

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Pravin Muthu

Date

Data-Driven Engineering of Therapeutic Enzymes

By

Pravin Muthu
Doctor of Philosophy

Chemistry

Stefan Lutz, Ph.D.
Advisor

R. Brian Dyer, Ph.D.
Committee Member

James T. Kindt, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Data-Driven Engineering of Therapeutic Enzymes

By

Pravin Muthu

B.S., Johns Hopkins University, 2010

M.S., Johns Hopkins University, 2011

Advisor: Stefan Lutz, Ph.D.

An Abstract of

A dissertation submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2015

Abstract

Data-Driven Engineering of Therapeutic Enzymes

By Pravin Muthu

While advances in biotechnology have enabled enzymes to make significant contributions to industrial catalysis and synthetic biology, this dissertation focuses on an often overlooked application: enzymes as therapeutics. As pharmaceutical agents, the catalytic aspect of enzymes enables novel modes of action not possible with conventional treatments. Separately, enzyme biosensors have advantages over traditional analytical techniques, in particular: selectivity, affordability and ease-of-use. Natural enzymes are often not ideal for therapy, but enzyme design can introduce catalytic properties far beyond their native function, enabling new frontiers in medicinal and diagnostic chemistry. The first chapter describes recent efforts to adapt advances in biotechnology to therapeutic enzymes, providing context for the original work presented. The following two chapters focus on the development of a non-invasive reporter gene and use different data-driven approaches, improving function through iterative data generation. A fully realized reporter system will provide clinical data for future cell and gene therapies by monitoring transgene expression and migration. The second chapter describes the use of structural calculations to design human deoxycytidine kinase to have preferential activity for an unnatural L-ribose probe over the native D-deoxynucleosides. The third chapter continues development using an alternate statistical approach to optimize gene sequences. The resulting reporter gene candidates display exquisite in vitro performance and are currently being evaluated within cell and animal models. The fourth chapter shifts focus to therapeutic dosage monitoring (TDM) which is the clinical practice of measuring drug concentrations within a patient's bloodstream to optimize dosing regimens. Instead of a single enzyme, this data-driven algorithm deconvolutes

kinetic observations from several enzymes to quantify multiple native and drug metabolites directly from complex biological samples. This modular detection strategy can readily translate to a variety of analytical applications. The fifth and final chapter provides commentary about the work presented and closes with a perspective on the applied sciences.

Data-Driven Engineering of Therapeutic Enzymes

By

Pravin Muthu

B.S., Johns Hopkins University, 2010

M.S., Johns Hopkins University, 2011

Advisor: Stefan Lutz, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2015

Acknowledgements

The professors, students and staff of the Chemistry Department provided a friendly, professional and helpful atmosphere; I appreciate their efforts and help toward my professional development. In particular, I would like to thank my advisor Prof. Stefan Lutz; who provided invaluable experience during my graduate career. His curiosity, fortitude and tenacity continue to influence me as a scientist. I would like to thank my committee Prof. R. Brian Dyer and Prof. James T. Kindt for their time, encouragement and support. I thank Prof. Emily Weinert, Dr. Justin Burns, and Dr. Daniel Morton for their mentorship and advice. I would also like to thank Ann Dasher for going above and beyond to look out for me and every graduate student.

All of the members of the Lutz Lab, past and present, have not only made my graduate career productive, but also enjoyable. In particular, I would like to thank Dr. Ashley Bagwell Daugherty; her dedication to her family, friends and work continues to be inspiring. I would also like to thank the current members of my lab, Matt Jenkins, Leann Quertinmont and Samantha Iamurri who created a fantastic working environment. I would also like to thank Hannah Chen; an undergraduate who had an intellect and work ethic far beyond her years. My friends made Atlanta a wonderful place to explore, eat and live; full of unforgettable memories. In particular, I would like to thank Dr. John Mancini, Dr. Kevin Yehl, Dr. Erin Schuler, Dr. Yoshi Narui, Dr. Brian Hayes, Dr. Brandon Greene, Michael Sullivan, Xin Jia, Shawwna Joynt and Jessica Hurtak, who provided conversation, humor, and advice throughout.

Finally, I dedicate this dissertation to my parents Muthu and Uma. I owe everything to their sacrifice, hard work, and unconditional support.

Table of Contents

Chapter 1: Introduction	1
1.1 Overview of Therapeutic Enzymes	2
1.1.1 Pharmaceutical Proteins	4
1.1.2 Gene-Directed Enzyme Prodrug Therapy	6
1.1.3 Diagnostic Enzymes	8
1.2 Proteases	10
1.2.1 Plasminogen Activators	10
1.2.2 Thrombin	14
1.2.3 Procoagulant	15
1.2.4 Digestive Proteases	17
1.2.5 Other indications	18
1.3 Deoxynucleoside Kinases	19
1.3.1 Herpes Simplex Virus type 1 Thymidine Kinase	22
1.3.2 <i>Drosophila melanogaster</i> Deoxynucleoside kinase	24
1.3.3 <i>H. sapiens</i> Deoxycytidine Kinase	26
1.3.4 <i>H. sapiens</i> Thymidine Kinase Type 2	28
1.4 Cytosine Deaminases	29
1.5 Purine Nucleoside Phosphorylases	31
1.6 Nitroreductases	32
1.7 Cholinesterases	33
1.7.1 Butyrylcholinesterase	34
1.7.2 Acetylcholinesterase	35
1.7.3 Other Esterases	37
1.8 References	39

Chapter 2: Computational Design of Deoxycytidine Kinase	56
2.1 Introduction	57
2.2 Results and Discussion	59
2.3 Methods	70
2.3.1 Materials.....	70
2.3.2 Computer Models.....	71
2.3.3 Site Directed Mutagenesis.....	71
2.3.4 Protein Expression and Purification.....	72
2.3.5 Enzyme Kinetics	73
2.4 References	74
Chapter 3: Design-of-Experiments for Deoxycytidine Kinase	80
3.1 Introduction	82
3.2 Results and Discussion	85
3.2.1 Round 1	85
3.2.2 Round 2	91
3.2.3 Round 3	93
3.2.4 Experimental Deconvolution	96
3.3 Conclusion	99
3.4 Material and Methods	100
4.4.1 Library Design	100
3.4.2 Library Construction and Evaluation	101
3.4.3 Enzyme Characterization	102
3.5 References	103
Chapter 4: Multi-Biosensor Detection of Nucleoside Analogs	108
4.1 Introduction	110

4.2 Results and Discussion	112
4.2.1 Multi-biosensor Design	112
4.2.2 Data Processing	114
4.2.3 Single Component Evaluation.....	117
4.2.4 Binary Mixture Evaluation.....	119
4.2.5 Blood Plasma Quantification	122
4.2.6 Point-of-Care Detection	123
4.3 Conclusion	125
4.4 Methods	125
4.4.1 Chemicals and Reagents	125
4.4.2 Protein Expression and Purification.....	126
4.4.3 Enzyme Characterization	126
4.4.4 Datasets	127
4.4.5 Cell-Phone Based Detection	128
4.5. References	130
Chapter 5: Final Thoughts	136
5.1 Computational Enzyme Engineering	137
5.2 Design-of-Experiments	138
5.3 Reporter Systems	140
5.4 Enzyme Biosensors	141
5.5 General Perspective	143
5.6 References	145

List of Figures

Figure 1.1. Mode of Action for Gene Directed Enzyme Prodrug Therapy.	7
Figure 1.2. Protease Therapeutics for Cardiovascular Disorders.	11
Figure 1.3. Abbreviated Nucleoside Salvage Pathway and GDEPT Candidates.	20
Figure 2.4. Structures of native substrates and analogs.	58
Figure 2.5. Summary of Amino Acid Substitution in dCK.	61
Figure 2.6. Correlation of Predicted Scores with Experimental Data.	63
Figure 2.7. Molecular Dynamics Simulation of dCK Variants.	66
Figure 3.8. Substitution Matrices and Activity Histograms for Rounds 1-3.	87
Figure 3.9. Catalytic Efficiencies of Lead Variants.	88
Figure 3.10. Positions of dCK Variant Substitution.	90
Figure 3.11. Catalytic Efficiencies of Deconvoluted Variants.	97
Figure 4.12. Investigated Substrates and Enzymes.	113
Figure 4.13. Data Analysis to Compare Observed Measurements to Expected Outputs.	115
Figure 4.14. Probability Distribution of Single Component Samples.	118
Figure 4.15. Probability Distribution for Binary Component Mixtures.	120
Figure 4.16. Time-course Kinetics of Plasma Samples.	123
Figure 4.17. Cellphone Camera-based Nucleoside Analog Detection.	124

List of Tables

Table 1.1. Pharmaceutical Proteases Modified by Directed Evolution	13
Table 1.2. Nucleoside Kinase GDEPT Candidates Modified by Directed Evolution	21
Table 1.3. Other GDEPT Candidates Modified by Directed Evolution	30
Table 2.4. Summary of Kinetic Data for dCK Library A	63
Table 2.5. Summary of Kinetic Data for dCK Library B	64
Table 2.6. Kinetic Parameters for dCK Variants with Native Substrates and L-FMAU	69
Table 3.7. Steady State Kinetics of Lead Variants for Selected Substrates.....	90
Table 4.8. Steady State Michaelis-Menten Parameters for Enzymes and Substrates.....	114
Table 4.9. Analyte Probability for Single Component Analysis	118
Table 4.10. True Analyte Probability within Sample	121
Table 4.11. Estimated Composition of Spiked Plasma Samples	123

Abbreviations

5FC	5-Fluorocytosine
AChE	Acetylcholinesterase
ADP	Adenosine Diphosphate
APC	Activated Protein C
ATP	Adenosine Triphosphate
AZT	3'-azidothymidine, Zidovudine
BuChE	Butyrylcholinesterase
CB1954	5-(aziridin-1-yl)-2,4-dinitrobenzamide
CDR	Complementary Determining Region
ChE	Cholinesterases
D4T	3',3'-didehydro-deoxythymidine, Stavudine
dCK	Human Deoxycytidine Kinase
D-dA	2'-Deoxyadenosine
ddC	2',3'-dideoxycytidine, Zalcitabine
D-dC	2'-Deoxycytidine
D-dG	2'-Deoxyguanosine
DDT	dichlorodiphenyltrichloroethane

DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNK	<i>Drosophila melanogaster</i> Deoxynucleoside Kinase
dNTP	Deoxynucleoside Triphosphates
D-T	D-Thymidine
DTT	Dithiothreitol
FDA	United States Food and Drug Administration
FTC	2',3'-dideoxy-5-fluoro-3'-thiacytidine, Emtricitabine
FVII	Factor VII
FVIIa	Activated Factor VII
GCV	9-(1,3-dihydroxy-2-propoxymethyl)guanine, Ganciclovir
GDEPT	Gene Directed Enzyme Prodrug Therapy
GEM	2',2'-difluorodeoxycytidine, Gemcitabine
HIV	Human Immunodeficiency Virus
HSV-TK	Herpes Simplex Virus Type 1 Thymidine Kinase
IPTG	Isopropyl β -D-1-thiogalactopyranoside
k_{cat}	Turnover Number
K_M	Michaelis Constant

L-FMAU	L-2'-fluoro-5-methyluracil-arabinofuranosyl
L-T	L-Thymidine
MD	Molecular Dynamics
MePR	9-(6-deoxy- α -L-talofuranosyl)-6-methylpurine
NA	Nucleoside Analog
NADH	Nicotinamide Adenine Dinucleotide
NCBI	United States National Center for Biotechnology Information
OD ₆₀₀	Absorbance at 600 nm
PAI-1	Plasminogen Activator Inhibitor 1
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEP	Phosphoenolpyruvate
PET	Positron Emission Tomography
PNP	Purine Nucleoside Phosphorylase
PON1	Human Serum Paraoxonase
PTE	Phosphotriesterase
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid

SNAP	Synaptosomal Associated Protein
TDM	Therapeutic Dosage Monitoring
TF	Tissue Factor
TK1	<i>Thermotoga maritima</i> Thymidine Kinase 1
TK2	Human Mitochondrial Thymidine Kinase 2
t-PA	Tissue Plasminogen Activating factor
u-PA	Urokinase Plasminogen Activator factor
UV	Ultraviolet
UV-Vis	Ultraviolet-Visible

Chapter 1

Introduction

Abstract

Enzyme engineering involves the introduction of novel properties not found within the native protein. Traditionally, enzyme engineering has been applied to the design of industrial biocatalysts for improved activity, stability or other desired properties; however, these strategies have utility in other disciplines as well. This chapter describes recent efforts to apply similar methodologies to design improved enzymes for therapeutic applications. This chapter does not intend to describe techniques and protocols in detail, but rather to explore the trends within literature for therapeutic enzyme engineering. These outcomes can be broadly separated into three types. The first are pharmaceutical enzymes, in which the protein itself is the therapeutic agent. Second as a diagnostic enzyme that has advantages over traditional analytical methods. Finally, toward the development of gene directed enzyme prodrug therapy (GDEPT) as a targeted alternative to current chemotherapy. This discussion into recent developments of engineered therapeutic enzymes provide context for the original work described in later chapters.

1.1 Overview of Therapeutic Enzymes

Enzymes form the basis of metabolism for all living organisms, performing a variety of chemical transformations with impressive specificity and rate enhancement. These biological catalysts have inherent regioselective and stereoselective properties, often catalyzing reactions on extremely narrow ranges of substrate. This exquisite control enables reactions to be catalyzed to high yield in mild environments while also excluding side reactions. Enzymes are a \$5 billion global market, having widespread roles from the manufacturing of consumer, industrial and medicinal products to tools for research and development (1). Increasingly, scientists have a desire to extend applied enzymology to create enzymes with properties not found in nature. These past two decades have witnessed intense industrial and academic endeavors to overcome the limitations of naturally evolved catalysts and broaden the scope of enzyme-based catalysis.

Enzyme engineering is a common approach to tailor enzyme function and encompasses two broadly defined strategies (2, 3). Perhaps the simplest is rational design, reliant on previous knowledge between sequence, structure and function. Supported by an ever increasing bioinformatic knowledge, site-directed mutagenesis introduces amino acid substitutions to modify the wild-type sequence, guided by intuition of the enzyme system (4). Multiple beneficial substitutions are then recombined to further improve enzyme performance. While not all aspects of enzyme function can be predicted, rational design has historically been successful in the discovery of phenotypically desirable variants (5).

The second and more generalizable approach is directed evolution (i.e. random design), requiring no knowledge of enzyme mechanism, and can optimize proteins for an arbitrary function (6). These approaches include techniques such as error prone PCR to randomly introduce mutations in a region of interest (e.g. active site), or even across the entire reading frame (7). These unguided manipulations of enzyme properties are screened for a desired function, and the

top performing variants (often containing multiple amino acid substitutions) template a new round of randomized sequences. This iterative approach is repeated until a variant with the desired phenotype is obtained. Directed evolution has proven successful for a variety of protein engineering efforts (5). Despite their descriptions, rational and random design are not inherently incompatible, and an emerging strategy is to combine the randomized methodology of directed evolution with elements of rational intuition (8-10). These data-driven approaches have encouraged a trend away from large, unguided libraries toward smaller, functionally enriched libraries, assisted by prior functional knowledge and iterative feedback (5, 9, 11). Commercial technologies employing data-driven design, such as ProSAR and ProteinGPS, have demonstrated the ability to evolve enzyme variants (harboring up to 30 amino acid substitutions) to meet a variety of design criteria, substantially increasing reaction rates and improved properties (12-15).

These engineering strategies have enabled the design of biocatalysts for specific conditions (temperature, pH, etc.), unnatural substrate selectivity and other desired properties toward the synthesis of bulk chemicals, food ingredients, and pharmaceuticals/agrochemical compounds (16). While enzyme engineering encompasses a broad variety of applications, an often overlooked application is the design of therapeutic enzymes.

By definition, therapeutic enzymes are unnaturally introduced enzymes that catalyze a reaction of clinical benefit. These enzymes have unique design criteria, making them distinct from general drug and protein design. Not only must enzymes bind targets with exquisite selectivity, but also physiologically convert the target molecule to the desired product. This catalytic aspect makes enzymes specific and potent drugs, capable of therapies not possible with traditional small molecules, thus permitting the treatment of a variety of disorders. However, protein modifications for clinical applications must also produce enzymes that are active under physiological conditions (e.g. blood plasma), have suitable pharmacokinetic profiles, and elicit no

immunogenic response from patient. These additional constraints increase the complexity and challenge of engineering therapeutic enzymes.

