the alkaline phosphatase reaction mechanism. In one possible modification, substrate could be acted upon by phosphorylated enzyme (Figure 4-1, State 4) and could proceed through the catalytic cycle from this altered entry point. These findings expand upon the mechanistic repertoire for the broadly conserved alkaline phosphatase scaffold.

Methods

Protein expression, purification and crystallization

Genomic DNA for B. cereus ATCC 14579D was purchased from ATCC. DNA encoding PPM (accession No. Q818Z9.1) was amplified by PCR using 5′-GCTAGCAATAAATATAAACGTATATTCCTAGTCG-3′ as the sense primer and 5′-CTCGAGCTATTTCTTTAGCTCGTTTAAG-3′ as the antisense primer. PPM was ligated into the plasmid pET28a (Novagen) between the NheI and XhoI sites to include an N-terminal hexahistidine tag and thrombin cleavage site. The resulting plasmid was electroporated into *E. coli* BL21-Gold(DE3) (Stratagene). Cells were grown in LB supplemented with 100 mg/l kanamycin at 37 °C with continuous shaking at 250 rev/min until the A_{600} was between 0.4 and 0.6. Protein expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside to

0.3 mM and was continued with shaking at 37 °C for 4 h. Cells were harvested by centrifugation at 5 000 g for 15 min at 4 °C and were resuspended in 25 mL buffer A (50 mM Tris–HCl pH 7.5, 200 mM NaCl, 10 mM imidazole) enriched with 1 μg/mL DNase I, 1 μg/mL leupeptin and 1 μg/mL pepstatin A. Cells were lysed by sonication on ice for 10 min (60 Sonic Dismembrator, Fischer Scientific) and then clarified by centrifugation at 30 000 g for 1 h at 4 °C in an SS-34 rotor (Sorvall). The supernatant was filtered through a 0.45 μm filter prior to chromatography at 4 °C. The filtered lysate was passed over a HisTrap HP column (GE) equilibrated with buffer A. Nonspecifically bound protein was removed by washing with ten column volumes of buffer A supplemented with 50 mM imidazole. Purified protein was eluted by adding buffer A supplemented with 300 mM imidazole. The eluant was concentrated to 500 mL using an Amicon Ultra 30 000 molecular-weight cutoff centrifugal filter (Millipore) spun in a swinging-bucket rotor at 5 000 g at 4 °C. The hexahistidine tag was removed by exchanging into buffer B (50 mMTris–HCl pH 7.5 and 200 mM NaCl) and cleaving overnight with 50 units of thrombin. The resultant protein contained three additional residues, Gly-Ser-His, at the N-terminus. PPM was separated from protein aggregates and the cleaved peptide tag by size-exclusion chromatography using a Superdex 200 10/300 GL column (GE) equilibrated with 25 mM Tris–HCl pH 7.5 and 1 mM $MnCl_2$. The protein was concentrated to between 50 and 75 mg/mL as determined by the Bio-Rad Protein Assay (Bio-Rad) and was stored at 4 °C prior to use.

The T85A variant was generated from pET28a-PPM using the Stratagene QuikChange Lightning Site-Directed Mutagenesis Kit with oligonucleotide primers (5´-CTGGTAAAGATGCAATGACAGGTCACTGG-3´ and 5´-CCAGTGACCTGTCATTGCATCTTTACCAG-3´). This variant was purified using the protocol developed for the wild-type enzyme. Purified PPM was concentrated and stored in aliquots at -80 °C in 25 mM Tris–HCl, 1 mM $MnCl_2$, pH 8.0. Xanthine oxidase was prepared from raw milk using the standard procedure[14] and human purine nucleoside phosphorylase was purified as previously reported[15] and in chapter 2.

The structure of PPM was determined alone, co-crystallized with ribose 5-phosphate, co-crystallized with glucose 1,6-bisphosphate, and following activation with glucose-1,6-bisphosphate. Crystals of PPM (12 mg/mL buffered in 25 mM Tris–HCl, pH 7.4, 1 mM $MnCl_2$) were grown at 18 ºC as described previously[16] using the hanging drop vapor diffusion method. Protein solution (1 µL) and 1 µL of crystallization reservoir solution were equilibrated over a reservoir containing 100 mM Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane (BisTris), pH 5.5, 50 mM $MnCl_2$, 14% polyethylene glycol 3350, and 75 mM NH_4CH_3COO. Co-crystals of PPM with ribose 5-phosphate were prepared by soaking fully-formed crystals in a solution containing all of the crystallization components and 10 mM ribose 5-phosphate for 1 hour at 18 °C. Co-crystals of PPM with glucose 1,6-bisphosphate were prepared with the same procedure, but were soaked with 5 mM glucose 1,6-bisphosphate. Crystals of activated PPM were grown from purified protein pre-incubated with 5 mM glucose 1,6-bisphosphate for 30 minutes at 22 °C before crystallization over a reservoir solution containing 100 mM BisTris pH 5.5, 50 mM $MnCl_2$, 14% polyethylene glycol 3350, and 50 mM $CH_3CO_2NH_4$. Prior to data collection, crystals were cryoprotected in a solution containing the crystallization components and 30% glycerol, then flash cooled in liquid nitrogen. Crystal quality was assessed by diffraction-based feedback at both the Advanced Photon Source (APS) beamline 21-ID-G and the Stanford Synchrotron Radiation Lightsource beamline 9-2. All x-ray diffraction datasets were collected at APS 21-ID-G using a temperature of -173 °C, a wavelength of 0.979 Å and a MAR 225 CCD detector. Data were processed using the HKL2000[17] and CCP4[18] program suites. Crystals formed in the monoclinic space group $P2_1$ with unit cell dimensions listed in Table 1. The structure of native PPM was determined by molecular replacement with the program PHASER[19] using the unpublished crystal structure of a putative PPM from *Streptococcus mutans* (PDB ID 2I09; New York SGX Research Center for Structural Genomics) as the search model. Residues 133 to 145 and 199 to 208 were absent from the search model and were built into $2|F_o|-|F_c|$ omit maps

using COOT[20]. A careful procedure of omit mapping during refinement was used to avoid model bias.

Model refinement was performed in the programs CNS[21] and REFMAC5[22] with Translation/Libration/Screw (TLS)[23] groups determined in the TLS Motion Determination server[24]. Superposition of residues 2–393 of the final model of *B. cereus* PPM with the equivalent residues (3–402) of the putative PPM from *S. mutans* resulted in an overall root mean squared (RMS) deviation of 1.0 Å for the core domain (residues 2–99 and 219–393) and an RMS deviation of 0.7 Å for the cap domain (residues 102–216).

Since all crystals of the *B. cereus* PPM were isomorphous, the refined model of native PPM was subjected to rigid body refinement in CNS[25] and used as a starting model for the remaining structures. Restraint libraries for ribose 5-phosphate and Thr(P) were generated using the PRODRG server[26]. The R_{free} test set of data was composed of 5847 (5%) reflections randomly-selected in CNS[25]. The reflections of the R_{free} test set were selected to be identical in the remaining data sets with additional reflections randomly-selected in the higher-resolution structure of PPM in complex with glucose 1,6-bisphosphate (Table 1). Refinement statistics and PDB accession codes for each model are listed in Table 1. Root mean squared (RMS) deviations between structures were calculated using the program LSQKAB[27]. Figures 4-2–4-4 were prepared using PyMol[28]. The electrostatic surface calculation shown in Figure 2E was prepared using the PDB2PQR[29] server set to use the CHARMM force field[30]. Surface maps were calculated using the APBS[31] tools plug-in for PyMol[28].

Table 4-1. Data collection and refinement statistics

	PPM	Ribose-5-phosphate soaked PPM crystals	Glucose-1,6-bisphosphate soaked PPM crystals	Activated PPM
PDB entry ID	3M8W	3M8Z	3OT9	3M8Y
Data Collection:				
Wavelength (Å)	0.979	0.979	0.979	0.979
Beamline	APS 21-ID-G	APS 21-ID-G	APS 21-ID-G	APS 21-ID-G
Resolution (Å) (last	50-1.85 (1.92-1.85)	20-1.90 (1.86-1.80)	50-1.75 (1.81-1.75)	50-2.1 (2.18-2.10)

shell)[a]				
Unit-cell dimensions (Å)	a = 91.2 b = 76.6 c = 106.6 α = 90° β = 108.9° γ = 90°	a = 90.6 b = 76.8 c = 107.1 α = 90° β = 108.7° γ = 90°	a = 91.9 b =76.6 c = 107.1 α = 90° β = 108.7° γ = 90°	a = 91.2 b = 76.7 c = 107.8 α = 90° β = 108.9° γ = 90°
Total reflections	361062	469857	424368	269649
Unique reflections	116902	122525	126882	75473
Reflections in test set	5847	5835	6361	5280
I/σ	13.3 (3.0)	25.7 (2.5)	23.7 (2.8)	14.2 (3.6)
Completeness (%)	98.6 (99.1)	98.8 (90.3)	94.0 (69.6)	91.8 (71.6)
R_{sym}[b]	0.080 (0.369)	0.063 (0.380)	0.053 (0.298)	0.084 (0.257)
Model Refinement:				
R_{cryst}[c]	0.175	0.177	0.162	0.166
R_{free}	0.216	0.208	0.190	0.213
RMSD Bonds (Å)	0.018	0.008	0.007	0.009
RMSD Angles (°)	1.388	1.112	1.076	1.194
Ramachandran Analysis:				
Most favored (%)	90.3	90.4	90.5	90.0
Additional allowed (%)	9.7	9.6	9.5	10.0
Generously allowed (%)	0.0	0.0	0.0	0.0
Disallowed (%)	0.0	0.0	0.0	0.0

a. Values in parenthesis are those from the highest resolution shell.
b. $R_{sym}=\sum |I_i-I_{mean}|/I_{mean,}$
c. $R_{cryst}=\sum||F_{obs}|-|F_{calc}||/\sum|F_{obs}|$

Phosphothreonine Western analysis

Either wild-type or the T85A variant of PPM (at a concentration of 2 µM and buffered in 50 mM Tris–HCl, pH 8.0 and 0.1 mM $MnCl_2$) were incubated with 0, 1, 2.5, 5, 7.5, 10, 25, 50, 75, 100 and 1 000 µM glucose 1,6-bisphosphate (wild-type) or 0 and 1 000 µM glucose 1,6-bisphosphate (T85A) at 23 °C for 30 minutes. SDS sample buffer was added, and each reaction was separated by SDS-PAGE and transferred to a 0.45 µm nitrocellulose membrane. Total protein was estimated with Ponceau S staining, which was subsequently removed by washing in water. The membrane was then blocked with Odyssey blocking buffer (LI-CORP Biosciences)

overnight and incubated with an anti-phosphothreonine antibody (Cell Signaling) diluted 1:1 000 in TBS containing 0.5 % BSA and 0.1 % Tween 20 (BSA-TTBS) for four hours at room temperature. Binding of the primary antibody was detected by incubation with an AlexaFluor680-conjugated goat anti-rabbit antibody (Invitrogen) diluted 1:10 000 in BSA-TTBS for 1 hour at room temperature and visualized using an Odyssey Infrared Imaging System. Both total protein and anti-phosphothreonine band intensities were determined by plotting lane intensities and then measuring the area of the peak corresponding to the band at 44 kDa in the program ImageJ[32].

Identification of phosphorylation by LC/MS/MS

Purified PPM (10 µg) was subjected to proteolytic digestion by trypsin, elastase, and subtilisin in parallel using a modification of a published procedure[33]. Data from each digest of the sample were acquired with 3 separate 90-minute LC/MS/MS data acquisitions using a Thermo-Fisher LTQ-orbitrap equipped with a nanospray source coupled to an Eksigent 1D+ nano-LC pump. Peptide tandem mass spectra were extracted from the instrument files (Scansifter) and searched against a database containing *B. cereus* proteins using SEQUEST[34], considering possible phosphorylation events on serine, threonine, and tyrosine. Search results were collated and filtered using IDPicker[35].

Assay for activity

Enzyme turnover was calculated by coupling the PPM catalyzed formation of ribose 1-phosphate with human purine nucleoside phosphorylase in the presence of hypoxanthine to form inosine. Inosine formation at endpoints was determined via correlating to hypoxanthine consumption with a typical hypoxanthine detection assay[36] in which xanthine oxidase

stoichiometrically converts hypoxanthine into to uric acid, H_2O_2, and superoxide ion, O_2^-. Superoxide in turn was determined via chemical reduction of iodonitrotetrazolium chloride to a formazan chromophore, which has a maximal absorbance at 546 nm. The amount of human purine nucleoside phosphorylase required for catalytic excess was determined empirically for reactions via a concentration series. In assays containing 50 nM PPM and 750 μM hypoxanthine, 1.1 μM human purine nucleoside phosphorylase was determined to be the concentration above which no greater rate of product formation was observed. A three-fold excess of human purine nucleoside phosphorylase (3.3 μM) was used in all experiments to assure catalytic excess. At specified time points, corresponding to <15% ribose 5-phosphate consumption, reactions were quenched by either heat denaturation or acidification, as described below. Hypoxanthine concentrations were determined by addition of Developer mix and comparison of ΔA_{546} to a standard curve run in parallel to biochemical assays. Developer mix contained 25 mM Tris–HCl, pH 8.0, 0.356 % v/v Triton X-100, 7.13 mM iodonitrotetrazolium chloride, and 2.4 mg/mL xanthine oxidase. Details of individual assay formats are provided below in each study. This assay meets the requirements for a valid tandem assay, which are: 1) the first reaction must be zero-order with respect to substrate; and 2) the second stage reaction must not be rate-limiting[37].

Endogenous metal ions were removed from affinity-purified PPM by treatment of 100 μM stock solutions with 1 mM EDTA for 1 hour at 4 °C. The protein was then exchanged into 25 mM Tris–HCl, pH 8.0 using a HiTrap desalting column. To determine metal dependence of PPM, chelated preparations were incubated at 23 °C with 1 mM of $MnCl_2$, $ZnCl_2$, $NiCl_2$, $MgCl_2$, or $CoCl_2$ for 3 min prior to biochemical assay. Relative turnover assays (100 μl) contained 1 μM glucose 1,6-bisphosphate, 10 nM reconstituted PPM, 25 mM Tris–HCl, pH 8.0, 750 μM hypoxanthine, and 3.3 μM purine nucleoside phosphorylase, and were initiated by the addition of 500 μM ribose 5-phosphate. PPM and purine nucleoside phosphorylase activities were quenched after an 8 minute reaction by heat denaturation at 95 °C for 5 min. Product formation

was measured via the hypoxanthine consumption assay, performed by adding 20 µL of Developer Mix to 75 µL of the turnover reaction. The absorbance at 546 nm was monitored at 15 s intervals until no further absorption change was observed, which typically occurred within 10 min.

The initial velocity of PPM was measured in the presence of increasing concentrations of glucose 1,6-bisphosphate in microtiter plates. PPM (10 µl of a 200 nM stock solution) was activated by incubating with 80 µL of a solution containing 0, 0.1, 0.25, 0.5 and 1 µM glucose 1,6-bisphosphate in 25 mM Tris–HCl, pH 8.0, 111 µM $MnCl_2$, 937.5 µM hypoxanthine and 4.1 µM human purine nucleoside phosphorylase. Turnover was initiated by the addition of 10 µL of 5 mM ribose 5-phosphate resulting in a final concentration of 0.5 mM. Reactions were quenched after 4 min by acid denaturation with 5 µL of a solution containing 1M HCl and 1M $CaCl_2$. Reactions were neutralized by the addition of 5 µL 1M NaOH immediately prior to determination of the hypoxanthine concentration. Hypoxanthine consumption was measured by removing a 75 µl aliquot of quenched/neutralized assay mixture and combining with 20 µL Developer Mix. The absorbance at 546 nm was monitored at 15 sec intervals until no further absorption change was observed, which usually occurred within 10 min.

Biochemical data for determination of kinetic parameters was determined in 96-well microtiter plates. Enzyme Mix was freshly prepared by pre-activating 200 nM PPM in 25 mM Tris–HCl, pH 8.0, with 5 µM glucose 1,6-bisphosphate and 100 µM $MnCl_2$ at 23 °C for 10 min and maintained at 4 °C until assayed. In the 96-well plate assay, 100 µl reactions contained 10 µL of Enzyme Mix and 80 µL of Assay Mix comprising 25 mM Tris–HCl, pH 8.0, and 111 µM $MnCl_2$, 937.5 µM hypoxanthine, and 4.1 µM human purine nucleoside phosphorylase. Reactions were initiated by the addition of 10 µL of the appropriate 10x substrate stock solutions and quenched after 4 minutes by addition of 5 µL of a solution containing 1M HCl and 1M $CaCl_2$. Reactions were neutralized by the addition of 5 µL 1M NaOH immediately prior to determination of the hypoxanthine concentration. Correspondingly, initial velocities of the activated enzyme

were measured in a final concentration of 25 mM Tris–HCl, pH 8.0, 100 μM $MnCl_2$, 0.5 μM glucose 1,6-bisphosphate, 20 nM activated PPM, and final substrate concentrations of 50, 75, 100, 170, 220, 270, 350, 500, and 700 μM ribose 5-phosphate. Hypoxanthine consumption was measured by combining 75 μl of quenched assay mixture with 20 μl Developer mix as described in the preceding glucose 1,6-bisphosphate dependence assay. Hypoxanthine concentrations were determined by comparison of ΔA_{546} to a standard curve run in parallel and initial velocities were fit to the Michaelis-Menten equation using non-linear regression analyses in the GraphPad Prism software package version 5.01 for Windows.

Phosphatase activity of PPM was measured during ribose 5-phosphate turnover using the Phosphate Sensor based assay (Invitrogen), which monitors the change in fluorescence of a fluorophore-labeled *E. coli* phosphate-binding protein upon phosphate binding. The reaction mixture contained 0.1 μM PPM, 0.1 mM $MnCl_2$, 25 mM Tris–HCl, pH 8.0, and 0.5 μM Phosphate Sensor. The reaction was initiated with 5 μM ribose 5-phosphate and 5 μM glucose 1,6-bisphosphate, and the fluorescence was monitored for 10 min at room temperature in a Cary Eclipse Spectrofluorimeter with λ_{ex} = 430 nm and λ_{em} = 470 nm. The concentration of phosphate formed by PPM was determined by comparison to a standard curve of potassium phosphate in 0.1 μM PPM, 0.1 mM $MnCl_2$, and 25 mM Tris–HCl.

Isotope relay assays

Isotopically labeled ribose 5-phosphates were synthesized using ribokinase with ribose or [$^{13}C_5$]ribose and ATP or [γ-$^{18}O_4$]ATP (Cambridge Isotope Laboratories). Ribokinase was expressed and purified as previously described[38] and stored at -80 °C in 25 mM Tris–HCl, pH 8.0 and 5 mM $MgCl_2$. Preparative reactions contained a final concentration of 1 mM ribose (ribose or [$^{13}C_5$]ribose), 1 mM ATP ([γ-$^{18}O_4$]ATP or ATP), 25 mM Tris–HCl, pH 8, 30 mM KCl, 0.5 mM $MnCl_2$, 0.5 mM $MgCl_2$ and 0.5 μM ribokinase. Reactions were incubated at 25 °C and quenched after 5 min by heat denaturation of ribokinase at 94 °C for 5 min. After centrifugation

for 2 min at 15 000 g at 4 °C, the enzymatically-synthesized ribose 5-phosphate species were stored at 4 °C until used.

Labeled ribose 5-phosphates, prepared as described above, were mixed in a 1:1 ratio on ice. Reactions in a final volume of 30 µL were initiated by addition of activated PPM to a final concentration of 2 µM at 25 °C. PPM (20 µM) was activated with 20 µM glucose 1,6-bisphosphate in the presence of 0.5 mM MnCl2 and 25 mM Tris–HCl, pH 8.0. After 30 minutes, the reactions were quenched with an equivalent amount of ice-cold methanol. In control reactions, functional activity of PPM was verified by observing conversion of ribose 1-phosphate to inosine by adding 10 µM human purine nucleoside phosphorylase and 1.5 mM hypoxanthine.

Ribose phosphate species were separated from contaminating salts and other reaction components using a Hypercarb porous graphite column (3 x 50 mm, ThermoFisher Scientific) and an isocratic flow of 0.2 mL/min of 20 mM CH_3COONH_4, pH 6, and 0.1% diethylamine. A ThermoPal autosampler was used to inject 10 µL of the quenched reaction and a divert valve was operational for the first 2.5 min to avoid introducing salts into the ion source. Ribose phosphate species typically eluted in the range of 3.5–4.5 min. Isotopologue peak intensities were averaged over the 3.5–4.5 min range for the data shown. Ribose phosphate species were analyzed on a triple quadrupole ESI–LC/MS (ThermoScientific Quantum Access) equipped with an ESI interface outfitted with a 100 µm internal diameter deactivated fused silica capillary. Nitrogen was used as the sheath gas and argon as the auxiliary gas, at 54 and 10 (arbitrary units), respectively. The mass spectrometer was operated in the negative mode with the following parameters: spray voltage, 3 kV; capillary temperature, 200 °C; vaporizer temperature, 280 °C; tube lens offset, -103 V; skimmer offset, -15 V; capillary offset, -35 V. The parameters were tuned by direct infusion of a 500 µM ribose 5-phosphate standard at 10 µL/min. Data acquisition and spectral analyses were conducted with Thermo Xcalibur Software, version 2.1.

Results

Structure of the B. cereus PPM with and without substrates

B. cereus PPM folds into two distinct domains: a core domain (residues 2–99 and 219–393) and a cap domain (residues 102–216) which are illustrated in Figure 4-2A. The core domain is organized around an alkaline phosphatase fold (Figure 4-2B,C), and superposition with the *E. coli* alkaline phosphatase (PDB ID 1ALK[13]) results in a RMS deviation of Cα atoms of 2.9 Å. The cap domain contains three helices surrounding a 5-stranded mixed β-sheet (Fig. 4-2A,D) and a homology search using the EMBL DaliLite server[39] established that the fold is unique to prokaryotic PPMs.

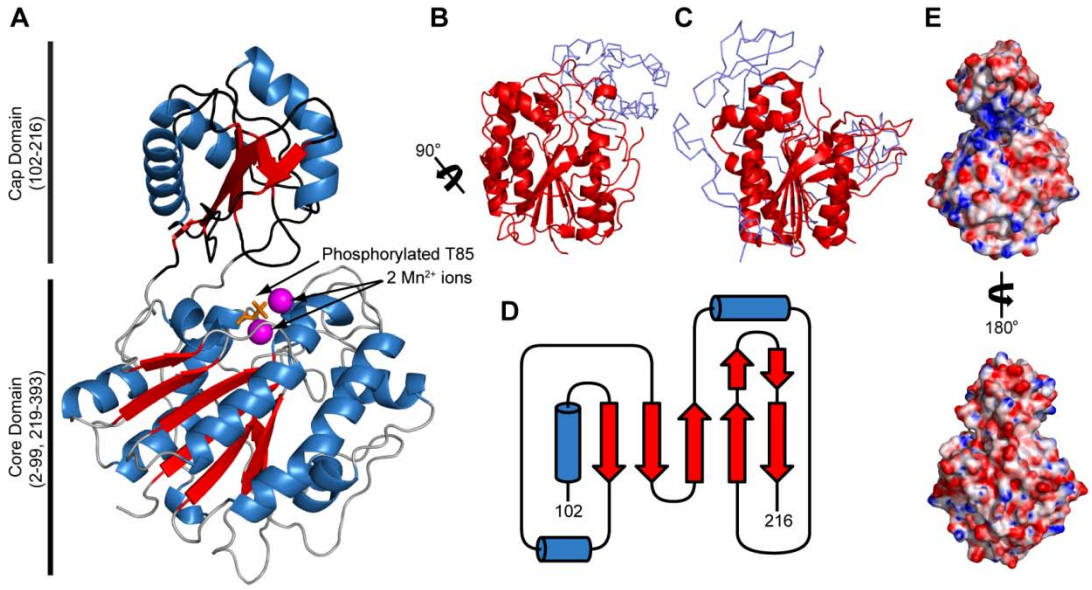

Figure 4-2. Structural overview of *B.cereus* PPM. Panel A shows the overall fold of *B. cereus* PPM. PPM is composed of a core domain and a cap domain. α-Helices are colored in blue and β-strands are colored red. The two Mn^{2+} ions are shown as a CPK representation and are colored magenta. The side chain of Thr(P)-85 is shown as a stick representation in orange. A comparison of the fold of the *B. cereus* PPM core domain to that of *E. coli* alkaline phosphatase (PDB ID 1ALK[13]) is shown in panel B,C. The structures of PPM (panel B, rotated 90° from the view in panel a about the black line) and alkaline phosphatase (panel C), were aligned by their homologous metal coordinating residues in PyMol[28]. Panel D shows a topology diagram of the cap domain (residues 102–216). The coloring scheme is the same as panel a. Two views of electrostatic surface potential mapped onto the surface of PPM are shown in panel E with negatively-charged surfaces colored red, positively-charged surfaces colored blue, and

contoured from -25 kT/e to +25 kT/e. This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

An electropositive cleft at the interface between the core and cap domains (Fig. 4-2E), houses the active site (Fig. 4-3A) and contains electron density consistent with two metals ions. Collection of diffraction data at the Mn-edge (λ=1.89 Å) revealed an anomalous signal for both metals at this wavelength (not shown). To be consistent with the nomenclature adopted for the alkaline phosphatase superfamily, the metals are referred to as Mn-1 and Mn-2. Comparison of the active site with *E. coli* alkaline phosphatase (Fig. 4-3B) and other structurally-characterized members of the alkaline phosphatase superfamily revealed that the coordinating ligands to this dimetallo center are structurally conserved with the exception of an additional ligand to Mn-1 that appears to be unique to prokaryotic PPMs (Asp-156 Oδ1).

Among the coordinating ligands of the dimetallo center is Thr-85, which is structurally homologous to the catalytic nucleophile, Ser-102, in the *E. coli* alkaline phosphatase (Fig 4-3B). In accordance with the alkaline phosphatase reaction mechanism (Figure 4-1), Ser-102 is not phosphorylated at the start of the catalytic cycle (Figure 4-1, Step 1), but becomes transiently phosphorylated as the catalytic intermediate (Figure 4-1, States 3,4,5). This transient phosphorylation was challenging to trap for crystallographic studies of the *E. coli* alkaline phosphatase[40]. As a result, observation of electron density consistent with phosphorylation of Thr-85 in the *B. cereus* PPM was unexpected (Figure 4-3C). The electron density at Thr-85 was weaker than expected for a phosphate modification, but could not be explained by a water molecule. This suggested that the crystals formed from a pool of protein containing a mixture of both phosphorylated and unphosphorylated enzyme. Crystals grown from protein pre-incubated with the reported activator glucose-1,6-bisphosphate[6] appeared to have stronger electron density for the phosphate modification of Thr-85, as shown in Figure 4-3D, suggesting that the

activator transfers a phosphoryl group to Thr-85. In the phosphorylated enzyme, two phosphate oxygens form a coordinating bridge between Mn-1 and Mn-2 (Figure 4-3E).

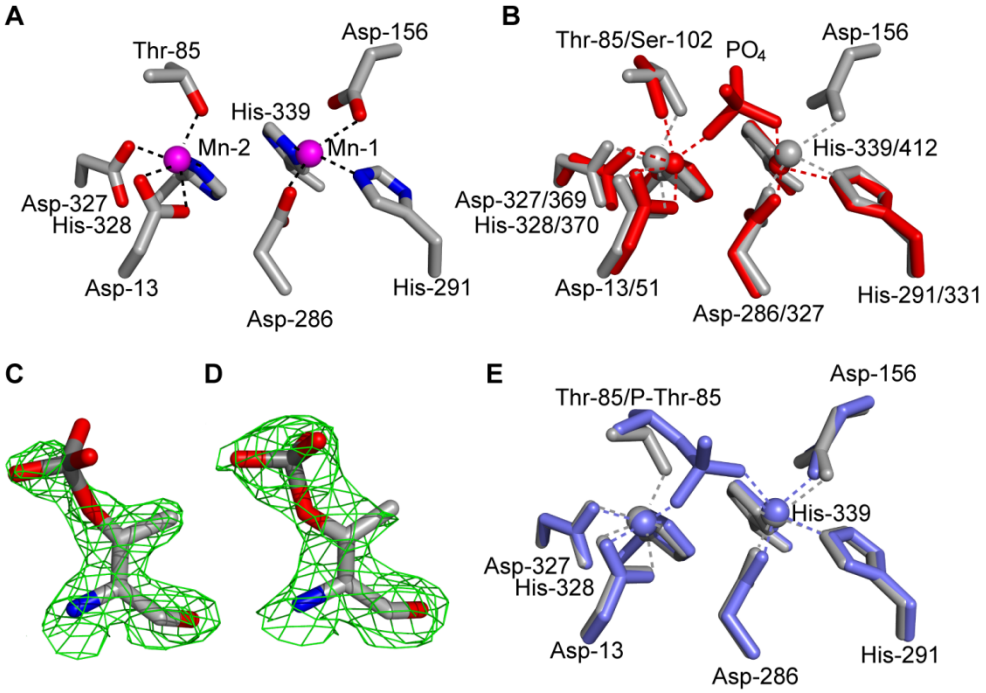

Figure 4-3. Active site architecture. *B. cereus* PPM purifies in a mixed state, with some percentage of the protein phosphorylated and some unphosphorylated. Figure panels depicting unphosphorylated protein use coordinates from the unphosphorylated population. A. Mn^{2+} coordination in unphosphorylated *B. cereus* PPM. Coordinating interactions are shown with dotted lines. B. Comparison of metal coordination in B. cereus PPM (gray) and *E. coli* alkaline phosphatase (red, PDB ID 1ALK). Comparison was performed with a local alignment in PyMol[28]. Residues are labeled with PPM numbering first and alkaline phosphatase numbering second. C & D. Occupancy of the phosphate modification of Thr-85 following incubation of *B. cereus* PPM with glucose 1,6-bisphosphate. $2|F_o|-|F_c|$ simulated annealing omit electron density maps calculated in CNS[25] after the removal of Thr-85 from the structure are contoured to 1.25 σ (green mesh). C. PPM that was not incubated with glucose 1,6-bisphosphate and D. PPM activated with glucose 1,6-bisphosphate. E. The structure of unphosphorylated *B. cereus* PPM (gray) overlaid with the structure of glucose 1,6-bisphosphate activated PPM (blue). A small rotation in the χ-1 angle of Thr-85 allows the phosphate modification to bridge the two Mn^{2+} ions. This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

Soaking pre-formed crystals of *B. cereus* PPM with the substrate, ribose 5-phosphate, resulted in the appearance of new electron density within the active site that was not observed

in either native crystals or crystals grown from protein pre-incubated with the activator (Fig. 4-4A,B). This electron density (Figure 4-4C) is not fully explained by the modeling of a single ribose 5-phosphate molecule. Since substrate reorientation is a likely event during catalysis, multiple binding positions are anticipated. Indeed, the most reasonable explanation of this electron density is that it reflects a mixed state, with ribose 5-phosphate binding in two, mutually-exclusive positions.

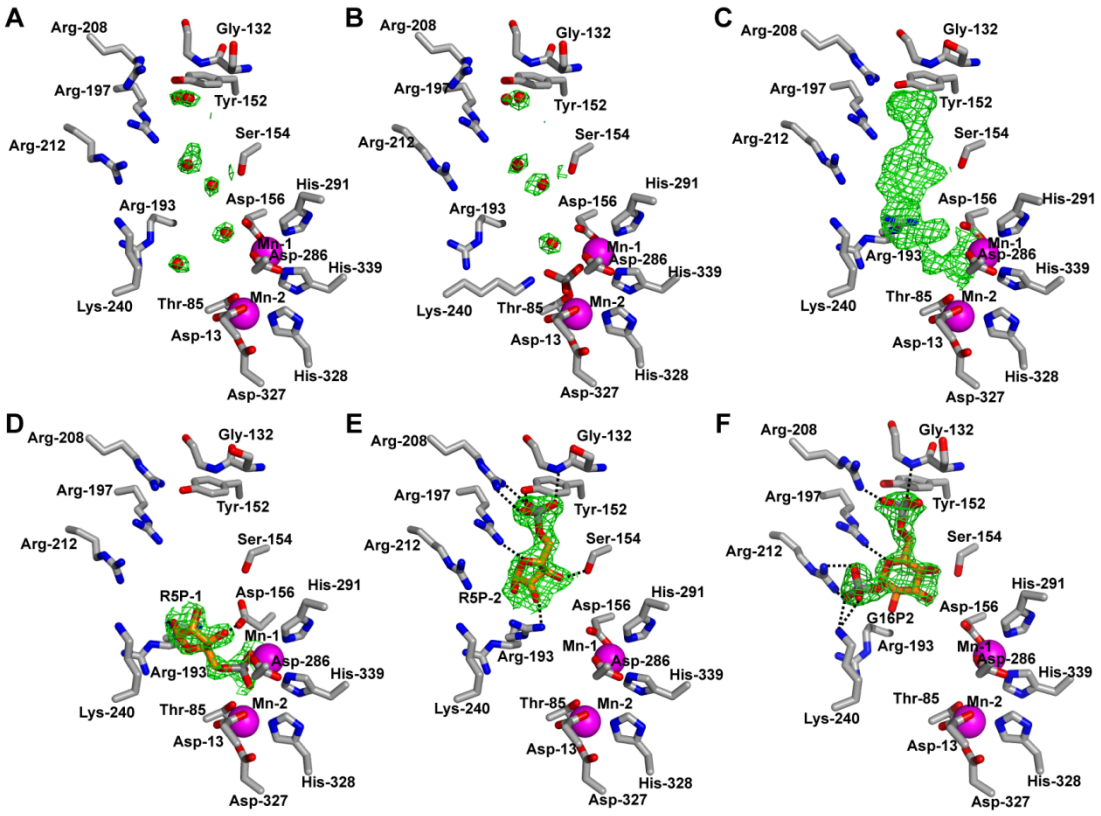

Figure 4-4. Binding sites for substrate and activator in the *B. cereus* PPM active site. In A–E, electron density maps (green mesh) are $2|F_o| - |F_c|$ simulated annealing omit maps contoured at 1.0 σ and calculated in CNS[25] after the removal of non-protein ligands from the active site cavity. A. Active site of purified PPM shows the location of water molecules within the cavity. B. Active site of glucose 1,6-bisphosphate activated PPM shows that the location of ordered water molecules within the active site does not change upon activation. C. Electron density appearing in the active site of crystals of B. cereus PPM soaked with ribose 5-phosphate. D & E. Interpretation of the active site density in ribose 5-phosphate soaked crystals. D. Ribose 5-phosphate (R5P-1; orange sticks) bound in the coordinating position. Only electron density consistent with substrate binding at this position is shown. Putative hydrogen bonds are denoted with dotted lines. E. Ribose 5-phosphate (R5P-2; orange sticks) bound in the distal position.

Only electron density consistent with substrate binding at that position is shown. F. Interpretation of electron density of B. cereus PPM crystals soaked with glucose 1,6-bisphosphate (G16P2; orange sticks). This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

In the first binding position of ribose 5-phosphate (Fig. 4-4D), Thr-85 is unphosphorylated, and the 5-phosphate of the substrate bridges Mn-1 and Mn-2. As a result, this position is termed the coordinating position. In the coordinating position, the substrate is additionally stabilized by two hydrogen bonds between Arg-193 Nη and the 1-OH group of the furanose ring, and between Asp-286 Oγ and the 3-OH of the furanose ring, and by two water-mediated interactions with the protein. A similar binding position has previously been observed in structures of alkaline phosphatase family members crystallized in complex with products.

The second binding position of ribose 5-phosphate has markedly clearer electron density (Figure 4-4E). While substrate bound in this position is oriented with the 1-OH group of the furanose ring toward the phosphate of Thr(P)-85, the distance between Thr(P)-85 and the 1-OH group is 8.5 Å, and no part of the molecule directly interacts with the dimetallo catalytic center. As a result, this position is referenced as the distal position. In the distal position, ribose 5-phosphate is stabilized by seven putative hydrogen-bonds to the cap domain of the protein, and four water-mediated interactions. Four of the hydrogen-bonds and two of the water-mediated interactions are to the 5-phosphate, which binds within a well-defined pocket. The remaining interactions are between the Nη1 and Nη2 atoms of Arg-193 and the 1-OH group and 2-OH group of the furanose ring, between Ser-154 Oγ and the 3-OH of the furanose ring, and two water-mediated interactions to the 1-OH group of the furanose ring. A similar binding position has not previously been observed in structures of alkaline phosphatase superfamily members.

Soaking of pre-formed crystals of *B. cereus* PPM with glucose 1,6-bisphosphate resulted in the appearance of clear electron density within the active site consistent with a single binding site for glucose 1,6-bisphosphate (Figure 4-4F). In this binding position, the pyranose ring and

the 6-phosphate superimpose closely with the furanose ring and 5-phosphate of ribose 5-phosphate in the distal position. The 1-phosphate binds within a second phosphoester binding pocket at the interface of the core and cap domains and includes contacts to the Nζ atom of Lys-240 and the Nη1 and Nη2 atoms of Arg-212. This position of glucose 1,6-bisphosphate is stabilized by seven direct contacts to the protein, only one of which is to the pyranose ring (between Ser-154 Oγ and the 3-OH of the pyranose ring), and three water-mediated contacts. Interestingly, the hydrogen-bonding contacts to the 6-phosphate of glucose 1,6-bisphosphate slightly differ from those to the 5-phosphate of ribose 5-phosphate, with one hydrogen bond shifting from a direct contact to a water-mediated contact. Steric differences between the phosphate orientations on a 5- versus 6-membered sugar ring may underlie this alteration in phosphate binding.

Verification of Thr-85 phosphorylation by Western and mass spectral analysis

Since the phosphorylation of the catalytic nucleophile, Thr-85, was unanticipated, and post-translational modifications cannot be unambiguously identified with even the best crystallographic data, the phosphorylation of Thr-85 was confirmed with both mass spectral and Western analyses. LTQ-orbitrap analysis of *B. cereus* PPM digested with trypsin, elastase, or subtilisin resulted in 82% sequence coverage. Of the peptides analyzed, eight unique peptides contained Thr-85, one of which (residues 80–95, STGKDTMTGHWEIMGL, mass of 1 876.76 Da) deviated from its predicted *m/z* ratio in a manner consistent with phosphorylation. The results of tandem mass spectrometry are shown in Figure 4-5A and confirmed both the presence and location of the site of phosphorylation in this peptide. This spectrum shows not only the loss of an *m/z* of 98 (H_3PO_4) from the parent ion but also from several of the fragment ions (*e.g.* b6, b10, and b12). Localization of the phosphate modification to Thr-85 is evidenced specifically with the b5, b6, y10 and y11 fragments flanking the site of phosphorylation. The identification of a mixture of phosphorylated and unphosphorylated peptides is consistent with

the crystal structure of PPM having only partial occupancy of the phosphate modification (Figure 4-3C).

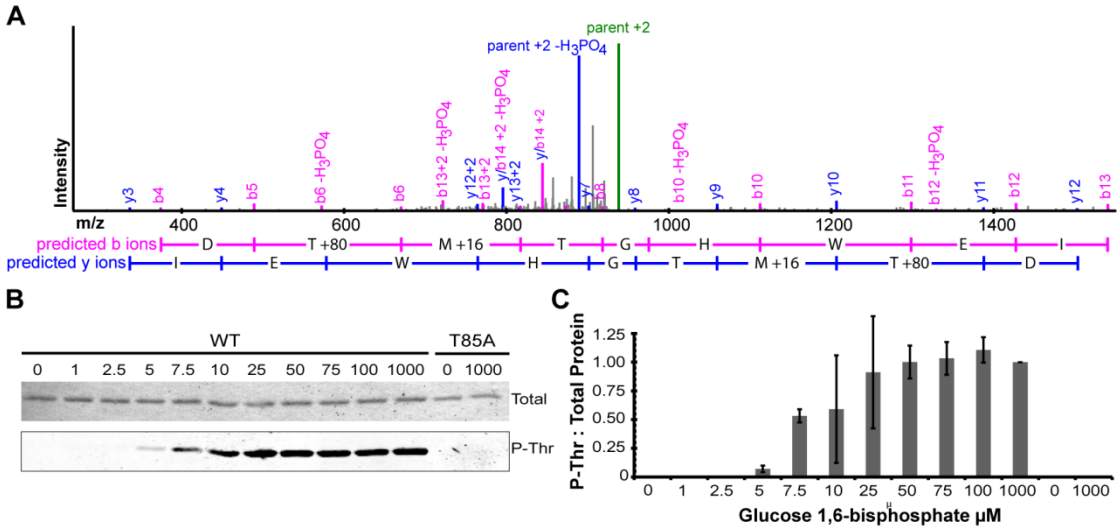

Figure 4-5. Phosphorylation of Thr-85. A. Mass spectrum of the phosphorylated peptide STGKDTMTGHWEIMGL. A peptide with a mass of 1 876.76 Da was isolated from the pool of peptides resulting from elastase digestion of PPM. The spectrum of the doubly charged phosphopeptide parent ion (*m/z* of 938.38) is shown in *green*. The b and y ions for the phosphopeptide are shown in *blue* and *magenta*, respectively. T +80, mass of phosphorylated threonine residue; M +16, mass of an oxidized methionine residue. B. Western analysis of phosphorylation of *B. cereus* PPM following incubation with glucose 1,6-bisphosphate. The upper image of Ponceau stained nitrocellulose verifies that equivalent amount of total protein was loaded in each lane. The lower image shows the Odyssey image of the same nitrocellulose membrane following incubation with a phosphothreonine-specific primary antibody and an Alexafluor-labeled secondary antibody. The concentration of glucose 1,6-bisphosphate (in µM) is indicated above each lane. C. Quantitation of data shown in panel B. Band intensities were calculated with the program ImageJ[32] and were normalized to the intensity of the band at 1 000 µM glucose-1,6-bisphosphate in each image. The values are the average of ratios from three independent experiments. This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

Western analysis was used to further verify both the phosphorylation of Thr-85 and an increase in phosphorylation upon the addition of glucose 1,6-bisphosphate. Both wild-type and T85A PPM were analyzed using a phosphothreonine-specific antibody after incubation with increasing amounts of glucose 1,6-bisphosphophosphate (Figure 4-5B,C). This antibody only recognized the wild-type enzyme, consistent with Thr-85 being the only site of phosphorylation.