1.1.1 Pharmaceutical Proteins

The trend toward engineered enzymes over native proteins requires an understanding of drug failures from the past. The justification for engineering biological agents can be broadly separated into three categories: 1) improved activity 2) optimized pharmacokinetics 3) reduced immunogenicity. While the focus of this work is toward therapeutic enzymes, this section draws broadly from protein engineering as both catalytic and non-catalytic proteins have similar challenges as pharmaceutical agents (17, 18). Therapeutic binding proteins have a longer historical precedent than their catalytic counterparts, and serve as examples to discuss concepts within therapeutic engineering.

Perhaps the simplest justification for engineering pharmaceutical drug is to improve its activity. Within the past decade, advances in directed evolution and high-throughput technology have enabled the design of drug candidates with substantially improved activity. Without previous knowledge, targeted mutagenesis to regions of interest randomly produce protein variants harboring several amino acid substitutions from the wild-type. These unguided modifications to protein sequence can improve binding affinity; however, they require the screening of a large number of variants. The advent of phage display, linking surface displayed fusion proteins to the genes which encode them, enabled the screening of enormous libraries of variants ($>10^{10}$) for *in vitro* binding (19). This high-throughput approach has led to the development of several therapeutic antibodies, such as Humira® (Abbvie). In laboratory assays, the engineered antibody demonstrated high affinity (picomolar) to inhibit TNF α receptors, implicated in the inflammatory response of autoimmune disorders, such as Crohn's disease (20). Within clinical trials, 80% of patients prescribed Humira® demonstrated therapeutic benefit, and

the engineered protein has annual sales at \$13 billion (as of date), making it the best selling drug worldwide (21, 22). Similarly, the use of high-throughput technologies, such as *in vitro* compartmentalization, have enabled the screening of large libraries for catalytic activity, intended for therapeutic applications (23). The desire for improved catalytic activity is the primary motivation for nearly every enzyme class and will be discussed in greater detail in following sections.

Apart from *in vitro* activity, the rate of pharmacokinetic absorption is critical to the function of any drug, and protein engineering is a useful tool to improve bioavailability. For example, insulin was the first approved protein drug. The hormone is natively stored as an inert hexamer, and is active as a monomer (24). Traditionally, the release and activity of recombinant insulin is modulated by formulation chemistry, however Humalog® (Eli Lilly) has a simple modification to dramatically alter pharmacokinetics. The engineered protein has the proline and lysine residues reversed at positions 28 and 29 respectively, resulting in a shift of the oligomeric equilibrium to exclusively form monomers (25). The observed clinical outcome was a greater absorption upon administration and rapid initiation of action. Another more substantial protein engineering effort was performed on tissue plasminogen activator (t-PA), as an improved thrombolytic drug. The native sequence effectively clears blood clots, however has a short 6 minute half-life due to downstream proteolysis (26). Lanoteplase, developed by Wyeth, lacks several protein binding domains and demonstrated an increased 45-minute biological half-life (27). Lanoteplase is an example of enzyme engineering for pharmacokinetics, and its clinical outcome will be described in greater detail in following sections.

The most common reason for drug failure, especially at the later stages, is due to immunogenicity, and can be overcome by protein engineering. Historically, the majority of pre-clinical research has been done with non-human proteins, however early clinical trials of

these proteins were almost always terminated due to an observed immune response (28). In some cases, simply altering the protein origin to the human homolog was sufficient. For example, *in vitro* and animal models for DNase I, a therapeutic agent to treat cystic fibrosis, were done with the bovine amino acid sequence (29). Pulmozyme® (Genentech) is the homologous sequence within humans and displayed the intended therapeutic effect without observed immunogenicity (30).

Unfortunately, the reduction of immune response toward proteins is not straightforward and remains a significant challenge. For antibody drugs, the ‘humanizing’ of mouse antibody sequences has demonstrated some success (31). For example, Herceptin® (Genentech), an effective therapeutic against early stage breast cancer consisting of the grafted complementarity determining region (CDR) sequence from the mouse antibody into the human antibody framework (32). An alternative approach has been the random replacement of surface residues and screening against known neutralizing antibodies. While these approaches have demonstrated limited success (33, 34), a general solution to protein immunogenicity has yet to be identified.

Protein engineering has been a clinically successful strategy to overcome the challenges of drug development over the past decades. While therapeutic enzymes are relatively recent, these drugs have the same challenges of low activity, poor pharmacokinetics and immunogenicity. Enzyme engineering is a proven toolkit and enables the design of improved catalytic drugs for current and future generations.

1.1.2 Gene-Directed Enzyme Prodrug Therapy

A more recent application of therapeutic enzymes is toward the development of improved chemotherapies through the introduction of foreign genes. Chemotherapy has been established for decades as the standard of treatment for cancer patients (35). However, the high systemic toxicity of the therapeutic agents remains a dose-limiting constraint. Moreover, available drugs are not

cancer specific, having off-target effects to noncancerous cells. Gene-directed enzyme prodrug therapy (GDEPT) involves the intratumoral delivery and expression of so-called ‘suicide-genes’, encoding for enzymes which by themselves are not toxic, yet elicit cytotoxic effects in combination with a prodrug (36).

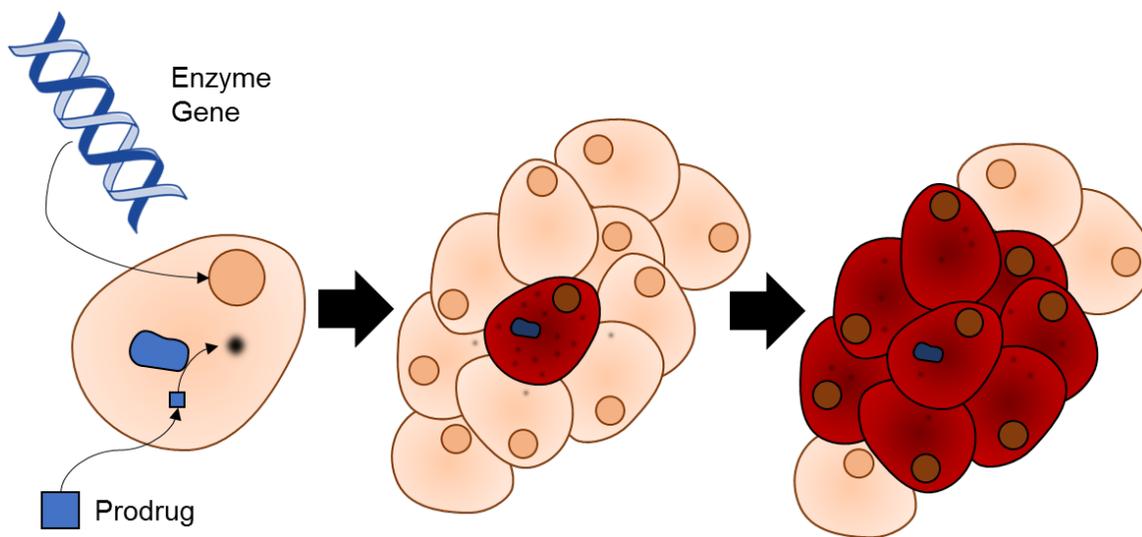


Figure 1.1. Mode of Action for Gene Directed Enzyme Prodrug Therapy. In a general case, the gene encoding for the prodrug activating gene (i.e. suicide gene) is delivered to the target cell and is subsequently expressed. Once the administered prodrug enters the cell, the ectopic enzyme activates the prodrug, inducing its cytotoxic effect. Prior to cell death, the activated drug is transferred to neighboring cells by mechanisms such as cell-to-cell junction. This bystander effect facilitates the destruction of surrounding cells as well, resulting in local tumor regression

Mootlen et al. introduced the GDEPT concept three decades ago, and while several systems have been developed since then, none have developed into marketable drugs (37). This suggests there are limitations that must be overcome for GDEPT to become standard-of-care. First, the enzyme must be selectively expressed within the target tumor cells, depending on specific delivery methods. Second, the enzyme must have high activity toward the prodrug, producing its cytotoxic effect at low concentrations. Currently, the poor transfection efficiency cannot selectively introduce the suicide gene to all tumor cells, however GDEPT relies on the

bystander effect to transfer the activated cytotoxic agent to neighboring tumor cells through a variety of mechanisms (e.g. cell-cell junctions), resulting in an overall tumor regression. Finally, the enzyme/prodrug must not independently result in toxicity. This generalized strategy is described in Figure 1.1.

The wild-type sequences are often poor GDEPT candidates, having low activity for prodrugs, while simultaneously having rampant activity for native metabolites, potentially misregulating cell metabolism (38). Several approaches have been investigated to overcome native limitations, and a detailed discussion of each strategy is beyond the scope of this work, and is described on other works (39). Instead, the focus is the use of protein engineering to optimize potential candidates for enzyme-prodrug therapies (39, 40). Several enzymes have been proposed for GDEPT, and four classes in particular have been the focus of enzyme engineering. The first and most established class is based on deoxynucleoside kinases, using deoxynucleoside analogs as prodrugs. The second consist of cytosine deaminases, an enzyme class not found within humans, which utilizes prodrugs originally developed as fungicides. The third pertain to purine nucleoside phosphorylase, another nucleoside metabolizing enzyme, and the final class encompasses nitroreductases, which activate custom prodrugs.

1.1.3 Diagnostic Enzymes

Enzymes are routinely used as analytical reagents, selectively detecting target analytes as a substrate. Relative to traditional analytical methods (i.e. chromatography-mass spectrometry), enzyme-based sensors are simple, selective systems with rapid assay times and low cost. Enzymes have particular advantages within a clinical setting and are often capable of detecting biochemical markers directly from biological samples. For decades, enzymes have been used within clinical assays to detect a variety of metabolic products, such as urea, glucose, cholesterol and steroid hormones; however there is still room for improvement.

Apart from limited reaction environments (i.e. pH, temperature, solvent), enzymes have a limited dynamic range and detection for a restricted set of compounds. For this reason, enzymes for analytical purposes are a relatively smaller market compared to enzymes as pharmaceutical/industrial applications (annual sales \$0.1 versus \$5 billion respectively) (1). That being said, research within the past two decades have focused on overcoming the previously mentioned limitations of enzyme-based biosensors. Within the biosensors literature, enzyme stability is commonly improved by immobilization (41). This technique however, has limited capability to improve other properties. Another strategy involves the use of enzyme engineering to introduce amino acid substitutions for the design of nearly any arbitrary purpose.

Interestingly, early work in directed evolution started with analytical enzymes as a model system. Established analytical enzymes, such as peroxidases, inherently have picomolar sensitivity, and are generally adaptable for high-throughput assays. Peroxidases in particular have historical precedent as ‘proof-of-concept’ in enzyme engineering and several groups over the past decade have been evolved peroxidases with increased stability, activity and selectivity (42-44) .

Beyond the ‘proof-of-concept’, enzyme engineering has been applied to cholinesterases, a family of enzymes critical to neurotransmission. Small molecules, such as organophosphates and carbamates, are potent inhibitors of these enzymes, and these compounds are both drugs for neurodegenerative disorders or neurotoxins (45). The main advantage of this enzyme class is their high activity, allowing for detection activity at low concentrations, either directly from electrochemical measurements or established coupled assays using colorimetric detection. Initial work focused on acetylcholinesterases (AChE), a neurotransmitter which is the primary molecular target of the aforementioned small molecules and butyrylcholinesterase (BuChE), a serum enzyme with detoxification properties (46). These cholinesterases have traditionally not been amenable to high-throughput methodologies, however recent advances using *in vitro*

compartmentalization have encouraged the evolution of high activity esterases to not only detect, but rapidly detoxify nerve agents (47).

1.2 Proteases

Originally thought to have a simple digestive role, proteases have since been discovered to have ubiquitous roles in biology. In fact, approximately 2% of the human genome encode for proteases, which are intricately involved in intracellular signaling and activating other proteins, such as growth factors and cytokines (48). In their own right, proteases are considered molecular targets for established drug classes, but the enzyme itself is not often considered a drug.

The focus of this section is to discuss the major applications of proteases within clinical therapies and improvements achieved by enzyme engineering. This discussion begins with the use of proteases to treat cardiovascular disorders, the first and most established use of enzymes for therapy. The following sections focus on other established applications of proteases, such as digestive and neurodegenerative disorders.

1.2.1 Plasminogen Activators

Urokinase plasminogen activator (u-PA) was the first enzyme approved by the FDA in 1978, and not only ushered proteases as a class of drug for thrombolytics, but demonstrated the commercial viability of enzymes as drugs (49). u-PA is a protease with broad substrate specificity, and its therapeutic mechanism as a thrombolytic is to cleave plasminogen to its active form plasmin, found at the fibrin laden surface of a blood clot (50). The site is then exposed for other factors, such as tissue plasminogen activator (t-PA), amplifying downstream signaling to instigate clot breakdown (51). u-PA is an alternative to surgical removal of the embolism, clearing blood clots through exposure to the agent. Due to the low cost and efficacy, u-PA retains its clinical application for non-targeted fibrinolysis to this day, such as catheter clearing. Unfortunately, further mechanistic studies discovered u-PA has relatively low affinity for fibrin

(compared to other factors such as t-PA) and cannot be intravenously administered to treat localized fibrin deposits.

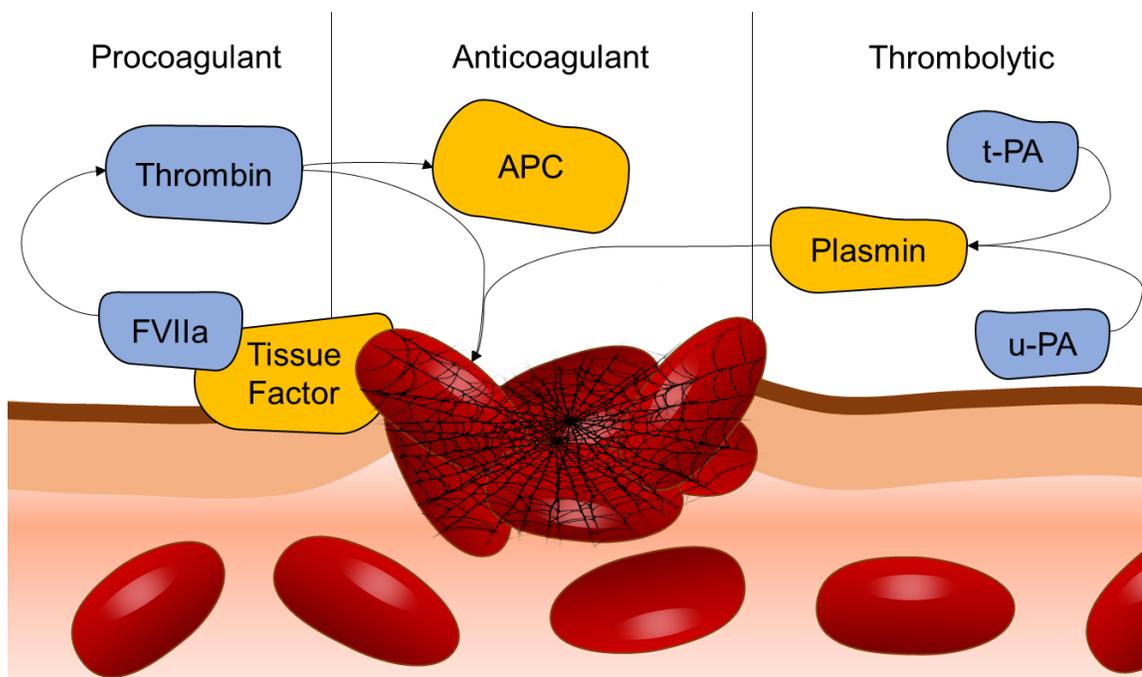


Figure 1.2. Protease Therapeutics for Cardiovascular Disorders. The signaling cascade in response to a damaged tissue is largely mediated by proteases, which can be modified for altered therapeutic benefit. Damaged tissue elicits the formation of a blood clot (thrombus) laden with the protein fibrin. For a thrombolytic effect, plasminogen is activated proteases, such as t-PA and u-PA, to plasmin clearing the blood clot for trauma bleeding, such as in a stroke. Thrombin has the opposite effect, and promotes blood clotting. Upon tissue exposure, cells display tissue factor on their surface, activating signaling proteases such as FVII to their activated form FVIIa. These converted proteins then further activate thrombin for its therapeutic effect. Engineered FVIIa have demonstrated utility to treat bleeding disorders, such as hemophilia. Thrombin also activates APC, as part of a feedback mechanism, to promote anticoagulation. Modulating thrombin for preferential APC activity has been used in anticoagulation therapies, curiously opposite of the protein's native function.