The phosphorylation of the wild-type enzyme increased upon the addition of the activator, glucose 1,6-bisphosphate, consistent with the observation that pre-incubation of PPM with glucose 1,6-bisphosphate appeared to increase the percentage of protein with phosphorylated Thr-85 in the crystal structure (Figure 4-3D).

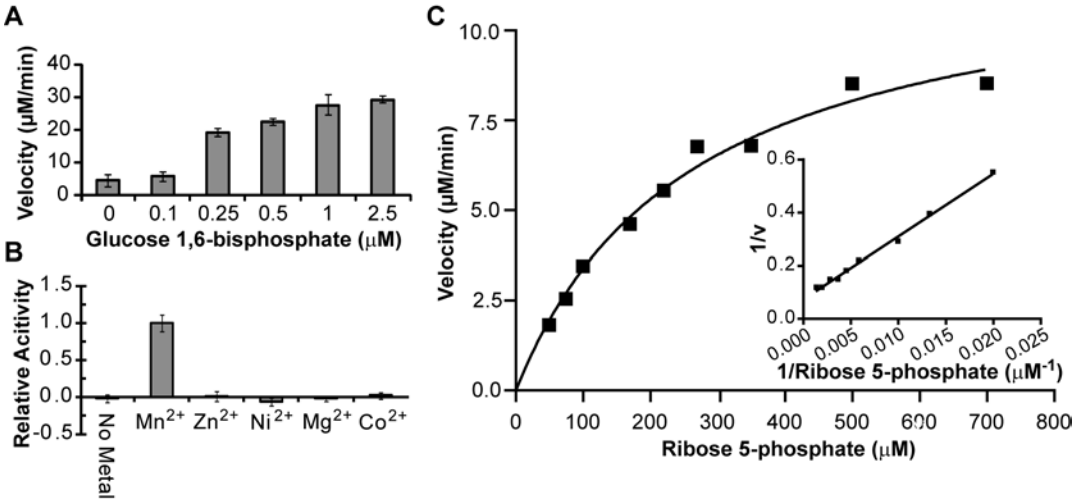

Figure 4-6. Enzymatic activity of PPM. Panel A illustrates the effect of glucose 1,6-bisphosphate concentration on the activity of *B. cereus* PPM. The calculated velocity is the average of three independent experiments. The activity of EDTA-chelated *B. cereus* PPM in the presence of 1 mM cations is shown in panel B. Activity is the average of three independent measurements and error bars represent the standard deviation of the measurement. Activity was normalized to the activity of *B. cereus* PPM measured after addition of 1 mM $MnCl_2$. In panel C, initial velocities of PPM are plotted as a function of ribose 5-phosphate concentration. Data are the average of three assays, and a double reciprocal plot of the data is shown in the inset. The calculated K_m, V_{max}, and k_{cat} are 263 ± 34 μM, 12.3 ± 0.7 μM min^{-1} and 10.25 ± 0.6 s^{-1}. This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

Activity of PPM

The crystallographic, Western, and mass spectrometric analyses demonstrated that the addition of the glucose 1,6-bisphosphate increased the phosphorylation of Thr-85 in wild-type enzyme. To assess the effect of increased phosphorylation on activity, *B. cereus* PPM was pre-incubated with increasing concentrations of glucose 1,6-bisphosphate and assessed the catalytic turnover (Figure 4-6A). A concentration-dependent enhancement of PPM activity was

observed between 0.1 µM and 2.5 µM glucose 1,6-bisphosphate, at which point maximal activity was reached. The metal-dependence of *B. cereus* PPM was confirmed by measuring the activity of chelated protein reconstituted with Mn^{2+}, Zn^{2+}, Ni^{2+}, Mg^{2+}, and Co^{2+}. Only the Mn^{2+} incorporated enzyme showed significant enzyme activity (Figure 4-6B). In these optimized conditions, the V_{max}, k_{cat}, and K_m were determined to be 12.3 ± 0.7 µM min^{-1}, 10.2 ± 0.6 s^{-1} and 263 ± 34 µM, respectively (Figure 4-6C).

Having established basic kinetic parameters for the conversion of ribose 5-phosphate to ribose 1-phosphate, the presence of intrinsic phosphatase activity, in which the phosphoryl group from Thr-85 is hydrolyzed, was determined. While statistically significant phosphatase activity was measured, the rate of hydrolysis is near the detection limit of the assay and appeared to be ~10 000-fold lower than the rate of ribose 1-phosphate and ribose 5-phosphate interconversion (data not shown).

Reaction of PPM with isotopically labeled substrate

Phosphomutases that are unphosphorylated when active commonly catalyze *intra*molecular phosphoryl transfer (Figure 4-7A), while those that are phosphorylated commonly catalyze *inter*molecular phosphoryl transfer (Figure 4-7B) with respect to substrate[41]. Accordingly, the observed correlation between enzyme phosphorylation and activity in PPM suggests that it catalyzes *inter*molecular phosphoryl transfer; however, this is counter to what is anticipated from the alkaline phosphatase general mechanism. In order to unambiguously demonstrate *inter*molecular transfer, an isotope relay assay was performed. In this assay, a 1:1 stoichiometric mixture of ribose 5-[$^{18}O_3$]phosphate and [$^{13}C_5$]ribose 5-phosphate was incubated with PPM under standard reaction conditions (Figure 4-7C). These substrates are labeled with heavy-atom isotopes in the phosphoryl group (+6 Da) and the ribose ring (+5 Da), respectively, rendering them, and all possible reaction products, distinguishable by mass spectrometric analysis. In the event of *intra*molecular transfer (mutase mechanism) the isotopic distribution of

the molecular ions for the two compounds should remain unchanged. Conversely, in the event of *inter*molecular transfer (transferase mechanism), in which the transferred phosphoryl group is derived from an antecedent substrate, the isotopic distribution for the molecular ions should scramble into four species of unique masses. It is unambiguous from these results (Figure 4-7D) that *inter*molecular phosphoryl group transfer is occurring under the PPM reaction conditions used in this study, with four isotopologues represented at roughly equal total ion current intensities. Moreover, when human purine nucleoside phosphorylase and hypoxanthine were added to this reaction, consumption of the ribose phosphates was apparent. Notably, newly formed isotopologues were consumed more rapidly than isotopologue masses corresponding to precursors, suggesting that the new isotopologues corresponded to predominantly ribose 1-phosphate.

To support the isotope relay results observed in the preceding study, an identical experiment was performed with a 1:1 mixture of [$^{13}C_5$]ribose 5-[$^{18}O_3$]phosphate and ribose 5-phoshate. As in the prior case, the isotopologue distribution of this mixture should remain unchanged in the event of an *intra*molecular reaction, whereas an *inter*molecular reaction should engender isotope relay of [$^{18}O_3$]phosphate to unlabeled ribose and relay of unlabeled phosphate to [$^{13}C_5$]ribose (Figure 4-7E). Again, the incubation of substrates with PPM led to the generation of isotopologues in equal ratios and the addition of human nucleoside phosphorylase to reactions resulted in the disappearance of these newly formed isotopologues (Figure 4-7F).

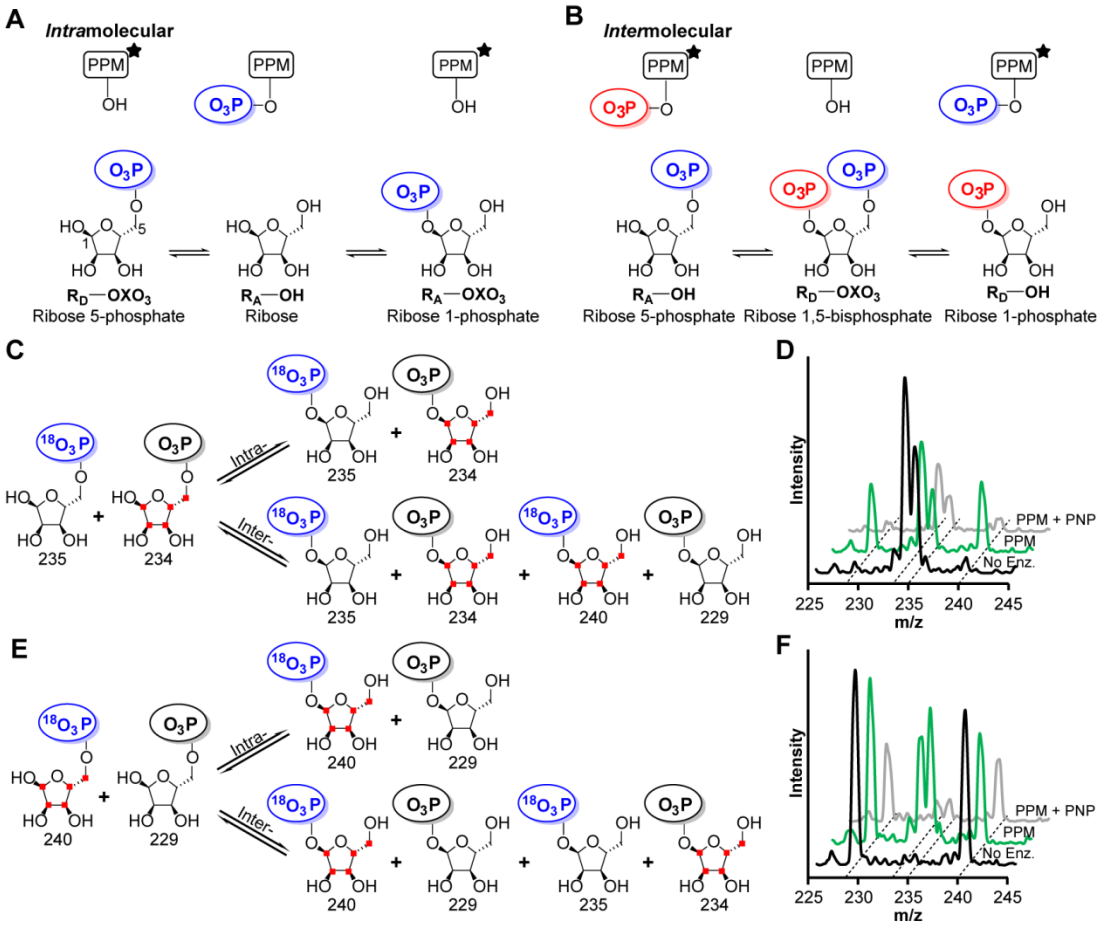

Figure 4-7. Isotope relay assays. A,B. Comparison of phosphoryl transfer scenarios. A schematic of *intra*molecular phosphoryl transfer is shown in panel A. In *intra*molecular transfer, the active form of the enzyme is dephosphorylated (filled star), corresponding to Figure 4-1, State 1. Panel B is a schematic of *inter*molecular phosphoryl transfer. In *inter*molecular transfer, the active form of the enzyme is phosphorylated (filled star), corresponding to Figure 4-1, State 4. C-F. Mixtures of ribose 5-phosphate synthesized from [U-$^{13}C_5$]ribose (rings with red squares) or ribose and [γ-$^{18}O_4$]ATP (blue) or ATP were incubated with PPM for 30 min and the masses of the products were measured by LC-MS. C. Theoretical masses for starting materials and the predicted products of either *intra*molecular or *inter*molecular phosphoryl transfer. D. Mass spectra of the experiment outlined in Panel C. In the absence of enzyme (black line) only the *m/z* 234 and 235 peaks corresponding to the starting materials are observed. Following incubation with PPM, peaks at *m/z* 229 and 240 appear (green line). The addition of human purine nucleoside phosphorylase to this reaction, which removes the product ribose 1-phosphate, decreases the intensity of these new peaks (grey line). E. Theoretical masses for starting materials and the predicted products of either *intra*molecular or *inter*molecular phosphoryl transfer. F. Mass spectra of the experiment outlined in E. In the absence of enzyme, only the *m/z* 229 and 240 peaks corresponding to the starting materials are present (black line). Incubation of the starting materials with PPM resulted in the appearance of peaks at *m/z* 234 and 235 (green line). These new peaks decreased in intensity following incubation with human purine nucleoside phosphorylase (grey line), which removes ribose 1-phosphate. This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

To determine if ribose could be a discrete intermediate or reversible shunt product in the PPM catalyzed reaction, the PPM reaction was performed as described above, with the addition of 1 mM [U-$^{13}C_5$]ribose. In these reactions, the initial isotopologue distribution of the reaction mixture remained unchanged (data not shown), ruling out ribose as a diffusible intermediate.

Discussion

Implications for the catalytic mechanism

Extrapolation of the general alkaline phosphatase reaction mechanism to PPM predicts that the enzyme should be active when unphosphorylated (Figure 4-1, State 1), and that catalysis should proceed via an *intra*molecular mutase mechanism. In such a mechanism, the reaction would proceed through an intermediate in which the enzyme is phosphorylated, but the substrate is dephosphorylated (Figure 4-1, State 4). Indeed, this is exactly what was observed for cofactor independent phosphoglycerate mutase[9, 10], which is the only other characterized phosphomutase in the alkaline phosphatase family.

Surprisingly, these results are inconsistent with this precedent. It was demonstrated that glucose 1,6-bisphosphate both increased enzyme activity (Figure 4-6A) and resulted in increased phosphorylation of Thr-85 (Figures 4-3C&D, 4-5B&C), suggesting that glucose 1,6-bisphosphate activates PPM by phosphorylating the catalytic nucleophile. These data are therefore consistent with substrates being acted upon by the phosphorylated enzyme, and generally proscribe an *intra*molecular sequence analogous to phosphoglycerate mutase.

To unambiguously demonstrate PPM proceeds via *intra*molecular or *inter*molecular transfer, a series of isotope relay experiments was performed. Under all assayed reaction conditions, the phosphoryl group of the product is indeed derived from an antecedent substrate (Figure 4-7D,F). Notably, analogous isotopic labeling experiments performed for

phosphoglycerate mutase of the alkaline phosphatase superfamily revealed no relay of labeled phosphoryl groups (or glyceryl groups), demonstrating the opposite result[9, 10].

Having ruled out a purely *intra*molecular process, there are several general scenarios that may result in the *inter*molecular transfer of phosphoryl groups (a full range of scenarios has been outlined by Britton and coworkers for phosphoglycerate mutase[41]). As shown in Figure 4-7B, one limiting case assumes a primed (phosphorylated) enzyme. In this scenario, the phosphoenzyme would transfer its phosphoryl group to the 1-position of ribose 5-phosphate, resulting in intermediary ribose 1,5-bisphosphate. A phosphoryl group would be subsequently transferred from the 5-position of this intermediate to the enzyme nucleophile, concomitantly generating the product ribose 1-phosphate, and priming the enzyme for the next reaction cycle.

Although there are alternatives to this reaction mechanism, a dephosphorylated starting enzyme must be invoked, which is inconsistent with the data outlined here. Briefly, to catalyze *inter*molecular transfer with a dephosphorylated enzyme, the dephosphorylation of ribose 5-phosphate by PPM must be followed by diffusion of the ribose intermediate at a rate competitive with phosphoryl transfer over the course of the assay. Despite the evidence that the active enzyme is phosphorylated, the possibility of 'leaky' mutase activity was assayed by performing the PPM reaction with labeled and unlabeled [$^{18}O_3$]ribose 5-phosphates with the addition of a molar excess of [$^{13}C_5$]ribose. No transfer of labeled or unlabeled phosphoryl groups to [$^{13}C_5$]ribose was observed under these conditions, thereby eliminating the possibility of a diffusible ribose intermediate in the context of a leaky mutase mechanism. Given the correlation of increased enzyme activity with phosphorylation, and the observation of *inter*molecular phosphoryl transfer between substrates, the most parsimonious interpretation supports the mechanism depicted in Figure 4-7B.

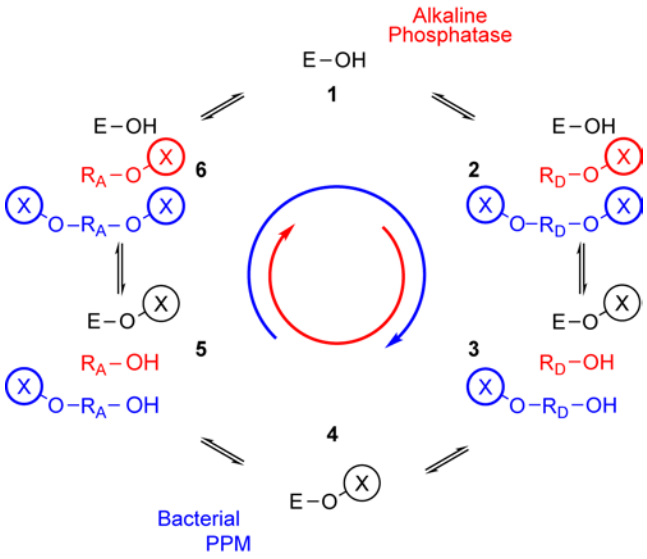

Figure 4-8. Modification of the alkaline phosphatase mechanism for *inter*molecular phosphoryl transfer. Alkaline phosphatase follows the reaction cycle from State 1 through State 6 (red path, red substrates), starting with unphosphorylated enzyme (State 1) and phosphorylated substrate (State 2), and proceeding through a transient intermediate where the enzyme is phosphorylated (State 4) and the substrate is dephosphorylated (States 3,4,5) before generating a phosphorylated product (State 6). The *inter*molecular transfer catalyzed by PPM can occur within the context of this reaction mechanism (blue path, blue substrates) if phosphorylated enzyme (State 4) acts on phosphorylated substrate (State 5). PPM could then follow a similar path through the same catalytic cycle by proceeding through an intermediate where the enzyme is dephosphorylated (State 1) and the substrate is doubly phosphorylated (States 6, 1, 2). To complete the reaction cycle, the bisphosphate intermediate could transfer its phosphoryl group back to the enzyme, resulting in a singly phosphorylated product (State 3) and a phosphorylated, active enzyme (State 4). This image copyright of the Journal of Biological Chemistry. Reproduced with permission. Originally published by Panosian, *et al*.[1]

To reconcile this *inter*molecular transferase mechanism with the general alkaline phosphatase mechanism (Figure 4-1,8), it is possbile that bacterial PPMs may have adapted to favor substrate entry in the catalytic cycle at an alternate point. In this scenario, phosphorylated PPM (Figure 4-8, State 4) binds the substrate ribose 5-phosphate (Figure 4-8, State 5), then transfers a phosphoryl group from Thr(P)-85 to ribose 5-phosphate, forming a ribose 1,5-bisphosphate intermediate and a dephosphorylated enzyme (Figure 4-8, State 6). The reaction cycle is completed with transfer of the 5-phosphate of ribose 1,5-bisphosphate to Thr-85 (Figure 4-8, States 2,3), priming the enzyme for the next turnover cycle (Figure 4-8, State 4).

The distinct entry point into the catalytic cycle requires that PPM differ from characterized enzymes within the alkaline phosphatase superfamily with respect to the stability of the phosphorylated catalytic nucleophile. In alkaline phosphatase, the nucleophile, Ser-102, is only transiently phosphorylated during the catalytic cycle, and efforts to stabilize this reaction intermediate proved challenging. The eventual crystal structure of phosphorylated alkaline phosphatase required both that the Zn^{2+} ions of the dimetallo center be replaced with catalytically inactive Cd^{2+} and that Ser-102 be mutated to threonine[40]. In peptides, phosphoserine has been shown to be more labile than phosphothreonine under alkaline conditions[42]. This cursorily implies that the use of a threonine as the active site nucleophile in PPM contributes to the improved longevity of the phosphoenzyme. However, the folded protein has a profound influence on the stability of phosphate modifications, and phosphoserines in phosphoryl transfer enzymes have measured $t_{1/2}$ values on the order of years[43]. While the origins of the improved stability of the phosphorylated nucleophile in PPM remain unclear, the substrate nevertheless likely encounters phosphorylated enzyme under physiological conditions.

Comparison of the structures of the B. cereus and S. mutans PPMs

The structure of *B. cereus* PPM was determined by molecular replacement using the unpublished coordinates for a putative PPM from *S. mutans* (PDB ID 2I09; New York SGX Research Center for Structural Genomics, unpublished) as a search model. While the backbone accuracy was sufficient for use as a molecular replacement search model, the *S. mutans* structure contains numerous errors. For example, four substantial errors at the active site would prevent the development of a structure-based reaction mechanism using the *S. mutans* structure as a guide: first, a short frame-shift (residues 218–220) places incorrect residues at the active site; second, no metals are included; third, the side chains of three active site residues including the catalytic threonine are mis-oriented; fourth, the post-translational

phosphorylation of the active site threonine is modeled as a water molecule. It is possible to verify that these are model building errors and not differences between PPM from these two organisms since both the coordinates and the structure factors were deposited for the *S. mutans* PPM. After the submission of this manuscript, the coordinates for the *S. mutans* PPM were superceded with a newly refined model (PDB ID 3MV7). While this included metal ions and corrected some of the errors, incorrectly positioned side chains in the active site would still preclude the development of a structure-based mechanism from these updated coordinates. A careful procedure of omit mapping was used to prevent these errors from propagating into the present study.

Comparison to phosphomutases in the α-phosphohexomutase and haloacid dehalogenase superfamilies

Phosphomutases have long been grouped into cofactor-independent enzymes that catalyze *intra*molecular transfer, and bisphosphate dependent enzymes that catalyze *inter*molecular transfer. The *inter*molecular transferase mechanism of PPM is shared with several well-studied cofactor-dependent phosphomutases within the α-phosphohexomutase and haloacid dehalogenase superfamilies[44-47]. Perhaps unsurprisingly, numerous parallels exist between PPM and phosphomutase enzymes within both of these superfamilies. For example, PPM and the phosphomutases within the α-phosphohexomutase and haloacid dehalogenase superfamilies are multi-domain enzymes with an active site located in a positively-charged cleft between domains[45, 48-50]. Activity is dependent upon divalent cations, the reactions proceed through bisphosphate intermediates[44, 51], and the enzymes require the continued presence of bisphosphate activators[46, 47] to counter the slow inactivation that occurs when the reaction intermediate prematurely disassociates from the enzyme, or when the phosphoryl group is transferred to water[2]. Taken together, these broad mechanistic parallels suggest that the details

of catalysis identified for phosphomutases within the α-phosphohexomutase and haloacid dehalogenase superfamilies may guide future studies on PPM.

Conclusions

The results presented here include the first published crystal structure and structure-based functional study of a prokaryotic PPM. Although prokaryotic PPMs belong to the alkaline phosphatase superfamily, the data presented here provide evidence that the catalytic cycles of these enzymes are distinct. This study serves as a reminder that caution should be used when extrapolating reaction mechanisms, even in closely related systems.

Acknowledgements

This research was originally published in The Journal of Biological Chemistry. Panosian, T. D.; Nannemann, D. P.; Watkins, G. R.; Phelan, V. V.; McDonald, W. H.; Wadzinski, B. E.; Bachmann, B. O.; Iverson, T. M. *Bacillus cereus* phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle. *Journal of Biological Chemistry.* 2011; 286:8043-8054. © the American Society for Biochemistry and Molecular Biology. TDP conducted all crystallography studies and the measurement of PPM phosphorolysis rate. DPN designed and conducted biochemical experiments for cofactor dependence and concentration response, substrate kinetic studies and isotope relay assays. TDP and DPN prepared samples for Western blotting. GRW performed the Western blots. contributed in the analysis of all data. VVP assisted in the design and execution of isotope relay experiments. WHM conducted the proteomics analysis of PPM. BEW advised GRW and assisted in the design of Western blot experiments. TDP, DPN, BOB, and TMI contributed to analysis of all data and wrote the manuscript.

References

1. Panosian, T. D.; Nannemann, D. P.; Watkins, G. R.; Phelan, V. V.; McDonald, W. H.; Wadzinski, B. E.; Bachmann, B. O.; 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, (10), 8043-54.

2. Knowles, J. R., Enzyme-catalyzed phosphoryl transfer-reactions. *Annu. Rev. Biochem.* **1980,** 49, 877-919.

3. Allen, K. N.; Dunaway-Mariano, D., Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **2004,** 29, (9), 495-503.

4. Tozzi, M. G.; Camici, M.; Mascia, L.; Sgarrella, F.; Ipata, P. L., Pentose phosphates in nucleoside interconversion and catabolism. *FEBS J.* **2006,** 273, (6), 1089-1101.

5. Horzempa, J.; Carlson, P. E.; O'Dee, D. M.; Shanks, R. M. Q.; Nau, G. J., Global transcriptional response to mammalian temperature provides new insight into *Francisella tularensis* pathogenesis. *BMC Microbiol.* **2008,** 8, 172.

6. Hammer-Jespersen, K.; Munch-Petersen, A., Phosphodeoxyribomutase from Escherichia coli: Purification and some properties. *Eur. J. Biochem.* **1970,** 17, 397-407.

7. Maliekal, P.; Sokolova, T.; Vertommen, D.; Veiga-Da-Cunha, M.; Van Schaftingen, E., Molecular identification of mammalian phosphopentomutase and glucose 1,6-bisphosphate synthase, two members of the α-D-phosphohexomutase family. *J. Biol. Chem.* **2007,** 282, (44), 31844-31851.

8. Galperin, M. Y.; Bairoch, A.; Koonin, E. V., A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Sci.* **1998,** 7, (8), 1829-1835.

9. Breathnach, R.; Knowles, J. R., Phosphoglycerate mutase from wheat-germ - studies with labeled-O-18 substrate, investigations of phosphatase and phosphoryl transfer activities, and evidence for a phosphoryl-enzyme intermediate. *Biochemistry* **1977,** 16, (14), 3054-3060.

10. Gatehouse, J. A.; 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, (14), 3045-53.

11. Galperin, M. Y.; Jedrzejas, M. J., Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes. *Proteins: Struct., Funct., Genet.* **2001,** 45, (4), 318-324.

12. Coleman, J. E., Structure and mechanism of alkaline-phosphatase. *Annu. Rev. Biophys. Biomol. Struct.* **1992,** 21, 441-483.

13. Kim, E. E.; Wyckoff, H. W., Reaction mechanism of alkaline phosphatase based on crystal structures - Two-metal ion catalysis. *J. Mol. Biol.* **1991,** 218, (2), 449-464.

14. Ball, E. G., Xanthine oxidase: purification and properties. *J. Biol. Chem.* **1939,** 128, (51), 51-67.

15. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010,** 23, (8), 607-616.

16. Panosian, T. D.; Nannemann, D. P.; Bachmann, B. O.; 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, 811-4.

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

18. Collaborative Computational Project Number, The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994,** 50, (5), 760-763.

19. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** 40, 658-674.

20. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004,** 60, 2126-2132.

21. Brunger, A. T., Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2007,** 2, (11), 2728-2733.

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

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

24. Painter, J.; Merritt, E. A., TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **2006,** 39, 109-111.

25. Brunger, A. T., Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2007,** 2, (11), 2728-2733.

26. Schuttelkopf, A. W.; van Aalten, D. M. F., PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004,** 60, 1355-1363.

27. Kabsch, W., Solution for best rotation to relate 2 sets of vectors. *Acta Cryst.* **1976,** A32, (SEP1), 922-923.

28. Delano, W. L. *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA, USA.: 2002.

29. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** 32, W665-W667.

30. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM - a program for macromolecular energy, minimization, and dynamics. *J. Comput. Chem.* **1983,** 4, (2), 187-217.

31. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, (18), 10037-10041.

32. Rasband, W. S. *ImageJ*, U.S. National Institutes of Health: Bethesda, Maryland, 2009.

33. MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** 99, (12), 7900-7905.

34. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994,** 5, (11), 976-989.

35. Ma, Z. Q.; Dasari, S.; Chambers, M. C.; Litton, M. D.; Sobecki, S. M.; Zimmerman, L. J.; Halvey, P. J.; Schilling, B.; Drake, P. M.; Gibson, B. W.; Tabb, D. L., IDPicker 2.0: improved protein assembly with high discrimination peptide identification filtering. *J. Proteome Res.* **2009,** 8, (8), 3872-3881.

36. Degroot, H.; Degroot, H.; 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, (1), 255-260.

37. Segel, I. H., *Biochemical Calculations: How to Solve Mathematical Problems in General Biochemistry, 2nd Edition*. Wiley: New York, 1976.

38. Scism, R. A.; Bachmann, B. O., Five-component cascade synthesis of nucleotide analogues in an engineered self-immobilized enzyme aggregate. *ChemBioChem* **2010,** 11, (1), 67-70.

39. Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A., Searching protein structure databases with DaliLite v.3. *Bioinformatics* **2008,** 24, (23), 2780-2781.

40. Wang, J.; Kantrowitz, E. R., Trapping the tetrahedral intermediate in the alkaline phosphatase reaction by substitution of the active site serine with threonine. *Protein Sci.* **2006,** 15, (10), 2395-2401.

41. Britton, H. G.; Carreras, J.; Grisolia, S., Mechanism of action of 2,3-diphosphoglycerate-independent phosphoglycerate mutase. *Biochemistry* **1971,** 10, (24), 4522-4533.

42. Kemp, B. E., Relative alkali stability of some peptide ortho-phosphoserine and ortho-phosphothreonine esters. *FEBS Lett.* **1980,** 110, (2), 308-312.

43. Ray, W. J. J.; Long, J. W., Thermodynamic and structural differences among catalytically active complexes of phosphoglucomutase - metal-ion effects. *Biochemistry* **1976,** 15, (18), 4018-4025.

44. Britton, H. G.; Clarke, J. B., Mechanism of phosphoglucomutase reaction - studies on rabbit muscle phosphoglucomutase with flux techniques. *Biochem. J.* **1968,** 110, (2), 161-179.

45. Lahiri, S. D.; Zhang, G. F.; Dunaway-Mariano, D.; Allen, K. N., Caught in the act: the structure of phosphorylated β-phosphoglucomutase from *Lactococcus lactis*. *Biochemistry* **2002,** 41, (26), 8351-8359.

46. Qian, N.; Stanley, G. A.; Hahnhagerdal, B.; Radstrom, P., Purification and characterization of two phosphoglucomutases from *Lactococcus lactis* subs. *lactis* and their regulation in maltose- and glucose-utilizing cells. *J. Bacteriol.* **1994,** 176, (17), 5304-5311.

47. Ray, W. J., Jr.; Roscelli, G. A., A kinetic study of the phosphoglucomutase pathway. *J. Biol. Chem.* **1964,** 239, (4), 1228-1236.

48. Dai, J. B.; Ray, W. J.; Konno, M., The crystal structure of muscle phosphoglucomutase refined at 2.7-angstrom resolution. *J. Biol. Chem.* **1992,** 267, (9), 6322-6337.

49. Regni, C.; Schramm, A. M.; Beamer, L. J., The reaction of phosphohexomutase from *Pseudomonas aeruginosa* - structural insights into a simple processive enzyme. *J. Biol. Chem.* **2006,** 281, (22), 15564-15571.

50. Silvaggi, N. R.; Zhang, C. C.; Lu, Z. B.; Dai, J. Y.; Dunaway-Mariano, D.; Allen, K. N., The X-ray crystal structures of human α-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a. *J. Biol. Chem.* **2006,** 281, (21), 14918-14926.

51. Dai, J. Y.; Wang, L. B.; Allen, K. N.; Radstrom, P.; Dunaway-Mariano, D., Conformational cycling in β-phosphoglucomutase catalysis: reorientation of the β-D-glucose 1,6-bisphosphate intermediate. *Biochemistry* **2006,** 45, (25), 7818-7824.

Chapter V

BIOCATALYTIC SYNTHESIS OF 2,3-DIDEOXYINOSINE THROUGH A DESIGNED, NON-NATURAL, *DE NOVO* PATHWAY

Introduction

the combination of synthetic biology and metabolic engineering protocols to form microbial factories has recently found success in biocatalytic production of high value compounds such as drugs and biofuels[1, 2]. For example, a precursor for the anti-malarial drug artemisinin has been produced in *E. coli*[3] and yeast[4] through co-expression of the non-mevalonate isopentenyl pyrophosphate biocatalytic pathway and enzymes from the native producer, *Artemisia annua*. Protein scaffolds were utilized to co-localize enzymes of the pathway and decrease loss of intermediates to diffusion[5]. Branched-chain alcohols, which are useful in diesel and jet fuels, have been generated through shunting of the amino acid biosynthesis pathways[6]. A *de novo* designed pathway for glucaric acid was introduced into *E. coli* which decreases the number of enzymatic steps from ten to five[7]. A second *de novo* pathway generates D- and L-1,2,4-butanetriol, a non-natural precursor to high-energy compounds, enantioselectively from D-arabinose or L-lyxose, respectively[8]. These pathways utilize enzymes from various and disparate pathways to generate noninnate high-value compounds in excellent yields.

Given the structural similarity of nucleoside analogs, dideoxynucleosides in particular, to natural products it is hypothesized that a biosynthetic pathway for their production can be generated by judicious selection of enzymes and/or enzyme engineering and pathway optimization. Nucleoside analogs have long been recognized for their utility in the treatment of viral infections and cancer. Furthermore, nucleoside analogs are utilized in DNA sequencing methodologies and other biotechnological techniques. As such a critical class of compounds the

synthesis and biological activity of nucleoside analogs is a major area of study. Chemical synthesis of nucleoside analogs, particularly ddI, is achieved from inosine and 2´-deoxyinosine through Barton deoxygenation[9, 10], Corey-Winters reduction[11, 12], the Mattocks reaction[13], and Eastwood olefination[14]. Alternatively, addition of nucleobase to synthesized unnatural sugars has proven useful in the synthesis of ddI[15] and can also be applied to the synthesis of a number of nucleoside analogs with sugar variations. A combination of expensive precursors, low yield steps, generation of regioisomeric and anomeric products which require chromatographic separation and large hazardous waste streams drive the price of manufactory. In fact, 55-99% of nucleoside analog drug prices can be attributed to the active pharmaceutical ingredient[16]. Thus, even modest improvements in process yields, reduction in waste stream or a switch to a cheaper starting material will likely facilitate broader distribution of dideoxyinosine.

Recently research has expanded into enzymatic synthesis of nucleosides and subsequently extended to generation of nucleoside analogs. This strategy utilizes the stereo- and regiochemical control of enzymes to set the orientation of the anomeric carbon along with the identity of the C-N bond and provides an advantage over chemical synthetic methods through application at aqueous process conditions at standard temperatures. Ribosyl transferases or pyrimidine and purine nucleoside phosphorylases (PNPs) have long been applied as biocatalysts to transfer unnatural sugars from a nucleoside analog with a facile synthetic route to form a second analog which is challenging to synthesize directly[17-19]. Chemical synthesis of the instable sugar-1-phosphate intermediate is avoided through *in situ* generation. Whole cells are frequently utilized to carry out this process, providing a cheap source of enzyme and nucleobase and stable environment for enzymatic function[20]. This strategy, however, still requires the chemical synthesis of a nucleoside analog precursor.

2-Deoxyribonucleosides have been formed through the condensation of D-glyceraldehyde-3-phosphate and acetaldeyde by 2-deoxyribose 5-phosphate aldolase to form 2-deoxyribose 5-phosphate[21, 22]. This product is enzymatically isomerized by

phosphopentomutase (PPM) to 2-deoxyribose 1-phosphate and incubated with nucleoside phosphorylases and an aromatic base to form the final product.

Nucleosides and nucleotides have also been formed from ribose, ATP and free base. Gross and Whitesides made nucleotides, specifically uridine monophosphate, by first generating PRPP using the activity of ribokinase and PRPPS. An ATP regeneration system and irreversible, CO_2-producing final step helped drive the reaction toward nucleotide product[23]. A similar protocol by Scism *et al* utilized a mutant HPRT to generate nucleotides with unnatural bases[24]. Chuvikovsky *et al* generated arabino, xylulo and lyxonucleosides with *E. coli* ribokinase. Thymine, ribothymidine, uridine, 2´-deoxyuridine, inosine and 2´-deoxyinosine were generated through reaction of sugar, ATP and nucleobase through tandem catalysis by ribokinase, PPM, and PNP[25].

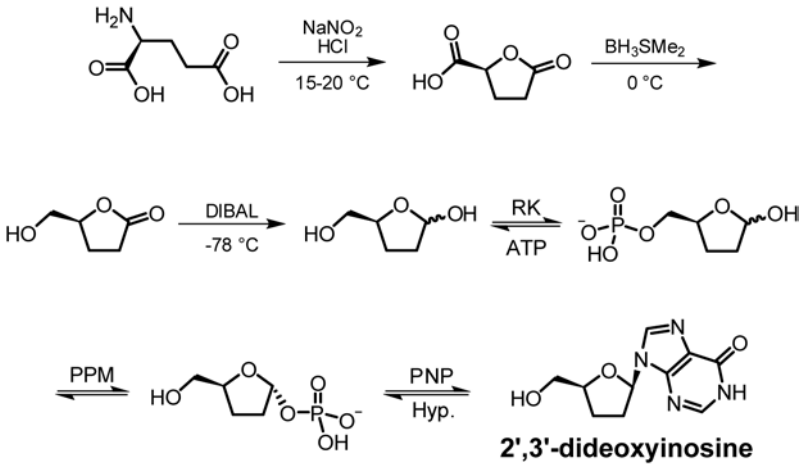

Figure 5-1. Chemo-enzymatic synthesis of 2´,3´-dideoxyinosine from glutamic acid proceeds chemically through 2´,3´-dideoxyribose and utilizes ribokinase (RK), *B. cereus* phosphopentomutase (PPM) and human purine nucleoside phosphorylase (PNP) to add hypoxanthine (Hyp.) to the unnatural sugar.

Here, the generation of a prototype pathway is described for the biocatalytic synthesis of dideoxyinosine from dideoxyribose, which in turn can be chemically synthesized from glutamic acid, resulting in a complete semi-synthetic route (Figure 5-1). The pathway brings together *E. coli* ribokinase, *Bacillus cereus* PPM and human PNP. Ribokinase functions as a component of

DNA, RNA and critical cofactors while also acting as an energy source[26]. PNP and phosphopentomutase act in nucleoside salvage enabling efficient recycling of nucleosides and allowing for catabolism of nucleosides in resource limited conditions[27, 28]. This pathway alleviates the requirement for natural nucleoside precursors, which can cost several hundred dollars per kilogram, or synthesis of donator analogs for transferases and provides an entirely regio- and enantioselective route for the synthesis of nucleoside analogs.

Low pathway throughput, despite extensive incubation time, is indicative of poor substrate turnover by enzymes in the pathway. Improvement of each enzyme for turnover of the unnatural substrate will likely increase the rate of ddI synthesis. With this in mind, PNP-46D6, a mutant improved in phosphorolysis of ddI, was tested for an ability to synthesize ddI and compared to wild type human PNP activity. Kinetic analysis of *B. cereus* PPM indicates slow turnover and poor binding of the unnatural substrate. Although low activity of PPM with dideoxyribose 5-phosphate has been reported in the literature, a mechanistic explanation has not yet been established. The crystal structure of *B. cereus* PPM has been recently reported with ribose 5-phosphate bound in the active site and found that a hydrogen bond is formed between Ser-154 and the 3-OH of substrate[29]. It is possible that the low activity of dideoxyribose 5-phosphate is due to altered substrate binding resulting in poor turnover of substrate. The crystal structure of PPM with the unnatural substrate in the active site is reported here and the potential of targeted mutagenesis at residue 154 is explored.

Finally, the foundation for engineering of ribokinase is laid by docking dideoxyribose into the active site. Ribose is oriented by ribokinase through formation of hydrogen bonds to each of the substrate hydroxyls. It is expected that loss of two hydroxyls will affect the ability of ribokinase to orient the substrate for phosphorylation. Mutation of the many hydrogen bond donors and acceptors in the binding pocket may establish an apolar surface for favorable interaction with the more hydrophobic dideoxyribose substrate.

Methods

Synthesis of dideoxyribose and dideoxyribose 5-phosphate

(*S*)-γ-Butyrolactone-γ-carboxylic acid is formed from L-glutamic acid following the method of Okabe *et al*[30]. Sodium nitrite (1.5 eq, 37.5 g, 545 mmol) dissolved in water and 5.6 N hydrochloric acid (1.4 eq, 90.6 mL, 508.6 mmol) were simultaneously added dropwise over 3-4 hours to a slurry of glutamic acid (1.0 eq, 53.3 g, 363.6 mmol) in water (170 mL) while maintaining the temperature at 15-20 °C. After complete addition of HCl and $NaNO_2$, the reaction is brought to ambient temperature and stirred overnight. The solution was concentrated *in vacuo* and further azeotroped with toluene. Ethyl acetate and anhydrous sodium sulfate were added and the mixture stirred for several hours. The precipitates were removed, washing with ethyl acetate. AG50W-X4 resin was added to the solution and stirred for 30 minutes to remove unreacted starting material. After filtration to remove the resin, the solution was concentrated *in vacuo* and recrystalized from dichloromethane. The crystals were separated by filtration, washed with DCM and dried to yield pure (*S*)-γ-butyrolactone-γ-carboxylic acid (18.2 g, 139.9 mmols, 39%). Recrystallization of the filtrate further increases the yield. ^{1}H NMR (400 MHz, DMSO-d6, δ): 4.97-4.93 (m, 1H, C*H*), 2.54-2.10 (m, 4H, CH_2-CH_2). ^{13}C NMR (100 MHz, DMSO-d6, δ): 177.8 (C5), 172.6 (C1), 76.4 (C4), 27.7 (C2) and 26.4 (C3).

To form (*S*)-γ-hydroxymethyl-γ-butyrolactone the unpurified (*S*)-γ-butyrolactone-γ-carboxylic acid from the previous step is reduced with borane dimethylsulfide[30]. The reactant (1.0 eq, 4.95 g, 38 mmol) was dissolved in 25 mL of anhydrous THF and cooled to 0 °C. Under argon, borane dimethylsulfide (10 M, 1.15 eq, 4.375 mL, 43.75 mmol) was added dropwise over a 60 min period. The solution was brought to room temperature, monitored by TLC and stirred for an additional two hours. Methanol (3 X 2.25 ml) was added to quench the excess reductant and the solution concentrated *in vacuo*, yielding (*S*)-γ-hydroxymethyl-γ-butyrolactone (3.8 g, 32.7 mmols, 86%) with sufficient purity for use without further purification. ^{1}H NMR (400 MHz,

$CDCl_3$, δ): 4.67-4.60 (m, 1H, C*H*), 3.62-3.93 (m, 2H, -CH_2OH), 2.67-2.10 (m, 5H, CH_2CH_2, -O*H*). ^{13}C NMR (100 MHz, $CDCl_3$, δ): 178.4 (C1), 81.3 (C4), 64.2 (C5), 28.9 (C2), 23.4 (C3).