Usage	Enzyme	Product/Reference	Modification	Observed Property
Thrombolytic	t-PA	Activase® (Approved 1987)	(none)	Local fibrinolysis at nascent blood clots for indications such as stroke
		Tenecteplase (Approved 2000)	T103N/N117Q/K296A/ H297A/R298A/R299A	Altered glycosylation and reduced inhibition, and increased half-life
		Lanoteplase (discontinued)	N117G, deletion of fibronectin and epidermal growth factor domain	Muted glycosylation and domain deletions improve half-life
Procoagulant	Thrombin	Recothrom® (Approved 2008)	(none)	Activates fibrinogen, promoting coagulation at damaged tissues for wound treatment
Anticoagulant	Thrombin	Gibbs et al. (1995)	E299A	Improved protein C activation, a feedback clotting factor, reversing native response
		Tsiang et al. (1995)	E217A	Selective improvement to protein C activation over fibrinogen
		Gruber et al. (2000)	W215A/E217A	Reduced fibrinogen activation, but maintains protein C activation
Procoagulant	Factor VIIa	NovoSeven® (Approved 1999)	(none)	Induced by tissue factor, a membrane signal protein, to promote coagulation for hemorrhage treatment
		NN1731 (discontinued)	V158D/E296V/M298Q	Protease forced in tissue factor induced conformation
		BAY 86-6150 (discontinued)	Y4 insertion, P10Q/K32E/D33F/A34E	Improved affinity for membranes and tissue factor induction
Digestive	Endoprotease	ALV003 (Phase II)	(none)	Combination of native glutamine and proline specific proteases to treat celiac diseases
		Ehren et al. (2008)	I581V/F459Y/M511L/I406V	Improve proline specific activity
		Gordon et al. (2012)	V119D/S262K/N291D/ D293T/G319S/D358G/ D368H	Improved specificity for -PQ- dipeptide and immunogenic peptides
Neurological	Botulinum toxin	Botox® (Approved 1989)	(none)	Cleaves SNAP-25 complex, inhibiting neurotransmission
		Chen and Barbieri (2009)	K224D	Altered specificity, instead cleaving SNAP-23 complex

Table 1.1. Pharmaceutical Proteases Modified by Directed Evolution

Plasminogen activating factor (t-PA) was the second enzyme to be FDA approved in 1987, developed by Genentech as Activase®, for myocardial infarction and eventually for stroke. This enzyme was the second recombinant protein drug to be marketed (the first being insulin in 1982). t-PA also activates plasminogen, but the protease preferentially acts on fibrin matrices, enabling a general administration to produce local fibrinolysis at indicated areas (52). However, t-PA is rapidly inactivated by endogenous inhibitors (specifically PAI-1) as part of a natural feedback regulation to prevent against rampant fibrinolysis (53). The recombinant t-PA has a biological half-life of a few minutes, requiring constant infusion during stroke/heart attack treatment.

Given the clinical potential, second generation t-PA drugs improved the properties of this protease. In 2000, the FDA approved TNKase® also developed by Genentech, demonstrating improved resistance to endogenous inhibitors. This improved variant contains three modifications from the wild-type t-PA sequence. The first modification was identified from the early investigations of Madison et al., employing alanine scanning to identify substitutions at positions 296-299 to limit interactions with the inhibitor PAI-I while retaining protease activity (54). From rational design studies, two additional modifications, T103N and N117Q, were included to alter the carbohydrate content of t-PA. The former introduces a new N-linked glycosylation site, improving resistance to other proteases, while the latter eliminates a non-human glycosylation site observed from recombinant expression. These sequence modifications resulted in a modest improvement of half-life, from 6 minutes to 18 minutes in clinical trials (26).

Unfortunately, not all attempts to improve t-PA bioavailability have been successful. Lanoteplase, developed by Wyeth, introduces the N117G single point mutation, removing a glycosylation site and deleting two recognition domains (finger and epidermal growth factor). The modified enzyme had a clinically prolonged high half-life of 45 minutes, improving clinical

utility by reducing the required dosing frequency. Tragically, the increased half-life had unintended consequences, resulting in severe intracranial hemorrhage due to unregulated thrombolysis (55). This example demonstrates the difficulty to engineer enzymes as drug therapies, as the prediction of both physiological and pharmacological consequences is often not possible.

1.2.2 Thrombin

Many other protease homologous to u-PA/t-PA have been redesigned for therapeutic benefit. One such protease, thrombin, has a role in maintaining the equilibrium between procoagulation and anticoagulation outcomes (56). Thrombin itself has a native procoagulation effect, proteolytically activating several downstream receptors. Interestingly, thrombin also activates protein C (APC), another trypsin-like protease, as a feedback mechanism. The secondary protein APC activates separate anticoagulant pathways. While the wild-type thrombin is marketed as a therapeutic for wound bleeding, researchers have since engineered thrombin for alternate enzyme properties to elicit therapeutic benefit.

Gibbs et al. created a rational library of 62 thrombin mutants, rationally selected to substitute solvent-exposed charged residues to alanine, and screened for anticoagulation properties, opposite to its native function (57). One such mutation, E299A, enhanced protein C activation 22-fold within *in vitro* experiments, however administration of the modified proteases within animal models did not affect blood clotting.

This early study did not account for the allosteric regulation of thrombin activity, as the protease is believed to have three physiological states, modulated by Na^+ availability within the blood (58). At high sodium concentrations, thrombin acts as a procoagulant, having high activity for its native substrate fibrin. In the absence of Na^+ , thrombin transitions between two states, one having little activity and the other having no detectable activity (59). Forcing thrombin to adopt

alternate forms through mutagenesis has been an extensively investigated strategy to elicit improved anticoagulation properties.

Tsiang et al. employed alanine scanning to 77 surface exposed positions, screening for improved activity in the presence of the sodium (60). The group identified the substitution E217A to have procoagulant properties, enhancing protein C activation 40-fold with reduced fibrin activity. Gruber et al. followed up this work by rational design of the double mutant W215A/E217A (61). The modified thrombin had 20,000 fold-lower activity for fibrin, yet still activated protein C in the presence of Na⁺. This modified enzyme has demonstrated efficacy in primate models for anticoagulation, and has been implicated in several other therapies.

1.2.3 Procoagulant

Factor VII (FVII) is another homologous clotting agents than has been investigated for procoagulation. Recombinant FVIIa, marketed as NovoSeven® by Novo Nordisk, was approved by the FDA for the treatment of hemophilia in 1999. The drug is gaining interest in trauma care for uncontrollable hemorrhage and various cardiovascular disorders. While the wild-type formulation has been demonstrated as a safe and effective treatment for procoagulation, there is a potential market for improved variants.

Persson et al. consolidated rationally identified substitutions from previous works, resulting in a triple mutant (V158D/E296V/M298Q) of FVIIa (62). The activity of the wild-type FVIIa is stimulated by association with tissue factor (TF), which is locally expressed during vascular injury. The triple mutant has amino acid substitutions that force the protease to adopt a conformation similar to the TF-induced structure. This modified FVIIa has 100-fold improved activity within *in vitro* studies, increasing resistance to both small molecule and protein inhibitors. The drug (labeled as NN1731) recently completed phase III clinical trials.

Unfortunately, several patients developed anti-drug antibodies eliciting a neutralizing response. Development of NN1731 has since been discontinued.

Harvey et al. also used protein engineering to improve FVIIa for a separate limitation (63). The wild-type FVIIa has low affinity for the membranes of platelets, reducing its *in vivo* efficacy. The group was guided by previous literature, suggesting mutagenesis to the Gla domain, membrane contact region, can improve affinity (64). Site-saturation mutagenesis of this 40 residue domain was done in combination with a light-scattering assay to evaluate protein binding to phospholipid vesicles. The resulting four substituted variant (P10Q/K32E/D33F/A34E) exhibited a 250-fold affinity improvement for membranes and 40-fold improvement in procoagulation activity compared to wild-type enzyme. Bayer pharmaceuticals continued development of the mutated FVIIa, termed BAY 86-6150, progressing to phase III clinical trial for the treatment of hemophilia. Unfortunately, the study was also halted, as patients developed anti-drug antibodies, neutralizing the drug effect, as with NN1731.

The clinical outcome of the modified FVIIa variants NN1371 and BAY 86-6150 highlight the difficulty of engineering enzymes as therapeutic agents. Both studies were halted as a precautionary measure due to an unforeseen immune response, resulting from genetic modification. By comparison, the standard of care treatment NovoSeven®, the recombinant expression of the wild-type sequence, has not been reported to elicit neutralizing antibodies. The development of non-immunogenic enzyme variants with alternate properties remains a consistent challenge for the engineering of pharmaceutical agents.

1.2.4 Digestive Proteases

One digestive target is celiac disease, a chronic illness in which patients elicit an immune response to gluten proteins. The molecular basis for celiac disease is gluten proteins, which predominantly contain proline and glutamine, making the peptides resistant to proteolysis by gastrointestinal endoproteases. The current treatment is the elimination of gluten from a patient's diet entirely, as no curative therapy exists (65). The condition is genetic, as 0.5-1% of the population lack proline- or glutamine-specific endoproteases greatly reducing the ability to detoxify dietary gluten (66). Currently, a combination of EP-B2 (a glutamine-specific protease expressed in barley) and PEP (a proline-specific protease expressed in bacteria), termed ALV003 by Alvine, is in Phase II clinical trials as an oral candidate (67). In addition to these native proteases, there have been recent attempts to apply directed evolution to evolve enzymes selective for gluten.

Ehren et al. focused on a proline-specific proteases derived from *S. capsulate* (14). Using a combination of sequence- and structure-based approaches with machine learning algorithms, the group identified several improved variants with clinical relevance. Through two rounds of iterative mutagenesis and analysis, lead variant (I581V/F459Y/M511L/I406V) displayed 20% enhanced specific activity at pH 4.5 and 200-fold greater resistance to other digestive proteases. The biological impact of these substitutions is not clearly understood.

In a somewhat orthogonal approach, Gordon et al. focused on the acidic protease kumaoslin from *A. senegalensis* to alter its specificity for a model gluten dipeptide (-PQ-) over native dipeptide cleavages (-PR-, -QP-, and -PE-) (68). The group used Rosetta, a computational protein design program, to re-engineer specificity toward the immunogenic element. The engineered enzyme (V119D/S262K/N291D/D293T/G319S/D358G/D368H) exhibited 100-fold greater proteolytic activity for a model gluten peptide over the native template enzyme, as well as

an 800-fold switch in substrate specificity toward immunogenic gluten peptides. Based on crystallography, these substitutions are thought to introduce a new hydrogen bond within the glutamine residue position (P1 pocket of protease) and general favorable interactions within the active site. The computationally engineered enzyme is resistant to proteolysis by digestive proteases and degraded over 95% of an immunogenic peptide implicated in celiac disease in under an hour.

1.2.5 Other indications

In the past several decades, proteases have provided clinical benefit to a variety of therapies beyond those discussed in this work and will continue to provide benefits with expanded use. The recombinant production of protease drugs has since replaced isolation from animal source and enables further optimization via genetic engineering. Since then, several examples of engineered proteases have improved activity, resistance to inhibitors, and even been evolved for proteolytic functions not naturally present. This ability to alter native protease function has the potential to redesign enzymes for treatment, even within unrelated metabolic pathways.

An expanding field of interest is within proteases for neurological disorders. The most well known example is Botox®, marketed by Allergan, to inhibit uncontrollable spasms, and more recently for cosmetic usage. The protease drug selectively cleaves a single isoform of a membrane signaling protein (SNAP25), preventing neuronal signal transduction (69). Chen and Barberi used rational engineering to alter substrate specificity of the protease to selectively inhibit SNAP23, implicated in non-neuronal secretory diseases (70). While not an exhaustive strategy of enzyme engineering, this work suggests the possibility to adapt existing therapies for novel indications.

By far the most common use of therapeutic enzymes is the treatment of orphan diseases, typically gene-replacement therapy for rare genetic disorders (71). To date, there are ten FDA approved enzyme designated as orphan drugs (none are engineered enzymes), however dozens of similar diseases with known enzyme-based targets have yet to be explored. Proteases are no exception and have tremendous potential for gene-replacement therapies. Puente et al. have reported 53 hereditary diseases caused by malfunctioning proteases, none of which are currently clinical drugs (72). Due to their orphan status, research into these established drugs are limited to the natural/recombinant purifications of the wild-type enzyme. However, just like any drug, there is substantial room for improvement for bioavailability and clinical effectiveness. With further understanding combined with improvements in protein engineering, enzymes as pharmaceutical agents will make a significant contribution to healthcare in the near future.

1.3 Deoxynucleoside Kinases

Nucleoside analogs are synthetic molecules designed to mimic their physiological counterparts to primarily treat viral infections, and more recently cancer. These analogs enter the cells through transporters and are activated by either native or viral deoxynucleoside kinases to their monophosphate forms (73). Each nucleoside kinase has selective activity for the native nucleosides, and by extension has selective activity for selective nucleoside analogs (74). The nucleoside monophosphates are then further phosphorylated to their active di- and triphosphate forms by native nucleoside monophosphate and nucleoside diphosphate kinases respectively. Once activated, these compounds directly incorporate into elongating DNA or inhibit intracellular enzymes to elicit their toxic effect (75).

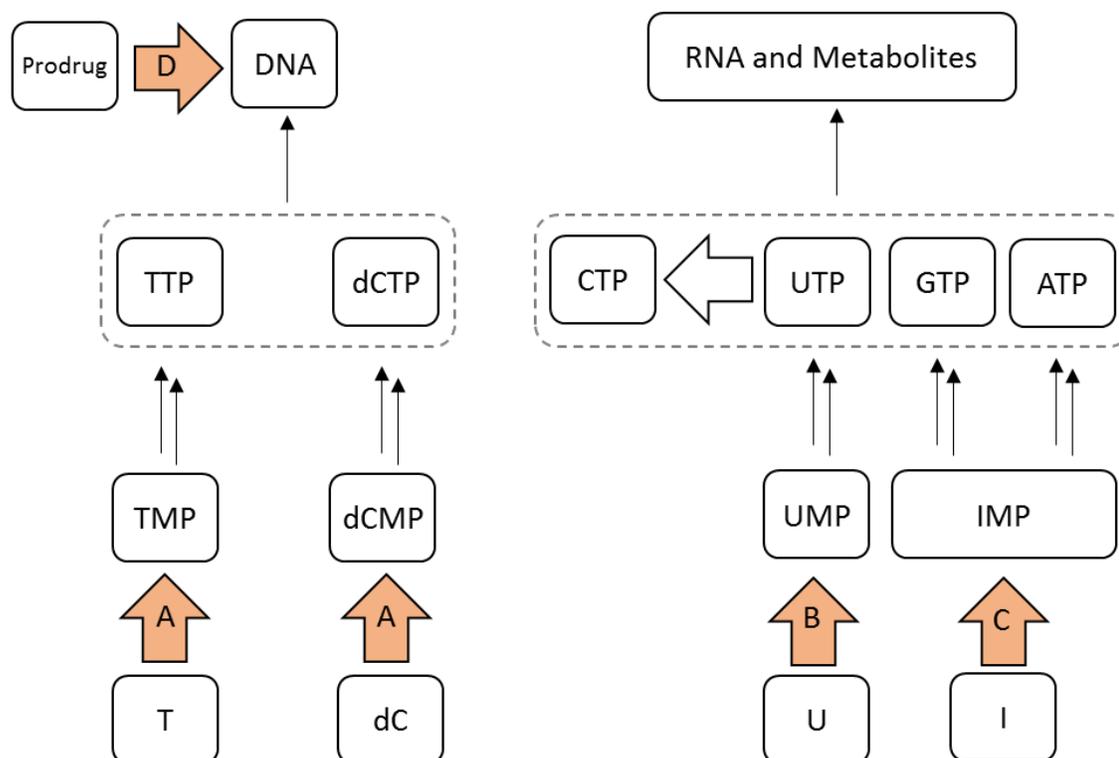


Figure 1.3. Abbreviated Nucleoside Salvage Pathway and GDEPT Candidates. The image above highlights the 4 classes of enzymes, shown in orange, discussed in this Chapter. In general, these candidate enzymes activate a non/less toxic prodrug from their integration into nucleoside metabolism eliciting their ultimate cytotoxic effect. (A) Describes the human deoxynucleoside kinases thymidine kinase (type 2) and deoxycytidine kinase having native activity for the native pyrimidine deoxynucleosides. These native kinases also have limited activity for nucleoside analog based prodrugs. (B) are cytosine deaminases, and enzyme not found within humans. These enzymes convert cytosine analogs to the corresponding uracil species, and incorporate into nucleoside metabolism. (C) Depicts the enzyme purine nucleoside phosphorylase, reversibly catalyzing the conversion the nitrogen base and ribose moiety to/from purine nucleosides. As such, certain analogs can readily incorporate into purine metabolism, specifically adenosine analogs, having amongst the most potent cytotoxic effects. The final class (D) are nitroreductases, again not found in humans, which directly activate DNA-chelating drugs, and do not interact with the nucleoside metabolism machinery.