2,3-Dideoxyribose was synthesized by reduction of (*S*)-γ-hydroxymethyl-γ-butyrolactone (1.0 eq, 106 mg, 0.9 mmol) with diisobutyl aluminum hydride (1.5 M in toluene, 1.1 eq, 0.67 ml, 1.35 mmol) in anhydrous dichloromethane (10.6 mL) under argon and at -78 °C[31]. The reaction was monitored by TLC and determined to be incomplete after 1.25 hours of stirring. Additional diisobutyl aluminum hydride (2 eq total, 0.55 mL, 0.45 mmol) was added and the solution stirred for an additional 30 min. Methanol (5 ml) was added and the solution brought to room temperature. The resulting suspension was purified by flash chromatography (9:1 dichloromethane/methanol). Dideoxyribose (95 mg, 0.8 mmol, 88%) was collected as an oil after concentration *in vacuo*. ^{1}H NMR (400 MHz, D_2O, δ): 5.78-4.97 (m, 1H, CH) 4.36-3.35 (m, 3H, -CH_2OH, CH), 2.17-1.47 (m, 4, CH_2-CH_2). ^{13}C NMR (100 MHz, D_2O, δ): 97.93, 97.55, 93.11 and 92.89 (CH), 80.30, 78.35, 63.95 and 63.42 (CH), 66.85, 66.27, 64.51 and 63.07 (CH_2), 32.53, 32.00, 27.43, 26.99, 25.55, 26.47, 24.2 and 24.12 (CH_2).

To form (*S*)-γ-Dibenzylphosphomethyl-γ-butyrolactone, (*S*)-γ-hydroxymethyl-γ-butyrolactone (1.0 eq, 740 mg, 6.4 mmol) was dissolved in anhydrous dichloromethane (5 mL) and anhydrous acetonitrile (1.5 mL). The solution was brought to 0 °C and tetrazole (3 % in acetonitrile, w/v, 2 eq, 37.7 mL, 12.75 mmol) was added followed by dibenzyl-(*N*,*N*)-diisopropyl-phosphoramadite (1.4 eq, 3.26 mL, 8.9 mmol) which was diluted in 18.5 mL of dichloromethane. After stirring for two hours, *tert*-butyl hydroperoxide solution (~5.5 M in decane, 1.2 eq, 1.4 mL, 7.7 mmol) were added and the reaction stirred for an additional hour. Half-saturated sodium bicarbonate (20 ml) was used to quench the remaining peroxide and the solution extracted with dichloromethane (2 x 20 ml). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo*[32]. The product was purified by flash chromatography (39:1 dichloromethane/methanol) and pure (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone (1.355 g,

3.6 mmol, 56.5%) was collected after two columns. ^{1}H NMR (400 MHz, $CDCl_3$, δ): 5.1-4.99 (m, 4H, CH_2), 4.64-4.55 (m, 1H, CH), 4.17-3.95 (m, 2H, CH_2), 2.04-1.93 (m, 2H, CH_2-CH_2). ^{13}C NMR (100 MHz, $CDCl_3$, δ): 176.7 (C1), 135.86 (d, J_{cp} = 2.32 Hz, CH), 135.8 (d, J_{cp} = 2.47 Hz, CH), 128.98 (d, J_{cp} = 5.36 Hz, CH), 128.39 (d, J_{cp} = 3.75 Hz, CH), 77.80 (d, J_{cp} = 8.06 Hz, CH), 69.95 (d, J_{cp} = 5.63 Hz, CH), 68.29 (d, J_{cp} = 5.48 Hz, CH), 28.3 (C2), 23.6 (C3). ^{31}P NMR (200 MHz, $CDCl_3$, δ): 0.26 (septet).

To form 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate, (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone (1.0 eq, 270 mg, 0.72 mmol) was dissolved in anhydrous dichloromethane (25 mL), placed under argon and cooled to -78 °C. Diisobutyl aluminum hydride (1.5 M in toluene, 4 eq, 1.91 ml) was added dropwise and the reaction stirred for two hours[31]. The reaction was quenched with methanol (12.5 ml) and brought to ambient temperature. The resulting solution was poured over Rochelle salts (25 mL) and stirred overnight. The mixture was then filtered with the solid dissolved in water and added to the supernatant. The aqueous phase was extracted with dichloromethane (3 x 25 ml), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. 2,3-The crude residue was purified by flash chromatography (ethyl acetate) to yield 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate (84 mg, 0.2 mmol, 31%). ^{1}H NMR (400 MHz, $CDCl_3$, δ): 7.34 (s, 10, Ar-H), 5.53-5.48 (m, 1H, CH), 5.12-4.99 (m, 4 H, CH_2), 4.39-4.18 (m, 1H, CH), 4.13-3.88 (m, 2H, CH2), 2.14-1.58 (m, 4H, CH_2-CH_2). ^{31}P NMR (200 MHz, $CDCl_3$, δ): 0.3-0.6.

The pathway intermediate 2,3-dideoxyribose 5-phosphate was formed by reductive deprotection of 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate[33]. 2,3-Dideoxyribose 5-(di-*O*-benzyl)phosphate (1.0 eq, 250 mg, 0.661 mmols) was dissolved in methanol (5 mL) followed by addition of 10% Pd/C (28 mg). The suspension was flushed with argon and hydrogen then stirred overnight with hydrogen in a double balloon. The solution was filtered through celite, washed with methanol, and concentrated *in vacuo*. The resulting compound was dissolved in water and 100.5 mg sodium carbonate added to neutralize the solution. The resulting methyl-

riboside was dissolved in water and hydrolyzed by addition of Dowex-H^+ (1.0 g) which had been washed with methanol and water. The pH was checked and confirmed to be ~2-3. The solution was stirred for 2-3 hours, filtered, concentrated *in vacuo*, and lyophilized resulting in pure 2,3-dideoxyribose 5-phosphate (40 mg, 0.2 mmol, 30 % yield). ^{1}H NMR (400 MHz, D_2O, δ): 5.58-5.48 (m, 1H, CH), 4.47-4.22 (m, 1H, CH), 4.02-3.77 (m, 2H, CH2), 1.76-2.20 (m, 4H, CH2-CH2). ^{13}C NMR (100 MHz, D_2O, 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 NMR (200 MHz, D_2O, δ): 1.79, 1.65, 1.58.

The concentration of dideoxyribose 5-phosphate solution was determined by a standard proton NMR experiment with an attenuated delay time (d1 = 15 s) using 7.5 mM dimethylformamide as an internal standard. Concentration was calculated by comparison of the peak area for the aldehyde of dimethylformamide (δ=~7.5-8) to the peak area of the anomeric hydrogens of dideoxyribose 5-phosphate (δ=~5.5-6), both of which occur in relatively isolated regions of the spectra[34].

Enzyme Expression and Purification

B. cereus PPM and human PNP were purified as previously described (as published[29, 35, 36] and in Chapters 2, 3 & 4). To express and purify *E. coli* ribokinase, pRAS1004[24] harboring the *rbsk* gene was transformed into electrocompetent *E. coli* BL21(DE3) cells by electroporation. pRAS1004 allows for expression of ribokinase with an N-terminal hexahistidine tag. Successful transformations were selected by plating on streptomycin-containing LB-agarose plates. A single colony was picked into 5 mL of LB broth with streptomycin and grown overnight at 37 °C with shaking. This culture was used to inoculate 500 mL of LB broth containing 50 µg/mL streptomycin for protein production. Production cultures were incubated at 37 °C with shaking in 2.8 L Fernbach flasks. Isopropyl-β-D-1-thiogalactopyranoside was added to a final concentration

of 1 mM after the culture had grown to an OD_{600} of 0.6. The cells were harvested by centrifugation after an additional 4 h incubation and the cell pellet stored at -80 °C. The frozen cell pellet was resuspended in 20 mL of Buffer A (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 mM Imidazole) supplemented with DNAse I prior to lysis by passage through a French pressure cell at 20 000 psi. The cell lysate was purified by centrifugation and the supernatant applied to a HisTrap FF crude (5 ml, GE Healthcare) Ni^{2+}-affinity column, which was pre-equilibrated with Buffer A. The column was washed with three column volumes of Buffer A followed by 10 column volumes of 90:10 v/v Buffer A : Buffer B (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 500 mM imidazole). Ribokinase was eluted by increasing the ratio of Buffer B to 60:40%, which is equivalent to ~300 mM imidazole. Fractions containing protein were combined and diluted 1:1 into Exchange buffer. After concentration with a Millipore centrifugal filter, the remaining imidazole was removed by application of the concentrated fractions to a HiTrap Desalting column which had been pre-equilibrated with Exchange Buffer (25 mM Tris-HCl, pH 8, 5 mM $MgCl_2$). Protein concentrations were established using the BCA Protein Assay Kit (Thermo Scientific) and aliquots were stored at -80 °C until use.

Mutation of Ser-154 to alanine was performed using the Quick Change Mutagenesis protocol (Stratagene) using the primers 5´-CAGGCTCTTTAATCGTTTATACTGCCGCTGATAGCGTATTGCAAATTGCAGC-3´ and 5´-GCTGCAATTTGCAATACGCTATCAGCGGCAGTATAAACGATTAAAGAGCCTG-3´.
Expression and purification of the S154A mutant of PPM followed the same protocol as wild type PPM.

Detection of Inosine and ddI by LC/ESI-MS/MS

Inosine and ddI were separated from contaminating salts and other reaction components using a Jupiter 5µ C_{18} column (2.0 mm x 150 mm, Phenomenex) and an isocratic flow of 0.2 mL/min of 95%:5% v/v H_2O:acetonitrile, 10 mM CH_3COONH_4, pH 6. A ThermoPal autosampler

was used to inject 10 µL of the sample onto the column. Under these conditions inosine and ddI typically eluted in the range of 2.5-3.5 minutes and 5.0-6.0 minutes, respectively. Samples were analyzed on a triple quadrupole ESI–LC/MS (ThermoScientific Quantum Access) equipped with an ESI interface outfitted with a 100 µm internal diameter deactivated fused Si capillary. The selection reaction monitoring mode was applied with a collision energy of 10 eV. Mass transitions of (267→137) or (237→137) corresponding to the fragmentation of inosine or ddI to hypoxanthine, respectively, were monitored. Nitrogen was used as the sheath gas and argon as the auxiliary gas, at 54 and 10 (arbitrary units), respectively. The mass spectrometer was operated in the positive mode with the following parameters: spray voltage, 4.5 kV; capillary temperature, 270 °C; vaporizer temperature, 0 °C; tube lens offset, +35 V; skimmer offset, 0 V; capillary offset, 35 V. The parameters were tuned by direct infusion of a 500 µM ddI standard at 10 µL/min. Data acquisition and spectral analyses were conducted with Thermo Xcalibur Software, version 2.1.

PNP forward synthesis rate

The viability of PNP in the synthesis direction was analyzed by the formation of inosine or ddI in the presence of *in situ*-generated ribose 1-phosphate or dideoxyribose 1-phosphate and hypoxanthine. Sugar 1-phosphate is generated by isomerization of ribose 5-phosphate or dideoxyribose 5-phosphate by PPM. PPM is first activated[29] for 10-15 minutes by the combination of 129 µL of 25 mM Tris-HCl, pH 8, 10 µL of 2mM $MnCl_2$, 20 µL of 100 µM glucose 1,6-bisphosphate, 5 µL of 400 µM PPM and 6 µL of 50 mM hypoxanthine. Upon addition of 20 µL of 50 mM sugar 5-phosphate the reaction is incubated at 37 °C for 6 h. The reaction is then divided into two 95 µL portions and brought to 25 °C. Addition of 5 µL of 1 µM purine nucleoside phosphorylase starts the synthesis of ddI or inosine. The final reaction conditions are as follows: 25 mM Tris-HCl, pH 8, 0.1 mM $MnCl_2$, 10 µM glucose 1,6-bisphosphate, 10 µM PPM, 5 mM sugar 5-phosphate, 1.5 mM hypoxanthine, and 50 nM wild type PNP or PNP-46D6[35], 25 °C.

Aliquots of the reactions were removed at 5, 10, 20, and 30 min and the enzyme was denatured by heating at 95 °C. Relative amounts of ddI and inosine formed in each reaction were compared by LC/ESI-MS/MS.

Initial velocity of PPM and PPM-S154A

Biochemical data for determination of kinetic parameters of PPM and PPM-S154A with dideoxyribose 5-phosphate were collected in microtubes. Activated phosphopentomutase was freshly prepared by pre-incubation of 10 μM PPM in 25 mM Tris-HCl, pH 8.0 with 40 μM glucose 1,6-bisphosphate and 100 μM $MnCl_2$ at 24 °C for 10 minutes and then placed on ice until assayed. 21 μL of a standard mix comprised of 25 mM Tris-HCl, pH 8.0, 2.14 mM hypoxanthine, 15.9 μM PNP-46D6, and 128 μM $MnCl_2$ was added to each microtube along with 6 μL of dideoxyribose 5-phosphate which had been dissolved in 500 mM Tris-HCl, pH 8.0. This mixture was brought to 24 °C prior to initiation of the reaction by addition of 3 μL of activated PPM. The final reaction conditions consisted of 30 μL of 140 mM Tris-HCl, pH 8.0, 0.1 mM $MnCl_2$, 4 μM glucose 1,6-bisphosphate, 1.5 mM hypoxanthine, 11.1 μM PNP-46D6 , 1 μM PPM and 0.5, 0.7, 1.0, 1.4, 2.0, 3.0, or 5.0 mM dideoxyribose 5-phosphate. Reactions were allowed to proceed for 15 min before denaturing the enzyme at 95 °C for 5 minutes. Samples were centrifuged and frozen at -80 °C until analysis.

The amount of dideoxyribose 5-phosphate turnover was monitored by the amount of ddI formed over the time course of each reaction. The concentration of ddI was established by HPLC using 2-deoxyguanosine as an internal standard. Hypoxanthine, 2-deoxyguanosine, and ddI were separated by a Nova-Pak C_{18} column (3.9 x 150 mm, Waters) and an isocratic flow of 1 mL/min of 95%:5% v/v H_2O:acetonitrile, 10 mM CH_3COONH_4, pH 6.0. A calibration curve for ddI was established by addition of 5 μL of 80 μM 2-deoxyguanosine to 20 μL of known ddI standards at concentrations of 3, 9, 18, 27, 36, 54, 72, 90, and 120 μM. Prior to analysis each sample was thawed and centrifuged. 5 μL of 80 μM 2-deoxyguanosine was added to 20 μL of

each sample and 20 µL of the sample injected onto the HPLC. Initial velocities were fit to the Michaelis-Menten equation using non-linear regression analyses in the GraphPad Prism software package version 5.01 for Windows.

Crystallization of PPM and PPM-S154A with dideoxyribose 5-phosphate

Crystals of wild type and S154A PPM were grown from a 12 mg/mL protein mixture in 25 mM Tris-HCl pH 7.4, 1 mM $MnCl_2$ using the hanging drop vapor diffusion method at 18 °C. The reservoir solution for wild type enzyme contained 100 mM BIS-TRIS pH 5.5, 50 mM $MnCl_2$, 75 mM Ammonium Acetate and 13% PEG 3350 and the reservoir solution for PPM-S154A was 100 mM BIS-TRIS pH 5.5, 50 mM $MnCl_2$, 50 mM ammonium acetate and 17% PEG 3350. Substrate was introduced into fully formed crystals by soaking in a solution that contained all components of the reservoir solution along with 10 mM 2,3-dideoxyribose 5-phosphate. Crystals were cryoprotected with a solution that was 70% v/v reservoir solution and 30% v/v glycerol before flash freezing in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source on beamline 21-ID-G and were processed and scaled using the HKL2000[37] and CCP4[38] suites of programs. Initial phases were determined by rigid body refinement of the structure of wild type *B. cereus* PPM (PDB ID 3M8W) in CNS[39] and the model was refined with automated refinement in CNS[39] and REFMAC[39] using TLS refinement with TLS groups[40] selected using the TLSMD server[41] and manual refinement in COOT[42]. A summary of crystallization and model refinement statistics is presented in Table 6.1. RMS deviations were calculated with lsqkab and figures depicting crystallographic models were generated with PyMOL.

Table 5-1: Data collection and refinement statistics

	Wild type PPM bound to 2,3-dideoxyribose 5-phosphate	S154A PPM bound to 2,3-dideoxyribose 5-phosphate
Data Collection:		
Wavelength (Å)	0.979	0.979

Beamline	APS 21-ID-G	APS 21-ID-G
Resolution (Å) (last shell)a	50-2.1 (2.18-2.10)	50-2.1 (2.18-2.10)
Unit-cell	a = 91.5	a = 92.3
dimensions (Å)	b = 76.5	b = 76.6
	c = 107.3	c = 107.2
	α = 90°	α = 90°
	β = 108.9°	β = 109.0°
	γ = 90°	γ = 90°
Total reflections	233118	274095
Unique reflections	79568	81866
Reflections in test set	3900	4039
I/σ	18.1 (3.2)	17.5 (3.1)
Completeness (%)	96.7 (88.7)	99.0 (95.6)
Rsymb	0.083 (0.283)	0.087 (0.369)
Model Refinement:		
Rcrystc	0.167	0.177
Rfree	0.211	0.215

a. Values in parenthesis are those from the highest resolution shell.
b. $R_{sym}=\sum |I_i-I_{mean}|/I_{mean,}$
c. $R_{cryst}=\sum||F_{obs}|-|F_{calc}||/\sum|F_{obs}|$

Initial velocity of PPM-S154A with ribose 5-phosphate

Initial velocity of PPM-S154A with ribose 5-phosphate as substrate was established using the assay reported for steady state kinetics in Chapter 4. PPM-S154A was assayed at 100 nM for turnover of 2 mM ribose 5-phosphate for 10 min. This initial velocity was compared to the wild type PPM turnover velocities established with 20 nM PPM and 700 μM ribose 5-phosphate over 4 minutes. Initial velocities for dideoxyribose 5-phosphate were established using the HPLC assay described for the kinetic analysis and 1 μM wild type PPM or PPM-S154A and 5 mM dideoxyribose 5-phosphate over a 15 minute time period.

Directed Biosynthesis of ddI

All reagents are dissolved in 25 mM Tris-HCl, pH 8.0. PPM was freshly activated by pre-incubation with glucose 1,6-bisphosphate. In a 1.5 mL microtube 14.8 µL of 25 mM Tris-HCl, pH 8.0, 2.4 µL of 50 mM hypoxanthine, 8 µL of 5 mM $MnCl_2$, 8 µL of 100 µM of glucose 1,6-bisphosphate, and 2 µL of wild type PPM were combined and allowed to incubate at room temperature for 10 min. An equimolar concentration (22.7 µM) of PPM and glucose 1,6-bisphosphate is present in this mixture. To this same microtube was added 4.8 µL of 500 mM KCl, 2.4 µL of PNP, 16 µL of 50 mM ribose or dideoxyribose, 8 µL of 50 mM ATP, and 16 µL of ribokinase. For reactions containing ribose, 245 µM wild type PNP was added to the reaction; for reactions containing dideoxyribose, 370 µM PNP-46D6 was added. Ribokinase was present in Exchange Buffer, which contains 5 mM $MgCl_2$. For an 80 µL reaction, the final reaction conditions were 25 mM Tris-HCl, pH 8.0, 1.5 mM hypoxanthine, 500 µM $MnCl_2$, 10 µM glucose 1,6-bisphosphate, 10 µM wild type PPM, 1 mM $MgCl_2$, 30 mM KCl, 7.35 µM wild type PNP or 11.1 µM PNP-46D6, 10 mM ribose or dideoxyribose, 5 mM ATP, and 100 µM ribokinase. Reactions were incubated at 37 °C for 0, 6 or 12 hours followed by inactivation of the enzyme by heating at 95 °C. The denatured protein was pelleted by centrifugation and the supernatant frozen at -80 °C. Immediately prior to analysis, the samples were thawed and centrifuged. The relative amount of inosine or ddI formed at each time point was analyzed by LC/ESI-MS/MS, as outlined above.

Computational docking of ribokinase

Coordinates for *E. coli* ribokinase (PDB ID 1rkd) were obtained from the Protein Data Bank. The protein backbone and side-chains were energy minimized within the Rosetta energy function to make 100 models[43]; the best scoring model by total energy was carried forward into docking. α-D-Ribose and α-D-2,3-dideoxyribose conformations were generated in the Molecular Operating Environment version 2009.10 (MOE, Chemical Computing Group) as a fully protonated molecule and allowing for all ring conformations. The 22 resulting conformations for

ribose and 18 for dideoxyribose were docked using RosettaLigand[44] (Rosetta3 r. 39549). The docking procedure utilizes 1) a broad search sequence where ligand translation of up to 2 Å and 360° of rotational freedom is allowed, 2) translation of the ligand, if necessary, to bring it into contact with the protein, and 3) a refinement phase with small movement of the ligand (0.1 Å translation and 3° rotation) and full side chain and backbone flexibility within 7 Å of the ligand along with optimization of all hydrogen-oxygen bond orientations on the ligand. 2000 models were generated for ribose and dideoxyribose. Models were filtered according to the RosettaLigand interface binding energy and the top 200 models for each ligand were carried forward. Stable ligand conformations were identified through clustering of all-against-all rmsd calculations for the placement of ribose or dideoxyribose with the bcl::Cluster application[45]. An average inter-cluster distance was calculated as the average pair-wise distance between all objects in two clusters with a 2 Å intra-cluster linkage cutoff. Clusters dropping below ten members were removed from consideration.

Results

Ribokinase, phosphopentomutase, and purine nucleoside phosphorylase have been collected from *E. coli*, *B. cereus* and human, respectively, to form an unnatural pathway for ddI from dideoxyribose. The rate of ddI phosphorolysis by hPNP was previously increased through computational rational design and random mutagenesis by directed evolution. The improved variant, PNP-46D6[35] (see also chapters 2 and 3), was incorporated into the designed pathway. *E. coli* ribokinase, *B. cereus* phosphopentomutase and PNP-46D6 were expressed in *E. coli* BL21(DE3) cells and purified to homogeneity by affinity chromatography.

Chemical synthesis of dideoxysugar substrates

Dideoxyribose was synthesized from glutamic acid through nitrous acid deamination to first form (*S*)-γ-carboxy-γ-butyrolactone with full retention of configuration[30] (Figure 5-2). A yield

of 39% was sufficient to carry forward, however, yields as high as 85% have been reported and up to 93-100% for a two step reaction[30, 46]. Therefore, the low yield reported here is not an indicator of the overall yield potential of the proposed pathway. This compound is then reduced with borane dimethyl sulfide to form (*S*)-γ-hydroxy-γ-butyrolactone in 86% overall yield[30].

(*S*)-γ-Hydroxy-γ-butyrolactone serves as a branch point for the synthesis of dideoxyribose and dideoxyribose 5-phosphate (Figure 5-2). Dideoxyribose is formed with an 88% yield by reduction of the lactone with DIBAL[31], affording the substrate precursor in sufficient yields for biochemical testing. 2,3-Dideoxyribose 5-phosphate is synthesized by phosphitylation of the 5-OH group and oxidation[47] to form a benzyl-protected phospholactone intermediate. Reduction with DIBAL[31] and hydrogenation[32] to remove the protecting groups forms the final product. The overall modest yields afforded by this route were sufficient for biochemical analysis.

Figure 5-2. Chemical synthesis of dideoxyribose and dideoxyribose 5-phosphate

Forward rate of ddI synthesis by PNP-46D6

A 22-fold improved phosphorolysis efficiency of ddI by hPNP-46D6 has been previously reported[35]. However, in this study hPNP-46D6 is applied to the synthesis of ddI and it is not clear if these mutations will affect both directions of the reaction in the same way. Therefore, the

rate of product synthesis was evaluated through *in situ* generation of ribose 1-phosphate and dideoxyribose 1-phosphate by PPM isomerization of the associated 5-phosphate prior to addition of wild-type or mutant PNPs. The amount of nucleoside product formed at varying times was analyzed by LC/ESI-MS/MS. As shown in Figure 5-3, the PNP-46D6 turnover rate in the synthesis direction was improved for ddI, relative to the wild-type hPNP rate, while the mutations simultaneously decrease the inosine synthesis rate. The net effect was a 52-fold shift in the selectivity ratio of the mutant for synthesis of ddI (as measured at 20 min) which is comparable to a 46-fold shift in the selectivity ratio for phosphorolysis of ddI.

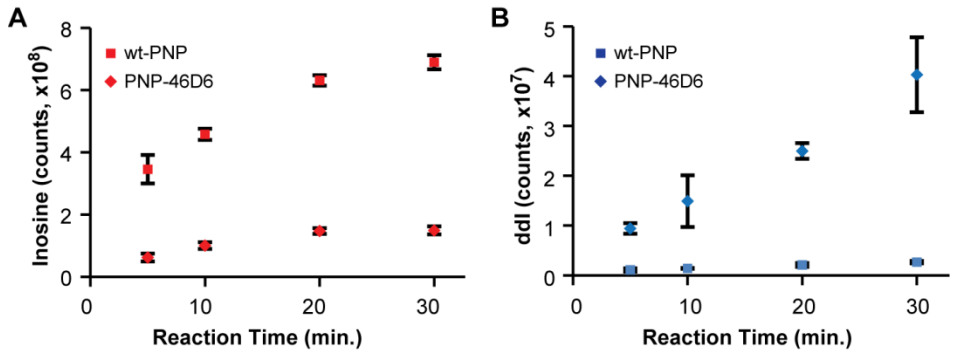

Figure 5-3. Comparison of wild-type and mutant PNP product synthesis rates. The amount of inosine (A) and dideoxyinosine (B) formed over time is compared upon *in situ* synthesis of the respective sugar 1-phosphate. Synthesis of ddI by PNP-46D6 occurs at a greater rate than that of wild type PNP. Error bars are the standard deviation of three independent measurements.

Binding and turnover of dideoxyribose 5-phosphate

In this study a PPM from *B. cereus* that will isomerize dideoxyribose 5-phosphate to dideoxyribose 1-phosphate was identified. Barbas and Wong previously reported the unsuccessful application of *E. coli* PPM for the desired rearrangement[33]. However, dideoxyribose 5-phosphate isomerization has been reported for a PPM cloned from *Bacillus stearothermophilus*, which has 88.9 % sequence similarity and 81% sequence similarity to *B.*

cereus[48]. Therefore, the PPM from *B. cereus* was cloned, expressed and characterized structurally and biochemically[29, 36] (Chapter 4). Turnover of dideoxyribose 5-phosphate was analyzed by linking activity to the generation of ddI from dideoxyribose 1-phosphate and hypoxanthine by PNP and subsequent analysis by HPLC. *B. cereus* PPM, like *B. stearothermophilus*, performs the phosphate transfer reaction on the targeted dideoxy sugar. Further analysis of enzyme turnover rates at multiple substrate concentrations established a k_{cat} and K_m of 7.5 ± 0.6 min^{-1} and 1 980 ± 350 μM (Figure 5-4).

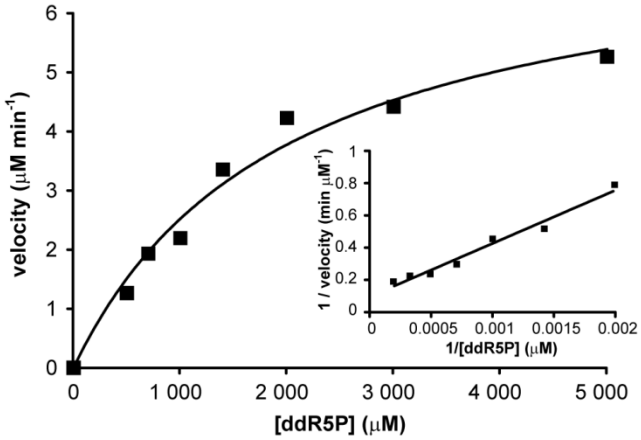

Figure 5-4. Dideoxyribose 5-phosphate turnover kinetics of phosphopentomutase. Initial velocities of PPM are plotted as a function of dideoxyribose 5-phosphate concentration. Data are the average of three assays, and a double reciprocal plot of the data is shown in the inset. Initial velocities were fit to the Michaelis-Menten equation using non-linear regression analyses in the GraphPad Prism software package version 5.01 for Windows. The calculated K_m, V_{max}, and k_{cat} are 1 985 ± 360 μM, 7.5 ± 0.6 μM min^{-1} and 7.5 ± 0.6 min^{-1}.

To identify a mechanistic cause for low dideoxy substrate turnover and aid in the identification of residues involved in orientation of the dideoxy sugar, fully formed crystals of PPM were soaked with dideoxyribose 5-phosphate and the crystal structure was determined. Density was identified in the PPM active site consistent with dideoxyribose 5-phosphate and analogous to the distal binding site of ribose 5-phosphate (Figure 5-5A). Electron density for the phosphate portion of substrate is strong and the molecule is primarily positioned by hydrogen

bonds formed with Nη1 of Arg-208, Nη2 of Arg-197, the hydroxyl of Tyr-152 and the backbone amide of Gly-132. In contrast, few direct interactions exist with the sugar portion of dideoxyribose 5-phosphate and electron density for the sugar ring is weak suggesting a large degree of flexibility in the positioning. The density is best interpreted with a rotation of the furanose ring relative to the position of ribose 5-phosphate which allows for interaction of the hydrophobic C2-C3 portion of dideoxyribose 5-phosphate with two exposed hydrophobic residues, Ile-195 and Val-158.

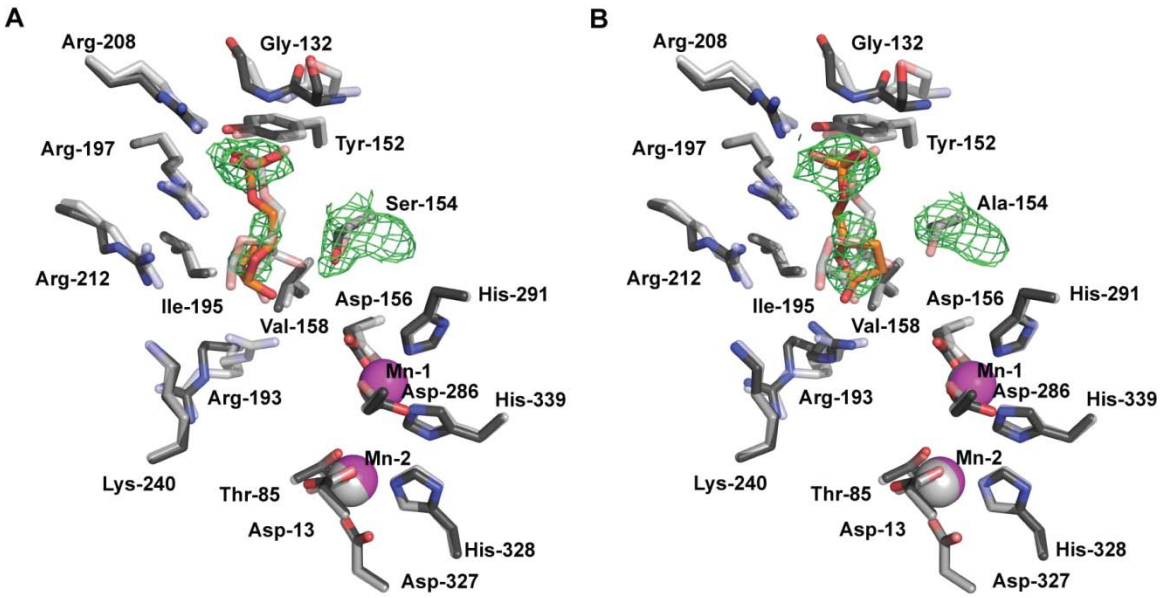

Figure 5-5. *B. cereus* phosphopentomutase co-crystallized with dideoxyribose 5-phosphate. A. Binding in wild type PPM. $2|F_O|-|F_C|$ electron density (green mesh contoured at 0.8 σ) calculated with residue 154 and the ligand omitted from the calculation shows the strong density for the phosphate of 2,3-dideoxyribose 5-phosphate (orange sticks) and weak density for the remaining portion of the molecule. Comparison of the structure bound to 2,3-dideoxyribose 5-phosphate (dark grey) to the structure bound to ribose 5-phosphate (PDBID 3M8W, light grey) shows that the rings are on different planes. B. Binding in S154A PPM. Electron density shown as in panel A for the active site of S154A shows that 2,3-dideoxyribose 5-phosphate (orange sticks) adopts a different orientation. Comparison of the S154A structure (dark grey) to the structure of wild type PPM bound to ribose 5-phosphate (PDBID 3M8W, light grey) shows that the rings of the two structures are in plane, but that the phosphates bind differently.

Ribose 5-phosphate is positioned by hydrogen bonds between the 3-OH and Oγ of Ser-154 and between the 2-OH group and Nη2 of Arg-193. The absence of these functional groups at the 2 and 3 position of dideoxyribose 5-phosphate likely increases the propensity for formation of a non-productive Michaelis complex and, as a result, successful turnover of the enzyme occurs much less often (7.5 min^{-1} with dideoxyribose 5-phosphate versus 10.2 s^{-1} with ribose 5-phosphate). It is hypothesized that the unfavorable polar/non-polar interaction between Oγ of Ser-154 and dideoxyribose 5-phosphate forces the non-natural substrate into the observed alternative conformation. Therefore, to generate a more favorable interaction the polarity of the active site was decreased by mutating Ser-154 to alanine which may allow dideoxyribose 5-phosphate to adopt a more favorable ring position by alleviating the polar strain.

Ser-154 was mutated to alanine by site-directed mutagenesis in an effort to decrease the polarity of this position and increase the likelihood of a native-like substrate orientation for dideoxyribose 5-phosphate. The structure of PPM-S154A was determined after soaking crystals with dideoxyribose 5-phosphate. The structure of the S154A mutant is largely superimposable with the structure of wild type PPM (RMS deviation for all Cα atoms is 0.8 Å). Electron density calculated from the refined model clearly demonstrates the presence of a molecule in the distal binding pocket (Figure 5-5B). This density is consistent with a molecule of dideoxyribose 5-phosphate. The phosphate is bound in an orientation typical of the distal binding pocket. The orientation of the deoxyfuranose ring, however, is different in PPM-S154A and wild type PPM. The C2-C3 portion of the ring is now oriented toward Ala-154 and a hydrogen bond is formed between Arg-193 and the 1-OH group of dideoxyribose 5-phosphate. The change in ring orientation supports the hypothesis that interactions with Ser-154 provide the impetus for an alternative binding conformation of dideoxyribose 5-phosphate in wild type PPM. Hydrogen bond formation between dideoxyribose 5-phosphate and Arg-193 is not observed in wild-type PPM where the primary interaction of Arg-193 is through a hydrogen bond to Asp-243;

interestingly, this interaction is one of two between the cap and core domain. In PPM-S154A, Arg-193 hydrogen bonds to the 1-OH group of dideoxyribose 5-phosphate whereas this interaction exists with the 2-OH group of ribose 5-phosphate. Despite persistent differences in the binding mode of dideoxyribose 5-phosphate in PPM-S154A and ribose 5-phosphate in wild type PPM many of the native interactions are recovered. Thus, formation of the S154A mutant may be a first step in recovery of the natural substrate orientation and possibly increased turnover.

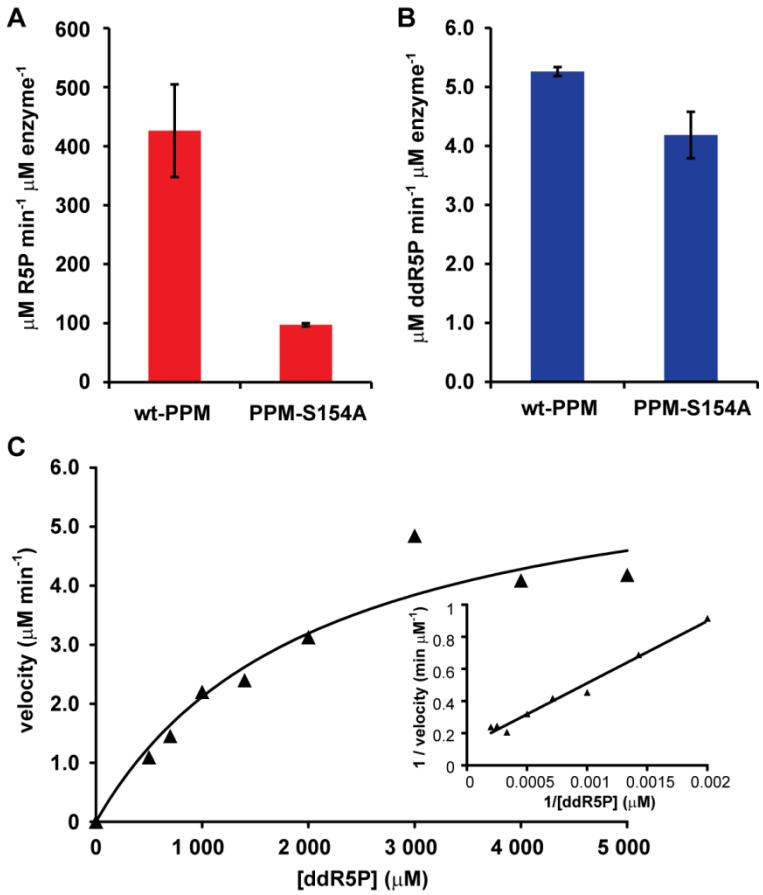

Figure 5-6. Comparison of wild-type PPM and PPM-S154A activity. A. Initial velocity of wild type PPM and PPM-S154A under near-saturating ribose 5-phosphate conditions. B. Initial velocity of wild type PPM and PPM-S154A under near-saturating dideoxyribose 5-phosphate conditions. Velocities in A and B are the average of three assays and error bars are the standard deviation.

C. Initial velocities of PPM-S154A are plotted as a function of dideoxyribose 5-phosphate concentration. Data are the average of three assays, and a double reciprocal plot of the data is shown in the inset. Initial velocities were fit to the Michaelis-Menten equation using non-linear regression analyses in the GraphPad Prism software package version 5.01 for Windows. The calculated K_m, V_{max}, and k_{cat} are 2070 ± 470 µM, 6.5 ± 0.7 µM min^{-1} and 6.5 ± 0.7 min^{-1}.

The activity of PPM-S154A with dideoxyribose 5-phosphate was tested to investigate the effect of the new furanose binding orientation. Dideoxyribose 5-phosphate turnover was measured in the presence of hypoxanthine and hPNP-46D6 and the product ddI concentration established by HPLC and compared to turnover rates of wild type and PPM-S154A established at fixed, near-saturating ribose 5-phosphate concentrations, and assayed as described in chapter 4. As expected, the loss of the interaction between Ser-154 Oγ and the 3-OH decreased turnover of ribose 5-phosphate by PPM-S154A. Preliminary concentration response experiments establish near-saturating conditions at 2 mM ribose 5-phosphate. When compared to turnover of wild type PPM at 700 µM ribose 5-phosphate PPM-S154A has a 4.4-fold reduced velocity per mole of enzyme (Figure 5-6A). In contrast, wild type PPM and PPM-S154A have similar velocities at 5 mM dideoxyribose 5-phosphate (Figure 5-6B). Enzyme turnover rates established at multiple dideoxyribose 5-phosphate concentrations reveal a k_{cat} and K_m of 6.5 ± 0.7 min^{-1} and 2 070 ± 470 µM for PPM-S154A (Figure 5-6C). These results indicate that although the turnover activity of PPM-S154A has not improved compared to wild type PPM the substrate selectivity has increased significantly.

In vitro biosynthesis of ddI

To demonstrate the viability of the pathway, dideoxyribose and hypoxanthine were combined with 11.1 µM PNP-46D6, 10 µM wild type PPM, 100 µM *E. coli* ribokinase, ATP and appropriate co-factors and incubated at 37 °C. Portions of the reaction were removed at 0, 6 and 12 h and the amount of ddI formed analyzed by LC/ESI-MS/MS. As shown in Figure 5-7,

ddI is formed in a time-dependent manner pointing to enzymatic activation of dideoxyribose by ribokinase and subsequent formation of ddI through the activities of PPM and PNP-46D6.

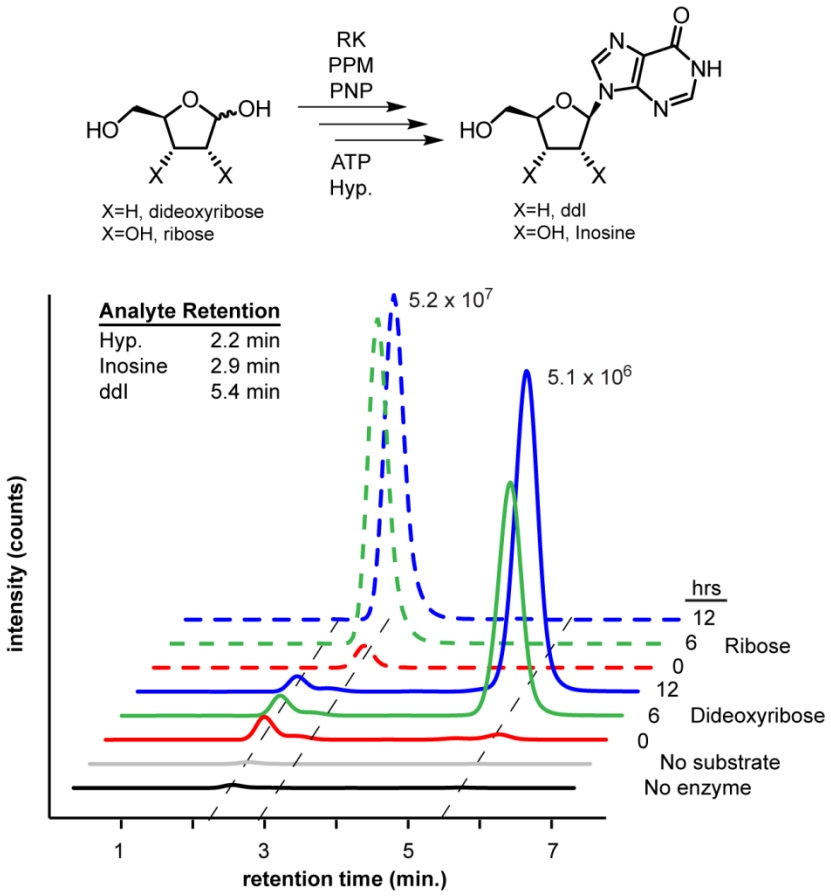

Figure 5-7. *In vitro* biocatalytic synthesis of ddI as analyzed by LC/ESI-MS/MS. Incubation of 10 mM substrate, 5 mM ATP, and 1.5 mM hypoxanthine (hyp) with *E. coli* ribokinase (RK), *B. cereus* phosphopentomutase (PPM) and human purine nucleoside phosphorylase (PNP) or PNP-46D6 forms nucleoside product. The relative amount of product is analyzed by LC/ESI-MS/MS in the positive mode with selection reaction monitoring for the fragmentation of nucleoside to hypoxanthine (inosine: 267 → 137; ddI: 237 → 137).

To check for chemical activation of dideoxyribose, activation through off-target activities of PPM or PNP, or contamination in the protein preps, reactions were run in the absence of ribokinase or without substrate. In each case mass correlating with product was not observed. Incubation of the enzyme mixture with ribose, hypoxanthine and ATP generated >10-fold more

inosine attesting to the viability of the enzyme and further pointing to the low rate of turnover of dideoxyribose by ribokinase.

Docking of dideoxyribose into ribokinase

To provide an explanation for the poor levels of dideoxyribose activation, the unnatural substrate was docked into ribokinase using RosettaLigand. Ribose was first docked into the binding site to provide a benchmark for the procedure. A plot of binding energy against the rmsd of the ligand shows a definitive energy funnel (Figure 5-8A). This indicates RosettaLigand recognizes models with a native-like configuration of ribose and scores these better than models with an alternative conformation. Clustering of the ribose models reveals a total of seven conformational groups (Table 5-2). Cluster 359 is populated by at least twice as many members as the others and has the best average interface energy between substrate and enzyme. An overlay of the best scoring member of cluster 359 is shown in Figure 5-8B. Together, these results provide some confidence in the results of docking dideoxyribose into ribokinase.

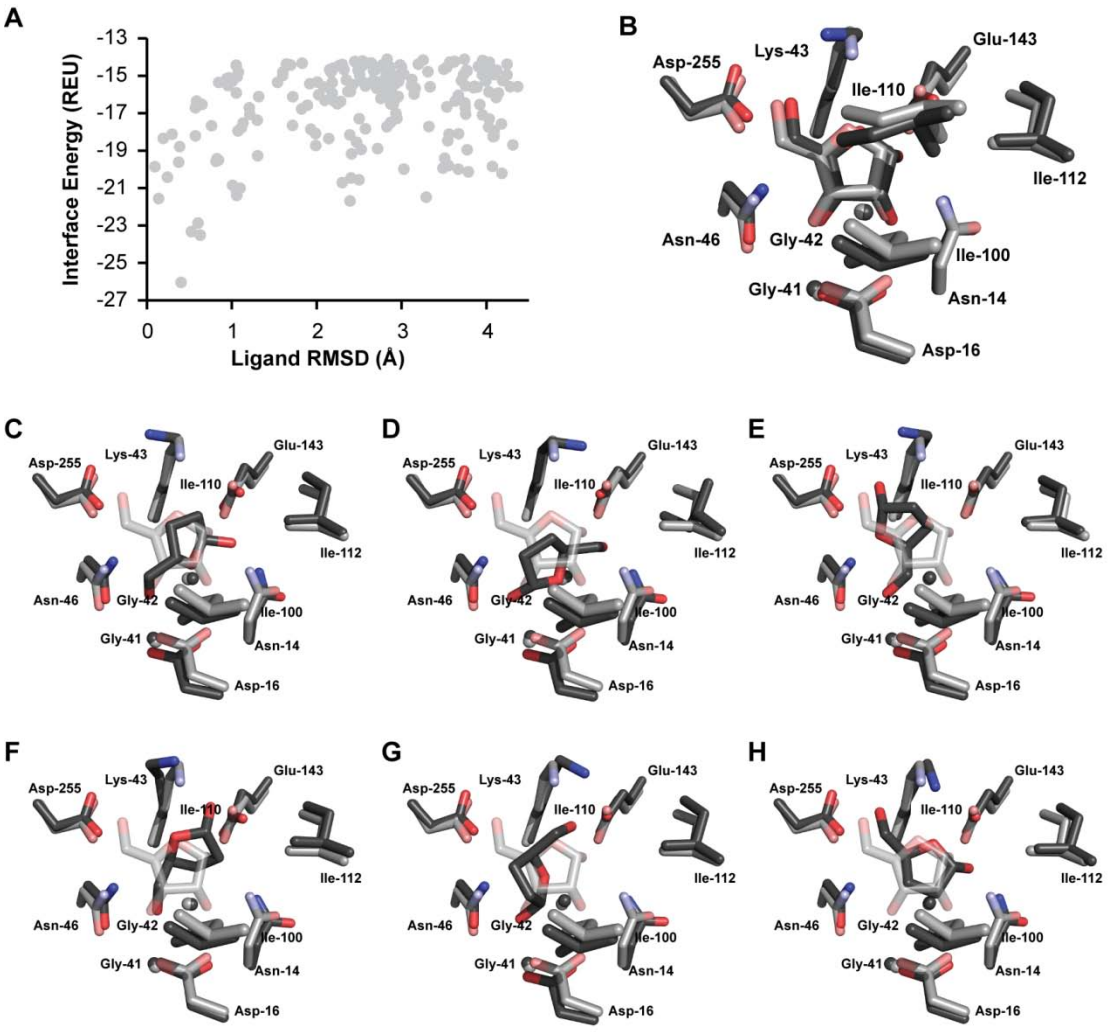

Figure 5-8. Computational docking of ribose and dideoxyribose into ribokinase with RosettaLigand. In (A), the interface binding score of ribose docked into ribokinase is plot against rmsd to the native binding conformation. Evidence of an energy funnel confirms that native-like models are recognized in the RosettaLigand energy function. B. The best scoring model of ribose (black sticks) is overlaid with the crystal structure (PDB ID 1rkd, gray sticks). C-H. Each cluster is overlaid with the crystal structure; C: 387, D: 377, E: 381, F: 376, G: 367, E: 368.

Dideoxyribose was then docked using the procedure established for docking of ribose. A total of 2 000 models were generated and the top 10% were identified based on interface binding energy. Clustering of these models reveals that six binding configurations are stably recognized upon docking of dideoxyribose to ribokinase. Two of the clusters are clearly larger than the others, possessing 56 and 40 members each (Table 5-3). Two other cluster size

ranges exist at ~20 and at <15. Each cluster was further analyzed by comparison of the average interface binding score and interface binding score of the best scoring cluster member. In general, the larger clusters have a lower average interface binding score and in most cases the best scoring models in the larger clusters have a better interface binding score than the smaller clusters. This is perhaps unsurprising as configurations which score well are likely to be enriched during the docking process. Furthermore, the lowest populated clusters have a poorer interface binding score than the more populated clusters indicating a preference for the other binding modes. However these binding modes must be sufficiently favorable as to be partially populated, albeit at low levels.

Table 5-2. Conformational clusters of ribose docked into ribokinase.

Cluster	Cluster Size	Linkage[a]	Cluster Interface Binding Score[b]	Best Interface Binding Score	Cluster RMSD to native[c]
359	40	1.26	-18.43 ± 2.77	-26.05	0.81 ± 0.35
378	24	1.79	-15.86 ± 1.32	-18.73	2.13 ± 0.46
383	24	1.95	-16.91 ± 2.37	-21.69	2.54 ± 0.25
368	19	1.49	-15.38 ± 1.34	-17.09	3.77 ± 0.17
367	16	1.49	-18.04 ± 2.14	-21.49	3.49 ± 0.21
382	16	1.93	-16.66 ± 2.06	-20.21	4.11 ± 0.11
374	10	1.68	-15.53 ± 1.24	-17.74	3.86 ± 0.19
384	10	1.96	-15.53 ± 0.80	-17.29	2.48 ± 0.23

a. Linkage is calculated as the average pair-wise distance between all objects in two sub-clusters
b. Reported values are the average interface binding score for all models in the cluster ± standard deviation
c. Reported values are the average rmsd to native for all models in the cluster ± standard deviation

Table 5-3. Table of conformational clusters of dideoxyribose docked into ribokinase by RosettaLigand.

Cluster	Cluster Size	Linkage[a]	Average Interface Binding Score	Best Interface Binding Score
387	56	1.61	-13.74 ± 1.47	-17.35
377	40	1.25	-13.29 ± 1.03	-15.81
381	20	1.39	-13.87 ± 1.52	-17.37
376	20	1.24	-13.41 ± 1.03	-16.70
367	13	1.02	-12.68 ± 0.56	-13.61
368	10	1.06	-12.80 ± 0.37	-13.52

a. Linkage is calculated as the average pair-wise distance between all objects in two sub-clusters
b. Reported values are the average interface binding score for all models in the cluster ± standard deviation

To assess the substrate binding mode prevalent in each cluster, clusters were represented by the cluster member with the best predicted binding energy (Figure 5-8C-H). Cluster 387 forms hydrogen bonds between the 5-OH group and both Asp-16 and Asn-46 while the 1-OH group is bound by Asn-14 and Glu-143 (Figure 5-8C). Cluster 377 resides in a conformation almost exactly opposite that of cluster 387 (Figure 5-8D), with the 1-OH group bound to Asp-16 and Asn-46 and the 5-OH group bound to Asn-14 and Glu-143. Cluster 381 (Figure 5-8E) and cluster 376 (Figure 5-8F) are found with the 5-OH group positioned in the Asp-16/Asn-46 pocket with the 1-OH group bound on either side of Lys-43 with Asp-255 or Glu-143, respectively. Cluster 367 (Figure 5-8G) binds opposite of cluster 376, bridging between Glu-143/Lys-43 at the 5-OH group and Asp-16/Asn-46 at the 1-OH. The smallest cluster, cluster 368, adopts a conformation most similar to ribose (Figure 5-8H). The 1-OH group is bound by Glu-143 and Asn-14 while the 5-OH group forms a hydrogen bond with Asp-255. The ring is further oriented by a hydrogen bond between Lys-43 and the furan oxygen.

Discussion

A chemo-enzymatic pathway for directed biosynthesis of nucleoside analogs has been outlined and its potential has been demonstrated through biocatalytic synthesis of dideoxyinosine (Figure 5-1). Dideoxyribose is synthesized in three steps from glutamic acid. A refined synthetic process will likely increase the yield of dideoxyribose to levels sufficient for manufacturing processes. The chemically synthesized precursor is transformed into ddI through tandem catalysis of ribokinase, PPM and PNP. However, the low levels of product formed over 24 h highlight the necessity for optimization of each enzyme in the pathway.

Toward this end human PNP was improved for phosphorolysis of dideoxyinosine resulting in the mutant PNP-46D6 which has a phosphorolysis efficiency 22-fold greater than wild type PNP. Previously, however, the viability of PNP-46D6 in the synthesis direction was not

confirmed because dideoxyribose 1-phosphate was not available. Here, the improved forward synthesis rate is confirmed by formation of the precursor through isomerization of dideoxyribose 5-phosphate by PPM. Because the concentration of dideoxyribose 1-phosphate in the reaction is not known and difficult to ascertain the present experiment does not establish a definitive forward rate. Instead, the ratio of turnover rates under these conditions is compared for wild type PNP and PNP-46D6. There is a marked improvement in synthesis of dideoxyinosine by the mutant and a concomitant decrease in synthesis of inosine (Figure 5-3). Overall, the mutations introduced to form PNP-46D6 provide a 52-fold improved substrate selectivity ratio in the synthesis direction. This value agrees well with the 46-fold improved selectivity ratio measured for phosphorolysis. Given the improved K_m for dideoxyinosine it is possible that PNP-46D6 also possesses an improved K_m for dideoxyribose 1-phosphate; this would be of great benefit in a tandem reaction sequence limited by poor turnover of enzymes earlier in the pathway as large precursor pools are not necessary for efficient turnover.

Dideoxyribose 1-phosphate is generated from dideoxyribose 5-phosphate by transfer of phosphate by PPM. It is interesting to note that *E. coli* PPM is reported not to accept dideoxyribose 5-phosphate as substrate despite a significant sequence similarity. Unlike with the study of PNP, the PPM reaction can be studied in the desired direction through synthesis of dideoxyribose 5-phosphate. The chemically synthesized pathway intermediate was utilized to establish kinetic constants for turnover of dideoxyribose 5-phosphate by PPM (Figure 5-4). The natural substrate is turned over 80-fold faster and has a 7.5-fold greater Michaelis constant.

To explain this disparity PPM was crystallized and dideoxyribose 5-phosphate was soaked into the binding site (Figure 5-5A). In wild type PPM the 3-OH of ribose plays a key role in positioning the substrate for catalysis through formation of a hydrogen bond with Ser-154. This handle is not available in dideoxyribose and has been exchanged for a much smaller hydrophobic methylene. To compensate for the largely polar active site of PPM the substrate is forced to adopt an alternate conformation. Rotation of the sugar moiety likely decreases the

interaction of the C2-C3 portion of dideoxyribose 5-phosphate with the multitude of polar residues in the active site.

To test this hypothesis Ser-154 was mutated to alanine, the activity compared to wild type PPM and the structure of dideoxyribose 5-phosphate-bound PPM-S154A solved. Removal of the polar residue caused a significant change in the conformation of the furanose ring (Figure 5-5B). Testing of the activity indicates that this mutation does not improve the turnover of dideoxyribose 5-phosphate, with PPM-S154A turnover on par with wild-type PPM turnover despite the ring migration (Figure 5-6B,C). Interestingly though, natural activity is reduced ~4-fold, increasing the substrate selectivity ratio of the mutant (Figure 5-4A).

The directed biosynthesis of ddI is demonstrated by phosphorylation of dideoxyribose by ribokinase and subsequent turnover by PPM and PNP (Figure 5-7). Overall, the reaction occurs at a rate which precludes its use in preparative syntheses. The ribokinase concentrations required for generation of detectable levels of ddI (100 μM in these experiments) are extreme and implicate ribokinase as the rate-limiting step in the pathway.

Docking studies with RosettaLigand were performed to identify interactions in the binding pocket for targeted engineering of the enzyme. Ribose and dideoxyribose were docked into the binding pocket. Successful recapitulation of the native binding pose provides confidence that the method will identify dideoxyribose configurations existing in nature. Ribose is oriented in the binding pocket by hydrogen bonds to each of the hydroxyls and the furanose oxygen, a hydrophobic overhang generated by Ile-100, Ile-110, Ile-112, and Ile-251, and the bottom of the pocket is lined by Gly-41 and Gly-42, which forms a hydrogen bond to the 3-OH of ribose with its backbone nitrogen (Figure 5-8B). A total of eight hydrogen bonds, occurring between Asn-14 and the 1- and 2-OH, Asp-16 and the 2- and 3-OH, Lys-43 and the furan oxygen, Asn-46 and the 3-OH, Glu-143 and the 1-OH, and Asp-255 and the 5-OH, are formed in the binding pocket. The large number of hydrogen bonds provide a strong binding force for the small molecule. Furthermore, the hydrophobic overhang forces the substrate into the polar residues which

reside in the lower part and bottom of the binding pocket while selecting for phosphorylation of the 5-OH group by enforcing the downward orientation of the 2- and 3-OH in the binding pocket. The predominant conformations generated during docking are formed through satisfaction of the many hydrogen bond donors and acceptors in the binding pocket. In phosphopentomutase, the hydrophobic portion of the tetrahydrofuranose ring controls the orientation of the substrate, in so far as it is free to move with the 5-phosphate anchor. Contrastingly, in ribokinase the C2-C3 portion is oriented to solvent in clusters 385 and 381 and toward polar side chains in clusters 386 and 376. These unfavorable interactions are tolerated in favor of formation of hydrogen bonds to the 1- and 5-OH group at the highly negative Asp-16 and Glu-143. Only cluster 371 and 360 orient this region toward hydrophobic residues. Cluster 367 abuts C2 and C3 to Ile-251 (not shown) while the native-like orientation of cluster 368 comes at the expense of unfavorable interactions with Asp-16. These results point to the formation of a stable, non-productive substrate binding complex which results in slow turnover of the enzyme.

Clearly, removal of polarity in the ribokinase binding pocket will increase the likelihood of native-like binding and decrease the impetus for mis-orientation of the residue. Targets for mutagenesis include Asp-16, Asn-46, Asn-14 and Gly-41. Because of the proximity of Asn-46 to the site of catalysis mutations are likely to have a pronounced affect. Similarly, Asn-14 is involved in orientation of the 1-OH. Mutation of Gly-41 may serve to fill space vacated by removal of the 2- and 3-OH of ribose.

Conclusions

In this study, a proof-of-principle for directed biosynthesis of nucleoside analogs is provided. Differences in substrate makeup clearly affect function of each enzyme in the pathway. Modifications in the binding pocket of PNP along with identification of sequence changes throughout the protein improved both synthesis and phosphorolysis rates of ddI; a similar strategy will likely improve the other enzymes in the pathway. Furthermore, it has been

shown that improvements made to reversible enzymes in one direction transfer, to some degree, to catalysis in the reverse direction.

Structural studies of PPM docked with dideoxyribose 5-phosphate provide a basis for subsequent engineering, identifying Ser-154, Ile-159 and Val-195 as targets. Mutation of Ser-154 to alanine has a marked effect on the orientation of the dideoxysugar but only a minimal effect on turnover of the unnatural substrate. However, because of the strong increase in the substrate selectivity ratio an engineered mutant should contain the S154A mutation. Further probing of the active site will identify the optimal amino acid identity at this position.

Ribokinase is likely the rate limiting step in the current set-up of the biocatalytic pathway. Docking of dideoxyribose into ribokinase implicates Asp-16 and Gly-41 as likely residues for mutagenesis. Because of their proximity in the binding pocket, mutations in these sites may be mutually exclusive. Strategies for a simultaneous search of the combinatorial sequence space include computational design or multi-site saturation mutagenesis.

It is possible that other nucleoside analogs could be generated with this pathway. Furthermore, analogs need not be limited to 6-oxopurines. For instance, a point mutation in hPNP alters the specificity to 6-aminopurines like adenosine and hPNP also shows off-target phosphorolysis activity for nucleosides with unnatural bases[49]. Additionally, nucleosidyl transferases can be incorporated in the pathway to switch the hypoxanthine base with pyrimidine bases[18].

Acknowledgements

This work was contributed to by David P. Nannemann, Timothy D. Panosian, William R. Birmingham, Vanessa V. Phelan, Tina M. Iverson, Jens Meiler, and Brian O. Bachmann. DPN designed and performed all biochemical experiments. TDP performed all crystallography experiments and generated PPM point mutants. WRB assisted in the collection of biochemical data and in the design of biochemical assays. VVP assisted in the design and collection of data

for LC/ESI-MS/MS experiments. TMI designed crystallography experiments and analyzed data. JM assisted in the analysis of computational experiments. BOB assisted in the design and analysis of all experiments.

References

1. Carothers, J. M.; Goler, J. A.; Keasling, J. D., Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.* **2009,** 20, (4), 498-503.

2. Steen, E. J.; Chan, R.; Prasad, N.; Myers, S.; Petzold, C. J.; Redding, A.; Ouellet, M.; Keasling, J. D., Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* **2008,** 7, 36.

3. Pfleger, B. F.; Pitera, D. J.; Smolke, C. D.; Keasling, J. D., Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **2006,** 24, (8), 1027-32.

4. 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.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** 440, (7086), 940-3.

5. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **2009,** 27, (8), 753-U107.

6. Atsumi, S.; Hanai, T.; Liao, J. C., Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008,** 451, (7174), 86-89.

7. Moon, T. S.; Yoon, S. H.; Lanza, A. M.; Roy-Mayhew, J. D.; Prather, K. L., Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* **2009,** 75, (3), 589-95.

8. Niu, W.; Molefe, M. N.; Frost, J. W., Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **2003,** 125, (43), 12998-12999.

9. Mitsuya, H.; Broder, S., Inhibition of the *in vitro* infectivity and cytopathic effect of human lymphotrophic-T virus type III/lymphadenopathy-associated virus (HTLV-III/LAV) by 2',3'-dideoxynucleosides. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** 83, (6), 1911-1915.

10. Torii, T.; Izawa, K.; Cho, D. H.; Jang, D. O., Synthesis of 2',3'-dideoxyinosine via radical deoxygenation. *Nucleosides, Nucleotides Nucleic Acids* **2007,** 26, (8-9), 985-988.

11. Chu, C. K.; Bhadti, V. S.; Doboszewski, B.; Gu, Z. P.; Kosugi, Y.; Pullaiah, K. C.; Vanroey, P., General syntheses of 2',3'-dideoxynucleosides and 2',3'-didehydro-2',3'-dideoxynucleosides. *J. Org. Chem.* **1989,** 54, (9), 2217-2225.

12. Saito, Y.; Zevaco, T. A.; Agrofoglio, L. A., Chemical synthesis of C-13 labeled anti-HIV nucleosides as mass-internal standards. *Tetrahedron* **2002,** 58, (47), 9593-9603.

13. Shiragami, H.; Amino, Y.; Honda, Y.; Arai, M.; Tanaka, Y.; Iwagami, H.; Yukawa, T.; Izawa, K., Synthesis of 2',3'-dideoxypurinenucleosides via the palladium catalyzed reduction of 9-(2,5-di-O-acetyl-3-bromo-3-deoxy-β-D-xylofuranosyl)purine derivatives. *Nucleosides Nucleotides* **1996,** 15, (1-3), 31-45.

14. Huryn, D. M.; Okabe, M., AIDS-driven nucleoside chemistry. *Chem. Rev. (Washington, DC, U. S.)* **1992,** 92, (8), 1745-1768.

15. Farina, V.; Benigni, D. A., A new synthesis of 2',3'-dideoxynucleosides for AIDS chemotherapy. *Tetrahedron Lett.* **1988,** 29, (11), 1239-1242.

16. Pinheiro, E.; Vasan, A.; Kim, J. Y.; Lee, E.; Guimier, J. M.; Perriens, J., Examining the production costs of antiretroviral drugs. *Aids* **2006,** 20, (13), 1745-1752.

17. Christoffersen, S.; Serra, I.; Terreni, M.; Piskur, J., Nucleoside phosphorylases from *Clostridium perfringens* in the synthesis of 2',3'-dideoxyinosine. *Nucleosides, Nucleotides Nucleic Acids* **2010,** 29, (4-6), 445-448.

18. Kaminski, P. A.; Dacher, P.; Dugue, L.; 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, (29), 20053-20059.

19. Lewkowicz, E. S.; Iribarren, A. M., Nucleoside phosphorylases. *Curr. Org. Chem.* **2006,** 10, (11), 1197-1215.

20. Rogert, M. C.; Trelles, J. A.; Porro, S.; Lewkowicz, E. S.; Iribarren, A. M., Microbial synthesis of antiviral nucleosides using *Escherichia coli* BL21 as biocatalyst. *Biocatal. Biotransform.* **2002,** 20, (5), 347-351.

21. Horinouchi, N.; Kawano, T.; Sakai, T.; Matsumoto, S.; Sasaki, M.; Mikami, Y.; Ogawa, J.; Shimizu, S., Screening and characterization of a phosphopentomutase useful for enzymatic production of 2 '-deoxyribonucleoside. *New Biotechnol.* **2009,** 26, (1-2), 75-82.

22. Horinouchi, N.; Ogawa, J.; Kawano, T.; Sakai, T.; Saito, K.; Matsumoto, S.; Sasaki, M.; Mikami, Y.; Shimizu, S., Biochemical retrosynthesis of 2 '-deoxyribonucleosides from glucose, acetaldehyde, and a nucleobase. *Appl. Microbiol. Biotechnol.* **2006,** 71, (5), 615-621.

23. Gross, A.; Abril, O.; Lewis, J. M.; Geresh, S.; Whitesides, G. M., Practical synthesis of 5-phospho-D-ribosyl α-1-pyrophosphate (PRPP) - enzymatic routes from ribose 5-phosphate or ribose. *J. Am. Chem. Soc.* **1983,** 105, (25), 7428-7435.

24. Scism, R. A.; Bachmann, B. O., Five-component cascade synthesis of nucleotide analogues in an engineered self-immobilized enzyme aggregate. *ChemBioChem* **2010,** 11, (1), 67-70.

25. Miroshnikov, A. I.; Esipov, R. S.; Muravyova, T. I.; Konstantinova, I. C.; Fateev, I. V.; Mikhailopulo, I. A., A new strategy for the synthesis of nucleosides: one-pot enzymatic transformation of D-pentoses into nucleosides. *Open Conf. Proc. J.* **2010,** 1, 98-102.

26. Sigrell, J. A.; Cameron, A. D.; Jones, T. A.; 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, (2), 183-193.

27. Rashid, N.; Imanaka, H.; Fukui, T.; Atomi, H.; Imanaka, T., Presence of a novel phosphopentomutase and a 2-deoxyribose 5-phosphate aldolase reveals a metabolic link between pentoses and central carbon metabolism in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* **2004,** 186, (13), 4185-4191.

28. Tozzi, M. G.; Camici, M.; Mascia, L.; Sgarrella, F.; Ipata, P. L., Pentose phosphates in nucleoside interconversion and catabolism. *FEBS J.* **2006,** 273, (6), 1089-1101.

29. Panosian, T. D.; Nannemann, D. P.; Watkins, G. R.; Phelan, V. V.; McDonald, W. H.; Wadzinski, B. E.; Bachmann, B. O.; 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, (10), 8043-54.

30. Okabe, M.; Sun, R. C.; Tam, S. Y. K.; Todaro, L. J.; Coffen, D. L., Synthesis of the dideoxynucleosides ddC and CNT from glutamic acid, ribonolactone, and pyrimidine bases. *J. Org. Chem.* **1988,** 53, (20), 4780-4786.

31. Smith, A. B.; Friestad, G. K.; Barbosa, J.; Bertounesque, E.; Hull, K. G.; Iwashima, M.; Qiu, Y. P.; Salvatore, B. A.; Spoors, P. G.; Duan, J. J. W., Total synthesis of (+)-calyculin A and (-)-calyculin B: asymmetric synthesis of the C(9-25) spiroketal dipropionate subunit. *J. Am. Chem. Soc.* **1999,** 121, (45), 10468-10477.

32. Humphries, M. J.; Ramsden, C. A., A fresh AIR synthesis. *Synthesis* **1999,** 1999, (6), 985-992.

33. Barbas, C. F.; Wong, C. H., Overexpression and substrate-specificity studies of phosphodeoxyribomutase and thymidine phosphorylase. *Bioorg. Chem.* **1991,** 19, (3), 261-269.

34. Gerritz, S. W.; Sefler, A. M., 2,5-dimethylfuran (DMFu): An internal standard for the "traceless" quantitation of unknown samples via H-1 NMR. *J. Comb. Chem.* **2000,** 2, (1), 39-41.

35. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010,** 23, (8), 607-616.

36. Panosian, T. D.; Nannemann, D. P.; Bachmann, B. O.; 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, 811-4.

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

38. Collaborative Computational Project Number, The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994,** 50, (5), 760-763.

39. Brunger, A. T., Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2007,** 2, (11), 2728-2733.

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

41. Painter, J.; Merritt, E. A., TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **2006,** 39, 109-111.

42. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004,** 60, 2126-2132.

43. Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D., High-resolution structure prediction and the crystallographic phase problem. *Nature* **2007,** 450, (7167), 259-264.

44. Davis, I. W.; Baker, D., RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009,** 385, (2), 381-392.

45. Alexander, N. S.; Woetzel, N.; Meiler, J. In *bcl::Cluster : a method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System*, Conference on Computational Advances in Bio and medical Sciences (ICCABS), Orlando, Florida, Feb. 3-5, 2011, 2011; Orlando, Florida, 2011; p in press.

46. Meira, P. R. R.; Moro, A. V.; Correia, C. R. D., Stereoselective Heck-Matsuda arylations of chiral dihydrofurans with arenediazonium tetrafluoroborates: an efficient enantioselective total synthesis of (-)-isoaltholactone. *Synthesis* **2007,** 2007, (15), 2279-2286.

47. Graham, S. M.; Pope, S. C., Selective phosphitylation of the primary hydroxyl group in unprotected carbohydrates and nucleosides. *Org. Lett.* **1999,** 1, (5), 733-736.

48. Hamamoto, T.; Noguchi, T.; Midorikawa, Y., Phosphopentomutase of *Bacillus stearothermophilus* TH6-2: the enzyme and its gene ppm. *Biosci., Biotechnol., Biochem.* **1998,** 62, (6), 1103-1108.

49. Stoeckler, J. D.; Poirot, A. F.; Smith, R. M.; Parks, R. E.; Ealick, S. E.; Takabayashi, K.; Erion, M. D., Purine nucleoside phosphorylase. 3. Reversal of purine base specificity by site-directed mutagenesis. *Biochemistry* **1997,** 36, (39), 11749-11756.

CHAPTER VI

DISSERTATION SUMMARY AND FUTURE DIRECTIONS

Synopsis

Microbial factories have provided an environmentally friendly means to synthesize complex organic molecules. Historically, native producers were engineered to increase the yield of innate products; more recently, however, heterologous expression methodologies applied by metabolic engineers and synthetic biologists have enabled the generation of microbial factories for non-innate molecules, including pharmaceuticals, commodities, fuels and other fine chemicals. Copying or replicating pathways permits simple manipulation of the pathway and host DNA and host metabolism and bypasses control of the pathway by the producing organism. The combination of sub-pathways from multiple source organisms also allows for control of pathway flux and frequently generates a fully integrated system for biosynthesis of the product from simple feedstocks. Non-natural molecules have been generated by directing alternative substrates through known pathways. Metabolic flux through these appropriated pathways has been increased through the generation of mutant enzymes with increased activities on and selectivity for the alternate compounds. The generation of novel pathways from disparate and unrelated enzymes, or *de novo* pathway design, allows for long pathways to be re-routed down more economical routes and for biosynthesis of molecules for which there is no naturally-available pathway.

A *de novo* pathway has been designed for ddI, a reverse transcriptase class nucleoside analog drug, which will allow for directed biosynthesis of the compound within a hosted system. The pathway decreases to three the number of required enzymes from the >10 enzyme steps required for biosynthesis of natural nucleosides.

Furthermore, ddI is not known to be synthesized in nature. A bioretrosynthetic analysis, in which the product is broken down to simpler molecules in a reverse step-wise manner, was employed in the design of the pathway. Ribokinase, phosphopentomutase, and purine nucleoside phosphorylase were chosen for directed biosynthesis of ddI from 2,3-dideoxyribose, a five carbon dihydroxypentanal analog of ribose. Enzymes from *E. coli*, *B. cereus* and human, respectively, were initially chosen for incorporation into the pathway.

In addition to a pathway design paradigm, a bioretrosynthetic approach was adopted for incorporation and optimization of enzymes into the pathway. Bioretrosynthesis minimizes to one the number of assays required for optimization of the pathway. Despite the use of multiple analytical techniques (UV/Vis, HPLC, LC/MS) during the study of this pathway, a single assay linked to the production of ddI or consumption of hypoxanthine was used for the ddI pathway.

Chemical synthesis of the pathway precursor 2,3-dideoxyribose from glutamic acid and heterologous expression of each enzyme allowed for *in vitro* demonstration of the pathway functionality. Turnover of the pathway was extremely low, generating only a few picograms of product per milliliter in a 12 hour time period. In comparison, an order of magnitude more inosine is formed in only an hour of incubation. Clearly, pathway flux is significantly affected by the change in substrate structure. This poor turnover is mirrored upon characterization of the individual enzymes in the pathway with the dideoxy substrate congener. Rational mutagenesis, guided by protein structure/function studies or computational prediction, and directed evolution were utilized to improve the function of enzymes in the pathway.

Mutation of active site residues in the first or second shell has a significant impact on specificity and substrate selectivity. Active site residues within PNP which control sugar selectivity were identified through calculation of substrate transition state

binding energies for inosine and ddI using the RosettaLigand small molecule docking algorithm. Computational design of the binding site and analysis of biochemical data narrowed the list of likely mutations to a single site, Y88. The complete complement of amino acid mutations were made at this position and kinetically characterized for comparison to predicted binding energies. Successful ranking of mutant activity was achieved within each substrate using the native RosettaLigand scoring function. Wild type PNP and PNP-Y88F were each correctly predicted as the most active mutants for inosine and ddI, respectively. Inaccuracies are observed in the relative prediction of substrate activity for each mutant and in the prediction of binding energies for mutants with charged residues. Re-weighting the score terms in the Rosetta energy function corrects the gross errors in prediction of substrate turnover, but at the expense of ranking ability. In particular, the reweighting results point to poor handling of electrostatic interactions, accounted for by the addition of a Generalized Borne solvation term, and difficulties in ranking of highly similar substrates. Despite this, the success in ranking is indicative of the methods applicability in systems where design is limited to a small number of amino acid positions. hPNP-Y88F has a 9-fold improved k_{cat} and 2-fold K_M over the wild type enzyme.

To further improve the ddI turnover rate, hPNP-Y88F became the template for a directed evolution study. A high-throughput screen was developed to improved variants from a library of mutant enzymes, which were generated by error-prone PCR (epPCR). Upon transformation of the library into *E. coli* BL21(DE3) for protein expression, individual colonies were picked into 96-well plates and grown for 24 hours in LB media. The cells were replicated, pelleted and lysed by addition of lysozyme and a single freeze/thaw cycle. The cell-free extract was transferred to a 384-well plate and assayed for ddI phosphorolysis activity. Hits were re-grown from glycerol stocks and validated. A secondary screen normalized the turnover rate to cell density and aided in the removal

of false positives. Kinetic characterization of the mutants identified those with improved binding and substrate turnover number. The mutant with the most improved catalytic efficiency was carried into the next round as the template for mutagenesis and selection.

After three rounds of directed evolution a mutant was identified with 36-fold improved turnover in *E. coli* extracts when compared to wild type hPNP, and 3-fold compared to hPNP-Y88F. Of note is a 2-fold improvement in K_m identified after the first round of directed evolution. Subsequent rounds of directed evolution did not improve the catalytic efficiency despite increases in turnover in cell free extracts. This, and the location of acquired mutations, indicates that directed evolution identified mutants with a greater concentration of active enzyme, likely brought about by increased expression levels, protein stabilization or stabilization of mRNA transcripts. Overall, the final mutant is 22-fold improved in phosphorolysis and 36-times more active in cell free extracts.

It is not necessarily the case that improvements in phosphorolysis efficiency will translate to improvements in the rate of nucleoside analog synthesis. To test this scenario dideoxyribose 1-phosphate was synthesized *in situ* through isomerization of dideoxyribose 5-phosphate with PPM and the rate of ddI synthesis compared for wild type PNP and PNP-46D6. The mutant enzyme possesses a 52-fold improved selectivity ratio in the synthesis direction. Since the selectivity ratio is 46-fold improved in the phosphorolysis direction, in this case improvements in one direction did replicate to the reverse direction.

Enzyme engineering studies are aided by basic knowledge of an enzymes structure and mechanism. In particular, a map of the interactions between substrate and product, coming in the form of substrate-bound crystal structures, is desirable. Structural studies of both ribokinase and PNP are abundant and substrate-bound crystal structures are available for each. In contrast, very little literature exists on the structure of PPMs Based on sequence comparison, the enzyme was classified as a member of the alkaline

phosphatase superfamily and early studies with *E. coli* PPM reinforced the classification, as the enzyme is dependent on divalent metal cations for activity. Like the PPM from *E. coli*, *B. cereus* PPM activity is dependent on Mn^{2+} and crystallization of the enzyme confirmed that the protein contains the core domain shared by alkaline phosphatase super-family members. In addition to the core domain a cap domain is formed by residues 102-216 and is unrelated to any known fold. Soaking of *B. cereus* PPM crystals in a solution containing ribose 5-phosphate revealed the presence of two substrate binding modes. The first utilizes a phosphate binding pocket distal to the catalytic center made up of cap domain residues while the second is coordinates the divalent metion ions and is positioned for reaction with Thr-85. The two positions, along with the orientation of monomers in the asymmetric unit, indicate a significant potential for domain re-orientation throughout catalysis. Significant plasticity would be required to allow for positional rearrangement of the bisphosphate intermediate during enzyme turnover. Hydrogen bond contacts between the enzyme and ribose sugar are limited to the upper substrate orientation and occur between the 3-OH of ribose and Ser-154 and between the furanose oxygen and Arg-193.

Of note is the requirement for phosphorylation of the catalytic nucleophile prior to turnover of ribose phosphates. This requirement is in stark contrast to all other alkaline phosphatase superfamily members which have been described to date. Together, the activation by sugar bisphosphates and presence of a stable phosphothreonine intermediate indicate the use of an *inter*molecular transferase mechanism whereas cofactor-independent phosphoglycerate mutase, a member of the same superfamily, proceeds through *intra*molecular transfer. Isotopically labeled ribose 5-phosphates were synthesized enzymatically from universally labeled [$^{13}C_5$]ribose and γ-[$^{18}O_4$]ATP to test the transferase mechanism. These studies confirmed an *inter*molecular catalytic cycle for bacterial phosphopentomutases.

Crystallization of PPM with ribose 5-phosphate allowed for identification of Arg-193 and Ser-154 as likely participants in orientation of substrate. To supplement this data PPM was crystallized with dideoxyribose 5-phosphate bound in the active site. These structures reveal an inability of PPM to orient the sugar for catalysis, as evidenced by decreased electron density for the sugar portion of the molecule. In addition, a rotation of the sugar toward Ile-195 and Val-158 occurs, which serves to decrease interactions between polar sidechains, specifically Ser-154, and the hydrophobic methylenes of dideoxyribose 5-phosphate. This alternative binding mode of dideoxyribose 5-phosphate represents a non-productive interaction resulting in an 82-fold decrease in k_{cat} and a 7.5-fold decrease in K_m.

Mutation of Ser-154 to alanine causes a shift in orientation of the dideoxyribose 5-phosphate furanose ring. Although turnover activity for dideoxyribose 5-phosphate did not correspondingly increase with the change in substrate orientation, ribose 5-phosphate activity decreased 4-fold. This change in activity increases the selectivity of the mutant for the unnatural substrate and serves as a first step in engineering of PPM.

Finally, to explain the poor phosphorylation of dideoxyribose by ribokinase dideoxyribose was docked into the ribokinase active site with RosettaLigand. These results indicate the importance of hydrogen bonding to substrate orientation. Dideoxyribose adopts a number of conformations in an effort to satisfy the many hydrogen bond donors and acceptors. In particular, Asp-16 is involved in misorientation of dideoxyribose 5-phosphate through formation of hydrogen bonds to the 1- and 5-OH group and by disfavoring orientations that favor turnover through unfavorable polar/non-polar interactions with the C2- and C3-methylenes. Mutation of Asp-16, and additionally Gly-41, may decrease the size of the binding pocket, thereby selecting for the smaller substrate and orienting it more appropriately for phosphorylation.

Significance

Pathway and enzyme engineering

This work demonstrates the potential for directed biosynthesis of nucleoside analogs, in particular dideoxyinosine. Biocatalytic production of ddI has social implications in its potential to decrease the manufacturing price of ddI. Synthesis of nucleoside analogs via biocatalysis benefits from the inherent stereo- and regiospecificity of enzymes. Classical synthetic methods for nucleoside analogs require many protecting group manipulations and are plagued by the generation of racemic or enantiomerically impure products which require chromatographic or other purifications. It also provides for processes which utilize environmentally friendly reaction conditions, such as ambient temperature and pressure and physiological pH, in an aqueous environment without the necessity of toxic or expensive catalysts. Altogether, these process improvements alleviate the large waste streams generated through classical chemical synthesis of pharmaceuticals. It is estimated that the pharmaceutical industry uses between 25 kg and 100 kg of solvent per kg of drug produced[1]. By reducing the amount of solvent waste generated in production, the pharmaceutical industry can save millions of dollars in solvent purchase and disposal expenses. Additionally, the proposed pathway utilizes a cheap starting material. Glutamic acid is a commodity material available in the dollar per kg range. Current manufacturing techniques likely utilize adenosine or inosine as the primary starting material. In comparison, adenosine and inosine have a median price of several hundred dollars per kg (data not shown). These cost savings can be passed to persons requiring antiretroviral treatment, many of which live in low-income countries.

This thesis utilizes a bioretrosynthetic analysis for design of the biocatalytic pathway. Additionally, bioretrosynthesis is introduced as an engineering paradigm. The work presented here provides a strong base for testing of this novel pathway

engineering mechanism. For example, subsequent work hypothesized that increased concentrations of PPM would be required to maximize pathway activity when PNP-46D6 is incorporated into the pathway (work by William R. Birmingham). In fact, the contrary is true! Because the affinity of PNP-46D6 is improved for the dideoxy substrate congener, dideoxyribose 1-phosphate precursor pools need not accumulate for similar activities to increase alleviating the requirement for increased concentrations of the penultimate protein. In terms of pathway flux optimization in a microbial factory the improved K_m will reduce the probability of intermediate diffusion and loss of starting materials to other metabolic processes.

It has been demonstrated that PPM is structurally homologous to alkaline phosphatase superfamily members. However, PPM was found to utilize a catalytic cycle that is different from other alkaline phosphatase superfamily members despite possessing the same core mechanism as the other superfamily members (covalent nucleophilic catalysis stabilized by two divalent cations). This observation has basic science implications. Understanding how enzymes have evolved a specific function has been a topic of study for decades. A common practice is the extrapolation of a common mechanism from a well-studied enzyme to evolutionarily related enzymes. This work serves as an example of a unique function evolved in an evolutionarily related superfamily of proteins. The PPM phosphotransfer cycle is most similar to that of α- and β-phosphoglucomutases which are unrelated in both sequence and structure and is an interesting case of convergent evolution.

Computational docking studies were a major part of Chapters 2 and 5. Very few enzyme re-design studies have been reported. The work presented in Chapter 2 benchmarks the RosettaLigand algorithm and provides confidence in the results of subsequent studies. Furthermore, computational studies enable quick prediction of enzyme-substrate interactions and analysis of mutant activities. Their utility lies in an

ability to quickly and cheaply identify positive research directions and also intractable lines of inquiry.