Enzyme	Reference	Modification	Observed Property
Thymidine kinase Herpes Simplex Type I	Black and Leob (1993)	(none) L159I/I160L/F161A/ A168Y/L169F	High promiscuous activity Improved activity for GCV, a purine analog
	Black et al. (2001)	L159I/I160F/F161L/ A168F/L169M	Improved activity for GCV
	Christians et al. (1999)	Chimeric Enzyme	Improved activity for AZT, a thymidine analog
	Hinds et al. (2000)	Q125E	Reduced pyrimidine activity
	Balzarini et al (2002)	A167Y	Reduced pyrimidine activity
Deoxynucleoside Kinase <i>D. melanogaster</i>		(none)	High promiscuous activity
	Knecht et al. (2000)	N45D/N64D	Increased activity for AZT/ddC, both pyrimidine analogs
	Knecht et al. (2007)	N45D/N64D/N210D/ L239P	General activity increase
	Knecht et al. (2002)	V84A/M88R/A110D	Activity increase for purine analogs
	Gerth et al. (2007)	Chimeric Enzyme	Improved activity for D4T, a thymidine analog
	Liu et al. (2009)	T85M/E172V/Y179F/ H193Y	Improved activity for ddT, a thymidine analog similar to D4T
Deoxycytidine Kinase <i>H. sapiens</i>		(none)	Selective deoxycytidine activity
	Sabini et al. (2007)	A100V/R104M/D133A	Expanded activity for thymidine analogs
	Iyidogan et al. (2008)	R104M/D133A	Improved activity for L-ribose analogs
	Muthu et al. (2014)	R104M/V130T/D133A/ L191A	
	Muthu and Lutz (2015)	E87Q/K88D/E90S/ R104M/D133N/L141M/ C146H/T153A/N60K/ W161F/R219K	Improved activity for L-FMAU, an L-ribose analog, while reducing activity for native substrates
Thymidine Kinase Type 2 <i>H. sapiens</i>		(none)	Selective thymidine activity
	Sabini et al. (2007)	Positions 1-18 Truncated	Selective accumulation of thymidine analogs upon overexpression
	Campbell et al. (2012)	N93D/L109F	Improved activity for L-FMAU

Table 1.2. Nucleoside Kinase GDEPT Candidates Modified by Directed Evolution

Within the pathway described in Figure 1.3, the first two phosphorylation steps are generally considered rate-determining step. The accumulation of the activated nucleoside analog is dependent on the catalytic efficiency of the metabolizing enzymes, and consequently the therapeutic effect. Due to the importance of this drug class, a significant body of literature exists to improve the activity of these kinases through protein engineering for improved therapies for both viral and cancer treatments and more recently for gene directed enzyme prodrug therapies (GDEPT) and positron emission tomography (PET) reporter genes.

1.3.1 Herpes Simplex Virus type 1 Thymidine Kinase

Herpes Simplex Virus type 1 thymidine kinase (HSV-TK) was the first, and is the most commonly used kinase for pre-clinical investigations. The enzyme candidate is the only GDEPT system clinically tested in human patients (76). The viral deoxynucleoside kinase is natively responsible for the phosphorylation of thymidine to replicate infection machinery (77). The enzyme has an alternate substrate specificity to native mammalian kinases and is a molecular target of various nucleoside analogs based drugs (78). The main advantage of HSV-TK is the viral enzyme exhibits 1,000-fold higher activity for nucleoside analogs over native human kinases. Of particular note, HSV-TK catalyzes the rate-limiting activation of ganciclovir, enabling the selective incorporation of the analog into nascent DNA of infected cells (79). The compromised genetic fidelity induces cell death within these compromised cells.

While HSV-TK demonstrates high turnover numbers (i.e., k_{cat}), the enzyme has poor affinity for prodrug analogs, generally preferring binding to native metabolites by 2-3 orders of magnitude, based on K_M . Because of competition for the active site, requisite doses are relatively high, significantly reducing the clinical value of HSV-TK for suicide gene therapies. Several directed evolution studies have been conducted to increase affinity toward analogs, effectively lowering the dose of prodrug required. From previous structural and kinetic studies, the positions 160-170 were known to have a role in substrate binding (80). Black and Loeb employed random mutagenesis to target this region, resulting in a million-member library, containing HSV-TK variants with up to five amino acid substitutions (46). The work utilized thymidine kinase deficient *E. coli*, and screened for affinity for ganciclovir, acyclovir and azidothymidine. In particular, amino acid substitutions to the F161 displayed 2-3 fold increased activity for both native thymidine and nucleoside analogs. The group continued this work with a more aggressive mutagenesis strategy, using a similar same selection strategy. This work focused on saturation of the active site positions 159-161 and 168-170 for random mutagenesis. Two variants, termed 30

(L159I/I160L/F161A/A168Y/L169F) and 75 (I160L/F161L/A168V/L169M), demonstrated 5-fold improvement in K_M for analogs and 50-fold improvement to overall activity.

Having more evidence in the active site positions responsible for substrate recognition, their work continued employed a semi-random library, targeting five active site residues. Black et al. identified a further optimized variant (L159I/I160F/F161L/A168F/L169M) (81). To date, this semi-random variant SR39 exhibits the greatest sensitivity to nucleoside analogs among identified HSV-TK variants, and has translated into a 10-fold reduction of ganciclovir concentrations compared to wild-type to inhibit tumor growth. The SR39 variant has repeatedly demonstrated its effectiveness as a suicide gene for several cancer models and remains a benchmark for GDEPT (82, 83).

Several groups have attempted sequence improvement to the HSV-TK sequence. Christians et al. used DNA shuffling to screening of several thousand variants, which demonstrated significant improvements to catalytic efficiency for azidothymidine, a potent viral prodrug (84). Hinds et al. focused on three variants Q125E, Q125N and Q125D and achieved a selective reduction in activity against the native pyrimidines (85). Computational modeling by Balzarini et al. targeted the A167Y and A168H, eliminating native pyrimidine activity, and slightly improving activity for nucleoside analogs (86). Unfortunately, these studies did not result in significant improvement over SR39.

The initial therapeutic indication of using kinases for GDEPT was to improve activation for nucleoside analogs. This first phosphorylation event is often considered a mechanistic bottleneck, preventing efficient phosphorylation to monophosphates (39). This recent decade has focused on synthetic alternatives for prodrug activation, specifically phosphate esters as masking groups (87). Instead of relying on deoxynucleoside kinases to generate a monophosphate species, prodrugs are delivered as phosphate esters, and the uncharged compounds diffuse across cell

membrane. Once inside the cell, these the masking groups are removed by chemical/enzymatic hydrolysis. The resulting charged phosphate is impermeable to the cell membrane is activated further to subsequent cytotoxic triphosphates. These synthetic alternatives have limited the utility of kinases for specific GDEPT applications.

As such, recent trends describe a shift from the activation of nucleoside analogs as prodrugs for GDEPT toward analogs as probes for reporter genes, in combination with positron emission tomography (PET) imaging (88). PET is versatile method for noninvasive visualization of biological processes in living subjects and is based on a two-component system. First is an isotopically labeled small-molecule reporter (typically ^{18}F). Second is the reporter gene expressed within target cells, which interacts and trap the radionuclide reporter whose accumulation can be detected via PET. Both wild type HSV-TK and SR39 have been used to monitor expression and migration of foreign genes within gene and somatic cell therapies in several preclinical applications (81, 83, 89). To date, HSV-TK is the only reporter that has been used to image therapeutic cells in human patients.

1.3.2 *Drosophila melanogaster* Deoxynucleoside kinase

Drosophila melanogaster deoxynucleoside kinase (Dm-dNK) is expressed in fruit fly embryos and was found to possess broad dNK specificity (90). This kinase also exhibits wide promiscuity, having activity across several classes of nucleoside analogs. The broad substrate scope of Dm-dNK has been used in a variety of biotechnological applications. Synthetic routes to generate nucleoside analog monophosphates typically have poor yield and purity, however Serra et al. used Dm-dNK for the biocatalytic transformation for a variety of analogs for improved synthesis (91). In an alternative application, our lab has investigated Dm-dNK as a potential biosensor. as part of therapeutic dosage monitoring (TDM), to dynamically optimize individual dosage regimens. This work incorporated Dm-dNK, among other deoxynucleoside kinases,

within an enzymatic array to algorithmically determine the concentrations of nucleoside-analog drugs directly from human plasma. This work potentially enables physicians to measure drug pharmacokinetics for individual patients to improve clinical outcome in future therapies.

Dm-dNK has also been the subject of several protein engineering studies to introduce novel kinetic properties for nucleoside analog activation. Among the first attempts was by Knecht et al, implementing several rounds of random mutagenesis to improve analog activity. Of particular interest, the double substituted MuD (N45D/N64D) demonstrated increased specificity toward pyrimidine prodrugs azidothymidine (AZT) and dideoxycytidine (ddC) (92). From subsequent crystallographic studies, these substitutions eliminate 3'-OH coordination enabling the alternate activity. In a separate work, Knecht et al. focused on purine analogs and employed a rational strategy based on structural comparisons of HSV-TK and Dm-dNK (93). The work identified a triple substituted variant (V84A/M88R/A110D) containing active site substitutions, which demonstrated improved kinetic performance for purine analogs, albeit not to the same extent as pyrimidine analogs. In their follow-up work, Knecht et al. performed DNA shuffling among previous lead variants, and identified further improvements to analog activation. Two variants B5 (V84A/N210D/L239P) and B10 (N45D/N64D/N210D/L239P) were further investigated within cellular models, and exhibited ~1,000-fold sensitivity to various purine analogs, such as cladribine (CdA) and (F-AraA) relative to wild-type transduced cells (94).

In an alternate directed evolution strategy, Gerth et al. implemented a non-homologous recombination techniques (ITCHY and SCRATCHY), generating hybrids of Dm-dNK and human thymidine kinase 2 (95). Of particular interest were the chimeras HD-12 and HD-16, exhibiting significantly improved activity for towards the anti-HIV prodrug 2',3'-didehydro-3'-deoxythymidine (d4T). At the time, this observed activity was two-orders of magnitude greater than any natural or engineered kinases, demonstrating the potential of directed evolution.

Liu et al. improved upon these results with the design of a FACS-based screening approach for Dm-dNK to increase substrate-specificity for the prodrug dideoxythymidine (ddT) (96). Through iterative rounds of random mutagenesis and DNA shuffling, this approach identified a variant R4.V3 (T85M/E172V/Y179F/H193Y), with a 20-fold increase in substrate preference for ddT over thymidine. Liu et al. followed up this work using Rosetta, a computational program, to redesign the enzyme active site and achieved a similar performing enzyme. However, this method did not rely on high-throughput screening and suggests the potential of employing computational approaches for more efficient design over traditional directed evolution approaches (97).

While both the Dm-dNK variants and HSV-TK variants are strong candidates for GDEPT, these non-human enzymes have a significant limitation, as immunocompetent hosts will reject transduced cells and recognize the foreign enzyme. Recent engineering efforts have focused on tailoring the properties of human kinases to overcome immunogenicity.

1.3.3 *H. sapiens* Deoxycytidine Kinase

Human deoxycytidine kinase (dCK) is among the latest kinases of clinical interest for suicide-gene therapy. The human-derived enzyme inherently has reduced immunogenicity and phosphorylates deoxycytidine and purines as its native function (88). Although the wild-type dCK has relatively poor activity compared to HSV-TK and Dm-dNK, dCK has sufficient activity to function as a molecular target, and is known to activate various nucleoside analog prodrugs (88).

Initial attempts to improve the catalytic efficiency of dCK were rational strategies based on crystallographic data. Sabini et al. transferred substitutions from the previously described triple mutant of Dm-dNK, resulting in the analogous dCK variant (A100V/R104M/D133A) (98), broadening substrate specificity, improved activity for the native deoxycytidine, and enabled

unprecedented activity for thymidine and several nucleoside analogs. In a separate study, Iyidogan and Lutz (2008) implemented a combination of rational and random design studies to explore substrate specificity, providing increased biochemical detail for substrate interactions (99). Of particular interest, was the variant epTK6 (D47E/R104Q/D133G/N163I/F242L), displayed a broader specificity and elevated turnover for all natural and nucleoside analog substrates. Hazra et al. expanded this investigation even further and performed site saturation at these positions to investigate substrate specificity (100). The double substituted, DMMA (R104M/D133A) and DMLA (R104L/D133A) demonstrated superior catalytic performance for analogs, and uniquely improved activity for the L-enantiomer of thymidine.

Similar to HSV-TK, engineered human dCK have seen an alternative application for PET reporter genes to track transgene migration and expression. Initial work focused on the native kinase within dCK-deficient cell lines, using the radioprobe 2-deoxy-2-fluoroarabinofuranosylcytosine to visualize activated lymphoid tissue *in vivo* (101). Interestingly, the previously engineered dCK variants, harboring several amino acid substitutions to expand the substrate specificity, demonstrated the ability to use thymidine-based radiotracers as well. These dCK mutants were capable of phosphorylating pyrimidine-based radiotracers at levels comparable to that of wild-type HSV-TK, using thymidine based probes (102).

Although the previously described dCK mutants display increased activity toward nucleoside analogs, these engineered kinases still favor the native metabolites, based on K_M values (98, 100). This limits their clinical value within cellular models, as competition for the active site reduces therapeutic effect while also potentially misregulating native nucleoside metabolism. This clinical demand for orthogonal kinases as next-generation reporters for high-contrast molecular PET imaging have encouraged several semi-rational design approaches for the generation of L-selective nucleoside kinases based on the dCK sequence.

As described in Chapter 2, our lab employed a semi-rational engineering strategy to create an L-ribose selective dCK variant (103). Their design strategy leveraged preexisting kinetic data to design a small focused library of 16 variants computationally determined to favor L-nucleosides over their native D-enantiomers. Experimental validation identified a lead variant B6-II (R104M/V130T/D133N/L191A) displaying a 10-fold preference for L-FMAU (a PET radioprobe). The engineered kinase also favored L-FMAU to D-dC, albeit to a lesser extent (2-fold). More recently, our group continued this work implementing a design-of-experiments like methodology (Chapter 3) and systematically evaluated approximately 200 distinct dCK sequences (harboring up to 11 substitutions from the wild type) to determine sequences which favored L-FMAU over D-dC. Several highly evolved kinases were identified, with the best variant, R3-37 (N60K/E87S/E90S/R104M/D133N/C146H/T153A/W161F/R219K), simultaneously displaying 10-fold increase in L-FMAU specific activity and a 100-fold preference over its native D-dC. This variant, and several other kinases identified from this work, are currently being evaluated for efficacy within animal models.