Best practices in directed evolution

Several significant conclusions can be drawn from the meta-analysis of directed evolution studies presented in Chapter 1. Directed evolution studies containing comparable kinetic data over the last 10 years were surprisingly sparse. However, despite a relatively small sample from which to draw conclusions, the summary provided by this study hints at some putative best practices to consider when implementing contemporary directed evolution methodologies:

Identify a progenitor enzyme acting on a substrate as similar as possible to the new substrate. This point is somewhat facile, but most of the new substrates identified by the retrospective analysis in Table 1 are closely structurally related to the parent enzyme's natural substrate. In many cases, natural enzymes performing a desired chemistry on a target substrate may be unavailable. In such cases, the next best option is to select a progenitor enzyme performing a desired functional group transformation.

As previously noted, saturation mutagenesis methods are over-represented in the high-fold improvement bins, pointing to a greater potential in targeting particular sites. When possible, it appears that structure or homology based site directed mutagenesis methods should be attempted first. When structural data for progenitor enzymes is unavailable, error prone PCR and gene shuffling are the next alternative.

Expect to use combinations of mutagenesis methods in a single campaign. While it is clear that mutation of active site residues provides a clear benefit, it is also clear that distal mutations serve to improve the activity (and/or stability) of enzymes.

Recombine identified mutations using site-directed mutagenesis and gene shuffling[2, 3]. While individual mutations are rarely 100% additive, recombination

frequently provides additional improvement in activity. Furthermore, recombination of the top mutants enables further mining of already-screened libraries.

The steep triage from over 1200 articles to only 81 qualifying cases is primarily reflective of a lack of uniformity in reporting standards for enzymes improved by directed evolution studies. To provide clearer data to guide improvements, it would also be helpful for practitioners of directed evolution studies to report complete biochemical kinetic data for progenitor enzymes and throughout each selection exercise.

Where biocatalysts for a particular reaction system are not available in the natural repertoire, directed evolution methodologies allow for generation of new enzymes with non-native reaction characteristics such as increased rates of catalysis or altered substrate and/or product profiles while functioning in a wide range of reaction conditions. As drug development continues to rely on the generation of chemically complex small molecules, the development of biocatalysts for regio- and stereoselective production of these complex entities will become an important tool for the medicinal and process chemist.

The retrospective assessment provided by this perspective aims to assist researchers interested in applying existing directed evolution methods for the development of new biocatalysts for transformation of non-native substrates. It is evident that directed evolution methodologies continue to contribute to significant improvements in enzyme catalysis, although a single and generalized route towards the most improved enzyme is frequently unclear. The numeric assessments provided here may be useful in expectations management for both investigators and reviewers of directed evolution studies. The disparity between the average and median fold improvements in parameters is striking, suggesting that the median value may represent a more equitable standard for performance improvements. However, numeric assessments fail to fully capture the successes and failures of directed evolution methodologies. Comprehensive

knowledge of the failures, which may or may not outnumber the reported successes, is lacking but regardless this would reduce the median fold improved values further. However, the success from a biocatalytic perspective must ultimately not be defined by fold improvements in biochemical parameters but by the attainment of goals for a given study. Based on this broad survey, it was clear that these criteria were met by the majority of campaigns reported in recent literature, underlining further the ongoing contribution of directed evolution methodologies to engineering of biocatalysts for new and existing substrates.

Future Directions

A chemo-enzymatic pathway for ddI production was constructed which generates the target molecule. However, the low levels of turnover leave much room for improvement. Rates of nucleoside phosphorolysis and synthesis were improved for hPNP but turnover rates are still significantly lower than those of inosine. This activity may be further improved through subsequent rounds of directed evolution. It is possible, though, that equivalent activities will not be reached with ddI: the hydrophobic nature of the molecule may repel approach by phosphate or purine base. Furthermore, the dideoxyribose moiety may be inherently less reactive.

The dataset provided here, and the plethora of binding and turnover data available for hPNP, hPNP mutants and homologs, may enable extension of computational prediction of beneficial mutations to residues distal to the binding pocket. Here, mutagenesis of residues >15 Å from the binding site affect substrate turnover preference. Similar datasets are available for binding of transition state analogs[4]. Furthermore, computational bench-marking to this rich data set will advance the understanding of RosettaLigand as it relates to complex, multimeric systems which utilize multiple substrates and generate multiple products.

The efficiency of PPM turnover of dideoxyribose 5-phosphate is 620-fold worse than ribose 5-phosphate. Crystallographic studies have suggested Ser-154, Arg-193, Ile-195 and Val-158 as targets for mutagenesis. Mutation of Ser-154 to alanine increased the selectivity for dideoxyribose 5-phosphate by decreasing turnover of the natural substrate without significantly affecting the turnover rate of the target molecule. Saturation mutagenesis at this position, and the others, will help to identify the optimal residue identity. It is possible that other amino acids may orient the furanose portion of dideoxyribose 5-phosphate in a more catalytically active conformation.

Dideoxyribose phosphorylation by ribokinase is currently the rate limiting step of the biocatalytic pathway. Docking studies with RosettaLigand pointed to Asp-16 and Gly-41 as potential targets for mutagenesis. Because of their proximity, amino acid substitutions may be mutually beneficial or mutually exclusive. Simultaneous saturation mutagenesis should be performed to identify the optimal combination of residues at these positions. A step-wise approach may lead to selection of an amino acid at the first position which precludes a better combination from occurring. Alternatively, computational design of the active site can be used to predict combinations of mutations which stabilize a productive binding orientation.

A primary advantage of microbial factories is the stable environment they provide for enzyme function. It is possible that low pathway turnover is a consequence of enzyme degradation during the long incubation time of the reaction. Incorporation of the pathway into *E. coli* and feeding of dideoxyribose will demonstrate production of the compound *in vivo* and is a major step toward high-volume production of ddI. Furthermore, extension of the pathway to a simpler precursor will improve the viability of the proposed chemo-enzymatic pathway. Currently, formation of dideoxyribose requires reduction at -78 °C, a temperature which is impractical and expensive on production scale, to stop reduction at the lactol. To alleviate this requirement, (*S*)-γ-butyrolactone-γ-

carboxylic acid could under reduction to (S)-1,4,5-pentanetriol, or dideoxyribitol. Subsequent formation of dideoxyribose could be accomplished by enzymatic oxidation with mannitol 1-dehydrogenase (EC 1.1.1.255). This enzyme has been previously been evolved to generate L-ribose from ribitol[5]. Further evolution would generate an enzyme useful in the current pathway.

Finally, this pathway can be applied to the synthesis of numerous other nucleoside analogs, such as azidothymidine, vidaribine, clofarabine, or apricitabine to name a few. Evolution of pyrimidine nucleoside phosphorylases or nucleosidyltransferases would provide access to each of the natural nucleobases. In addition to the natural bases, many nucleoside phosphorylases, nucleosidyltransferases, phosphoribosyltransferases possess off target activities for unnatural bases. It is likely that directed biosynthesis of these compounds will streamline the synthesis of these compounds, in addition to increasing their availabity.

References

1. Ritter, S. K., Greening up process chemistry. *Chem. Eng. News* **2010,** 88, (43), 45-47.

2. Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H., Evolutionary history of a specialized p450 propane monooxygenase. *J. Mol. Biol.* **2008,** 383, (5), 1069-80.

3. Park, S. H.; Park, H. Y.; Sohng, J. K.; Lee, H. C.; Liou, K.; Yoon, Y. J.; Kim, B. G., Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* **2009,** 102, (4), 988-94.

4. Saen-Oon, S.; Ghanem, M.; Schramm, V. L.; Schwartz, S. D., Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase. *Biophys. J.* **2008,** 94, (10), 4078-88.

5. Christ, T. N.; Deweese, K. A.; Woodyer, R. D., Directed evolution toward improved production of L-ribose from ribitol. *Comb. Chem. High Throughput Screening* **2010,** 13, (4), 302-308.

Appendix A

DIRECTED EVOLUTION METHODS

Library Generation and Mutagenesis

Random mutagenesis methods aim to identify mutations throughout the gene coding region and include error prone polymerase chain reaction (epPCR), chemical mutagenesis, UV irradiation, and mutating bacterial strains. The most common is epPCR,[1] of which there are several variations. Typical early experiments used an error prone, thermostable polymerase lacking proofreading ability, such as *Taq*, to amplify a gene of interest, incorporating random mutations along the entire length of the gene.[2] The error rate of the polymerase alone was titrated through addition of Mn^{2+} to stabilize mismatched base pairs, and by manipulation of the number of thermocycles and template and dNTP concentrations. More recently, polymerases have been engineered to possess a higher inherent mutation rate, with less bias (e.g. balancing transitions vs. transversions) than wild-type *Taq* polymerase; Mutazyme (Stratagene Inc.) is frequently used, commercially available engineered error-prone polymerase. Alternatively, the addition of unnatural nucleotide analogs to the PCR reaction gives a similar end-result. Once incorporated, alternative pairings are generated as the nucleotide analogs are capable of pairing with multiple canonical nucleotides.[3] Similarly, chemical mutagens may be employed to modify template DNA and foster mutation. Reaction of template DNA with nitrous acid, formic acid, hydrazine or ethyl methane sulfonate can result in chemical changes in nucleotide bases, altering their hydrogen bonding properties and increasing the propensity for the formation of non-canonical pairings.[4, 5] Mutator strains and whole cell mutation by UV irradiation have been used to generate mutations *in vivo*.[6] In this way, low yield ligation steps can be avoided but at the expense of possibly

deleteriously mutating the host genome, resistance genes, copy number control sequences or promoter regions of the heterologous plamids. To balance the bias of transitions (A→G, T→C) and transversions (A/G↔ C/T), some researchers have used multiple epPCR methods thereby increasing the mutational diversity in the library.[7-9] Wong and coworkers performed a statistical analysis of the mutational spectrum achieved through the application of 19 random mutagenesis experiments to three genes. Using this data they developed a web application, the Mutagenesis Assist Program (MAP), to aid researchers in choosing an optimal method for a given target gene.[2] The MAP analysis takes into account the codon usage of the template and, for each random mutagenesis method, provides data on the expected number of stop codons, number of available mutations from each codon, incorporation of destabilizing glycines or prolines, number of preserved amino acids and the propensity for mutation to amino acids with alternative chemical properties.

The preceding “random” mutagenesis methods can facilitate the identification of advantageous mutations or tunable sites located throughout a protein sequence and can be successfully applied in the absence of structural data. Saturation mutagenesis protocols test all 20 amino acids at a targeted site or sites and enable a controlled and comprehensive search of mutational sequence space via degenerate oligonucleotide primers. Although saturation mutagenesis can be applied to every position in a gene, as in Gene Site Saturation Mutagenesis,[10] the method is most commonly targeted to potentially beneficial sites using structural data or through homology modeling. Saturation mutagenesis can be applied to a single site, to multiple sites in an iterative fashion, or to multiple sites simultaneously. Cassette mutagenesis,[11] involves simultaneous saturation mutagenesis of multiple, proximal residues. In this category, Reetz and coworkers introduced a method dubbed the combinatorial active-site saturation test, or CASTing, in 2005.[12] Most often, CASTing has been applied to alter the

shape of binding sites for the purposes of changing the enantioselectivity of an enzyme.[13] Overall, saturation mutagenesis methods have the advantage of substantially reducing library size requirements by focusing on a subset of 'hotspots' either within the active site or as identified by broader sequence scans (i.e. random mutagenesis).

The term 'gene shuffling' refers to the *in vitro* recombination of template progenitor genes and is analogous to homologous recombination and sexual PCR. It ostensibly allows for generation of chimeric sequences more diverse than standard epPCR without resorting to high polymerase-dependent mutation rates which lead to the incorporation of deletory mutations in a high fraction of variants.[14] The progenitor genes may be comprised of either naturally occurring homologs of a gene family, or selected mutants of a single gene generated by other mutagenesis methods. The archetypal DNA shuffling method introduced by Stemmer and coworkers shuffles homologous DNA sequences by generating oligonucleotide fragments from selected progenitor genes by treatment with DNAse, followed by reassembly of the mixed oligonucleotides by PCR.[15] Since then, a number of other protocols have been developed for homologous recombination. In StEP (Staggered Extension Protocol), homologous DNA templates are mixed together with one or more primers, and subjected to very short annealing and extension steps.[16] Methods for recombination of non-homologous sequences include incremental truncation of hybrid enzymes (ITCHY and SCRATCHY) and sequence homology-independent recombination (SHIPREC).[17-19]

Frequently, beneficial mutations generated by one of the aforementioned methods (random mutagenesis, site-directed mutagenesis, gene shuffling) are subsequently combined into a single construct to further optimize enzyme function. By generating these combinations, unproductive and/or neutral mutations may also be identified and removed. This can be accomplished by gene shuffling, long used for this purpose,[15] or stepwise site-directed mutagenesis to deconstruct the role of each

mutation. It is apparent that most successful directed evolution campaigns utilize multiple forms of mutagenesis, both iteratively and sometimes concurrently in a given study, with some exceptional studies performing in excess of 20 rounds using a broad array of methods.[20]

Identification of improved variants

In selection-based methods, a desired enzymatic activity, such as amino acid synthesis, is linked to host cell survival through growth under selective conditions, such as in minimal medium lacking amino acid. Other typical examples include the synthesis of an essential amino acid[21-23] or nucleotide[18, 19] or degradation of an environmental toxin.[15, 24-31] Positive selections can be performed on agar plates or rely on enrichment of cells harboring enzymes of the desired activity in a liquid culture.[32] Negative selection methods, in which the desired activity results in death, often entail that cells must be sorted and replicated into both library reference and assay microtiter plates to facilitate backtracking and resurrection of the desired library cultures, thereby making them lower throughput.[33, 34] A notable exception is the evolution of tRNA synthetase activity for incorporation of unnatural amino acids, in which negative selection was performed to dial out undesired activities.[35, 36]

An ideal selection scheme involves engineering an organism to be dependent on reproduction via the presence of endogenously biosynthesized small molecule in a concentration dependent manner. A potential solution to this challenging problem was described by Schwimmer and coworkers[37] using a yeast two-hybrid system to evolve a receptor, capable of transcriptional activation of an essential gene, to bind a new small molecule ligand. Specifically, a nuclear receptor retinoid X receptor that naturally binds 9-cis-retinoic acid was engineered to bind a synthetic retinoid-like compound LG335 with an improvement in EC_{50} from 10 μM to 40 nM. In an alternative approach, ligand binding

protein switches have been generated by engineering protein hybrids with resistance genes.[38]

Challenges to the development of a successful positive selection scheme are not limited to the molecular recognition engineering problem. Of critical importance is the level of response to the ligand concentration. Once a ligand reaches a concentration sufficient for organism survival, selective pressure will diminish; therefore, it is more useful that the response occurs in a concentration dependent manner rather than as a switch. Moreover, ligands need to be constrained to the intracellular environment to be of maximal utility and must not be toxic to the host organism. Despite these challenges, *in vivo* methods allow for assessment of large numbers of variants. In fact, the throughput of selection methods is usually only limited by ligation and transformation efficiencies.

In screening-based methods, the rate of substrate turnover is determined via biochemical assay, often via colorimetric or fluorometric measurements. *In vitro* (cell-free) activity assays generally require the isolation of individual colonies into 96- or 384-well plates, followed by lysis and multiple reagent transfers. Thus, screening-based methods are limited by cost and time to less than 10^6 mutants (10^3-10^4 is more common). As an exception, *in vivo* screening-based methodologies can be applied to larger mutant libraries and it is the linking of the desired biochemical activity to a phenotype that limits the realistic library size. Examples of *in vivo* activity assay techniques include agar plate colony screening[9, 39-48], flow cytometric cell sorting[49-54] or phage display[55, 56], and result in larger libraries on the order of 10^7-10^{10} mutant constructs. Like selection methods, the throughput of these *in vivo* methods is solely limited by ligation and transformation efficiency.

Some colorimetric screens can be modified and applied to agar plate colony screening. In these methods, colonies displaying a desired color within a certain time

window are selected as active. Frequently, the colorimetric agar plate assay is used as a primary screen to identify enzyme variants with functional turnover. Subsequently, a secondary and sometimes tertiary activity assay is performed in a 96-well microtiter plate biochemical assay to confirm and identify the most active clones.[57-59] These cases benefit from the larger library sizes assayable via the initial agar plate screen before encountering the time and cost limitations of single colony isolation techniques.

Flow cytometry and phage display methods permit the assay of larger mutant libraries. Fluorescence activated cell sorting, or FACS, is a type of flow cytometry in which a mixture of cells is sorted, one cell at a time, based upon the specific light scattering and fluorescent characteristics of each cell. Systems amenable to FACS must operate intracellularly, which often requires that substrates of the screen be cell permeable and the products impermeable. To address this requirement, FACS methods utilizing *in vitro* compartmentalization (IVC) allow the compartmentalization of DNA, transcriptase, ribosomes and fluorescent substrates/products in artificial water-oil emulsions thereby linking and rendering sortable product formation, enzyme, and the enzyme encoding gene.[54]

In phage display screening, the DNA of a gene library is fused to a viral coat protein gene, and combined in a phagemid (a vector containing both plasmid and phage origins of replication). The phagemid DNA then becomes packaged in the phage capsid that displays the protein. The phage are subjected to chromatographic separation (affinity binding) to select for desired activity.[60] Very rarely is phage display connected directly to enzyme activity although an increase in a biosynthetic enzyme activity can be identified through tighter binding to a tethered transition state analog.[61] Other display methods, such as cell surface display, provide the advantage of a large library size in addition to facile subsequent screening capability.[62]

As it is challenging to link enzymatic reactions to cell viability it is also difficult to contrive reporter systems that induce a change in color or fluorescence *in vivo*, or in compartments. Such systems often require an enzymatic assay for activity in cell lysates from colonies isolated in microtiter plates. Although these assays require the growth of individual cultures for assay, there is no limit to the varieties of instrumental analysis that can be performed, and most assays can be implemented in a 96-well format and auto-sampling technologies. Examples of these methods include analytical HPLC,[63] mass spectrometry,[25] TLC, [64, 65] radioactivity,[66] and UV/Vis spectroscopy (e.g. NADH reduction)[67-69].

References

1. Wong, T. S.; Zhurina, D.; Schwaneberg, U., The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screening* **2006,** 9, (4), 271-288.

2. Wong, T. S.; Roccatano, D.; Zacharias, M.; Schwaneberg, U., A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol.* **2006,** 355, (4), 858-71.

3. Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E., An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **1996,** 255, (4), 589-603.

4. Lai, Y. P.; Huang, J.; Wang, L. F.; Li, J.; Wu, Z. R., A new approach to random mutagenesis *in vitro*. *Biotechnol. Bioeng.* **2004,** 86, (6), 622-627.

5. Myers, R. M.; Lerman, L. S.; Maniatis, T., A general-method for saturation mutagenesis of cloned DNA fragments. *Science* **1985,** 229, (4710), 242-247.

6. Balashov, S.; Humayun, M. Z., Specificity of spontaneous mutations induced in mutA mutator cells. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2004,** 548, (1-2), 9-18.

7. Patrick, W. M.; Matsumura, I., A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **2008,** 377, (2), 323-336.

8. Roodveldt, C.; Tawfik, D. S., Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **2005,** 44, (38), 12728-12736.

9. Rowe, L. A.; Geddie, M. L.; Alexander, O. B.; Matsumura, I., A comparison of directed evolution approaches using the β-glucuronidase model system. *J. Mol. Biol.* **2003,** 332, (4), 851-60.

10. 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.; Burk, M. J., Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.* **2003,** 125, (38), 11476-11477.

11. Reidhaar-Olson, J. F.; Sauer, R. T., Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* **1988,** 241, (4861), 53-57.

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

13. Reetz, M. T., Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chem., Int. Ed.* **2010,** doi: 10.1002/anie.201000826.

14. Crameri, A.; Raillard, S. A.; Bermudez, E.; Stemmer, W. P., DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **1998,** 391, (6664), 288-291.

15. Stemmer, W. P. C., DNA shuffling by random fragmentation and reassembly - *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** 91, (22), 10747-10751.

16. Zhao, H. M.; Giver, L.; Shao, Z. X.; Affholter, J. A.; Arnold, F. H., Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* **1998,** 16, (3), 258-261.

17. Sieber, V.; Martinez, C. A.; Arnold, F. H., Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **2001,** 19, (5), 456-460.

18. Ostermeier, M.; Shim, J. H.; Benkovic, S. J., A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **1999,** 17, (12), 1205-1209.

19. Lutz, S.; Ostermeier, M.; Moore, G. L.; Maranas, C. D.; Benkovic, S. J., Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** 98, (20), 11248-53.

20. 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.; Huisman, G. W., Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **2007,** 25, (3), 338-344.

21. Akanuma, S.; Yamagishi, A.; Tanaka, N.; Oshima, T., Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. *Protein Sci.* **1998,** 7, (3), 698-705.

22. Rothman, S. C.; Kirsch, J. F., How does an enzyme evolved *in vitro* compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J. Mol. Biol.* **2003,** 327, (3), 593-608.

23. Yano, T.; Oue, S.; Kagamiyama, H., Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** 95, (10), 5511-5515.

24. Barak, Y.; Ackerley, D. F.; Dodge, C. J.; Banwari, L.; Alex, C.; Francis, A. J.; Matin, A., Analysis of novel soluble chromate and uranyl reductases and generation of an improved enzyme by directed evolution. *Appl. Environ. Microbiol.* **2006,** 72, (11), 7074-7082.

25. 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.; Lassner, M. W., Discovery and directed evolution of a glyphosate tolerance gene. *Science* **2004,** 304, (5674), 1151-1154.

26. Cho, C. M. H.; Mulchandani, A.; Chen, W., Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Appl. Environ. Microbiol.* **2004,** 70, (8), 4681-4685.

27. Claren, J.; Malisi, C.; Hocker, B.; Sterner, R., Establishing wild-type levels of catalytic activity on natural and artificial (β/α)(8)-barrel protein scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (10), 3704-3709.

28. Gulick, A. M.; Fahl, W. E., Forced evolution of glutathione-S-transferase to create a more efficient drug detoxication enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** 92, (18), 8140-8144.

29. Hoseki, J.; Yano, T.; Koyama, Y.; Kuramitsu, S.; Kagamiyama, H., Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus*. *J. Biochem.* **1999,** 126, (5), 951-956.

30. Landis, D. M.; Loeb, L. A., Random sequence mutagenesis and resistance to 5-fluorouridine in human thymidylate synthases. *J. Biol. Chem.* **1998,** 273, (40), 25809-25817.

31. Walter, K. U.; Vamvaca, K.; Hilvert, D., An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* **2005,** 280, (45), 37742-37746.

32. Zhang, K. C.; Li, H.; Cho, K. M.; Liao, J. C., Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** 107, (14), 6234-6239.

33. Black, M. E.; Newcomb, T. G.; Wilson, H. M.; Loeb, L. A., Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, (8), 3525-9.

34. Christians, F. C.; Scapozza, L.; Crameri, A.; Folkers, G.; Stemmer, W. P., Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* **1999,** 17, (3), 259-264.

35. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G., Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** 99, (17), 11020-4.

36. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of *Escherichia coli*. *Science* **2001,** 292, (5516), 498-500.

37. Schwimmer, L. J.; Rohatgi, P.; Azizi, B.; Seley, K. L.; Doyle, D. F., Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, (41), 14707-14712.

38. Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M., Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, (32), 11224-11229.

39. Bosma, T.; Damborsky, J.; Stucki, G.; Janssen, D. B., Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl. Environ. Microbiol.* **2002,** 68, (7), 3582-7.

40. Carter, B. T.; Lin, H.; Goldberg, S. D.; Althoff, E. A.; Raushel, J.; Cornish, V. W., Investigation of the mechanism of resistance to third-generation cephalosporins by class C β-lactamases by using chemical complementation. *ChemBioChem* **2005,** 6, (11), 2055-67.

41. Cheon, Y. H.; Park, H. S.; Kim, J. H.; Kim, Y.; Kim, H. S., Manipulation of the active site loops of D-hydantoinase, a (β/α)(8)-barrel protein, for modulation of the substrate specificity. *Biochemistry* **2004,** 43, (23), 7413-7420.

42. Delagrave, S.; Murphy, D. J.; Pruss, J. L. R.; Maffia, A. M.; Marrs, B. L.; Bylina, E. J.; Coleman, W. J.; Grek, C. L.; Dilworth, M. R.; Yang, M. M.; Youvan, D. C., Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. *Protein Eng.* **2001,** 14, (4), 261-267.

43. Iffland, A.; Gendreizig, S.; Tafelmeyer, P.; Johnsson, K., Changing the substrate specificity of cytochrome c peroxidase using directed evolution. *Biochem. Biophys. Res. Commun.* **2001,** 286, (1), 126-132.

44. Cihlar, T.; Ray, A. S., Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **2010,** 85, (1), 39-58.

45. Nakagawa, Y.; Hasegawa, A.; Hiratake, J.; Sakata, K., Engineering of *Pseudomonas aeruginosa* lipase by directed evolution for enhanced amidase activity: mechanistic implication for amide hydrolysis by serine hydrolases. *Protein Eng. Des. Sel.* **2007,** 20, (7), 339-346.

46. Nakazawa, H.; Okada, K.; Onodera, T.; Ogasawara, W.; Okada, H.; Morikawa, Y., Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Appl. Microbiol. Biotechnol.* **2009,** 83, (4), 649-657.

47. Park, S. H.; Park, H. Y.; Sohng, J. K.; Lee, H. C.; Liou, K.; Yoon, Y. J.; Kim, B. G., Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* **2009,** 102, (4), 988-94.

48. Zhang, Z. G.; Liu, Y.; Guengerich, F. P.; Matse, J. H.; Chen, J.; Wu, Z. L., Identification of amino acid residues involved in 4-chloroindole 3-hydroxylation by cytochrome P450 2A6 using screening of random libraries. *J. Biotechnol.* **2009,** 139, (1), 12-18.

49. Antipov, E.; Cho, A. E.; Wittrup, K. D.; Klibanov, A. M., Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (46), 17694-17699.

50. Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D. S., High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **2005,** 12, (12), 1281-1289.

51. Griswold, K. E.; Aiyappan, N. S.; Iverson, B. L.; Georgiou, G., The evolution of catalytic efficiency and substrate promiscuity in human theta class 1-1 glutathione transferase. *J. Mol. Biol.* **2006,** 364, (3), 400-410.

52. Griswold, K. E.; Kawarasaki, Y.; Ghoneim, N.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G., Evolution of highly active enzymes by homology-independent recombination. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, (29), 10082-10087.

53. Liu, L. F.; Li, Y. F.; Liotta, D.; Lutz, S., Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res.* **2009,** 37, (13), 4472-4481.

54. Mastrobattista, E.; Taly, V.; Chanudet, E.; Treacy, P.; Kelly, B. T.; Griffiths, A. D., High-throughput screening of enzyme libraries: *in vitro* evolution of a β-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* **2005,** 12, (12), 1291-1300.

55. Droge, M. J.; Boersma, Y. L.; van Pouderoyen, G.; Vrenken, T. E.; Ruggeberg, C. J.; Reetz, M. T.; Dijkstra, B. W.; Quax, W. J., Directed evolution of *Bacillus subtilis* lipase A by use of enantiomeric phosphonate inhibitors: crystal structures and phage display selection. *ChemBioChem* **2006,** 7, (1), 149-157.

56. Sunbul, M.; Marshall, N. J.; Zou, Y. K.; Zhang, K. Y.; Yin, J., Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. *J. Mol. Biol.* **2009,** 387, (4), 883-898.

57. Matsumura, I.; Wallingford, J. B.; Surana, N. K.; Vize, P. D.; Ellington, A. D., Directed evolution of the surface chemistry of the reporter enzyme β-glucuronidase. *Nat. Biotechnol.* **1999,** 17, (7), 696-701.

58. Schmidt-Dannert, C., Engineering novel carotenoids in microorganisms. *Curr. Opin. Biotechnol.* **2000,** 11, (3), 255-261.

59. Schmidt-Dannert, C.; Umeno, D.; Arnold, F. H., Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.* **2000,** 18, (7), 750-3.

60. Hida, K.; Hanes, J.; Ostermeier, M., Directed evolution for drug and nucleic acid delivery. *Adv. Drug Delivery Rev.* **2007,** 59, (15), 1562-78.

61. Fernandez-Gacio, A.; Uguen, M.; Fastrez, J., Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **2003,** 21, (9), 408-414.

62. Jose, J., Autodisplay: efficient bacterial surface display of recombinant proteins. *Appl. Microbiol. Biotechnol.* **2006,** 69, (6), 607-14.

63. Hibbert, E. G.; Senussi, T.; Costelloe, S. J.; Lei, W. L.; Smith, M. E. B.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase activity on non-phosphorylated substrates. *J. Biotechnol.* **2007,** 131, (4), 425-432.

64. Ohki, T.; Shibata, N.; Higuchi, Y.; Kawashima, Y.; Takeo, M.; Kato, D.; Negoro, S., Two alternative modes for optimizing nylon-6 byproduct hydrolytic activity from a carboxylesterase with a beta-lactamase fold: X-ray crystallographic analysis of directly evolved 6-aminohexanoate-dimer hydrolase. *Protein Sci.* **2009,** 18, (8), 1662-1673.

65. Wei, C. L.; Yang, Y. B.; Deng, C. H.; Liu, W. C.; Hsu, J. S.; Lin, Y. C.; Liaw, S. H.; Tsai, Y. C., Directed evolution of *Streptomyces clavuligerus* deacetoxycephalosporin C synthase for enhancement of penicillin G expansion. *Appl. Environ. Microbiol.* **2005,** 71, (12), 8873-8880.

66. Bhuiya, M.-W.; Liu, C.-J., Engineering monolignol 4-O-methyltransferases to modulate lignin biosynthesis. *J. Biol. Chem.* **2010,** 285, (1), 277-285.

67. Hsu, C. C.; Hong, Z.; Wada, M.; Franke, D.; Wong, C. H., Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. *Proc Natl Acad Sci U S A* **2005,** 102, (26), 9122-6.

68. Williams, G. J.; Domann, S.; Nelson, A.; Berry, A., Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** 100, (6), 3143-8.

69. Woodhall, T.; Williams, G.; Berry, A.; Nelson, A., Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics. *Angew. Chem., Int. Ed.* **2005,** 44, (14), 2109-2112.

Appendix B

TABLE OF STATISTICS FOR SELECTED DIRECTED EVOLUTION STUDIES

Table B-1. ■ Experimental details of the selected directed evolution studies included in directed evolution meta-analysis. Random mutagenesis – experiments which mutate the whole gene in the absence of guiding criteria. ■ Saturation mutagenesis – iterative and/or simultaneous application of saturation mutagenesis. ■ Gene Shuffling – Use of recombinatorial methods to generate diversity. ■ Step-wise method application – cases where random mutagenesis, saturation mutagenesis, or gene shuffling were used iteratively. ■ Saturation mutagenesis of "hotspots" –probing of positions known to affect function, identified through stochastic methods. ■ Combination of improved variants – recombination of variants through site-directed mutagenesis or gene shuffling. ▲ Selection or plate-based screens use an agar plate to enable high throughput screening of enzyme variants ▲ Assays which require growth of individual variants in liquid cultures ▲ Experiments utilizing fluorescence activated cell sorting ▲ Enrichment includes liquid phase selection of improved variants and enrichment by phage display ▲ Colorimetric agar plate-based screens which require a secondary or tertiary assay to distinguish the most. ■ Studies which enable small-molecule synthesis ■ Evolutionary or other scientific studies ■ Generation of enzymes for detoxification of harmful agents

Ref.	EC	Enzyme	Natural Substrate	New Substrate	Mut. Protocol	Screen. Method	Intent of Study	Fold Improved Rate	Fold Improved K_m	Fold Improved Catalytic Efficiency
[1]	1.1	Choline oxidase	choline	tris-(2-hydroxyethyl)-methylammonium methylsulfate	■ ■ ■ ■	▲	■	2.69	3.24	8.72
[2]	1.1	Pyranose-2-oxidase	glucose	galactose	■	▲	■	0.15	21.98	3.33
[3]	1.1	Vanillyl-alcohol oxidase	vanillyl alcohol	creosol	■	▲	■	7	1.54	10.77
[4]	1.1	Galactose Oxidase	galactose	guar	■ ■	▲	■	4.88	3.94	19.2
[5]	1.1	β-Isopropylmalate dehydrogenase	NAD	NADP	■	▲	■	1.36	61.31	83.45
[6]	1.11	Cytochrome C peroxidase	cytochrome c	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)	■ ■	▲	■	9.45		
[7]	1.11	Bromoperoxidase A2	*para*-nitropheylacetate	*para*-nitrophenyl caprylate	■ ■	▲	■	140.23		
[8]	1.11	Lignin peroxidase	lignin	2,4-dichlorophenol	■ ■ (■)	▲	■	8.52	0.48	4.05
[9]	1.11	Horseradish peroxidase	tyrosinol	L-tyrosinol	■ ■ ■ (■)	▲	■	32.69		
[9]	1.11	Horseradish peroxidase	tyrosinol	D-tyrosinol	■ ■ ■ (■)	▲	■	6.16		
[10]	1.13	4-Methyl-5-nitrocatechol monooxygenase	4-methyl-5-nitrocatechol	4-nitrophenol	■ ■	▲	■	3.21	3.5	11.25

11	1.13	Extradiol dioxygenase	1,2-dihydroxynaphthalene	3,4-dihydroxybiphenyl	■ ■ ■	▲	■	2.38	315.79	750
12	1.14	Toluene dioxygenase	toluene	4-picoline	■ ■ (■)	▲	■	5.6		
13	1.14	Deacetoxycephalosporin synthase	penicillin N	penicillin G	■ ■	▲	■	3.09	13.58	41.91
14	1.14	Aniline dioxygenase	aniline	2,4-dimethylaniline	■ ■ (■)	▲	■	3.5		
15	1.14	P450 1A2	heterocyclic aromatic amines	2-amino-3,5-dimethylimidazo[4.5.f]quinoline	■	▲	■	4.5	2.6	11.7
16	1.14	P450 BM-3	fatty acids and alkanes	12-*p*-nitrophenyl carboxylic acid	■	▲	■	novel	novel	novel
17	1.14	P450 2A6	indole	4-chloroindole	■ ■	▲	■	0.68	7.52	5.15
18	1.14	P450 2B1	NADPH	H_2O_2	■	▲	■	6		
19	1.4	Phenylalanine dehydrogenase	phenylalanine	DL-propargylglycine	■ ■ (■)	▲	■	1.27	5.78	7.4
20	1.4	D-Amino acid oxidase	D-alanine	D-aspartate	■	▲	■	1.78	3.38	6
20	1.4	D-Amino acid oxidase	D-alanine	D-arginine	■	▲	■	1.38	1.17	1.6
21	1.4	Glutamate Dehydrogenase	2-ketoglutarate	2-ketoisovalerate	■	▲	■	4.53	1.05	4.75
22	1.6	Chromate reductase	chromium (VI)	uranium (VI)	■	▲	■	11.41	0.14	1.6
23	2.1	(Iso)eugenol 4-O-methyltransferase	(Iso)eugenol	coniferyl alcohol	■	▲	■	8.83	8.02	70.8
24	2.2	Transketolase	xylulose-5-phosphate / ribose-5-phosphate	β-Hydroxypyruvate / glycolaldehyde	■	▲	■	4.83		
25	2.2	Transketolase	Xylulose-5-phosphate / ribose-5-phosphate	β-Hydroxypyruvate / propionaldehyde	■	▲	■	8.92	2.55	22.7
26	2.3	γ-Glutamyltranspeptidase	glutamine	glutaryl-7-ACA	■ ■	▲	■	18.28	2.64	48.19
27	2.3	Simvastatin synthase	LovF-ACP-α-S-methyl butyrate	α-dimethylbutyryl-S-methyl-mercaptoproprionate	■ ■ ■	▲	■	7.27	0.97	7.1
28	2.3	N-Acetyltransferase	unknown	glyphosate	■	▲	■	416	24	9984
29	2.4	Cyclodextrin glucanotransferase	short polysaccharides	soluble starch	■ ■ ■ ■	▲	■	10.9		
30	2.4	Kanamycin/Vancomycin glycosyl transferases (chimera)	kanamycin or vancomycin / dTDP glucose	2-deoxystreptamine / TDP-glucose	■	▲	■			4
31	2.4	Glycosynthase XynB2 (E335G mutant)	α-D-xylopyranosyl-fluoride	α-D-xylopyranosyl-fluoride	■	▲	■	35.15	0.06	2.1

32	2.4	Glutamine phosphoribosylpyrophosphate amidotransferase	5-phosphoribosyl-1-pyrophosphate	phosphoribosylanthranilate	■	▲	■	30		
33	2.4	Cellobiose phosphorylase	cellobiose	lactose	■ ■	▲	■	10.18	0.69	7.04
34	2.4	Purine nucleoside phosphorylase	inosine	dideoxyinosine	■ ■ (■)	▲	■	4.99	4.38	21.91
35	2.5	Glutathione transferase	1-menaphthyl sulfate/dichloromethane	7-amino-4-chloromethyl coumarin	■ ■ ■ ■ (■)	▲	■	844.44	22.71	19176
36	2.5	Glutathione transferase	herbicides	flurodifen	■	▲	■	29.25		
37	2.5	Glutathione transferase	1-menaphthyl sulfate/dichloromethane	7-amino-4-chloromethyl coumarin	■	▲	■	258.43	2.89	748
38	2.6	Aspartate aminotransferase	aspartate	tyrosine	■	▲	■			130.91
39	2.6	Aspartate aminotransferase	hydroxyphenylpyruvate	hydroxyphenylpyruvate	■ ■	▲	■	1.49	3.1	4.6
40	2.7	Allokinase	allose	glucose	■	▲	■	1.3	4.46	5.8
40	2.7	N-Acetyl-D-mannosamine kinase	N-acetyl-D-mannosamine	glucose	■	▲	■	3.82	3.13	11.9
41	2.7	Deoxythymidine kinase	deoxythymidine	azidothymidine	■	▲	■	1.44	1.15	1.7
42	2.7	Phosphopantetheinyl transferase	Coenzyme A	3′-dephospho-Coenzyme A	■	▲	■			329.33
43	2.7	Deoxynucleoside kinase	thymidine	fluorescent dideoxythymidine	■ ■ ■ (■)	▲	■	0.56	9.93	5.5
44	3	Phosphotriesterase-homology protein	unknown	2-naphthylacetate	■ ■ ■	▲	■			44.29
45	3	E. coli Ebg	unknown	fluorescein di-β-D-galactopyranoside	■ ■	▲	■			3941
46	3.1	Organophosphorous hydrolase	paraoxon	chlorpyrifos	■	▲	■	418.13	1.86	777
44	3.1	Phosphotriesterase	Paraoxon	2-naphthylacetate	■ ■ ■	▲	■			12.92
47	3.1	Lipase A	(R)-1,2-O-isopropylidene-sn-glycerol butyrate	(S)-1,2-O-isopropylidene-sn-glycerol butyrate	■	▲	■	0.96	1.75	1.7
48	3.1	Serum paraoxonase	L-homocysteine thiolactone	γ-thiobutyrolactone	■	▲	■			101.42
49	3.1	Phosphotriesterase	organophasphates	ethyl paraoxon	■	▲	■	104	5.36	557
50	3.1	Organophosphorous hydrolase	paraoxon	methyl parathion	■ (■)	▲	■	17.22	1.97	34
51	3.1	Lipase	2-naphthyl oleate	N-(2-naphthyl)oleamide	■ ■	▲	■	2.2		
52	3.1	Lipase	2-naphthyl oleate	N-(2-naphthyl)oleamide	■ ■	▲	■			27.59

53	3.1	Carboxylesterase	unknown	6-aminohexanoate-linear-dimer					5.25	19.55	840.97
54	3.2	Endo-β-1,4-glucanase	β-1,4-glucosidic bonds of amorphous cellulose	carboxymethyl-cellulose					2.55		
55	3.2	Endoglucanase	β-1,4-glucosidic bonds of amorphous cellulose	carboxymethyl-cellulose	(■)				1.32	0.5	0.66
56	3.2	β-Galactosidase	*para*-nitrophenyl-β-D-galactopyranoside	*para*-nitrophenyl-β-D-fucopyranoside					215	0.27	58.89
57	3.2	β-Glucuronidase	*para*-nitrophenyl-β-D-glucopyranoside	*para*-nitrophenyl-β-D-galactopyranoside							15.65
58	3.2	β-Glucuronidase	glucoronide	*para*-nitrophenyl--D-xylopyranoside					63	3.27	206.2
59	3.2	β-Glucosidase	cellotetraose	cellobiose					1.44	0.99	1.44
60	3.4	DD-Transpeptidase	D-ala-D-ala	cefotaxime	(■)				17822.97		
61	3.5	D-Hydantoinase	hydantoin	D,L-hydroxyphenylhydantoin					11.11	0.42	4.6
62	3.5	β-Lactamase	β-lactams	cefotaxime					3083	0.0381	1175
63	3.5	Glutaryl acylase	glutaryl-7-ACA	adipyl-7-ADCA					5.8	6.2	36
64	3.5	Glutaryl acylase	glutaryl-7-ACA	adipyl-7-ADCA	(■)				2.02	1.41	2.85
65	3.5	Glutaryl acylase	glutaryl-7-ACA	adipyl-7-ADCA					1.2	6.86	8.2
66	3.5	Mandelamide hydrolase	mandelamide	lactamide	(■)				3.84	0.6	2.29
67	3.5	Metallo-β-lactamase	β-lactams	Cephalexin					12.41	0.6	7.42
68	3.5	Glutaryl acylase	glutaryl-7-ACA	cephalosporin C					4.29	0.9	3.86
69	3.8	Haloalkane dehalogenase	haloalkanes	1,2,3-trichloropropane	(■)				3.5	2.2	7.7
70	4.1	Sialic acid aldolase	sialic acid	ammonium (4R,5R,6R)-6-(dipropylcarbamoyl)-2,4,5-trihydroxytetrahydro-2H-pyran-2-carboxylate					1.76	28.21	49.55
71	4.1	KDPGal Aldolase	glyceraldehyde-3-phosphate	D-erythrose 4-phosphate					5.21	11.63	60.64
72	4.1	Sialic acid aldolase	D-sialic acid	3-deoxy-manno-2-octulosonic acid					9.45	2.68	25.3
73	4.1	Tagatose-1,6-bisphosphate aldolase	product: tagatose bisphosphate	product: fructose bisphosphate					8.78	8.13	71.34
74	4.2	Carbonic Anhydrase II	HCO_3^-	2-naphthylacetate	(■)						39
75	5.3	Phosphoribosyl anthranilate isomerase	PROFAR	PRA					800		
76	5.3	Phosphoribosylanthranilate isomerase	PROFAR	PRA					novel	novel	novel

77	5.5	Muconate lactonizing enzyme	cis-cis-muconate	2-hydroxysuccinyl-2,4-cyclohexadiene carboxylate	■	▲	■	123456.79

References

1. Ribitsch, D.; Winkler, S.; Gruber, K.; Karl, W.; Wehrschutz-Sigl, E.; Eiteljorg, I.; Schratl, P.; Remler, P.; Stehr, R.; Bessler, C.; Mussmann, N.; Sauter, K.; Maurer, K. H.; Schwab, H., Engineering of choline oxidase from *Arthrobacter nicotianae* for potential use as biological bleach in detergents. *Appl. Microbiol. Biotechnol.* **2010,** 87, (5), 1743-1752.