1.3.4 *H. sapiens* Thymidine Kinase Type 2

An alternative candidate for non-immunogenic enzyme design is the human mitochondrial thymidine kinase type 2 (TK2). TK2 phosphorylates thymidine, deoxycytidine, and deoxyuridine, as well as several antiviral and anticancer nucleoside analogs (104). While native expression of TK2 is low compared to the other deoxynucleoside kinases, TK2 is the only pyrimidine kinase among non-proliferating cells. TK2 displays neutral enantioselectivity as both L-thymidine and L-deoxycytidine are efficiently phosphorylated. Unfortunately, the structure of the mitochondrial enzymes has yet to be solved. The poor recombinant expression and lower *in vitro* activity, relative to other kinases, has historically limited protein engineering of TK2. However, recent trends toward human-derived kinases for therapy has developed renewed interest.

The native TK2 kinase displays activity for variety of nucleoside radiolabels and has demonstrated utility as a PET reporter, although with lower performance compared to HSV-TK (105). Campbell et al. used enzyme engineering to overcome the poor performance and develop a reporter gene with orthogonal activity to the wild type enzyme, having activity for preferential activity for L-ribose nucleoside analogs, such as L-FMAU. (106). Guided by structural homology models, the doubly substituted variant N93D/L109F demonstrated improved activity for L-FMAU and had reduced activity for the native D-thymidine. Their work demonstrated efficacy in various preclinical animal models, and remains a candidate therapy.

1.4 Cytosine Deaminases

Cytosine Deaminases are another family of nucleoside metabolizing enzymes studied by directed evolution for GDEPT. The enzyme converts cytosine into uracil and ammonia, and is the mode of action for several fungicides such as 5-fluorocytosine (5FC) (107). The product directly inhibits thymidine synthase, disrupting metabolism and misincorporation into RNA/DNA synthesis. The activated drug has an alternate application as a potent antitumor agent within humans, as mammalian deaminases are considered insensitive to prodrugs. The key advantage of cytosine deaminase systems is the therapeutic prodrugs are smaller, having increased permeability to cell membranes, even to the blood-brain barrier.

The most extensive mutagenesis work been done on the bacterial variant of cytosine deaminase, with the goal to increase activity toward analogs. Initial work leverages structural data, as Mahan et al. targeted residues from 310-320 (which undergoes conformational rearrangement during catalysis) with alanine scanning (108). The D314A variant demonstrated a 20-fold decrease for native cytosine activity and a simultaneous 2-fold increase for 5-fluorocytosine activity. In a follow up study, the group used an *E. coli* negative selection assay to select a randomized library throughout the enzyme reading frame (109). The D314G substitution

greatly enhance sensitivity for analogs, and the study was subsequently continued with site saturation at this position. Every substitution (except D314S) had significantly reduced activity for cytosine, exhibiting at least a 4-fold shift in substrate preference and later confirmed *in vivo*.

Enzyme	Reference	Modification	Observed Property
Cytosine Deaminase <i>E. coli</i>		(none)	Activates various bactericide/fungicide prodrugs
	Mahan et al. (2004)	D134G	Selective increase to 5FC, a cytosine analog
	Fuchita et al. (2009)	V152A/F316C/D317G	Selective increase to 5FC and decrease to cytosine activity
Cytosine Deaminase <i>S. cerevisiae</i>		(none)	High activity for fungicides, however poor stability
	Krokengian et al. (2005)	A23L/I140L/V81I	Improved stability
	Stolworthy et al. (2008)	D92E/M93E/I98L	Improved activity to 5FC
Purine nucleoside phosphorylase <i>E. coli</i>		(none)	Activates adenosine analog prodrugs
	Bennet et al. (2003)	M64V	Improved activity for MePR, an adenosine analog
Nitroreductase <i>E. coli</i>		(none)	Molecular target for several bactericides
	Grove et al. (2003)	F125K	Improved activity for CB1954, an effective DNA crosslinker
	Jaberipour et al. (2010)	T41L/F70A	Improved activity for CB1954
	Barack et al. (2006)	Y230A/Y239N/T160N/Q175L)	Improved activity for CB1954

Table 1.3. Other GDEPT Candidates Modified by Directed Evolution

More recent attempts have improved the selectivity of cytosine deaminase even further. Fuchita et al. expanded targeted mutagenesis to the regions 149-159 and 310-320, screening a million-member library through *E. coli* negative selection (110). The lead candidate bCD₁₅₂₅ (V152A/F316C/D317G), displayed the greatest sensitivity with an approximate 20-fold shift in substrate preference toward 5FC. Subsequent *in vivo* work confirmed this shift in substrate preference with a similar fold-reduction in IC₅₀ values.

In addition to the bacterial deaminase, a separate body of literature has focused on the yeast cytosine deaminase, which has higher overall activity, but is limited by reduced enzyme stability, having a half-life of a few hours. To this end, Korkegian et al. took a computational route to improve the stability of the fungal deaminase (111). The group used Rosetta Design to predict amino acid substitutions to improve stability. The initial experiments identified two

interacting substitutions, A23L and I140L, with an additional distant V108I, thought to improve hydrophobic packing within the protein core. The resulting variant significantly improved half-life on the order of days.

Stolworthy et al. sought to improve activity toward 5FC by employing a million member library using random mutagenesis to select residues based on sequence homology (112). The group identified a triple substituted variant (D92E/ M93L/I98L) having the highest sensitivity through negative selection assays. *In vivo* testing displayed the most efficient cytosine deaminase variant to date. Unlike deoxynucleoside kinases, the deaminases have not progressed to human testing. There are several immunogenicity concerns regarding ectopic expression of a microbial enzyme within patients. Additionally, the prodrugs themselves (potent bactericides/fungicides) eliminate intestinal flora, potentially resulting in unintended side effects.

1.5 Purine Nucleoside Phosphorylases

Purine nucleoside phosphorylase (PNP), another class of nucleoside metabolizing enzymes. The hydrolase cleaves nucleosides to the corresponding nitrogen base and sugar base. Its therapeutic benefit derives from the ability to activate adenosine prodrugs (113). The key advantage is the extreme potency of the adenine prodrugs: 1000-fold more therapeutically active than other classes of anticancer therapy, however is coupled with high toxicity as a significant limitation (114). As such, there have been recent efforts to engineer a PNP enzyme to catalyze the conversion of substrates ignored by human enzymes.

Initial work focuses on 9-(6-deoxy- α -L-talofuranosyl)-6-methylpurine (MePR) interactions with an *E. coli* PNP, which has poor activity with the native sequence. Molecular modeling suggested MePR has a significant steric clash between the M64 and the substrate C6-methyl position. Bennett et al. performed site saturation at this position, and M64V had the highest activity, with a 140-fold improved activity for MePR (115). Unfortunately, even with the

improved activity, *in vivo* data demonstrated lack of clinical efficacy. Animal models demonstrate the intestinal flora efficiently cleave the prodrug, increasing the cytotoxic range. PNP-based systems are currently marred by the choice of prodrug; however, work continues within synthetic chemistry to develop improved small molecules.

Another prodrug of interest is fludarabine, which is converted to the active form 2-fluoroadenine. To date, work has focused on the native *E. coli* PNP, demonstrating efficacy in cell cultures and mouse models. More recently, an adenoviral vector expressing PNP was injected intratumorally within human patients to evaluate safety and efficacy in Phase I clinical trials (IND 14271). The therapy is used in combination with fludarabine to treat solid tumors. Tumor reduction was observed in a dose-dependent response, and is progressing to phase II studies.

1.6 Nitroreductases

Nitroreductases are an example of a non-nucleoside based GDEPT system. These enzymes are a common class of chemoprotector within bacteria, specifically reducing nitro groups of aromatic substituents. These flavin containing enzymes were originally discovered to activate antibiotics, and the enzyme class is entirely absent from mammalian metabolism. The key advantage is within its orthogonality, as an effective prodrug is biologically inert within normal cells. Most work has focused on 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB1954), a prodrug which serves as an effective DNA cross-linker when activated (116).

Initial work by Grove et al. used structural information and saturated six positions within the active site to improve activity for CB1954. The F124K mutant demonstrated a 5-fold improvement for the drug and achieved a similar effect within cell-based models (117). The group continued their work and Jaberipour et al. devised a small 53 member library of double mutants from the same six positions (118). The most effective mutants, T41L/N71S and T41L/F70A, demonstrated 15-fold improved activity and were confirmed in cellular models.

Interestingly, similar activity improvements were observed in a broad range of substrates, increasing potential avenues for therapeutic benefit. In an orthogonal approach, Barack et al. incorporated random mutations throughout the entire reading frame (119). The best variant Y6 (Y230A/Y239N/T160N/Q175L), had a 5-fold improvement toward tumor reduction over the wild-type.

While the four previously discussed enzyme classes have been the most extensively studied (at least in reference to directed evolution), this discussion is by no means an exhaustive list. To date, there are more than twenty GDEPT gene therapy/prodrug strategies described in the literature, twelve such strategies including enzymes that have been evaluated to at least cellular models and have not been subjected to directed evolution for a therapeutic target (39, 40). Since its first description in 1986, hundreds of clinical trials have explored various aspects of suicide gene therapies, and the recent decade has described extensive efforts for engineered GDEPT systems. Even within these four discussed enzymes classes, considerable gains in enzyme properties are yet to be achieved. No ideal system has been discovered, justifying further investigation into the discussed and other enzyme classes.

1.7 Cholinesterases

Cholinesterases refer to a broad family of enzymes that selectively hydrolyze choline esters. Within humans, there are two types: acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) (120). AChE is a critical enzyme found in the central and peripheral nervous systems, having neurotransmission functions at nerve endings. BuChE is more ubiquitously found, and its function is largely unknown. Both enzymes are of particular interest as organophosphates and carbamates are potent inhibitors. These small molecules are exposed to the body as either drugs or poisons, and understanding their interactions is the subject of a large body of literature. Initial mutagenesis work focused on probing the human cholinesterases to

explain and predict the consequences of the administration of these small molecules; however, recent work has shifted toward alternate applications (121). An engineered cholinesterase has potential therapeutic benefit as both as a potential drug candidates and as a next-generation biosensor.

1.7.1 Butyrylcholinesterase

While the physiological function of butyrylcholinesterase (BuChE) is not known, the serum enzyme displays broad detoxification properties, acting as a stoichiometric scavenger against toxins, such as organophosphates (122). Interestingly, BuChE also metabolizes cocaine, albeit less efficiently. An engineered enzyme has potential application for treatment for general cocaine addiction and acute toxicity. Unfortunately, the human BuChE is not yet amenable to high-throughput screening as variants require proper glycosylation and tetramer assembly, necessitating mammalian expression systems. However, BuChE has been previously engineered for therapeutic benefit using rational design.

Xie et al. investigated 24 variants: the A328Y substitution demonstrated a four-fold improvement over wild-type enzyme (123). Sun et al. continued BuChE engineering using molecular modeling to identify the double substituted A328W/Y332A variant, which demonstrated a 20-fold improvement over wild type (124). Pan et al. further improved on the computational framework, using molecular dynamics simulations to re-design the BChE sequence. Their work suggested the A199S/S287G/A328W/Y332G variant demonstrated a reduced energetic barrier for the reaction transition state. Experimental validation confirmed these theoretical results, demonstrating a 500-fold improved catalytic efficiency for cocaine (125). Several of these BuChE variants have been tested within animal models for acute cocaine toxicity and show promise for clinical efficacy (126).

1.7.2 Acetylcholinesterase

Acetylcholinesterase (AChE) is the molecular target of organophosphates and carbamates, the basis of most pesticides, as well as nerve agents. These esters influence enzyme activity (having essential nerve functions within insects and other animals), with organophosphates acting as irreversible inhibitors and carbamates as reversible inhibitors. These compounds can be designed to have acute toxicity to selective organisms depending on the pesticide functionalization, then rapidly hydrolyze, making them environmentally friendly alternatives to more persistent insecticides. While administration of exogenous AChE has been demonstrated as an effective stoichiometric scavenger, AChE has more utility as a biosensor to detect pesticides and related compounds, such as nerve agents (127, 128). The ability to detect single or multiple chemical agents directly from field samples is an unmet need within environmental safety.

Traditional detection of organophosphates and carbamates is based on gas-chromatography coupled to mass spectrometry (129). This technique requires time consuming sample preparation, specialized equipment and personnel. A proposed alternative is to use the molecular target of these pesticides, acetylcholinesterase (AChE), to detect low levels of analytes. This concept was first demonstrated by Bernabei et al., in which the group monitored AChE activity in the presence of different concentrations of organophosphate and carbamate pesticides (130). The characteristic reduction in enzyme activity is directly correlated to pesticide concentration, resulting in an impressive detection limit within the parts per billion (picomolar).

Villatte et al. expanded this concept, screening several AChE homologs for their varying sensitivity to pesticides (131). The group identified the *Drosophila melanogaster* homolog as the most active, understandably having sensitivity to insecticides. The group used structural information and identified a single point mutation (Y408F), which increased general activity by

10-fold. Encouraged by the functional gains through mutagenesis, several groups continued this work and used a variety of directed evolution strategies to obtain optimized variants. Boublik et al. (2002) used alanine scanning to identify critical active site positions, then site-saturated characteristic positions to improve activity (132). An optimized variant (E69Y/Y71D) demonstrated a selective 300-fold improvement against organophosphates (specifically paraoxon), suggesting the possibility to tailor enzymes for individual metabolites.

Most sensors are used independently, detecting a single analyte within a system of interest. Current research is trending toward the concept of sensor arrays, in which significantly more information is gathered from a system than a single reading. These enzyme-based arrays to identify complex mixtures was originally described by Bachmann and Schmid (133). The group was attempting to develop a multielectrode biosensor to identify mixtures of organophosphate and carbamate pesticides. In their initial attempt, the group used four homologs of acetylcholinesterase (AChE), the molecular target of pesticides, and trained the sensor using known binary mixtures. The group used machine learning to blindly determine each component within a mixture to a range of 0-20 $\mu\text{g/L}$ (nanomolar).

The group followed up the work with Bachmann et al., using a small library of variants, identified by the previously described literature, from *Drosophila melanogaster* AChE (WT, Y408F, G368L and F368H) (134). These variants were rationally selected to have divergent sensitivity for a variety of organophosphate and carbamate pesticides. The group was able to deconvolute various binary solutions of 0-5 $\mu\text{g/L}$ of distinct combinations of pesticides using the same machine learning approach. Since then, other groups have used similar systems to demonstrate the ability of this system to detect trace pesticides directly from food and field samples for environmental applications.

1.7.3 Other Esterases

As mentioned earlier, organophosphate-based pesticide and nerve agents are potent inhibitors of AChE, however the enzyme itself cannot detoxify these chemical agents. While stoichiometric bioscavengers based on BChE and AChE have been developed, the requisite mass ratio (mg of protein are needed for single exposure) is not an effective strategy. As such, a catalytic scavenger would serve as better subject for directed evolution. However, natural enzyme degradation is quite low, requiring substantial enzyme engineering efforts.

As a proof of concept, Griffiths and Tawfik selected a bacterial phosphotriesterase (PTE), naturally evolved since the introduction of organophosphate pesticides more than 50 years ago (47, 135). The significant innovation of this work was the implementation of an *in vitro* compartmentalization technique, developed by the group a few years earlier (23). The group screened a million member library of phosphodiesterase using an *in vitro* compartmentalization strategy coupled with FACS to obtain the h5 variant (I106T/F132L) with 50-fold improved activity against organophosphates (47).

The group later moved on to the serum paraoxonase (PON1), a native human esterase with promiscuous activity against organophosphates (136). Gupta et al. utilized a similar compartmentalization strategy to improve the activity of a human enzyme, the native form having limited detoxification of organophosphate nerve agents. The group screened a million member library to degrade a fluorometric analog, and identified the variant 4E9 (L69G/S111T/H115W/H134R/F222S/T332S) having a 10,000-fold improved activity to detoxify a potent nerve agent cyclosarin (137).

Overall, enzyme engineering and directed evolution in particular has established itself as a powerful strategy to overcome challenges within therapeutic enzymes, from the development of

enzyme-based drugs, analytical detection and detoxification. In the coming decades, we can expect engineered enzymes have increasing roles within clinical and diagnostic applications.

1.8 References

1. Adrio, J. L., and Demain, A. L. (2014) Microbial enzymes: tools for biotechnological processes, *Biomolecules* 4, 117-139.
2. Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) Evolving strategies for enzyme engineering, *Curr Opin Struct Biol* 15, 447-452.
3. Bornscheuer, U. T., and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design, *Curr Opin Chem Biol* 5, 137-143.
4. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene* 77, 51-59.
5. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. (2012) Engineering the third wave of biocatalysis, *Nature* 485, 185-194.
6. Lanio, T., Jeltsch, A., and Pingoud, A. (1998) Towards the design of rare cutting restriction endonucleases: using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude, *J Mol Biol* 283, 59-69.
7. Neylon, C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution, *Nucleic Acids Res* 32, 1448-1459.
8. Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, *Curr Opin Biotechnol* 16, 378-384.
9. Lutz, S. (2010) Beyond directed evolution--semi-rational protein engineering and design, *Curr Opin Biotechnol* 21, 734-743.
10. Turner, N. J. (2009) Directed evolution drives the next generation of biocatalysts, *Nat Chem Biol* 5, 567-573.

11. Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution, *ChemBiochem* 9, 1797-1804.
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., and Huisman, G. W. (2007) Improving catalytic function by ProSAR-driven enzyme evolution, *Nat Biotechnol* 25, 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., and Hughes, G. J. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture, *Science* 329, 305-309.
14. Ehren, J., Govindarajan, S., Moron, B., Minshull, J., and Khosla, C. (2008) Protein engineering of improved prolyl endopeptidases for celiac sprue therapy, *Protein Eng Des Sel* 21, 699-707.
15. Govindarajan, S., Mannervik, B., Silverman, J. A., Wright, K., Regitsky, D., Hegazy, U., Purcell, T. J., Welch, M., Minshull, J., and Gustafsson, C. (2015) Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase, *ACS Synth Biol* 4, 221-227.
16. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow, *Nature* 409, 258-268.
17. Kurtzman, A. L., Govindarajan, S., Vahle, K., Jones, J. T., Heinrichs, V., and Patten, P. A. (2001) Advances in directed protein evolution by recursive genetic recombination: applications to therapeutic proteins, *Curr Opin Biotechnol* 12, 361-370.
18. Vasserot, A. P., Dickinson, C. D., Tang, Y., Huse, W. D., Manchester, K. S., and Watkins, J. D. (2003) Optimization of protein therapeutics by directed evolution, *Drug Discov Today* 8, 118-126.

19. Clackson, T., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991) Making antibody fragments using phage display libraries, *Nature* 352, 624-628.
20. Thie, H., Meyer, T., Schirrmann, T., Hust, M., and Dubel, S. (2008) Phage display derived therapeutic antibodies, *Curr Pharm Biotechnol* 9, 439-446.
21. Breedveld, F. C., Weisman, M. H., Kavanaugh, A. F., Cohen, S. B., Pavelka, K., van Vollenhoven, R., Sharp, J., Perez, J. L., and Spencer-Green, G. T. (2006) The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment, *Arthritis Rheum* 54, 26-37.
22. Ecker, D. M., Jones, S. D., and Levine, H. L. (2015) The therapeutic monoclonal antibody market, *MAbs* 7, 9-14.
23. Sepp, A., Tawfik, D. S., and Griffiths, A. D. (2002) Microbead display by in vitro compartmentalisation: selection for binding using flow cytometry, *FEBS letters* 532, 455-458.
24. Wood, A. J., Holleman, F., and Hoekstra, J. B. (1997) Insulin lispro, *N Engl J Med* 337, 176-183.
25. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric Lys B28 Pro B29-human insulin, *Structure* 3, 615-622.
26. Semba, C. P., Sugimoto, K., Razavi, M. K., Society of, C., and Interventional, R. (2001) Alteplase and tenecteplase: applications in the peripheral circulation, *Tech Vasc Interv Radiol* 4, 99-106.
27. Bode, C., Smalling, R. W., Berg, G., Burnett, C., Lorch, G., Kalbfleisch, J. M., Chernoff, R., Christie, L. G., Feldman, R. L., Seals, A. A., and Weaver, W. D. (1996) Randomized

- comparison of coronary thrombolysis achieved with double-bolus reteplase (recombinant plasminogen activator) and front-loaded, accelerated alteplase (recombinant tissue plasminogen activator) in patients with acute myocardial infarction. The RAPID II Investigators, *Circulation* 94, 891-898.
28. McCafferty, J., and Glover, D. R. (2000) Engineering therapeutic proteins, *Curr Opin Struct Biol* 10, 417-420.
29. Shak, S., Capon, D. J., Hellmiss, R., Marsters, S. A., and Baker, C. L. (1990) Recombinant human DNase I reduces the viscosity of cystic fibrosis sputum, *Proc Natl Acad Sci U S A* 87, 9188-9192.
30. Quan, J. M., Tiddens, H. A., Sy, J. P., McKenzie, S. G., Montgomery, M. D., Robinson, P. J., Wohl, M. E., Konstan, M. W., and Pulmozyme Early Intervention Trial Study, G. (2001) A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities, *J Pediatr* 139, 813-820.
31. Roguska, M. A., Pedersen, J. T., Keddy, C. A., Henry, A. H., Searle, S. J., Lambert, J. M., Goldmacher, V. S., Blättler, W., Rees, A. R., and Guild, B. C. (1994) Humanization of murine monoclonal antibodies through variable domain resurfacing, *Proc Natl Acad Sci U S A* 91, 969-973.
32. Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., and Burchmore, M. (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer, *J Clin Oncol* 20, 719-726.
33. Meyer, D. L., Schultz, J., Lin, Y., Henry, A., Sanderson, J., Jackson, J. M., Goshorn, S., Rees, A. R., and Graves, S. S. (2001) Reduced antibody response to streptavidin through site-directed mutagenesis, *Protein Sci* 10, 491-503.

34. Laroche, Y., Heymans, S., Capaert, S., De Cock, F., Demarsin, E., and Collen, D. (2000) Recombinant staphylokinase variants with reduced antigenicity due to elimination of B-lymphocyte epitopes, *Blood* 96, 1425-1432.
35. Corrie, P. G. (2008) Cytotoxic chemotherapy: clinical aspects, *Medicine* 36, 24-28.
36. Greco, O., and Dachs, G. U. (2001) Gene directed enzyme/prodrug therapy of cancer: historical appraisal and future prospectives, *J Cell Physiol* 187, 22-36.
37. Moolten, F. L. (1986) Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy, *Cancer Res* 46, 5276-5281.
38. Encell, L. P., Landis, D. M., and Loeb, L. A. (1999) Improving enzymes for cancer gene therapy, *Nat Biotechnol* 17, 143-147.
39. Zarogoulidis, P., Darwiche, K., Sakkas, A., Yarmus, L., Huang, H., Li, Q., Freitag, L., Zarogoulidis, K., and Malecki, M. (2013) Suicide Gene Therapy for Cancer - Current Strategies, *J Genet Syndr Gene Ther* 4.
40. Duarte, S., Carle, G., Faneca, H., De Lima, M. C. P., and Pierrefite-Carle, V. (2012) Suicide gene therapy in cancer: where do we stand now?, *Cancer letters* 324, 160-170.
41. Sassolas, A., Blum, L. J., and Leca-Bouvier, B. D. (2012) Immobilization strategies to develop enzymatic biosensors, *Biotechnol Adv* 30, 489-511.
42. Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A. H. (1999) Directed evolution of a fungal peroxidase, *Nat Biotechnol* 17, 379-384.
43. Lin, Z., Thorsen, T., and Arnold, F. H. (2000) Functional expression of horseradish peroxidase in E. coli By directed evolution, *Biotechnol Prog* 16, 680.
44. Morawski, B., Quan, S., and Arnold, F. H. (2001) Functional expression and stabilization of horseradish peroxidase by directed evolution in *Saccharomyces cerevisiae*, *Biotechnol Bioeng* 76, 99-107.

45. Pohanka, M. (2011) Cholinesterases, a target of pharmacology and toxicology, *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 155, 219-229.
46. Black, M. E., and Loeb, L. A. (1993) Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants in vitro, *Biochemistry* 32, 11618-11626.
47. Griffiths, A. D., and Tawfik, D. S. (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization, *EMBO J* 22, 24-35.
48. Rodriguez, D., Morrison, C. J., and Overall, C. M. (2010) Matrix metalloproteinases: what do they not do? New substrates and biological roles identified by murine models and proteomics, *Biochim Biophys Acta* 1803, 39-54.
49. Sasahara, A. A. (1973) *The Urokinase Pulmonary Embolism Trial; a National Cooperative Study*, American Heart Assn.
50. Andreasen, P. A., Kjoller, L., Christensen, L., and Duffy, M. J. (1997) The urokinase-type plasminogen activator system in cancer metastasis: a review, *Int J Cancer* 72, 1-22.
51. Medved, L., and Nieuwenhuizen, W. (2003) Molecular mechanisms of initiation of fibrinolysis by fibrin, *Thromb Haemost* 89, 409-419.
52. Hoylaerts, M., Rijken, D. C., Lijnen, H. R., and Collen, D. (1982) Kinetics of the activation of plasminogen by human tissue plasminogen activator. Role of fibrin, *J Biol Chem* 257, 2912-2919.
53. Andreasen, P. A., Egelund, R., and Petersen, H. H. (2000) The plasminogen activation system in tumor growth, invasion, and metastasis, *Cell Mol Life Sci* 57, 25-40.
54. Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M. J., and Sambrook, J. F. (1989) Serpin-resistant mutants of human tissue-type plasminogen activator, *Nature* 339, 721-724.

55. Bhana, N., and Spencer, C. M. (2000) Risperidone: a review of its use in the management of the behavioural and psychological symptoms of dementia, *Drugs Aging* 16, 451-471.
56. Vu, T. K., Hung, D. T., Wheaton, V. I., and Coughlin, S. R. (1991) Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation, *Cell* 64, 1057-1068.
57. Gibbs, C., Coutre, S., Tsiang, M., Li, W., Jain, A., Dunn, K., Law, V., Mao, C., Matsumura, S., and Mejza, S. (1995) Conversion of thrombin into an anticoagulant by protein engineering. *Nature* 378, 413-416.
58. Bah, A., Carrell, C. J., Chen, Z., Gandhi, P. S., and Di Cera, E. (2009) Stabilization of the E* form turns thrombin into an anticoagulant, *J Biol Chem* 284, 20034-20040.
59. Bah, A., Garvey, L. C., Ge, J., and Di Cera, E. (2006) Rapid kinetics of Na⁺ binding to thrombin, *J Biol Chem* 281, 40049-40056.
60. Tsiang, M., Jain, A. K., Dunn, K. E., Rojas, M. E., Leung, L. L., and Gibbs, C. S. (1995) Functional mapping of the surface residues of human thrombin, *J Biol Chem* 270, 16854-16863.
61. Gruber, A., Cantwell, A. M., Di Cera, E., and Hanson, S. R. (2002) The thrombin mutant W215A/E217A shows safe and potent anticoagulant and antithrombotic effects in vivo, *J Biol Chem* 277, 27581-27584.
62. Persson, E. (2004) Variants of recombinant factor VIIa with increased intrinsic activity, *Semin Hematol* 41, 89-92.
63. Harvey, S. B., Stone, M. D., Martinez, M. B., and Nelsestuen, G. L. (2003) Mutagenesis of the gamma-carboxyglutamic acid domain of human factor VII to generate maximum enhancement of the membrane contact site, *J Biol Chem* 278, 8363-8369.
64. Neuenschwander, P. F., and Morrissey, J. H. (1994) Roles of the membrane-interactive regions of factor VIIa and tissue factor. The factor VIIa Gla domain is dispensable for

- binding to tissue factor but important for activation of factor X, *J Biol Chem* 269, 8007-8013.
65. Tack, G. J., Verbeek, W. H., Schreurs, M. W., and Mulder, C. J. (2010) The spectrum of celiac disease: epidemiology, clinical aspects and treatment, *Nat Rev Gastroenterol Hepatol* 7, 204-213.
66. Gujral, N., Freeman, H. J., and Thomson, A. B. (2012) Celiac disease: prevalence, diagnosis, pathogenesis and treatment, *World J Gastroenterol* 18, 6036-6059.
67. Tye-Din, J. A., Anderson, R. P., Ffrench, R. A., Brown, G. J., Hodsman, P., Siegel, M., Botwick, W., and Shreeniwas, R. (2010) The effects of ALV003 pre-digestion of gluten on immune response and symptoms in celiac disease in vivo, *Clin Immunol* 134, 289-295.
68. Gordon, S. R., Stanley, E. J., Wolf, S., Toland, A., Wu, S. J., Hadidi, D., Mills, J. H., Baker, D., Pultz, I. S., and Siegel, J. B. (2012) Computational design of an alpha-gliadin peptidase, *J Am Chem Soc* 134, 20513-20520.
69. Blasi, J., Chapman, E. R., Link, E., Binz, T., Yamasaki, S., De Camilli, P., Sudhof, T. C., Niemann, H., and Jahn, R. (1993) Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25, *Nature* 365, 160-163.
70. Chen, S., and Barbieri, J. T. (2009) Engineering botulinum neurotoxin to extend therapeutic intervention, *Proc Natl Acad Sci U S A* 106, 9180-9184.
71. Melnikova, I. (2012) Rare diseases and orphan drugs, *Nat Rev Drug Discov* 11, 267-268.
72. Puente, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach, *Nat Rev Genet* 4, 544-558.
73. Baldwin, S. A., Beal, P. R., Yao, S. Y., King, A. E., Cass, C. E., and Young, J. D. (2004) The equilibrative nucleoside transporter family, SLC29, *Pflugers Arch* 447, 735-743.
74. Arner, E. S., and Eriksson, S. (1995) Mammalian deoxyribonucleoside kinases, *Pharmacol Ther* 67, 155-186.

75. Jordheim, L. P., Durantel, D., Zoulim, F., and Dumontet, C. (2013) Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases, *Nat Rev Drug Discov* 12, 447-464.
76. Trask, T. W., Trask, R. P., Aguilar-Cordova, E., Shine, H. D., Wyde, P. R., Goodman, J. C., Hamilton, W. J., Rojas-Martinez, A., Chen, S. H., Woo, S. L., and Grossman, R. G. (2000) Phase I study of adenoviral delivery of the HSV-tk gene and ganciclovir administration in patients with current malignant brain tumors, *Mol Ther* 1, 195-203.
77. Wigler, M., Silverstein, S., Lee, L. S., Pellicer, A., Cheng, Y., and Axel, R. (1977) Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells, *Cell* 11, 223-232.
78. Liu, Y., and Santi, D. V. (1998) A continuous spectrophotometric assay for thymidine and deoxycytidine kinases, *Anal Biochem* 264, 259-262.
79. Kokoris, M. S., and Black, M. E. (2002) Characterization of herpes simplex virus type 1 thymidine kinase mutants engineered for improved ganciclovir or acyclovir activity, *Protein Sci* 11, 2267-2272.
80. Munir, K. M., French, D. C., Dube, D., and Loeb, L. A. (1992) Permissible amino acid substitutions within the putative nucleoside binding site of herpes simplex virus type 1 encoded thymidine kinase established by random sequence mutagenesis, *J Biol Chem* 267, 6584-6589.
81. Black, M. E., Kokoris, M. S., and Sabo, P. (2001) Herpes simplex virus-1 thymidine kinase mutants created by semi-random sequence mutagenesis improve prodrug-mediated tumor cell killing, *Cancer Res* 61, 3022-3026.
82. Pantuck, A. J., Matherly, J., Zisman, A., Nguyen, D., Berger, F., Gambhir, S. S., Black, M. E., Belldegrun, A., and Wu, L. (2002) Optimizing prostate cancer suicide gene therapy using herpes simplex virus thymidine kinase active site variants, *Hum Gene Ther* 13, 777-789.