2. Spadiut, O.; Pisanelli, I.; Maischberger, T.; Peterbauer, C.; Gorton, L.; Chaiyen, P.; Haltrich, D., Engineering of pyranose 2-oxidase: Improvement for biofuel cell and food applications through semi-rational protein design. *J. Biotechnol.* **2009,** 139, (3), 250-257.

3. van den Heuvel, R. H. H.; van den Berg, W. A. M.; Rovida, S.; van Berkel, W. J. H., Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *J. Biol. Chem.* **2004,** 279, (32), 33492-33500.

4. Delagrave, S.; Murphy, D. J.; Pruss, J. L. R.; Maffia, A. M.; Marrs, B. L.; Bylina, E. J.; Coleman, W. J.; Grek, C. L.; Dilworth, M. R.; Yang, M. M.; Youvan, D. C., Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. *Protein Eng.* **2001,** 14, (4), 261-267.

5. Miller, S. P.; Lunzer, M.; Dean, A. M., Direct demonstration of an adaptive constraint. *Science* **2006,** 314, (5798), 458-461.

6. Iffland, A.; Gendreizig, S.; Tafelmeyer, P.; Johnsson, K., Changing the substrate specificity of cytochrome c peroxidase using directed evolution. *Biochem. Biophys. Res. Commun.* **2001,** 286, (1), 126-132.

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

8. Ryu, K.; Hwang, S. Y.; Kim, K. H.; Kang, J. H.; Lee, E. K., Functionality improvement of fungal lignin peroxidase by DNA shuffling for 2,4-dichlorophenol degradability and H2O2 stability. *J. Biotechnol.* **2008,** 133, (1), 110-115.

9. Antipov, E.; Cho, A. E.; Wittrup, K. D.; Klibanov, A. M., Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (46), 17694-17699.

10. Leungsakul, T.; Johnson, G. R.; Wood, T. K., Protein engineering of the 4-methyl-5-nitrocatechol monooxygenase from *Burkholderia* sp strain DNT for enhanced degradation of nitroaromatics. *Appl. Environ. Microbiol.* **2006,** 72, (6), 3933-3939.

11. Fortin, P. D.; MacPherson, I.; Neau, D. B.; Bolin, J. T.; Eltis, L. D., Directed evolution of a ring-cleaving dioxygenase for polychlorinated biphenyl degradation. *J. Biol. Chem.* **2005,** 280, (51), 42307-42314.

12. Sakamoto, T.; Joern, J. M.; Arisawa, A.; Arnold, F. H., Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate. *Appl. Environ. Microbiol.* **2001,** 67, (9), 3882-7.

13. Wei, C. L.; Yang, Y. B.; Deng, C. H.; Liu, W. C.; Hsu, J. S.; Lin, Y. C.; Liaw, S. H.; Tsai, Y. C., Directed evolution of *Streptomyces clavuligerus* deacetoxycephalosporin C synthase for enhancement of penicillin G expansion. *Appl. Environ. Microbiol.* **2005,** 71, (12), 8873-8880.

14. Ang, E. L.; Obbard, J. P.; Zhao, H. M., Directed evolution of aniline dioxygenase for enhanced bioremediation of aromatic amines. *Appl. Microbiol. Biotechnol.* **2009,** 81, (6), 1063-1070.

15. Kim, D.; Guengerich, F. P., Selection of human cytochrome P450 1A2 mutants with enhanced catalytic activity for heterocyclic amine N-hydroxylation. *Biochemistry* **2004,** 43, (4), 981-8.

16. Nazor, J.; Schwaneberg, U., Laboratory evolution of P450BM-3 for mediated electron transfer. *ChemBioChem* **2006,** 7, (4), 638-644.

17. Zhang, Z. G.; Liu, Y.; Guengerich, F. P.; Matse, J. H.; Chen, J.; Wu, Z. L., Identification of amino acid residues involved in 4-chloroindole 3-hydroxylation by cytochrome P450 2A6 using screening of random libraries. *J. Biotechnol.* **2009,** 139, (1), 12-18.

18. Kumar, S.; Chen, C. S.; Waxman, D. J.; Halpert, J. R., Directed evolution of mammalian cytochrome P4502B1. *J. Biol. Chem.* **2005,** 280, (20), 19569-19575.

19. Chen, S. H.; Engel, P. C., Efficient screening for new amino acid dehydrogenase activity: directed evolution of *Bacillus sphaericus* phenylalanine dehydrogenase towards activity with an unsaturated non-natural amino acid. *J. Biotechnol.* **2009,** 142, (2), 127-134.

20. Sacchi, S.; Rosini, E.; Molla, G.; Pilone, M. S.; Pollegioni, L., Modulating D-amino acid oxidase substrate specificity: production of an enzyme for analytical determination of all D-amino acids by directed evolution. *Protein Eng. Des. Sel.* **2004,** 17, (6), 517-25.

21. Zhang, K. C.; Li, H.; Cho, K. M.; Liao, J. C., Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** 107, (14), 6234-6239.

22. Barak, Y.; Ackerley, D. F.; Dodge, C. J.; Banwari, L.; Alex, C.; Francis, A. J.; Matin, A., Analysis of novel soluble chromate and uranyl reductases and generation of an improved enzyme by directed evolution. *Appl. Environ. Microbiol.* **2006,** 72, (11), 7074-7082.

23. Bhuiya, M.-W.; Liu, C.-J., Engineering monolignol 4-O-methyltransferases to modulate lignin biosynthesis. *J. Biol. Chem.* **2010,** 285, (1), 277-285.

24. Hibbert, E. G.; Senussi, T.; Costelloe, S. J.; Lei, W. L.; Smith, M. E. B.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase activity on non-phosphorylated substrates. *J. Biotechnol.* **2007,** 131, (4), 425-432.

25. Hibbert, E. G.; Tarik, S. A.; Mark, E.; Costelloe, S. J.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J. Biotechnol.* **2008,** 134, (3-4), 240-245.

26. Yamada, C.; Kijima, K.; Ishihara, S.; Miwa, C.; Wada, K.; Okada, T.; Fukuyama, K.; Kumagai, H.; Suzuki, H., Improvement of the glutaryl-7-aminocephalosporanic acid acylase activity of a bacterial gamma-glutamyltranspeptidase. *Appl. Environ. Microbiol.* **2008,** 74, (11), 3400-3409.

27. Gao, X.; Xie, X. K.; Pashkov, I.; Sawaya, M. R.; Laidman, J.; Zhang, W. J.; Cacho, R.; Yeates, T. O.; Tang, Y., Directed Evolution and Structural Characterization of a Simvastatin Synthase. *Chem. Biol.* **2009,** 16, (10), 1064-1074.

28. 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.; Lassner, M. W., Discovery and directed evolution of a glyphosate tolerance gene. *Science* **2004,** 304, (5674), 1151-1154.

29. Kelly, R. M.; Leemhuis, H.; Dijkhuizen, L., Conversion of a cyclodextrin glucanotransferase into an alpha-amylase: Assessment of directed evolution strategies. *Biochemistry* **2007,** 46, 11216-11222.

30. Park, S. H.; Park, H. Y.; Sohng, J. K.; Lee, H. C.; Liou, K.; Yoon, Y. J.; Kim, B. G., Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* **2009,** 102, (4), 988-94.

31. Ben-David, A.; Shoham, G.; Shoham, Y., A universal screening assay for glycosynthases: Directed evolution of glycosynthase XynB2(E335G) suggests a general path to enhance activity. *Chem. Biol.* **2008,** 15, (6), 546-551.

32. Patrick, W. M.; Matsumura, I., A study in molecular contingency: Glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **2008,** 377, (2), 323-336.

33. De Groeve, M. R. M.; De Baere, M.; Hoflack, L.; Desmet, T.; Vandamme, E. J.; Soetaert, W., Creating lactose phosphorylase enzymes by directed evolution of cellobiose phosphorylase. *Protein Eng. Des. Sel.* **2009,** 22, (7), 393-399.

34. Nannemann, D. P.; Kaufmann, K. W.; Meiler, J.; Bachmann, B. O., Design and directed evolution of a dideoxy purine nucleoside phosphorylase. *Protein Eng. Des. Sel.* **2010,** 23, (8), 607-616.

35. Griswold, K. E.; Aiyappan, N. S.; Iverson, B. L.; Georgiou, G., The evolution of catalytic efficiency and substrate promiscuity in human theta class 1-1 glutathione transferase. *J. Mol. Biol.* **2006,** 364, (3), 400-410.

36. Dixon, D. P.; McEwen, A. G.; Lapthorn, A. J.; Edwards, R., Forced evolution of a herbicide detoxifying glutathione transferase. *J. Biol. Chem.* **2003,** 278, (26), 23930-5.

37. Griswold, K. E.; Kawarasaki, Y.; Ghoneim, N.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G., Evolution of highly active enzymes by homology-independent recombination. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, (29), 10082-10087.

38. Rothman, S. C.; Kirsch, J. F., How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J. Mol. Biol.* **2003,** 327, (3), 593-608.

39. Rothman, S. C.; Voorhies, M.; Kirsch, J. F., Directed evolution relieves product inhibition and confers in vivo function to a rationally designed tyrosine aminotransferase. *Protein Sci.* **2004,** 13, (3), 763-72.

40. Larion, M.; Moore, L. B.; Thompson, S. M.; Miller, B. G., Divergent evolution of function in the ROK sugar kinase superfamily: Role of enzyme loops in substrate specificity. *Biochemistry* **2007,** 46, (47), 13564-13572.

41. Knecht, W.; Petersen, G. E.; Munch-Petersen, B.; Piskur, J., Deoxyribonucleoside kinases belonging to the thymidine kinase 2 (TK2)-like group vary significantly in substrate specificity, kinetics and feed-back regulation. *J. Mol. Biol.* **2002,** 315, (4), 529-540.

42. Sunbul, M.; Marshall, N. J.; Zou, Y. K.; Zhang, K. Y.; Yin, J., Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. *J. Mol. Biol.* **2009,** 387, (4), 883-898.

43. Liu, L. F.; Li, Y. F.; Liotta, D.; Lutz, S., Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res.* **2009,** 37, (13), 4472-4481.

44. Roodveldt, C.; Tawfik, D. S., Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily. *Biochemistry* **2005,** 44, (38), 12728-12736.

45. Mastrobattista, E.; Taly, V.; Chanudet, E.; Treacy, P.; Kelly, B. T.; Griffiths, A. D., High-throughput screening of enzyme libraries: *in vitro* evolution of a β-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* **2005,** 12, (12), 1291-1300.

46. Cho, C. M. H.; Mulchandani, A.; Chen, W., Altering the substrate specificity of organophosphorus hydrolase for enhanced hydrolysis of chlorpyrifos. *Appl. Environ. Microbiol.* **2004,** 70, (8), 4681-4685.

47. Droge, M. J.; Boersma, Y. L.; van Pouderoyen, G.; Vrenken, T. E.; Ruggeberg, C. J.; Reetz, M. T.; Dijkstra, B. W.; Quax, W. J., Directed evolution of *Bacillus subtilis* lipase A by use of enantiomeric phosphonate inhibitors: crystal structures and phage display selection. *ChemBioChem* **2006,** 7, (1), 149-157.

48. Aharoni, A.; Amitai, G.; Bernath, K.; Magdassi, S.; Tawfik, D. S., High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. *Chem. Biol.* **2005,** 12, (12), 1281-1289.

49. Hawwa, R.; Larsen, S. D.; Ratia, K.; Mesecar, A. D., Structure-based and random mutagenesis approaches increase the organophosphate-degrading activity of a phosphotriesterase homologue from *Deinococcus radiodurans*. *J. Mol. Biol.* **2009,** 393, (1), 36-57.

50. Cho, C. M. H.; Mulchandani, A.; Chen, W., Functional analysis of organophosphorus hydrolase variants with high degradation activity towards organophosphate pesticides. *Protein Eng. Des. Sel.* **2006,** 19, (3), 99-105.

51. Fujii, R.; Nakagawa, Y.; Hiratake, J.; Sogabe, A.; Sakata, K., Directed evolution of *Pseudomonas aeruginosa* lipase for improved amide-hydrolyzing activity. *Protein Eng. Des. Sel.* **2005,** 18, (2), 93-101.

52. Nakagawa, Y.; Hasegawa, A.; Hiratake, J.; Sakata, K., Engineering of *Pseudomonas aeruginosa* lipase by directed evolution for enhanced amidase activity: mechanistic implication for amide hydrolysis by serine hydrolases. *Protein Eng. Des. Sel.* **2007,** 20, (7), 339-346.

53. Ohki, T.; Shibata, N.; Higuchi, Y.; Kawashima, Y.; Takeo, M.; Kato, D.; Negoro, S., Two alternative modes for optimizing nylon-6 byproduct hydrolytic activity from a carboxylesterase with a beta-lactamase fold: X-ray crystallographic analysis of directly evolved 6-aminohexanoate-dimer hydrolase. *Protein Sci.* **2009,** 18, (8), 1662-1673.

54. Lin, L.; Meng, X.; Liu, P. F.; Hong, Y. Z.; Wu, G. B.; Huang, X. L.; Li, C. C.; Dong, J. L.; Xiao, L.; Liu, Z., Improved catalytic efficiency of Endo-β-1,4-glucanase from *Bacillus subtilis* BME-15 by directed evolution. *Appl. Microbiol. Biotechnol.* **2009,** 82, (4), 671-679.

55. Nakazawa, H.; Okada, K.; Onodera, T.; Ogasawara, W.; Okada, H.; Morikawa, Y., Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Appl. Microbiol. Biotechnol.* **2009,** 83, (4), 649-657.

56. Parikh, M. R.; Matsumura, I., Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of β-fucosidase from β-galactosidase. *J. Mol. Biol.* **2005,** 352, (3), 621-628.

57. Rowe, L. A.; Geddie, M. L.; Alexander, O. B.; Matsumura, I., A comparison of directed evolution approaches using the β-glucuronidase model system. *J. Mol. Biol.* **2003,** 332, (4), 851-60.

58. Geddie, M. L.; Matsumura, I., Rapid evolution of β-glucuronidase specificity by saturation mutagenesis of an active site loop. *J. Biol. Chem.* **2004,** 279, (25), 26462-8.

59. McCarthy, J. K.; Uzelac, A.; Davis, D. F.; Eveleigh, D. E., Improved catalytic efficiency and active site modification of 1,4-β-D-glucan glucohydrolase A from *Thermotoga neapolitana* by directed evolution. *J. Biol. Chem.* **2004,** 279, (12), 11495-502.

60. Peimbert, M.; Segovia, L., Evolutionary engineering of a β-Lactamase activity on a D-Ala D-Ala transpeptidase fold. *Protein Eng.* **2003,** 16, (1), 27-35.

61. Cheon, Y. H.; Park, H. S.; Kim, J. H.; Kim, Y.; Kim, H. S., Manipulation of the active site loops of D-hydantoinase, a (β/α)(8)-barrel protein, for modulation of the substrate specificity. *Biochemistry* **2004,** 43, (23), 7413-7420.

62. Carter, B. T.; Lin, H.; Goldberg, S. D.; Althoff, E. A.; Raushel, J.; Cornish, V. W., Investigation of the mechanism of resistance to third-generation cephalosporins by class C β-lactamases by using chemical complementation. *ChemBioChem* **2005,** 6, (11), 2055-67.

63. Otten, L. G.; Sio, C. F.; Reis, C. R.; Koch, G.; Cool, R. H.; Quax, W. J., A highly active adipyl-cephalosporin acylase obtained via rational randomization. *FEBS J.* **2007,** 274, 5600-5610.

64. Sio, C. F.; Riemens, A. M.; van der Laan, J. M.; Verhaert, R. M. D.; Quax, W. J., Directed evolution of a glutaryl acylase into an adipyl acylase. *Eur. J. Biochem.* **2002,** 269, (18), 4495-4504.

65. Otten, L. G.; Sio, C. F.; Vrielink, J.; Cool, R. H.; Quax, W. J., Altering the substrate specificity of cephalosporin acylase by directed evolution of the β-subunit. *J. Biol. Chem.* **2002,** 277, (44), 42121-7.

66. Wang, P. F.; Yep, A.; Kenyon, G. L.; McLeish, M. J., Using directed evolution to probe the substrate specificity of mandelamide hydrolase. *Protein Eng. Des. Sel.* **2009,** 22, (2), 103-110.

67. Tomatis, P. E.; Rasia, R. M.; Segovia, L.; Vila, A. J., Mimicking natural evolution in metallo-β-lactamases through second-shell ligand mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, (39), 13761-13766.

68. Pollegioni, L.; Lorenzi, S.; Rosini, E.; Marcone, G. L.; Molla, G.; Verga, R.; Cabri, W.; Pilone, M. S., Evolution of an acylase active on cephalosporin C. *Protein Sci.* **2005,** 14, (12), 3064-3076.

69. Bosma, T.; Damborsky, J.; Stucki, G.; Janssen, D. B., Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl. Environ. Microbiol.* **2002,** 68, (7), 3582-7.

70. Woodhall, T.; Williams, G.; Berry, A.; Nelson, A., Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics. *Angew. Chem., Int. Ed.* **2005,** 44, (14), 2109-2112.

71. Ran, N. Q.; 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, (19), 6130-6139.

72. Hsu, C. C.; Hong, Z. Y.; Wada, M.; Franke, D.; Wong, C. H., Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** 102, (26), 9122-9126.

73. Williams, G. J.; Domann, S.; Nelson, A.; Berry, A., Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** 100, (6), 3143-8.

74. Gould, S. M.; Tawfik, D. S., Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* **2005,** 44, (14), 5444-5452.

75. Claren, J.; Malisi, C.; Hocker, B.; Sterner, R., Establishing wild-type levels of catalytic activity on natural and artificial (β/α)(8)-barrel protein scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** 106, (10), 3704-3709.

76. Jurgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmanns, M.; Sterner, R., Directed evolution of a (β/α)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** 97, (18), 9925-30.

77. Schmidt, D. M. Z.; Mundorff, E. C.; Dojka, M.; Bermudez, E.; Ness, J. E.; Govindarajan, S.; Babbitt, P. C.; Minshull, J.; Gerlt, J. A., Evolutionary potential of (β/α)(8)-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* **2003,** 42, (28), 8387-8393.