83. Wiewrodt, R., Amin, K., Kiefer, M., Jovanovic, V. P., Kapoor, V., Force, S., Chang, M., Lanuti, M., Black, M. E., Kaiser, L. R., and Albelda, S. M. (2003) Adenovirus-mediated gene transfer of enhanced Herpes simplex virus thymidine kinase mutants improves prodrug-mediated tumor cell killing, *Cancer Gene Ther* 10, 353-364.
84. Christians, F. C., Scapozza, L., Cramer, A., Folkers, G., and Stemmer, W. P. (1999) Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling, *Nat Biotechnol* 17, 259-264.
85. Hinds, T. A., Compadre, C., Hurlburt, B. K., and Drake, R. R. (2000) Conservative mutations of glutamine-125 in herpes simplex virus type 1 thymidine kinase result in a ganciclovir kinase with minimal deoxypyrimidine kinase activities, *Biochemistry* 39, 4105-4111.
86. Balzarini, J., Liekens, S., Solaroli, N., El Omari, K., Stammers, D. K., and Karlsson, A. (2006) Engineering of a single conserved amino acid residue of herpes simplex virus type 1 thymidine kinase allows a predominant shift from pyrimidine to purine nucleoside phosphorylation, *J Biol Chem* 281, 19273-19279.
87. Hecker, S. J., and Erion, M. D. (2008) Prodrugs of phosphates and phosphonates, *J Med Chem* 51, 2328-2345.
88. Eriksson, S., Kierdaszuk, B., Munch-Petersen, B., Oberg, B., and Johansson, N. G. (1991) Comparison of the substrate specificities of human thymidine kinase 1 and 2 and deoxycytidine kinase toward antiviral and cytostatic nucleoside analogs, *Biochem Biophys Res Commun* 176, 586-592.
89. Fillat, C., Carrio, M., Cascante, A., and Sangro, B. (2003) Suicide gene therapy mediated by the Herpes Simplex virus thymidine kinase gene/Ganciclovir system: fifteen years of application, *Curr Gene Ther* 3, 13-26.

90. Munch-Petersen, B., Piskur, J., and Sondergaard, L. (1998) Four deoxynucleoside kinase activities from *Drosophila melanogaster* are contained within a single monomeric enzyme, a new multifunctional deoxynucleoside kinase, *J Biol Chem* 273, 3926-3931.
91. Serra, I., Conti, S., Piškur, J., Clausen, A. R., Munch-Petersen, B., Terreni, M., and Ubiali, D. (2014) Immobilized *Drosophila melanogaster* deoxyribonucleoside kinase (DmdNK) as a high performing biocatalyst for the synthesis of purine arabinonucleotides, *Adv Synth & Catal* 356, 563-570.
92. Knecht, W., Munch-Petersen, B., and Piskur, J. (2000) Identification of residues involved in the specificity and regulation of the highly efficient multisubstrate deoxyribonucleoside kinase from *Drosophila melanogaster*, *J Mol Biol* 301, 827-837.
93. Knecht, W., Sandrini, M. P., Johansson, K., Eklund, H., Munch-Petersen, B., and Piskur, J. (2002) A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines, *EMBO J* 21, 1873-1880.
94. Knecht, W., Rozpedowska, E., Le Breton, C., Willer, M., Gojkovic, Z., Sandrini, M. P., Joergensen, T., Hasholt, L., Munch-Petersen, B., and Piskur, J. (2007) *Drosophila* deoxyribonucleoside kinase mutants with enhanced ability to phosphorylate purine analogs, *Gene Ther* 14, 1278-1286.
95. Gerth, M. L., and Lutz, S. (2007) Non-homologous recombination of deoxyribonucleoside kinases from human and *Drosophila melanogaster* yields human-like enzymes with novel activities, *J Mol Biol* 370, 742-751.
96. Liu, L., Li, Y., Liotta, D., and Lutz, S. (2009) Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting, *Nucleic Acids Res* 37, 4472-4481.
97. Liu, L., Murphy, P., Baker, D., and Lutz, S. (2010) Computational design of orthogonal nucleoside kinases, *Chem Commun* 46, 8803-8805.

98. Sabini, E., Hazra, S., Konrad, M., Burley, S. K., and Lavie, A. (2007) Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase, *Nucleic Acids Res* 35, 186-192.
99. Iyidogan, P., and Lutz, S. (2008) Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity, *Biochemistry* 47, 4711-4720.
100. Hazra, S., Sabini, E., Ort, S., Konrad, M., and Lavie, A. (2009) Extending thymidine kinase activity to the catalytic repertoire of human deoxycytidine kinase, *Biochemistry* 48, 1256-1263.
101. Radu, C. G., Shu, C. J., Nair-Gill, E., Shelly, S. M., Barrio, J. R., Satyamurthy, N., Phelps, M. E., and Witte, O. N. (2008) Molecular imaging of lymphoid organs and immune activation by positron emission tomography with a new [18F]-labeled 2'-deoxycytidine analog, *Nat Med* 14, 783-788.
102. Likar, Y., Zurita, J., Dobrenkov, K., Shenker, L., Cai, S., Neschadim, A., Medin, J. A., Sadelain, M., Hricak, H., and Ponomarev, V. (2010) A new pyrimidine-specific reporter gene: a mutated human deoxycytidine kinase suitable for PET during treatment with acycloguanosine-based cytotoxic drugs, *J Nucl Med* 51, 1395-1403.
103. Muthu, P., Chen, H. X., and Lutz, S. (2014) Redesigning Human 2'-Deoxycytidine Kinase Enantioselectivity for L-Nucleoside Analogues as Reporters in Positron Emission Tomography, *ACS Chem. Biol* 9, 2326-2333.
104. Wang, L., Munch-Petersen, B., Herrstrom Sjoberg, A., Hellman, U., Bergman, T., Jornvall, H., and Eriksson, S. (1999) Human thymidine kinase 2: molecular cloning and characterisation of the enzyme activity with antiviral and cytostatic nucleoside substrates, *FEBS Lett* 443, 170-174.
105. Ponomarev, V., Doubrovin, M., Shavrin, A., Serganova, I., Beresten, T., Ageyeva, L., Cai, C., Balatoni, J., Alauddin, M., and Gelovani, J. (2007) A human-derived reporter gene for

- noninvasive imaging in humans: mitochondrial thymidine kinase type 2, *J Nucl Med* 48, 819-826.
106. Campbell, D. O., Yaghoubi, S. S., Su, Y., Lee, J. T., Auerbach, M. S., Herschman, H., Satyamurthy, N., Czernin, J., Lavie, A., and Radu, C. G. (2012) Structure-guided engineering of human thymidine kinase 2 as a positron emission tomography reporter gene for enhanced phosphorylation of non-natural thymidine analog reporter probe, *J Biol Chem* 287, 446-454.
107. Mullen, C. A., Kilstrup, M., and Blaese, R. M. (1992) Transfer of the bacterial gene for cytosine deaminase to mammalian cells confers lethal sensitivity to 5-fluorocytosine: a negative selection system, *Proc Natl Acad Sci U S A* 89, 33-37.
108. Mahan, S. D., Ireton, G. C., Stoddard, B. L., and Black, M. E. (2004) Alanine-scanning mutagenesis reveals a cytosine deaminase mutant with altered substrate preference, *Biochemistry* 43, 8957-8964.
109. Mahan, S. D., Ireton, G. C., Knoeber, C., Stoddard, B. L., and Black, M. E. (2004) Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy, *Protein Eng Des Sel* 17, 625-633.
110. Fuchita, M., Ardiani, A., Zhao, L., Serve, K., Stoddard, B. L., and Black, M. E. (2009) Bacterial cytosine deaminase mutants created by molecular engineering show improved 5-fluorocytosine-mediated cell killing in vitro and in vivo, *Cancer Res* 69, 4791-4799.
111. Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme, *Science* 308, 857-860.
112. Stolworthy, T. S., Korkegian, A. M., Willmon, C. L., Ardiani, A., Cundiff, J., Stoddard, B. L., and Black, M. E. (2008) Yeast cytosine deaminase mutants with increased thermostability impart sensitivity to 5-fluorocytosine, *J Mol Biol* 377, 854-869.

113. Bzowska, A., Kulikowska, E., and Shugar, D. (2000) Purine nucleoside phosphorylases: properties, functions, and clinical aspects, *Pharmacol Ther* 88, 349-425.
114. Bennett, E. M., Li, C., Allan, P. W., Parker, W. B., and Ealick, S. E. (2003) Structural basis for substrate specificity of *Escherichia coli* purine nucleoside phosphorylase, *J Biol Chem* 278, 47110-47118.
115. Bennett, E. M., Anand, R., Allan, P. W., Hassan, A. E., Hong, J. S., Levasseur, D. N., McPherson, D. T., Parker, W. B., Secrist, J. A., 3rd, Sorscher, E. J., Townes, T. M., Waud, W. R., and Ealick, S. E. (2003) Designer gene therapy using an *Escherichia coli* purine nucleoside phosphorylase/prodrug system, *Chem Biol* 10, 1173-1181.
116. Bridgewater, J., Springer, C., Knox, R., Minton, N., Michael, N., and Collins, M. (1995) Expression of the bacterial nitroreductase enzyme in mammalian cells renders them selectively sensitive to killing by the prodrug CB1954, *Eur J Cancer* 31, 2362-2370.
117. Grove, J. I., Lovering, A. L., Guise, C., Race, P. R., Wrighton, C. J., White, S. A., Hyde, E. I., and Searle, P. F. (2003) Generation of *Escherichia coli* nitroreductase mutants conferring improved cell sensitization to the prodrug CB1954, *Cancer Res* 63, 5532-5537.
118. Jaberipour, M., Vass, S. O., Guise, C. P., Grove, J. I., Knox, R. J., Hu, L., Hyde, E. I., and Searle, P. F. (2010) Testing double mutants of the enzyme nitroreductase for enhanced cell sensitisation to prodrugs: effects of combining beneficial single mutations, *Biochem Pharmacol* 79, 102-111.
119. Barak, Y., Thorne, S. H., Ackerley, D. F., Lynch, S. V., Contag, C. H., and Matin, A. (2006) New enzyme for reductive cancer chemotherapy, YieF, and its improvement by directed evolution, *Mol Cancer Ther* 5, 97-103.

120. Vellom, D. C., Radic, Z., Li, Y., Pickering, N. A., Camp, S., and Taylor, P. (1993) Amino acid residues controlling acetylcholinesterase and butyrylcholinesterase specificity, *Biochemistry* 32, 12-17.
121. Schwarz, M., Glick, D., Loewenstein, Y., and Soreq, H. (1995) Engineering of human cholinesterases explains and predicts diverse consequences of administration of various drugs and poisons, *Pharmacol. Ther.* 67, 283-322.
122. Lockridge, O. (2015) Review of human butyrylcholinesterase structure, function, genetic variants, history of use in the clinic, and potential therapeutic uses, *Pharmacol Ther* 148, 34-46.
123. Xie, W., Altamirano, C. V., Bartels, C. F., Speirs, R. J., Cashman, J. R., and Lockridge, O. (1999) An improved cocaine hydrolase: the A328Y mutant of human butyrylcholinesterase is 4-fold more efficient, *Mol Pharmacol* 55, 83-91.
124. Sun, H., Pang, Y.-P., Lockridge, O., and Brimijoin, S. (2002) Re-engineering butyrylcholinesterase as a cocaine hydrolase, *Mol Pharmacol* 62, 220-224.
125. Pan, Y., Gao, D., Yang, W., Cho, H., Yang, G., Tai, H. H., and Zhan, C. G. (2005) Computational redesign of human butyrylcholinesterase for anticocaine medication, *Proc Natl Acad Sci U S A* 102, 16656-16661.
126. Brimijoin, S., Gao, Y., Anker, J. J., Gliddon, L. A., Lafleur, D., Shah, R., Zhao, Q., Singh, M., and Carroll, M. E. (2008) A cocaine hydrolase engineered from human butyrylcholinesterase selectively blocks cocaine toxicity and reinstatement of drug seeking in rats, *Neuropsychopharmacology* 33, 2715-2725.
127. Kim, K., Tsay, O. G., Atwood, D. A., and Churchill, D. G. (2011) Destruction and detection of chemical warfare agents, *Chem Rev* 111, 5345-5403.

128. Lenz, D. E., Yeung, D., Smith, J. R., Sweeney, R. E., Lumley, L. A., and Cerasoli, D. M. (2007) Stoichiometric and catalytic scavengers as protection against nerve agent toxicity: a mini review, *Toxicology* 233, 31-39.
129. Alder, L., Greulich, K., Kempe, G., and Vieth, B. (2006) Residue analysis of 500 high priority pesticides: better by GC-MS or LC-MS/MS?, *Mass Spectrom Rev* 25, 838-865.
130. Bernabei, M., Chiavarini, S., Cremisini, C., and Palleschi, G. (1993) Anticholinesterase activity measurement by a choline biosensor: application in water analysis, *Biosens Bioelectron* 8, 265-271.
131. Villatte, F., Marcel, V., Estrada-Mondaca, S., and Fournier, D. (1998) Engineering sensitive acetylcholinesterase for detection of organophosphate and carbamate insecticides, *Biosens Bioelectron* 13, 157-164.
132. Boublik, Y., Saint-Aguet, P., Lougarre, A., Arnaud, M., Villatte, F., Estrada-Mondaca, S., and Fournier, D. (2002) Acetylcholinesterase engineering for detection of insecticide residues, *Protein Eng* 15, 43-50.
133. Bachmann, T. T., and Schmid, R. D. (1999) A disposable multielectrode biosensor for rapid simultaneous detection of the insecticides paraoxon and carbofuran at high resolution, *Anal Chim Acta* 401, 95-103.
134. Bachmann, T. T., Leca, B., Vilatte, F., Marty, J. L., Fournier, D., and Schmid, R. D. (2000) Improved multianalyte detection of organophosphates and carbamates with disposable multielectrode biosensors using recombinant mutants of *Drosophila* acetylcholinesterase and artificial neural networks, *Biosens Bioelectron* 15, 193-201.
135. Benning, M. M., Kuo, J. M., Raushel, F. M., and Holden, H. M. (1994) Three-dimensional structure of phosphotriesterase: an enzyme capable of detoxifying organophosphate nerve agents, *Biochemistry* 33, 15001-15007.

136. Amitai, G., Gaidukov, L., Adani, R., Yishay, S., Yacov, G., Kushnir, M., Teitlboim, S., Lindenbaum, M., Bel, P., Khersonsky, O., Tawfik, D. S., and Meshulam, H. (2006) Enhanced stereoselective hydrolysis of toxic organophosphates by directly evolved variants of mammalian serum paraoxonase, *FEBS J* 273, 1906-1919.
137. Gupta, R. D., Goldsmith, M., Ashani, Y., Simo, Y., Mullokandov, G., Bar, H., Ben-David, M., Leader, H., Margalit, R., Silman, I., Sussman, J. L., and Tawfik, D. S. (2011) Directed evolution of hydrolases for prevention of G-type nerve agent intoxication, *Nat Chem Biol* 7, 120-125.

Chapter 2

Computational Design of Deoxycytidine Kinase*

Abstract

* Reproduced with permission from Muthu P, Chen HX, Lutz S. "Redesigning Human 2'-Deoxycytidine Kinase Enantioselectivity for L-Nucleoside Analogues as Reporters in Positron Emission Tomography." ACS Chemical Biology. 2014;9 (10):2326-33. Copyright © 2014 American Chemical Society.

Recent advances in nuclear medicine have allowed for positron emission tomography (PET) to track transgenes in cell-based therapies using PET reporter gene/probe pairs. A promising example for such reporter gene/probe pairs are engineered nucleoside kinases that effectively phosphorylate isotopically labeled nucleoside analogues. Upon expression in target cells, the kinase facilitates the intracellular accumulation of radionuclide monophosphate, which can be detected by PET imaging. We have employed computational design for the semi-rational engineering of human 2'-deoxycytidine kinase to create a reporter gene with selectivity for L-nucleosides including L-thymidine and 1-(2'-fluoro-5-methyl- β -L-arabinofuranosyl) uracil. Our design strategy relied on a combination of preexisting data from kinetic and structural studies of native kinases, as well as two small, focused libraries of kinase variants to generate an in silico model for assessing the effects of single amino acid changes on favorable activation of L-nucleosides over their corresponding D-enantiomers. The approach identified multiple amino acid positions distal to the active site that conferred desired L-enantioselectivity. Recombination of individual amino acid substitutions yielded orthogonal kinase variants with significantly improved catalytic performance for unnatural L-nucleosides but reduced activity for natural D-nucleosides.