Appendix C

NMR SPECTRA

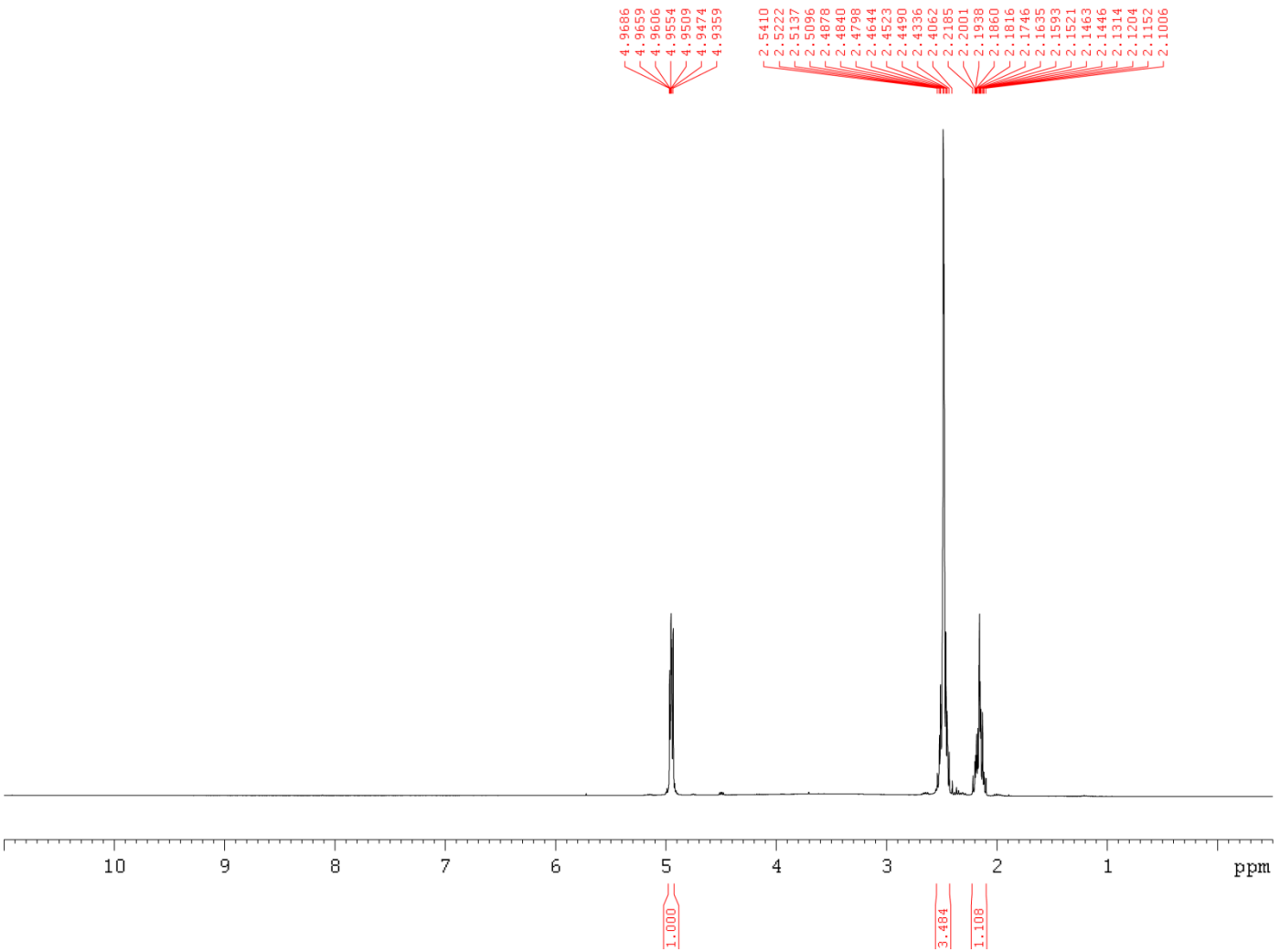

Figure C-1. ^{1}H NMR of (*S*)-γ-butyrolactone-γ-carboxylic acid

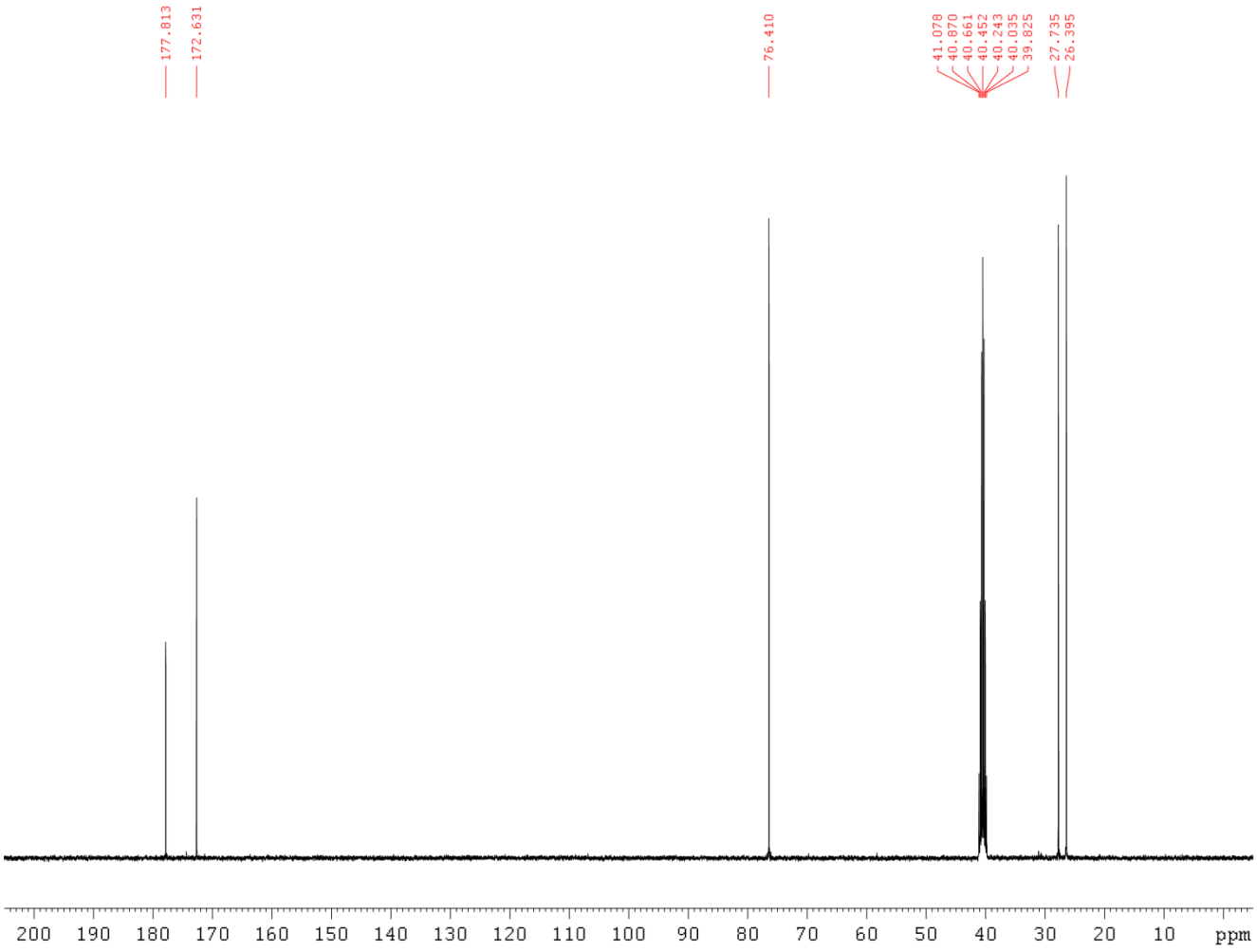

Figure C-2. ^{13}C NMR of (*S*)-γ-butyrolactone-γ-carboxylic acid

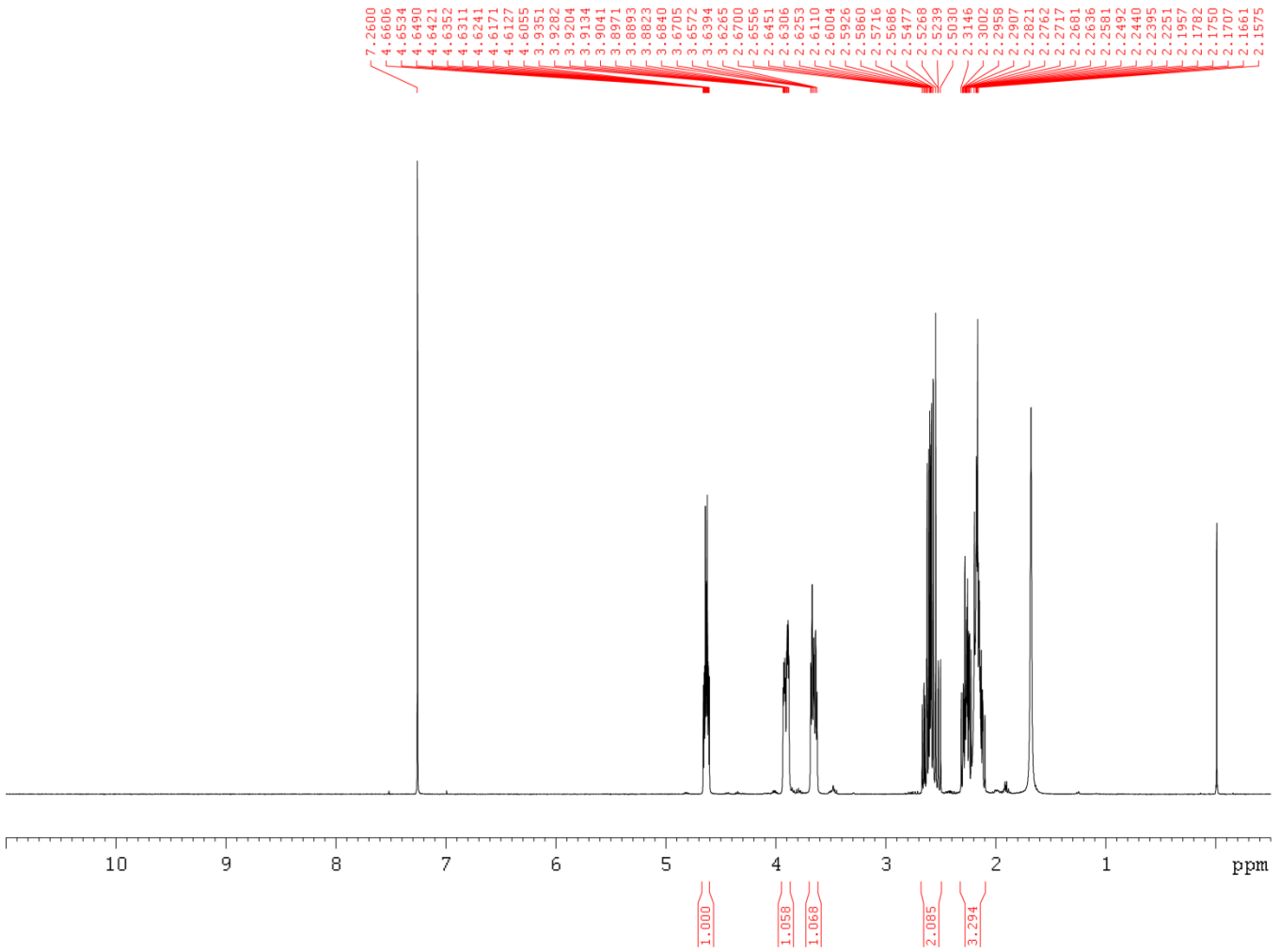

Figure C-3. ^{1}H NMR of (*S*)-γ-hydroxymethyl-γ-butyrolactone

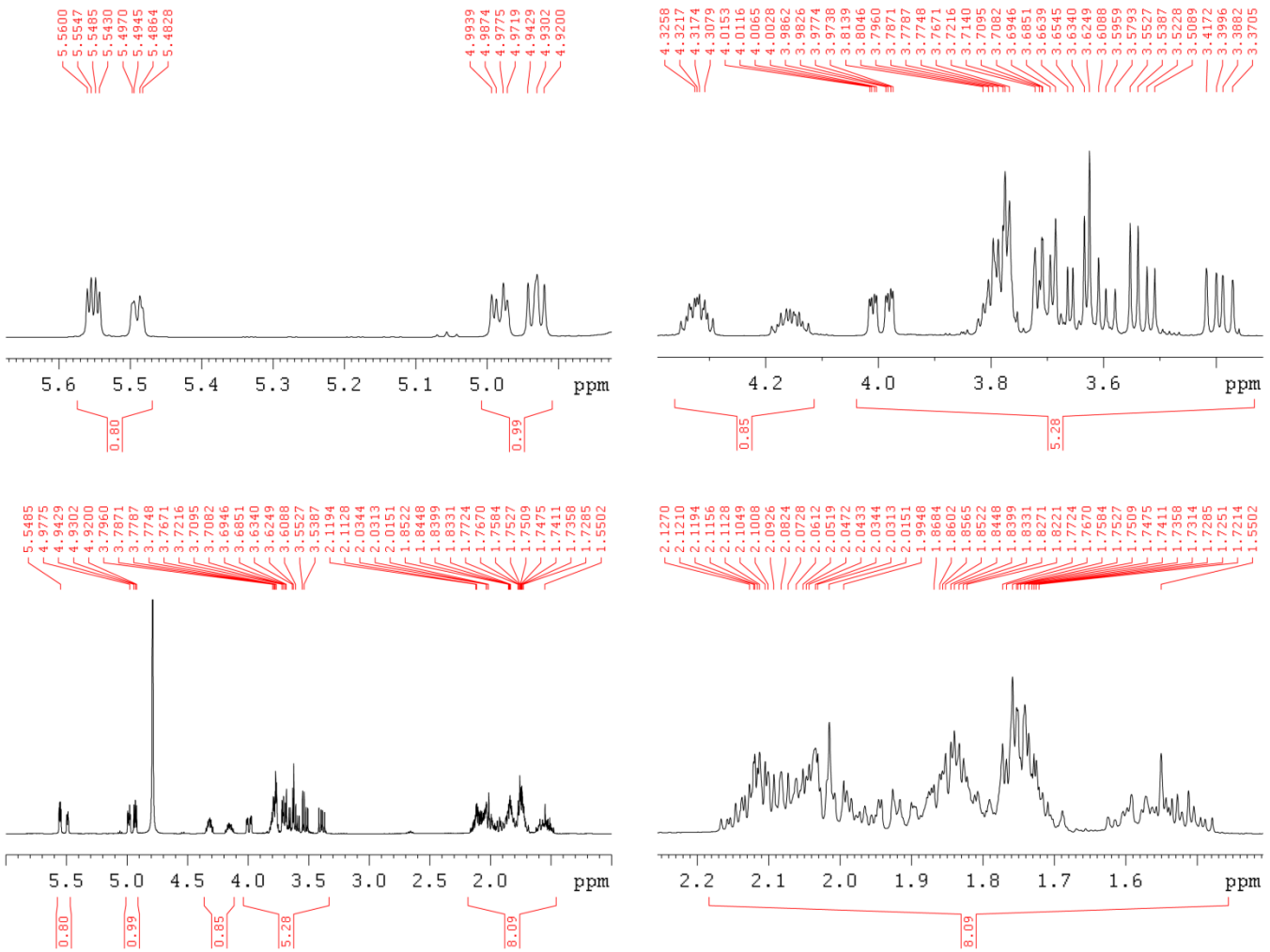

Figure C-4. ^{1}H NMR of (*S*)-γ-hydroxymethyl-γ-butyrolactone

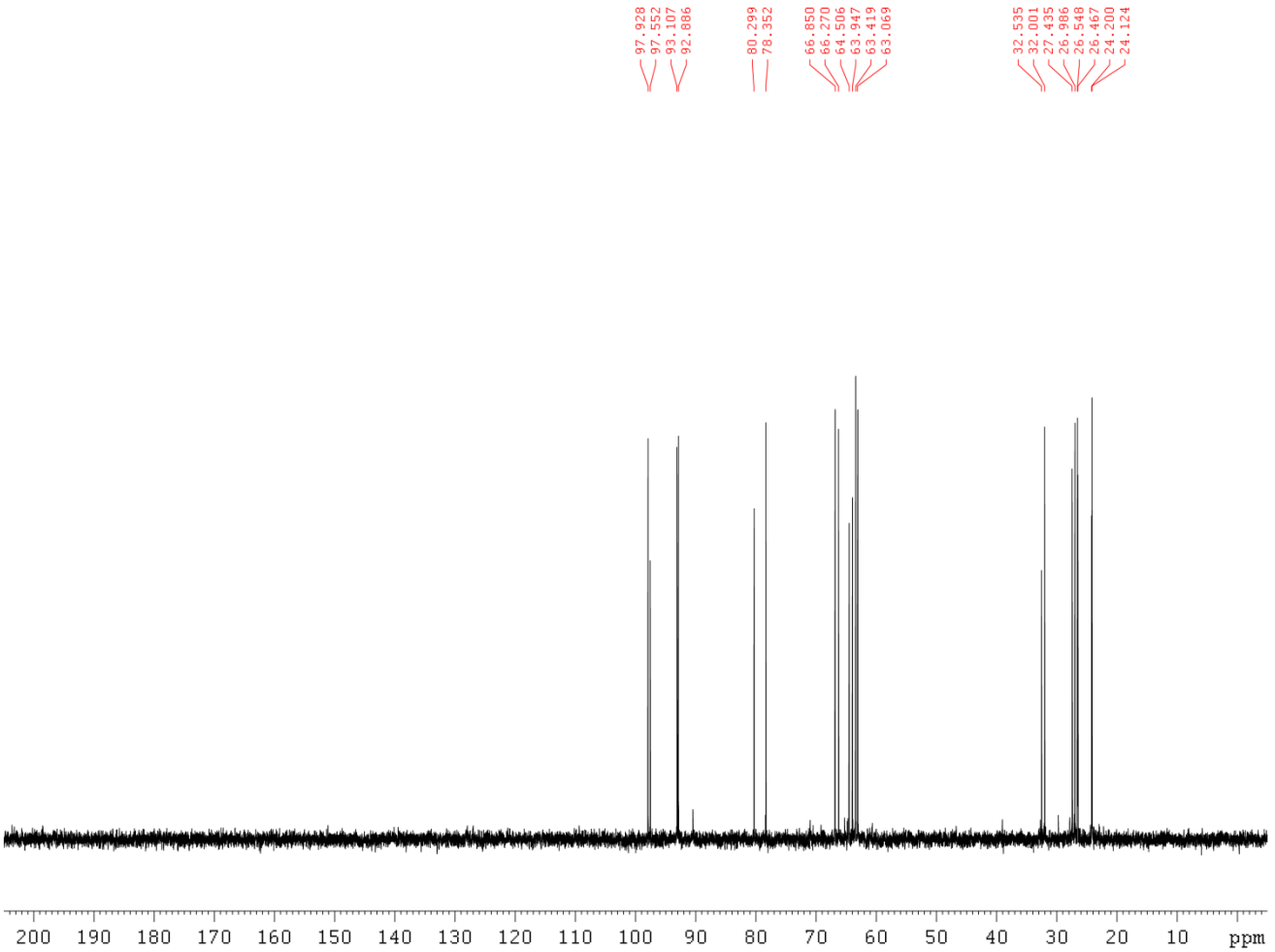

Figure C-5. ^{13}C NMR of (*S*)-γ-hydroxymethyl-γ-butyrolactone

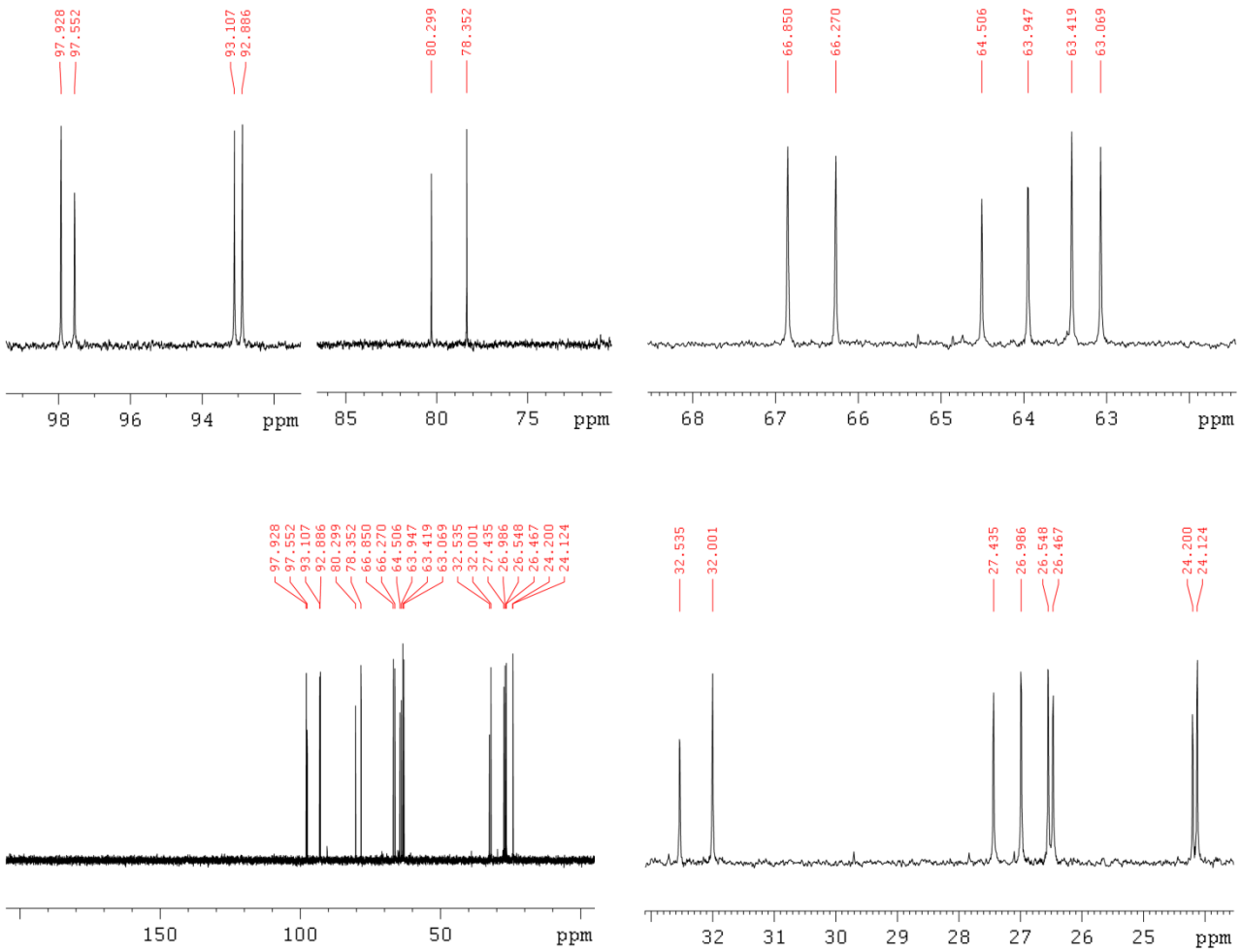

Figure C-6. ^{13}C NMR of (*S*)-γ-hydroxymethyl-γ-butyrolactone

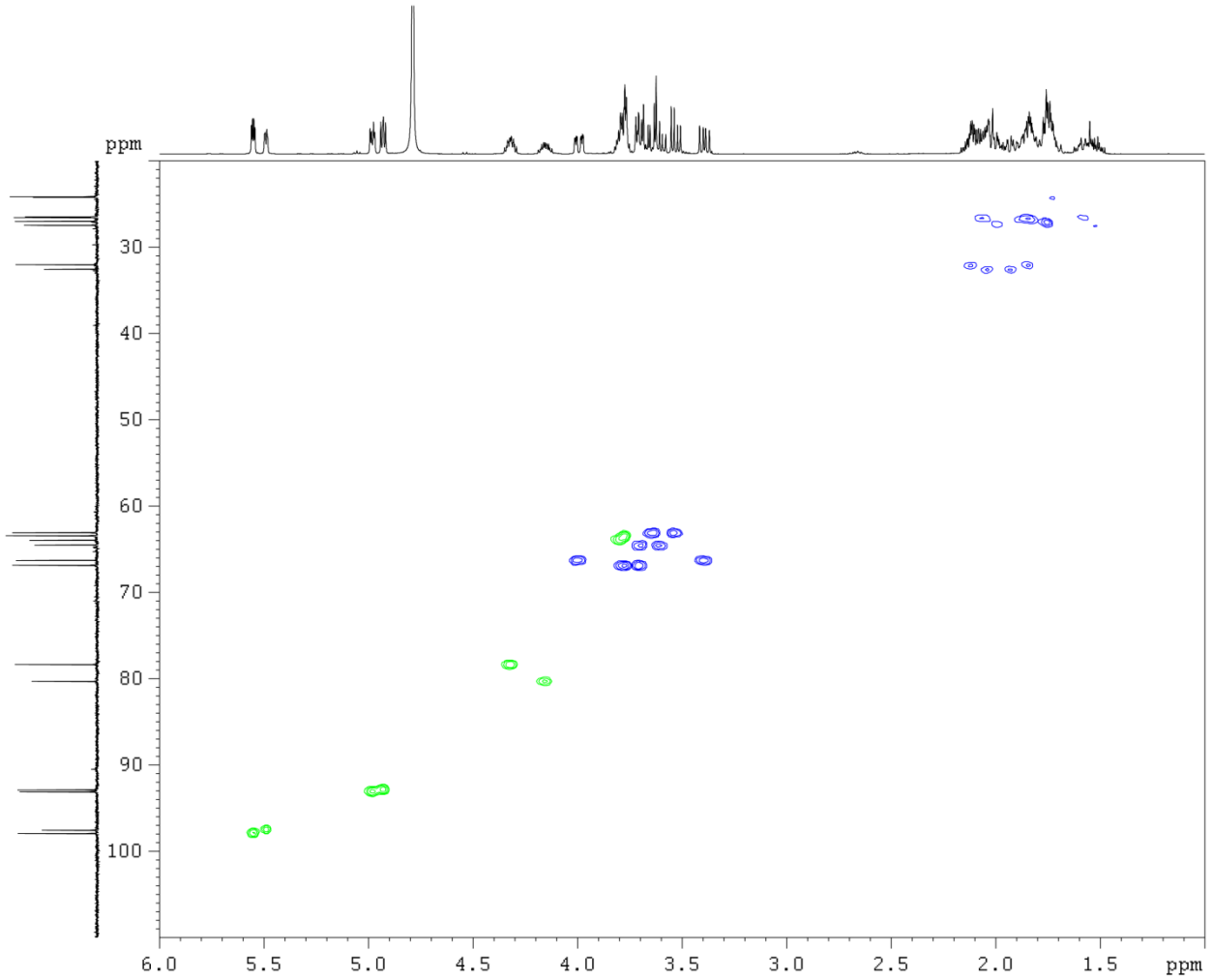

Figure C-7. HSQC of (*S*)-γ-hydroxymethyl-γ-butyrolactone

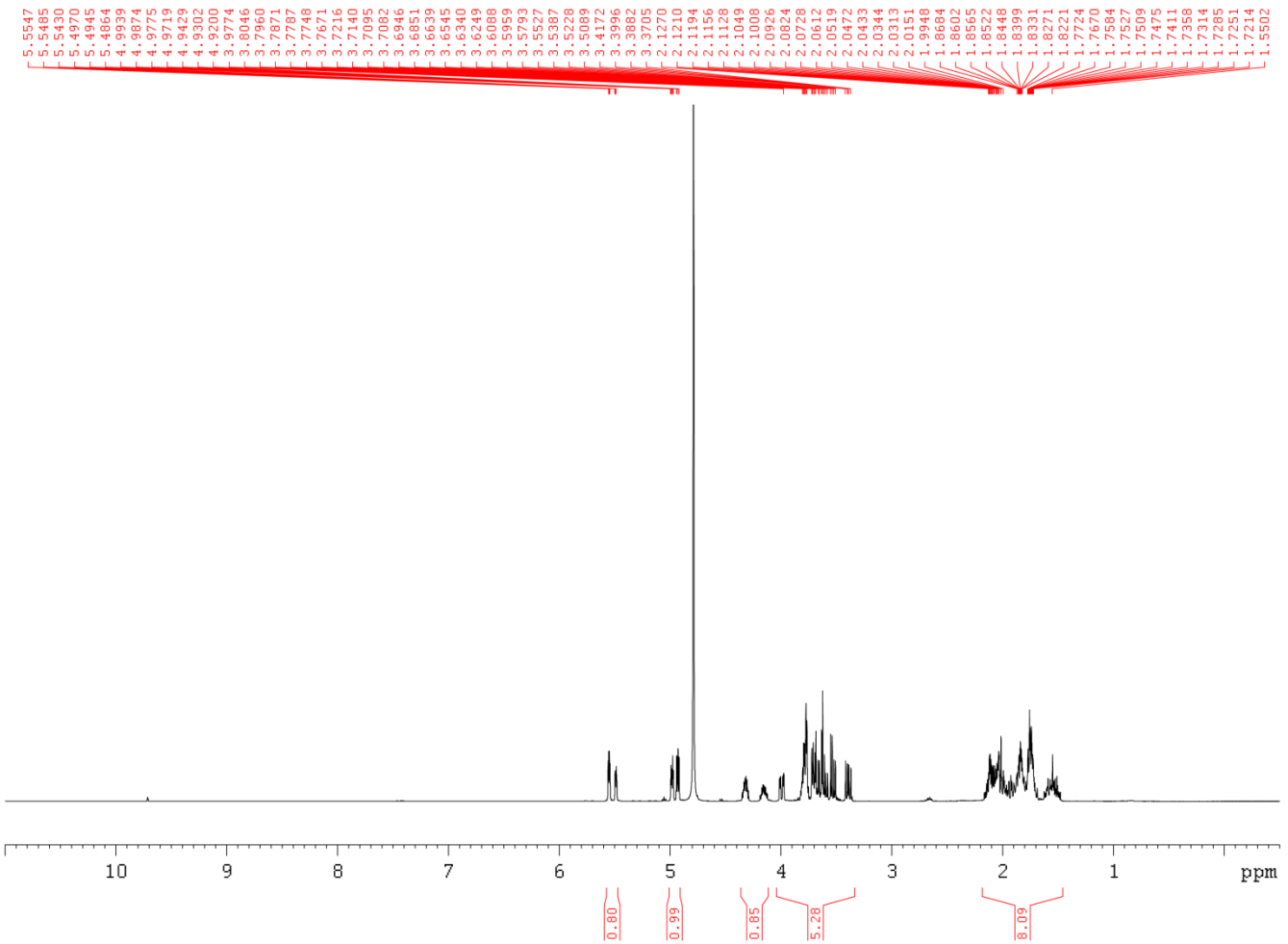

Figure C-8. ^{1}H NMR of 2,3-dideoxyribose

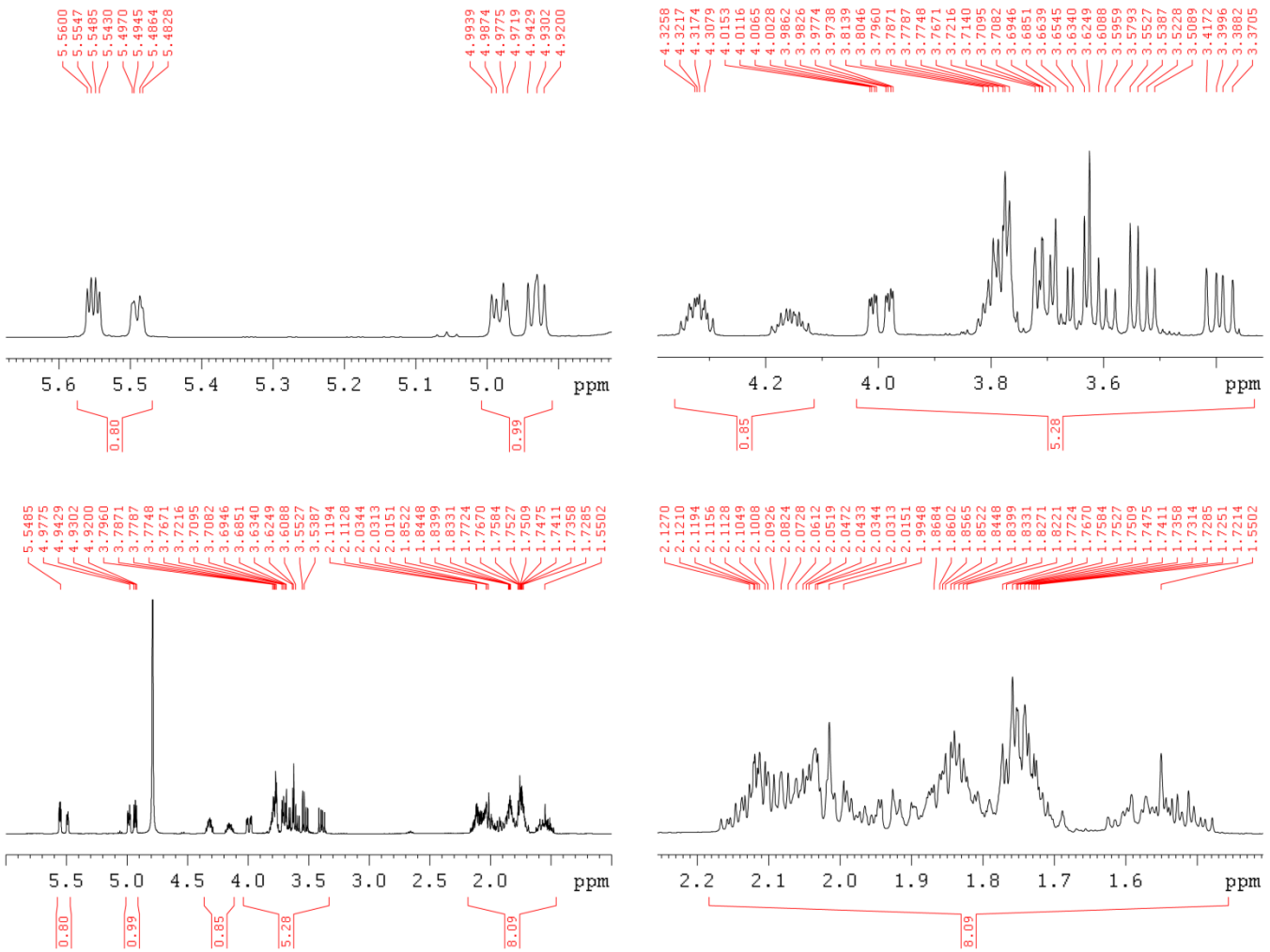

Figure C-9. ^{1}H NMR of 2,3-dideoxyribose

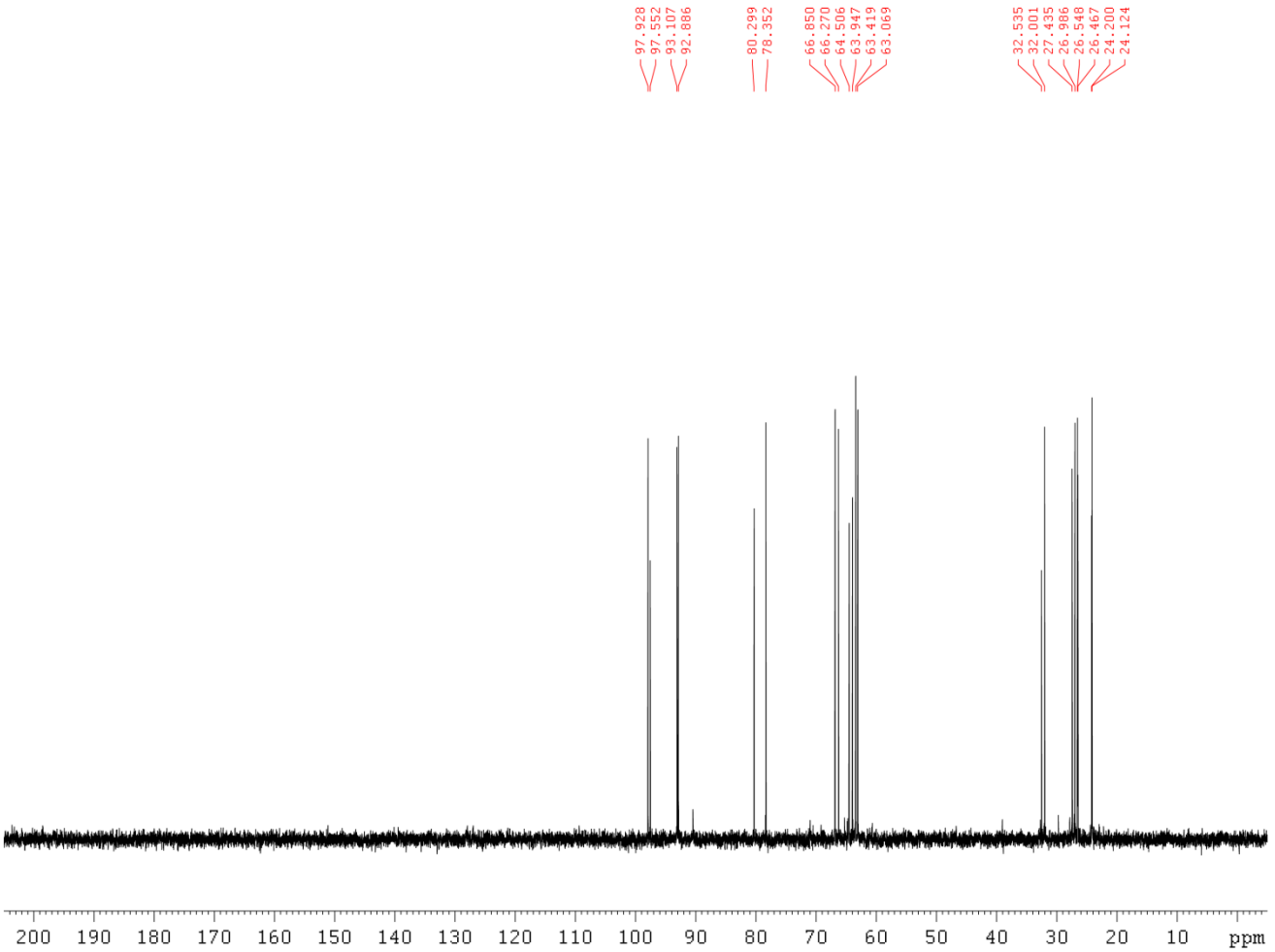

Figure C-10. ^{13}C NMR of 2,3-dideoxyribose

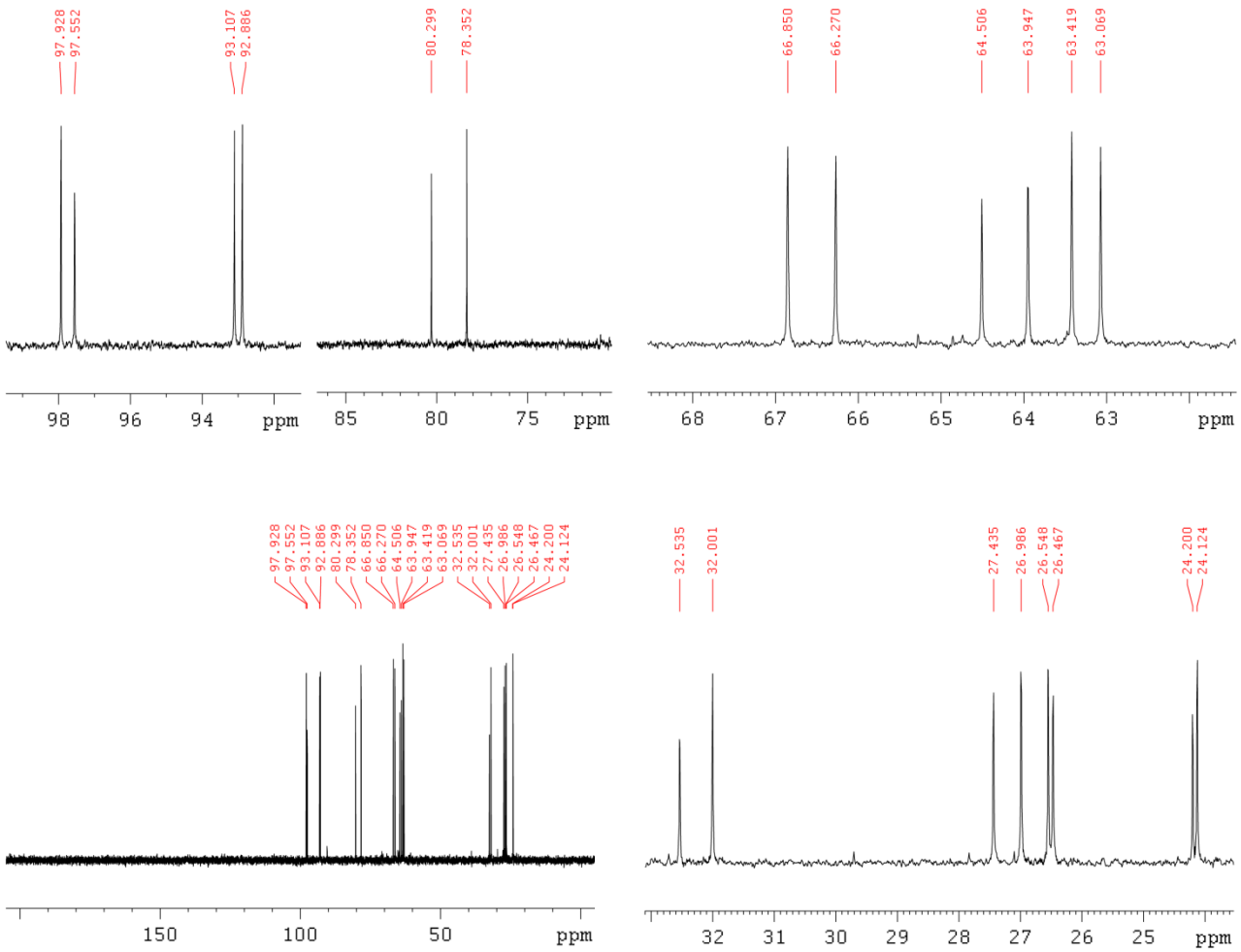

Figure C-11. ^{13}C NMR of 2,3-dideoxyribose

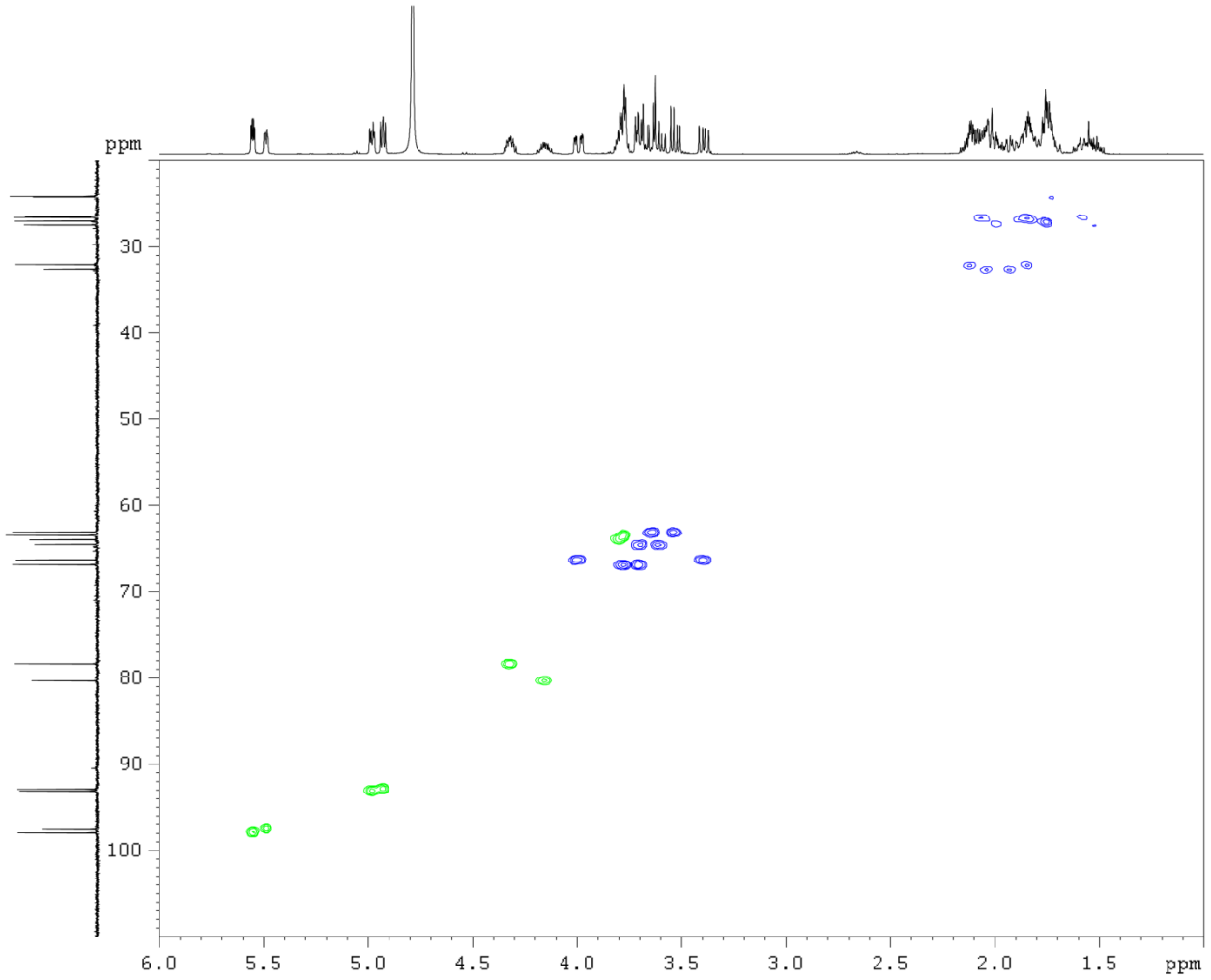

Figure C-12. HSQC of NMR of 2,3-dideoxyribose

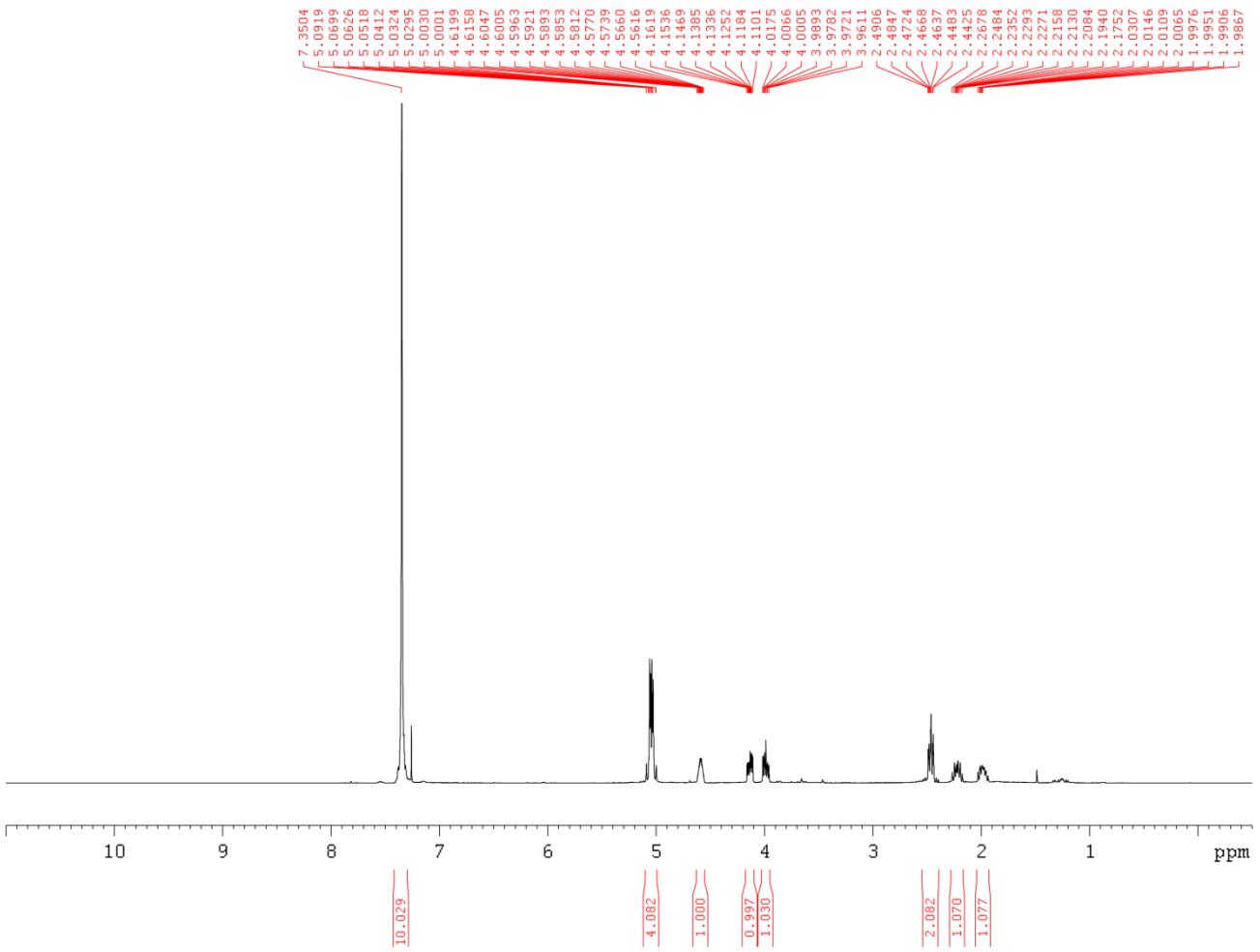

Figure C-13. ^{1}H NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

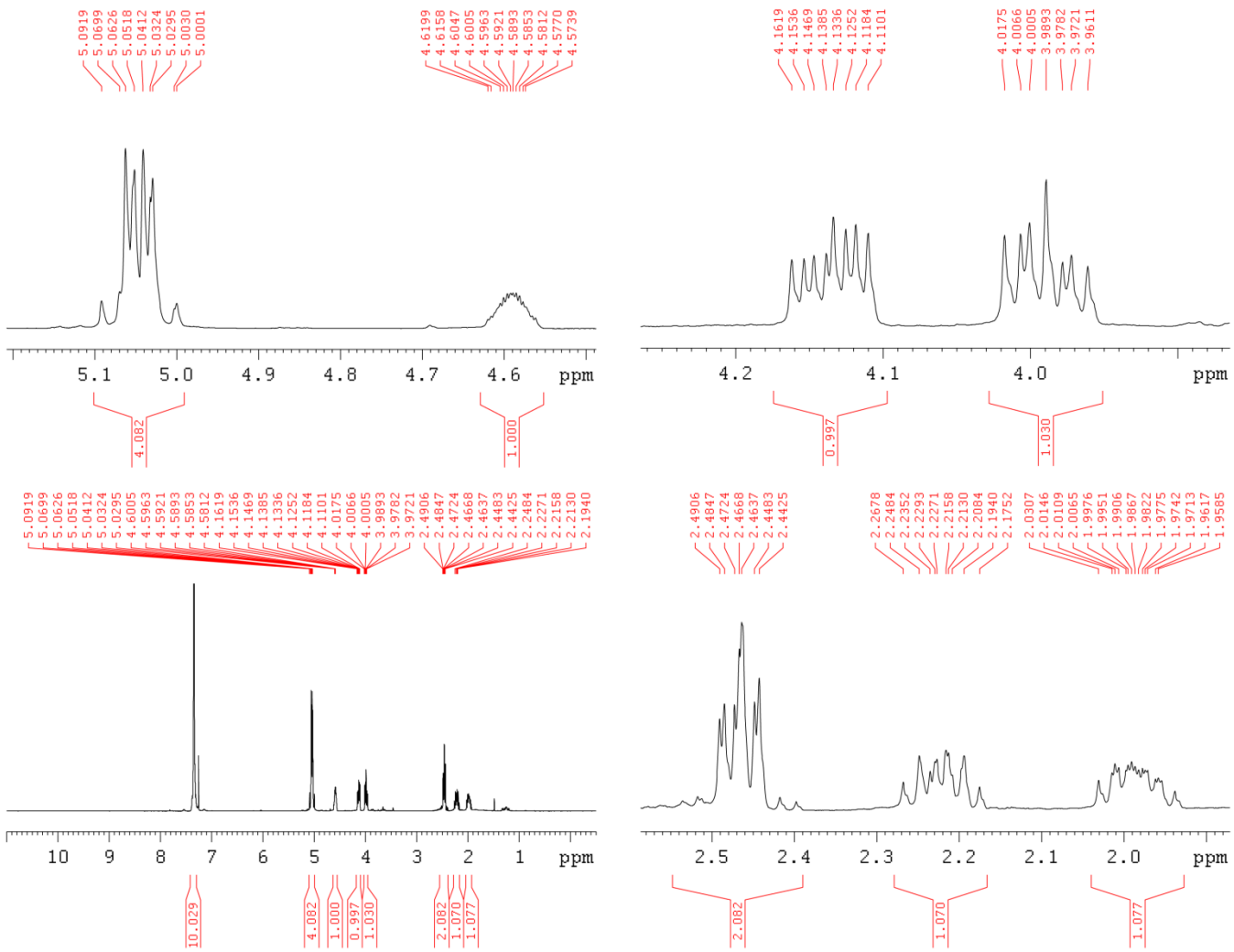

Figure C-14. ^{1}H NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

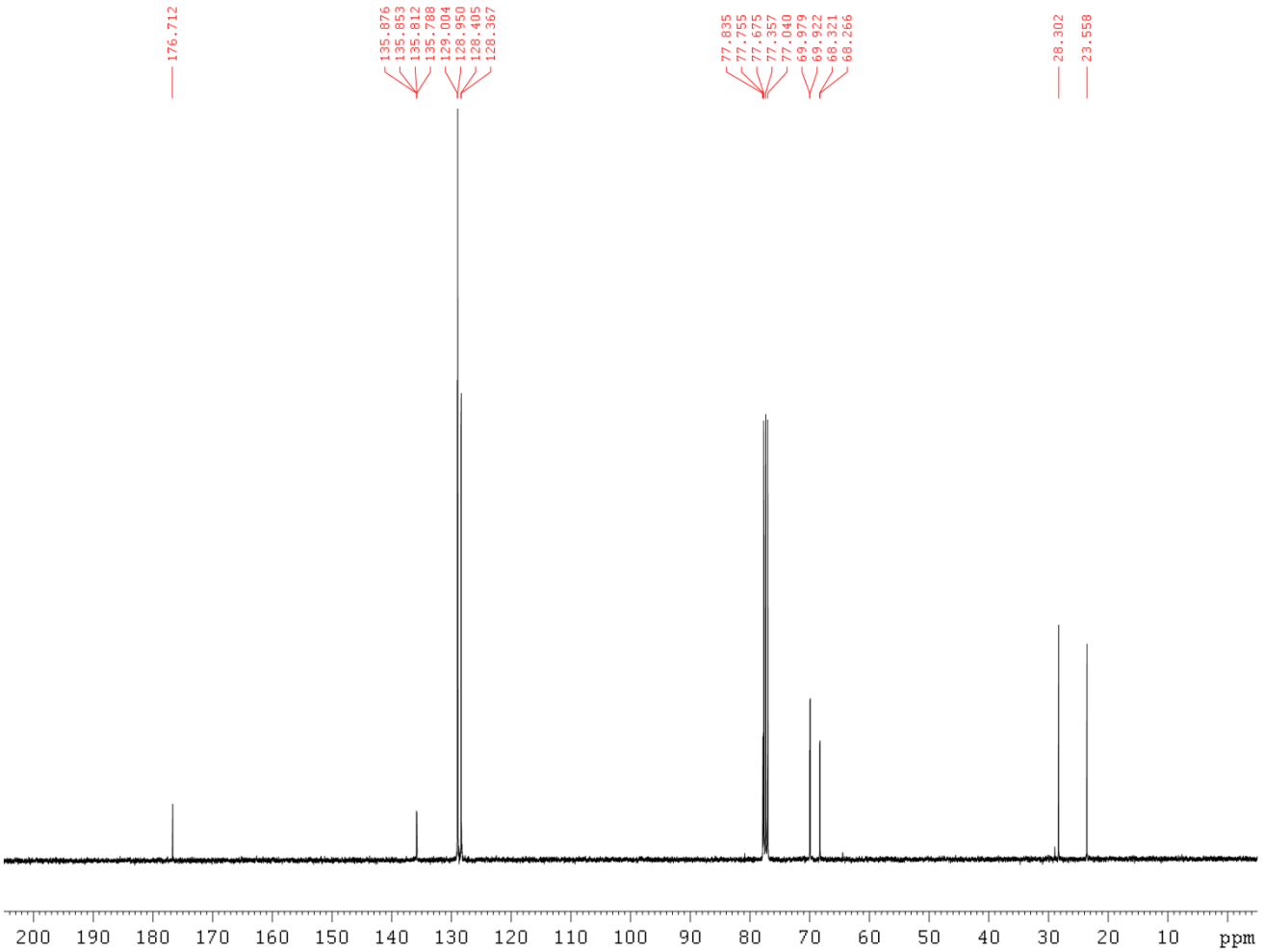

Figure C-15. ^{13}C NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

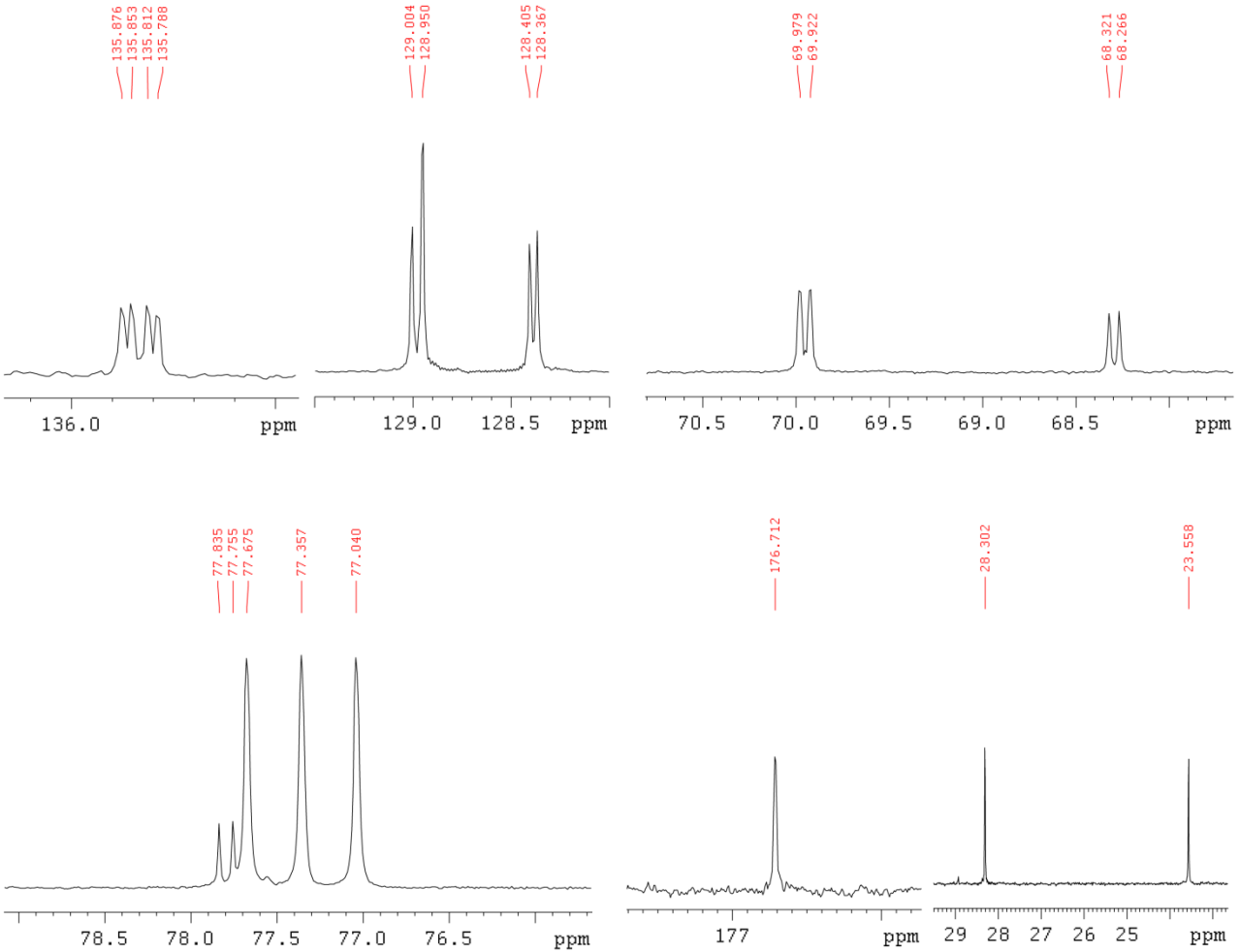

Figure C-16. ^{13}C NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

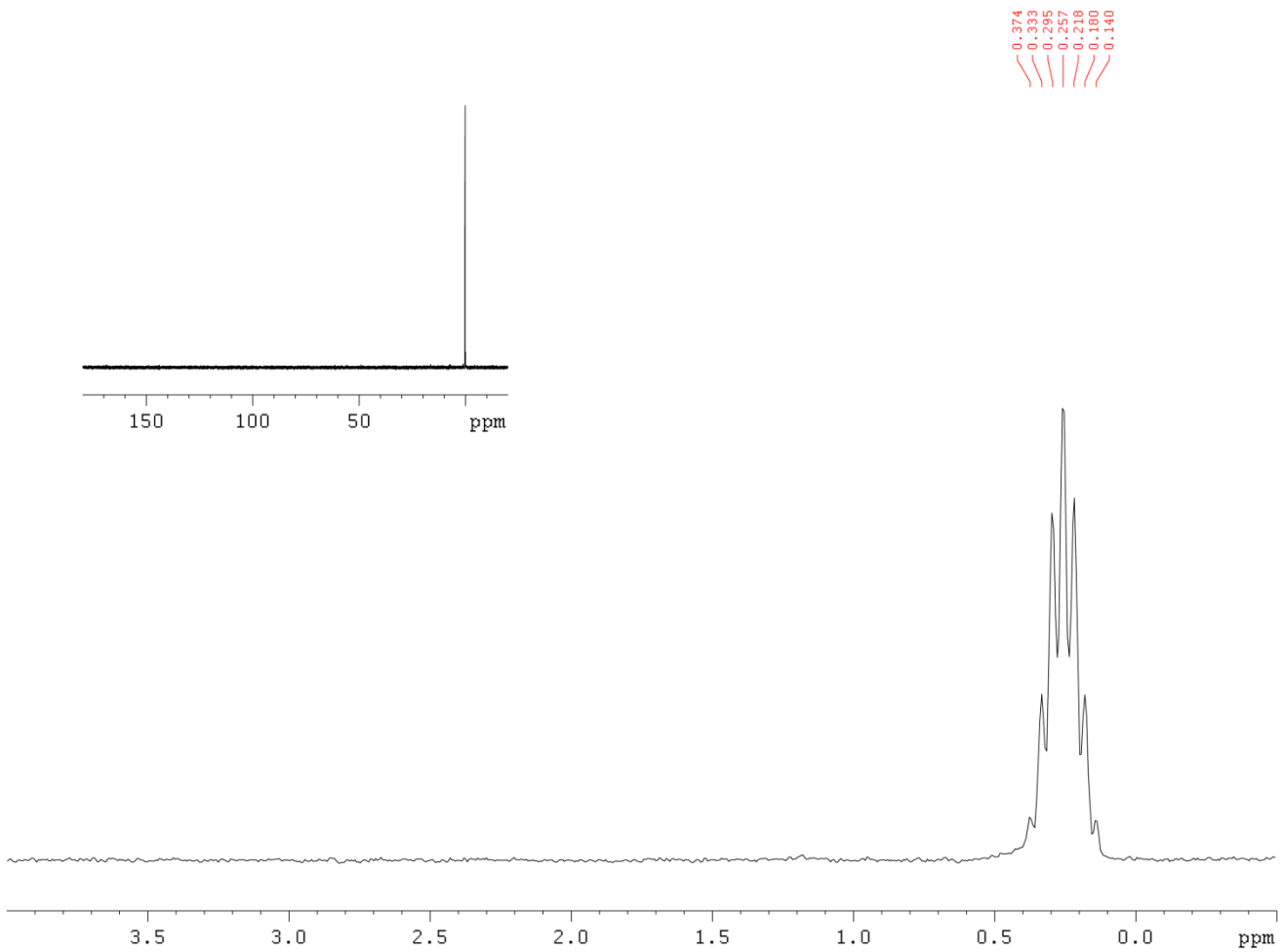

Figure C-17. ^{31}P NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

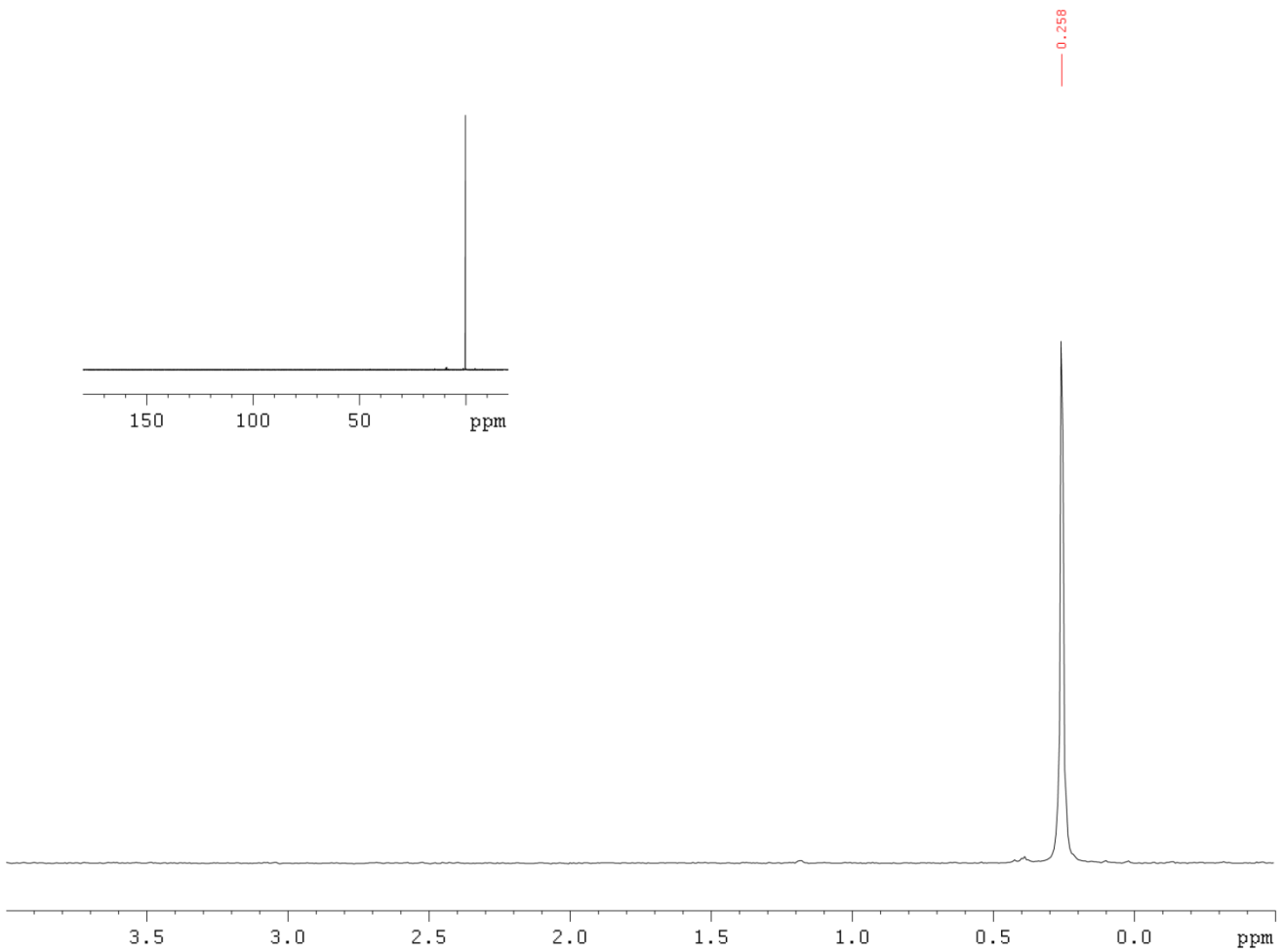

Figure C-18. ^{1}H-decoupled ^{31}P NMR of (*S*)-γ-dibenzylphosphomethyl-γ-butyrolactone

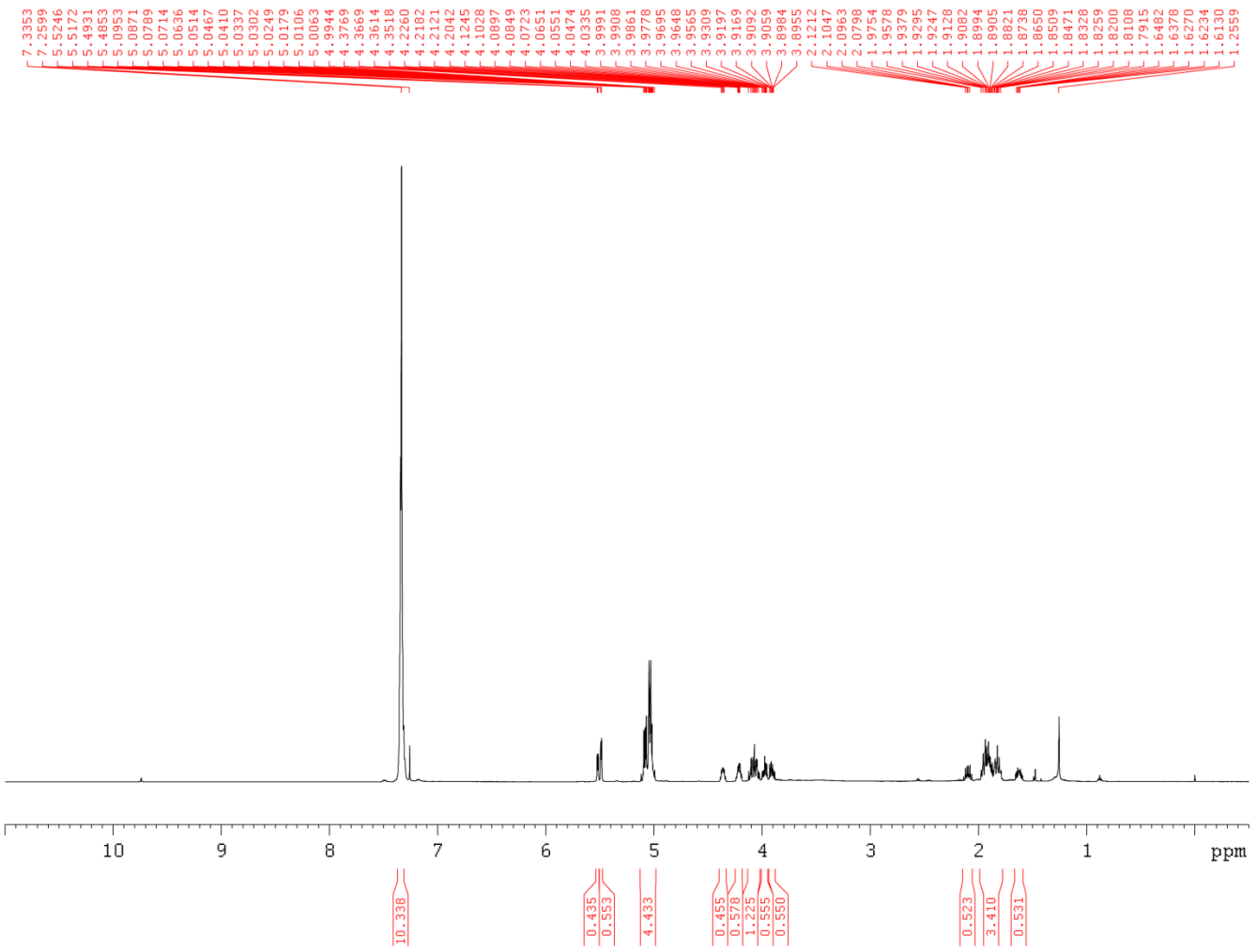

Figure C-19. ^{1}H NMR of 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate

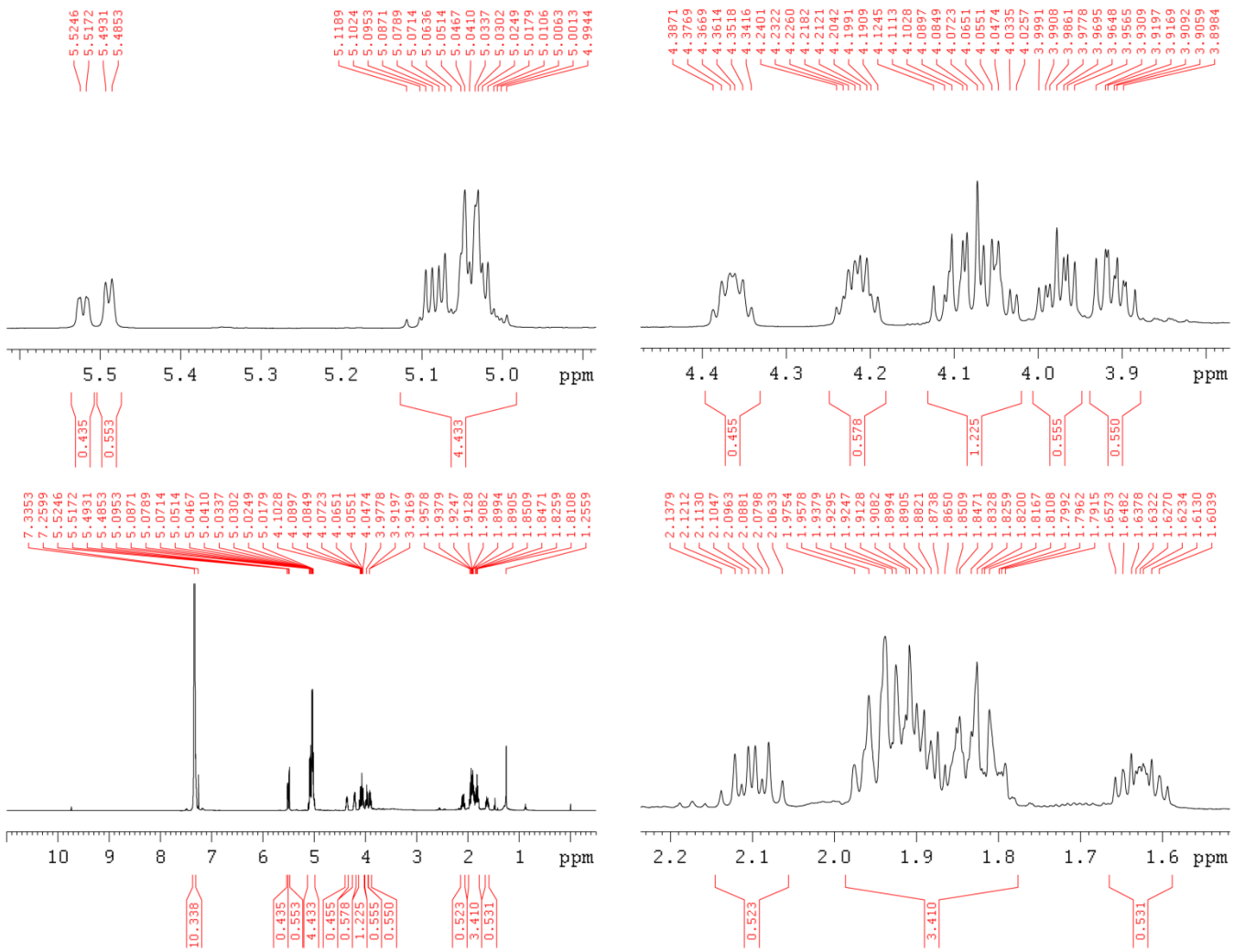

Figure C-20. ^{1}H NMR of 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate

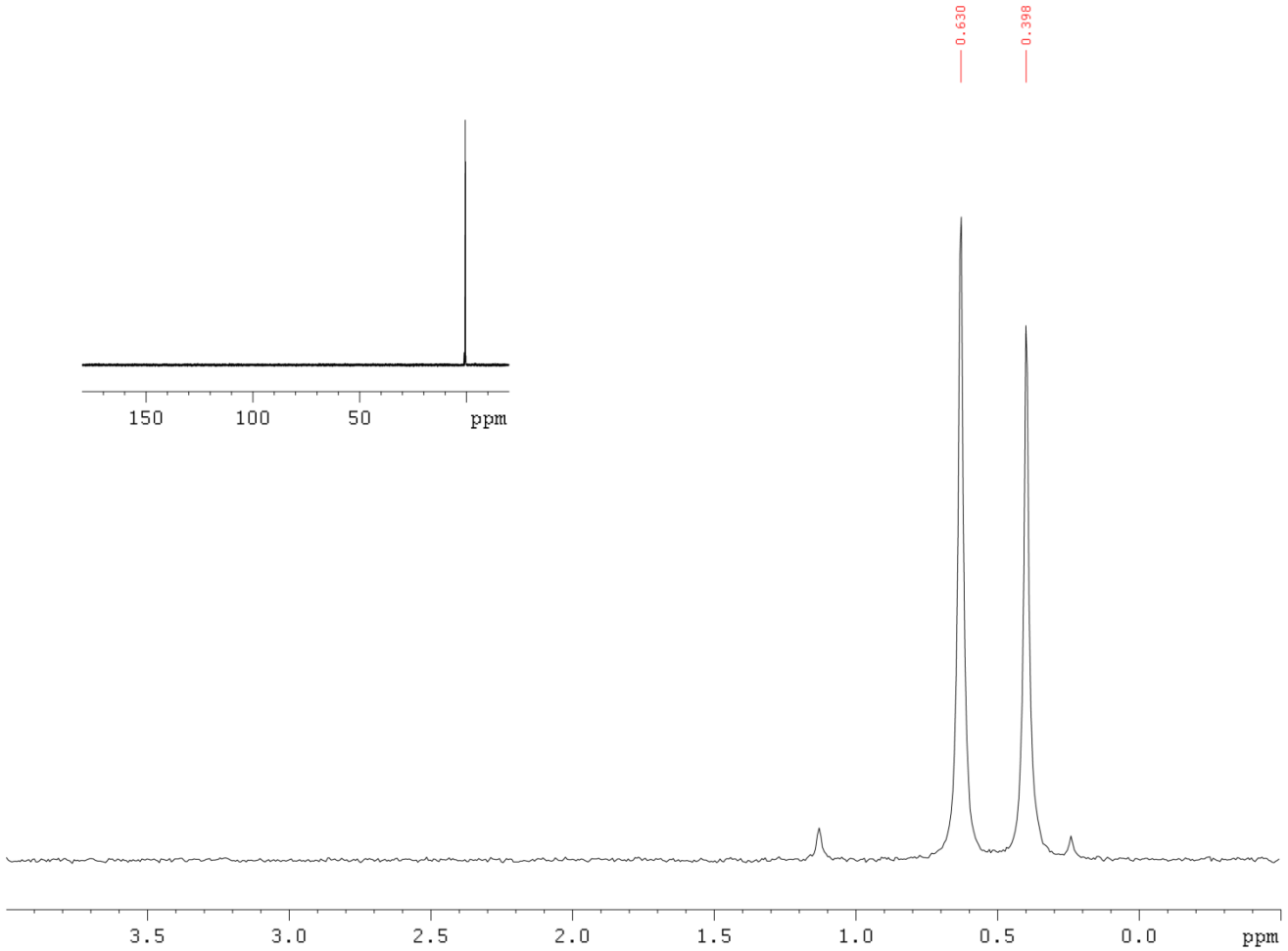

Figure C-21. ^{1}H-decoupled ^{31}P NMR of 2,3-dideoxyribose 5-(di-*O*-benzyl)phosphate

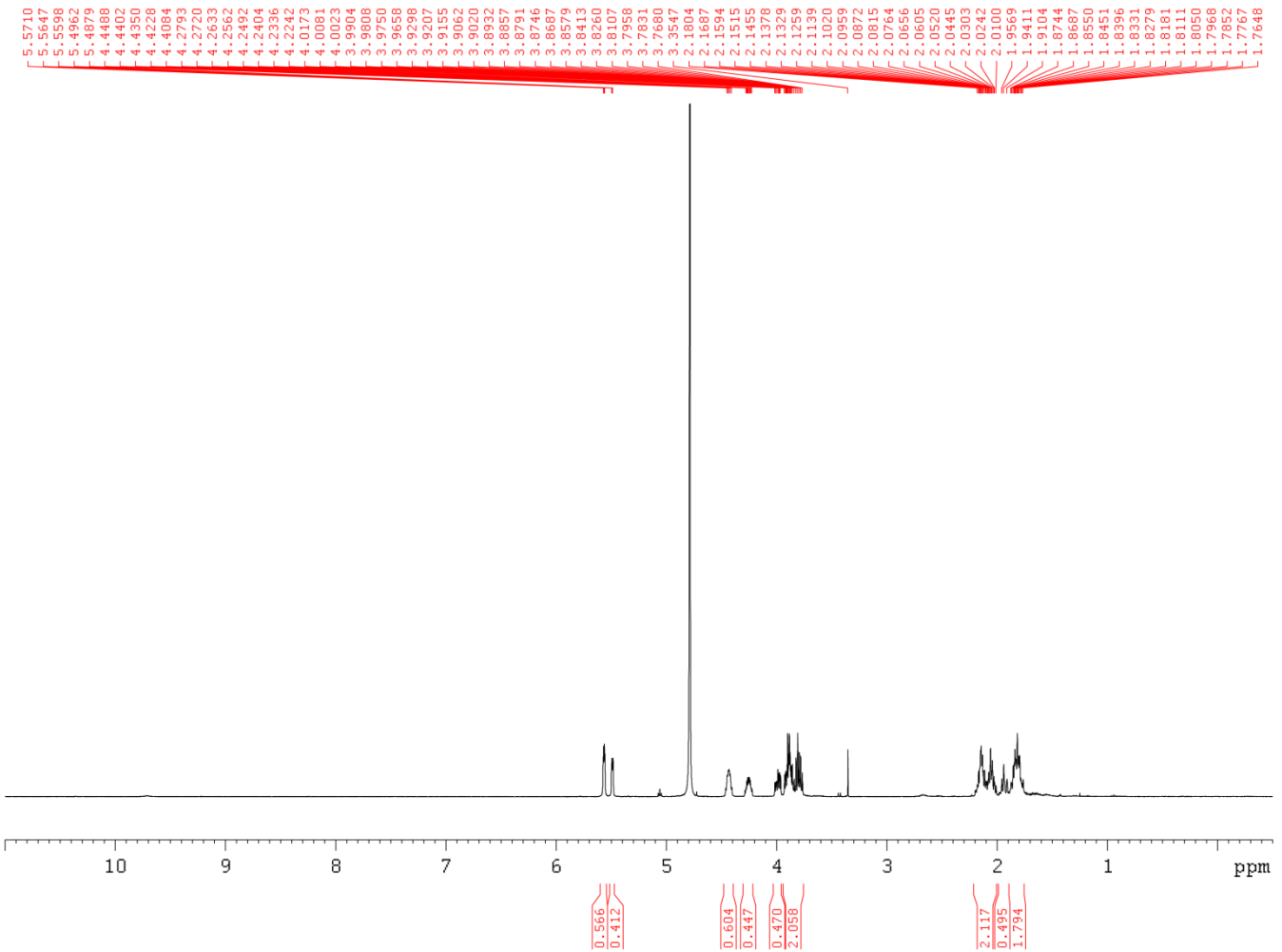

Figure C-22. ^{1}H NMR of 2,3-dideoxyribose 5-phosphate

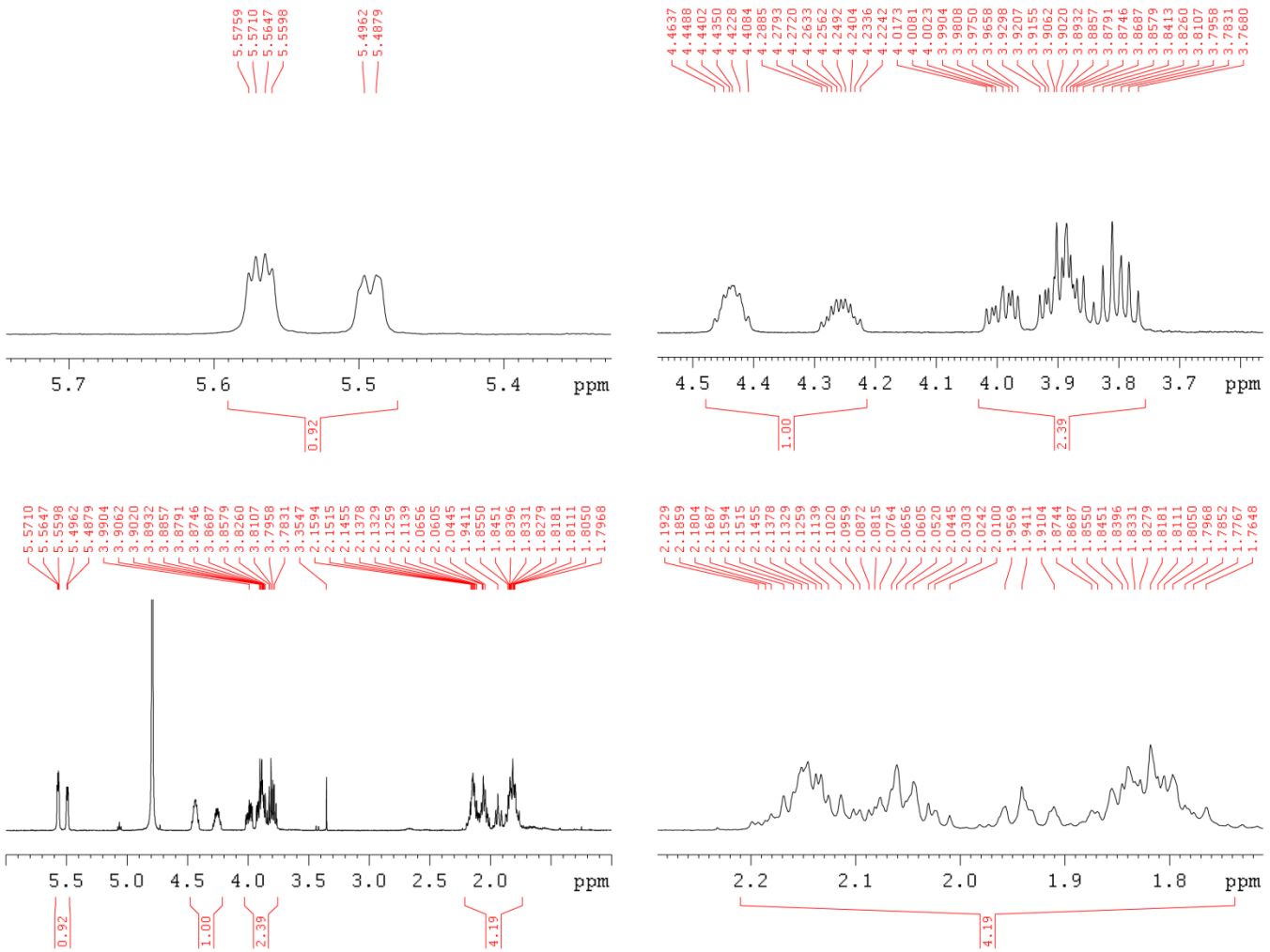

Figure C-23. ^{1}H NMR of 2,3-dideoxyribose 5-phosphate

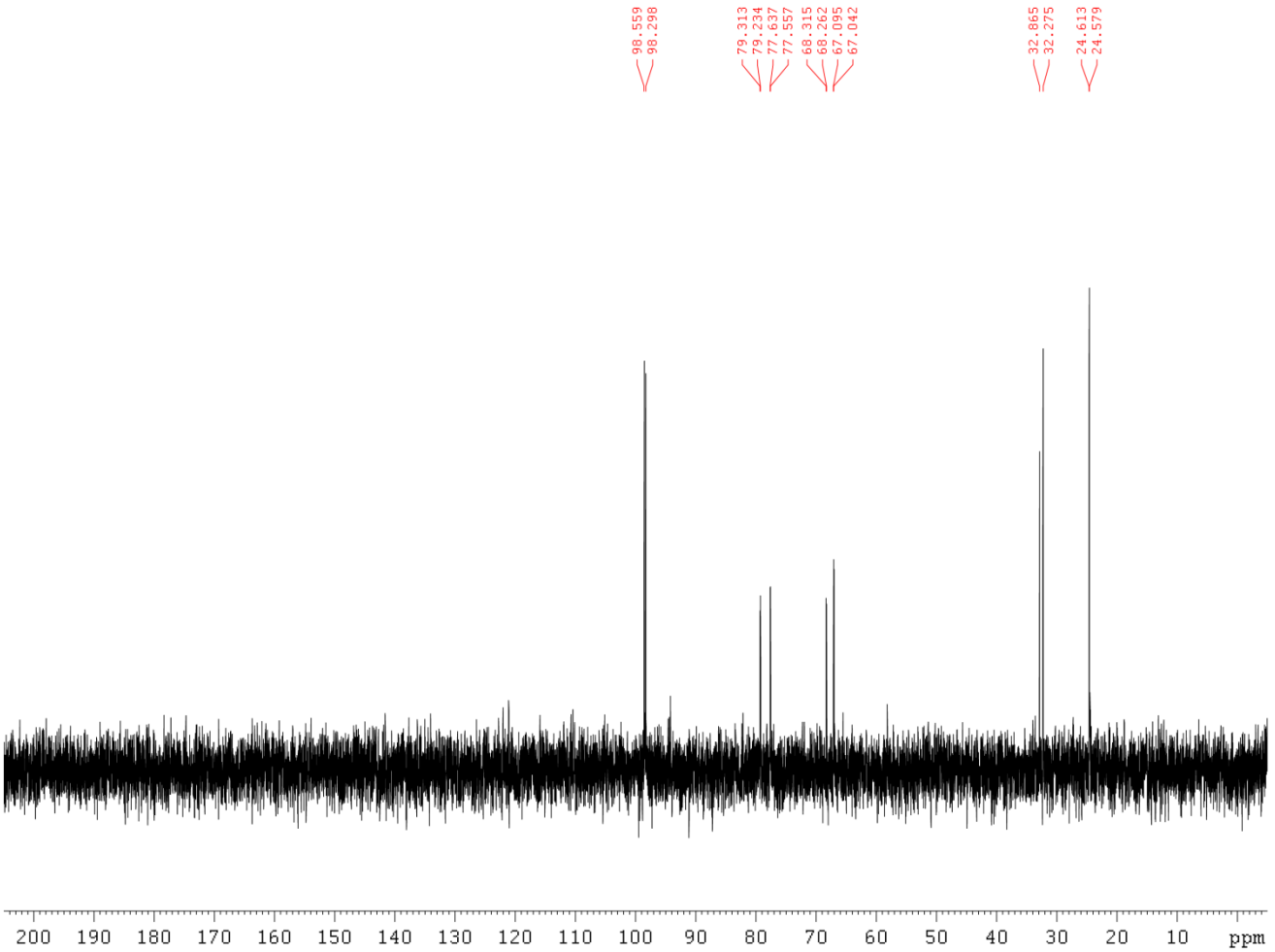

Figure C-24. ^{13}C NMR of 2,3-dideoxyribose 5-phosphate

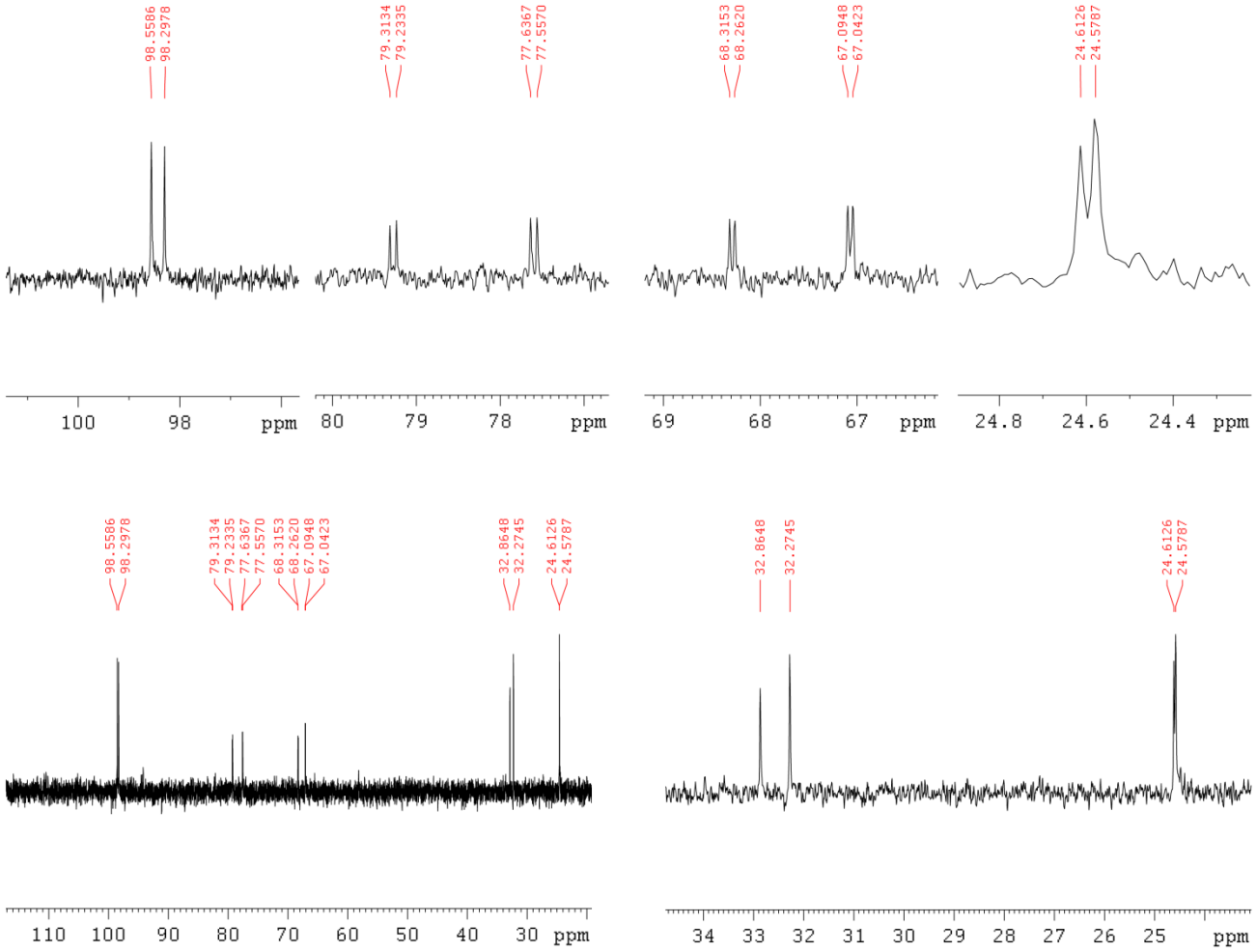

Figure C-25. ^{13}C NMR of 2,3-dideoxyribose 5-phosphate

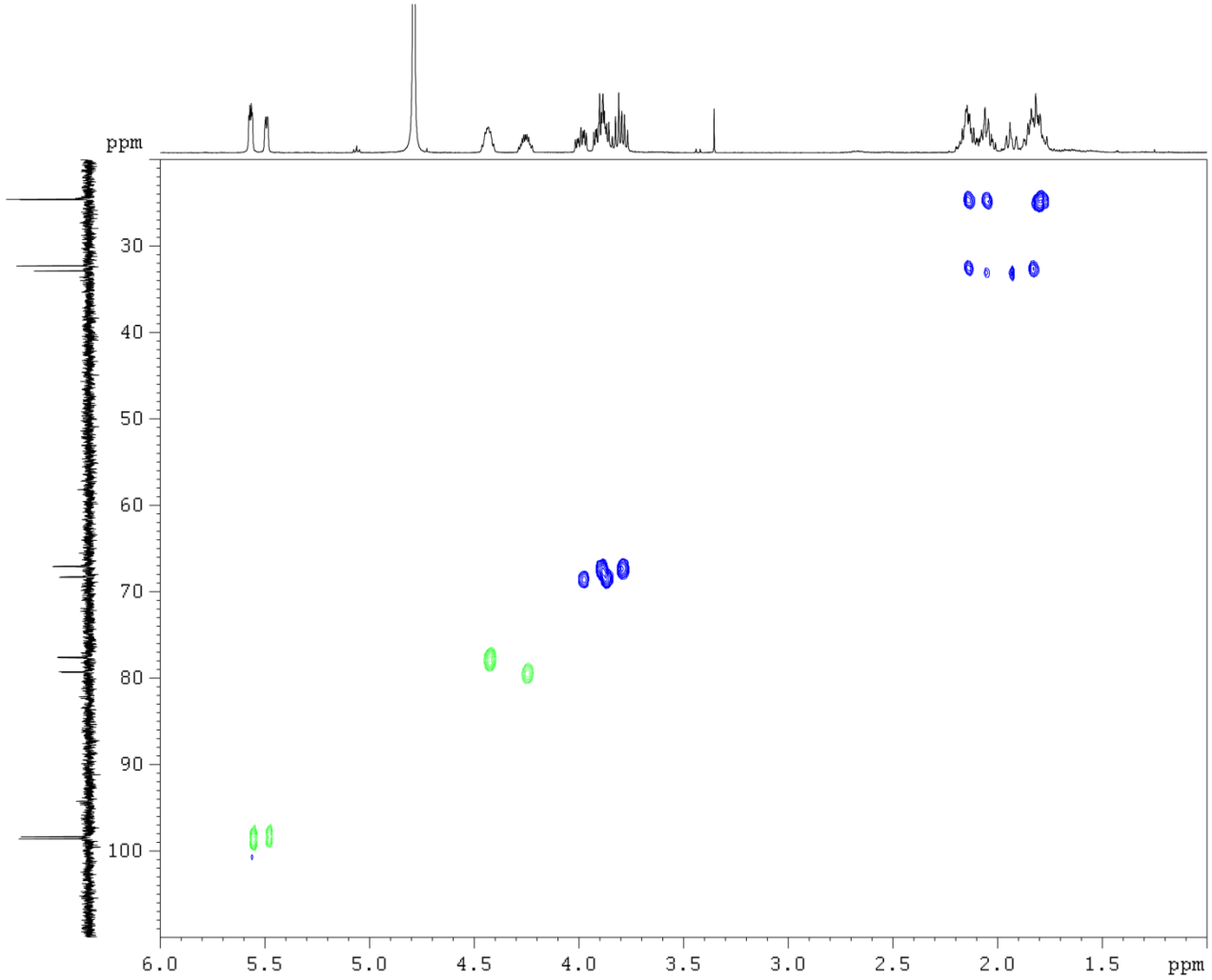

Figure C-26. HSQC of 2,3-dideoxyribose 5-phosphate

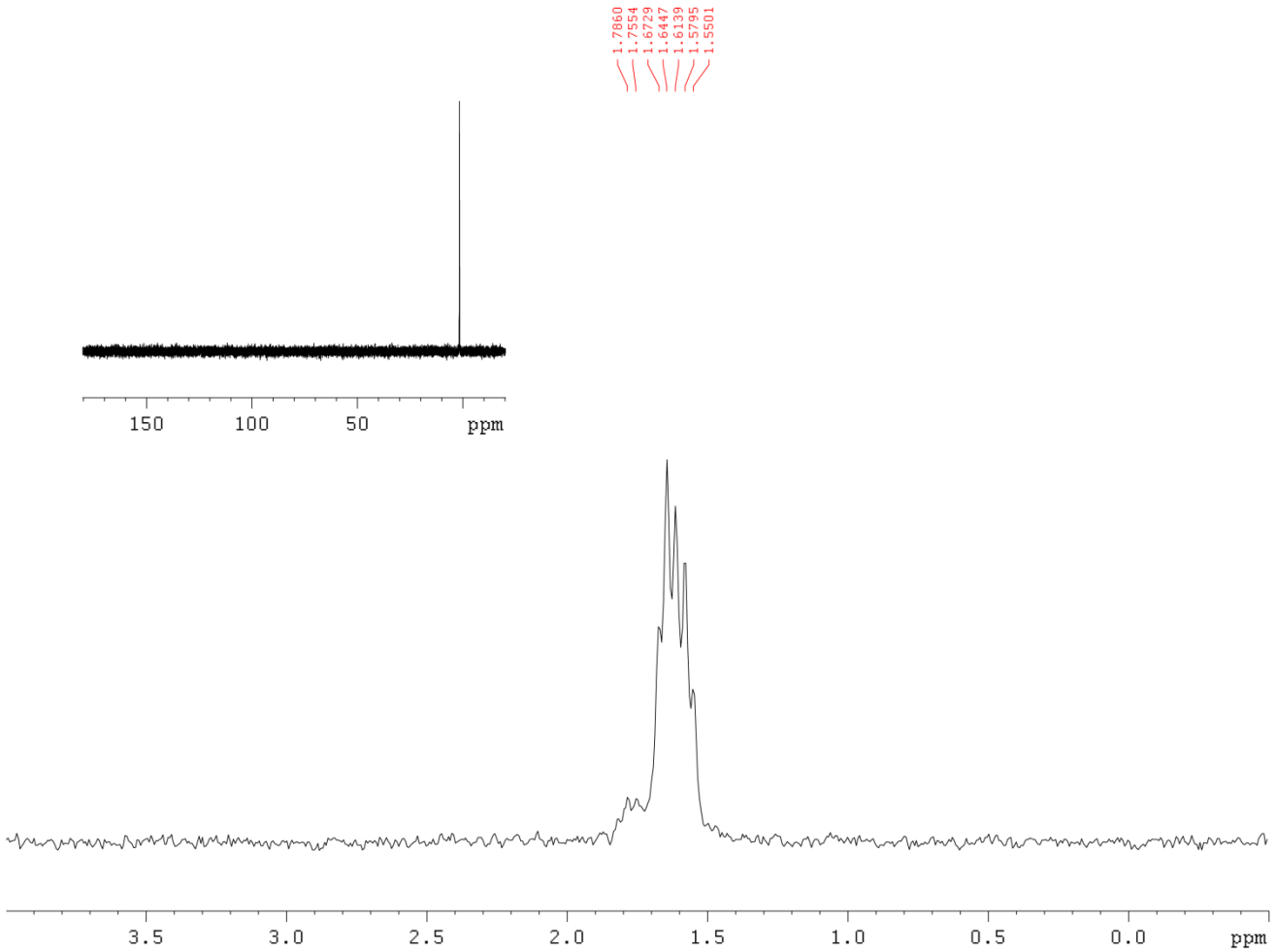

Figure C-27. ^{31}P NMR of 2,3-dideoxyribose 5-phosphate

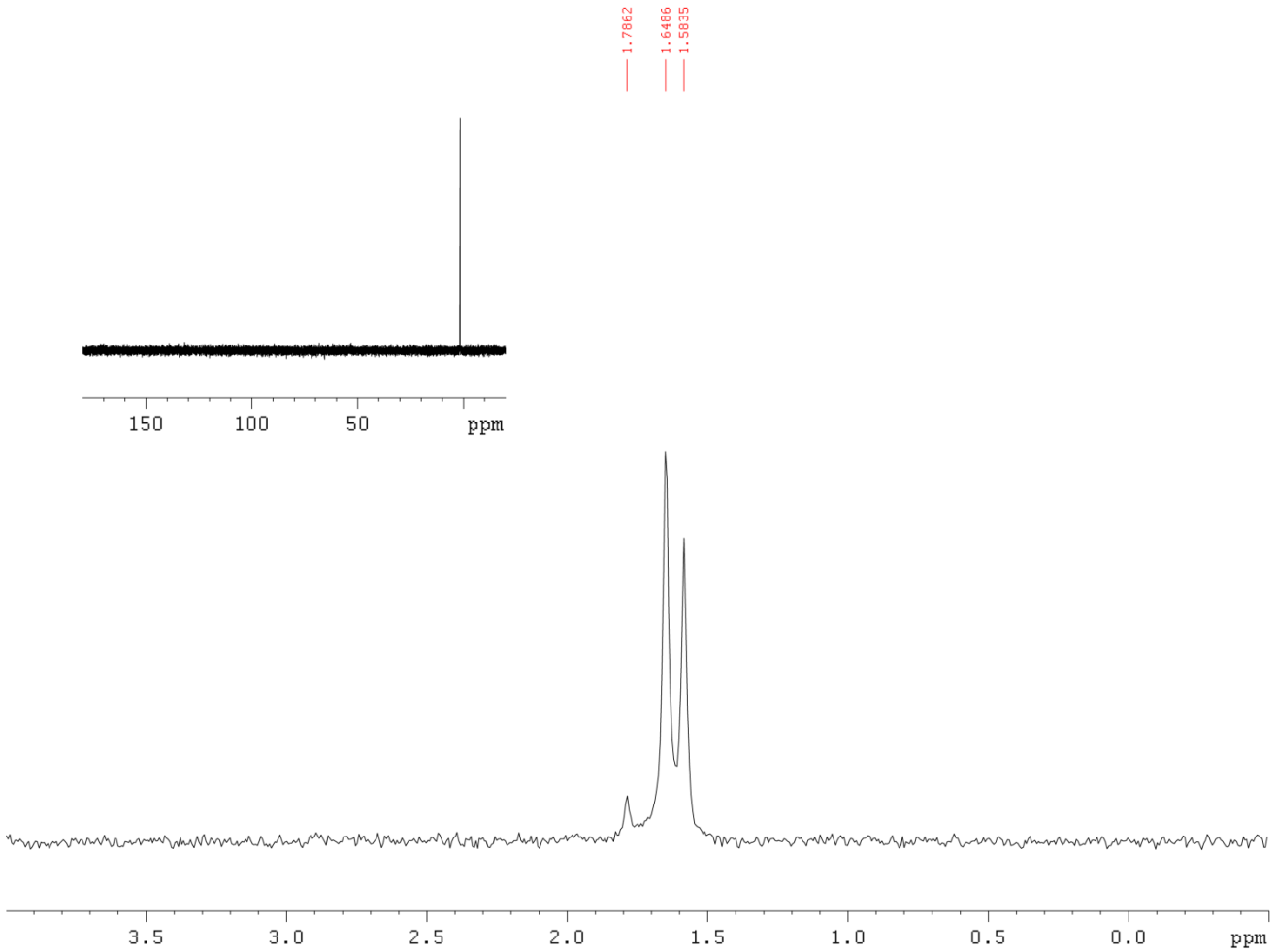

Figure C-28. ^{1}H-decoupled ^{31}P NMR of 2,3-dideoxyribose 5-phosphat