2.1 Introduction

Molecular imaging by positron emission tomography (PET) offers a powerful and versatile method for noninvasive visualization of biological processes in living subjects.(1) In preclinical and clinical studies, PET imaging is routinely used to monitor disease development, progression, and treatment. The technique is based on a two-component system, including an isotopically labeled small molecule reporter and a reporter gene.(2-4) The role of the reporter gene is to co-localize with target cells, as well as to interact and trap the radionuclide reporter whose accumulation can be detected via PET. Among the leading reporter systems are nucleoside kinases used in combination with ^{18}F - or ^{125}I -labeled nucleoside analogue (NA) reporters.

The phosphorylation of NAs for PET imaging was first implemented with thymidine kinase from Herpes Simplex virus (HSV-tk).(5) The enzyme's high substrate promiscuity enables effective activation of the small molecule reporter and has found widespread application in cell culture studies, as well as in translational work with animals and humans.(6-12) These studies are typically conducted with wild type Herpes enzyme or an engineered variant, HSV-sr39TK,(13) using various NAs including 1-(2' -fluoro-5-methyl- β -L-arabinofuranosyl) uracil (L-FMAU).(14) Ideally, these NAs are not phosphorylated by the endogenous human nucleoside kinases yet are readily activated by the Herpes enzymes. However, the experiments with HSV-tk have revealed significant limitations as this PET reporter gene suffers from problems with activity, specificity, and risk of adverse immune reactions. First, HSV-tk is promiscuous, which is important for NA activation, yet it is still primarily a thymidine kinase. The enzyme's high thymidine kinase activity, together with elevated expression levels can result in deregulation of the host cell's tightly controlled dNTP pool, which among other things causes declining DNA replication fidelity.(15, 16) Second, the catalytic activity of HSV-tk for NAs is far from optimal as reflected in inferior kinetic parameters compared to natural nucleosides.(17) Third, the clinical

use of viral kinases raises concerns over potential immunogenicity, especially upon its repetitive and long-term application.(18)

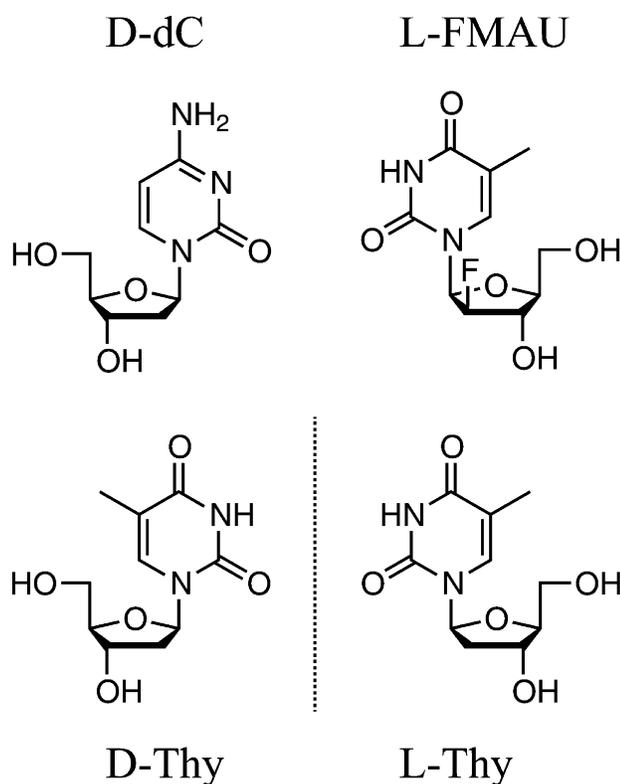


Figure 2.4. Structures of native substrates and analogs. Structures of native D-deoxycytidine and D-thymidine (D-dC and D-Thy), as well as the L-stereoisomeric form of Thy and the nucleoside analogue 2-fluoro-2-deoxy-1- β -L-arabinofuranosyl-5-methyluracil (L-FMAU).

Over the past decade, advances in protein engineering have supported scientists in tailoring kinases to address some of these limitations. Recent efforts have concentrated on the exploration of human nucleoside kinases as reporter genes. The application of human enzymes minimizes the risk of an immune response in clinical applications. A majority of these studies have focused on engineering human deoxycytidine kinase (dCK) and human thymidine kinase 2 (TK2) with the goal to broaden and change the substrate specificities of these enzymes.(19-25) Improvements in their kinetic parameters compared to the native enzymes have been significant,

yet these efforts have tended to identify generalists with broad substrate specificity for native nucleosides and NAs. More recently, a study by Lavie and co-workers specifically focused on tailoring human TK2 for selective L-nucleoside activation.(25) Guided by structural information, amino acid substitutions in two positions (N93D and L109F) resulted in a 2-fold decline in activity for D-Thy. The TK2 variants also showed increased turnover for L-FMAU by the same magnitude, resulting in enhanced catalytic activity. However, the enzyme's specificity constants for these two substrates remained largely unchanged as raised K_M values compensate for the activity gains.

In accordance with clinicians' demands for "ideal" orthogonal kinases as next-generation reporters for high-contrast, low-impact molecular PET imaging,(26) we herein report a semi-rational design approach for the generation of L-selective nucleoside kinases based on dCK. More specifically, two previously engineered dCK variants (ssTK1A and ssTK3)(22) were chosen as parental kinases due to the enzymes' broad substrate specificity including phosphorylation of numerous NA prodrugs and L-nucleosides.(22, 27) Kinetic data for these dCK variants and structural information from crystallographic studies of the wild type enzyme were used to generate a computational model for scoring the impact of amino acid changes on D- or L-Thy bound in the phosphoryl acceptor site. The predictive framework was then tested and refined through two rounds of mutagenesis, using small, focused libraries of dCK variants with single amino acid changes. Combination of individual beneficial mutations were mostly additive and resulted in functional gains, yielding two L-selective candidates with superior kinetic performance for the PET reporter L-FMAU over the natural substrates.

2.2 Results and Discussion

For our initial efforts to explore the contributing factors to enantioselectivity in dCK, we assembled a computational model for protein–ligand interactions. The predictive framework was

based on the crystal structures of wild type dCK bound with 5-methyl dC as phosphoryl acceptor and ADP (PDB code: 3KFX(28)), as well as the dCK variant R104M/D133A in complex with L-Thy and ADP (PDB code: 3HP1(23)). Besides the two amino acid changes at positions 104 and 133, the atomic composition of each protein structure was identical. The only significant difference between the modeled structures was the inverted ribose conformation of the two substrates to maintain the correct geometry of their 5' -OH groups for phosphoryl transfer. The impact of amino acid substitutions on the enantioselectivity of dCK was then modeled using the Rosetta macromolecular modeling suite.(29) Identical amino acid changes were made to both the D- and L-Thy bound dCK structures via fixed backbone side chain replacement followed by independent energy minimizations to generate a structure ensemble. The ensemble average scores were used to identify substitutions in individual amino acid positions favoring bound L-ribosyl nucleosides. The initial data suggested that the standard Rosetta score function did not provide sufficient resolution to capture experimental findings previously reported for dCK variants.(22) For this reason, we re-parameterized Rosetta's score function using 10 dCK crystal structures (PDB codes: 2NO1, 2NO7, 2NOA, 2NO6, 3KFX, 1P5Z, 2NO9, 1P62, 3HP1, and 2ZI4), along with their corresponding experimental data on catalytic activity.(19, 23, 28, 30, 31) The new score function significantly improved the correlation with experimental data and hence was used for designing a small test library (Library A).

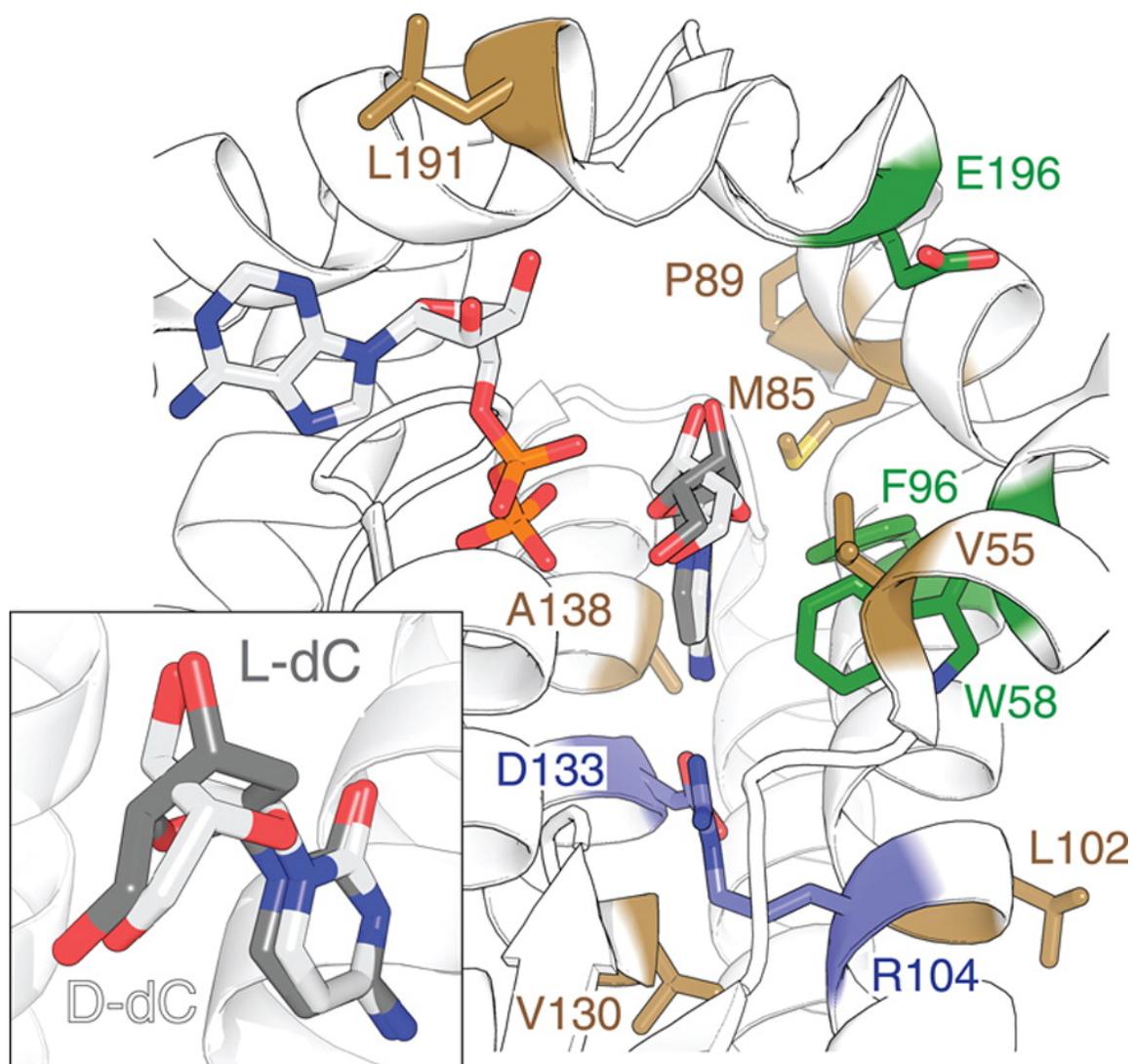


Figure 2.5. Summary of Amino Acid Substitution in dCK. Amino acid substitution in human dCK with bound ADP in the phosphoryl donor site, as well as D- and L-dC in the phosphoryl acceptor binding pocket (PDB access codes: 2NO1 and 2NO7(30)). Variant positions R104M and D133N in ssTK3 are marked in violet. The three positions probed in Library A are highlighted in green, while the seven residues varied in Library B are colored in brown. Insert: Overlay of the D- and L-dC bound in the active site shows the highly similar positioning of the pyrimidine moiety, as well as the 3- and 5-hydroxyl groups.

Library A consisted of eight members. The previously described dCK variant ssTK1A (A100V/R104M/D133S) was selected as scaffold (variant A0) due to its exquisitely high activity for D-dC and D-Thy.(22) Building on ssTK1A, the top three predicted variants (ssTK1A with

F96D, W58E, or E196L) were prepared. In addition, we included three variants with slightly more conservative amino acid changes in the same three positions (ssTK1A with F96Y, W58V, or E196A). Finally, ssTK3 (R104M/D133N) was added to Library A as a control. This variant was identified in the same study as ssTK1A and distinguishes itself by its broad substrate specificity for pyrimidine and purine 2'-deoxynucleosides. Following site-directed mutagenesis, cloning, expression, and purification, all eight Library A variants had their kinetic parameters determined for D- and L-Thy. The catalytic properties of the parental ssTK1A and control ssTK3 for the two enantiomers show some interesting differences. Both dCK variants exhibit similar K_M values for D- and L-Thy, while turnover rates differ by 2-fold. Overall, both enzymes are effective phosphoryl-transfer catalysts, either showing neutral enantioselectivity (ssTK3) or exhibiting a moderate preference for the D-isomer (ssTK1A). Among the six new variants of our survey library, the three lead candidates all resulted in inactive enzymes. The rather dramatic amino acid changes in the selected positions are likely responsible for the failure to detect catalytic activity for these variants. In contrast, the three alternatives carrying more conservative amino acid substitutions yielded dCK variants with detectable phosphoryl-transfer activity. Although the catalytic properties of all three enzymes declined relative to the parental ssTK1A, the desired change in enantioselectivity could be observed in two of the three variants. Variants A5 and A6 both showed approximately 3-fold selectivity improvement in favor of the L-enantiomer. While the effect in A5 is associated with worsening K_M and k_{cat} values for either substrate, the change in A6 is largely driven by differences in K_M values. The experimental data from Library A was used to update and refine our score function in an effort to obtain more accurate predictions on the effects of amino acid changes. Instead of using a traditional linear score function, we investigated the use of non-parametric functions, commonly used in machine learning/data mining applications. After trying various regression methods,(32-35) we found the

k-nearest neighbor algorithm best captured the results from the initial library and was used to design a second generation library (Library B).

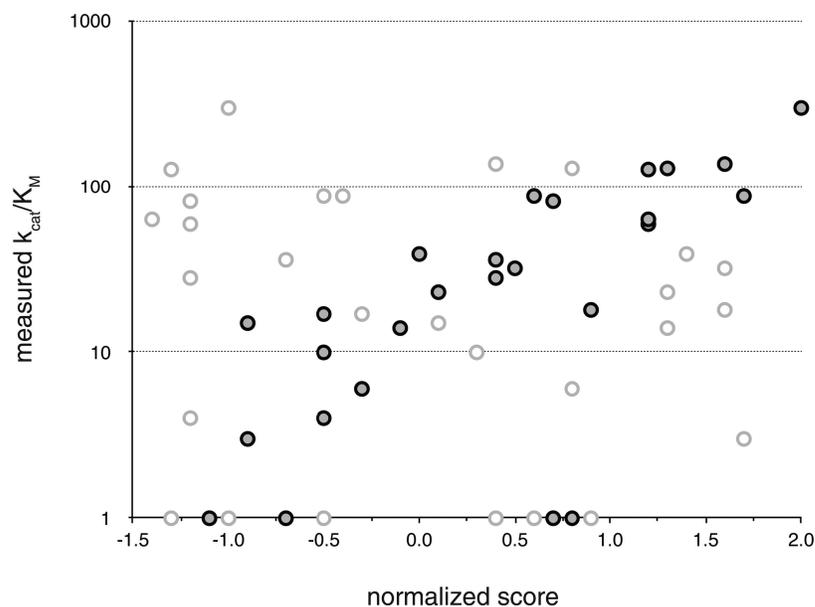


Figure 2.6. Correlation of Predicted Scores with Experimental Data. Data obtained with the standard Rosetta score function are shown as open circles ($R^2 = 0.02$, Pearson's $R = 0.08$) while results after re-parameterization are marked by gray circles ($R^2 = 0.62$, Pearson's $R = 0.82$).

Variant	Mutations	K_M	k_{cat}	k_{cat}/K_M	K_M	k_{cat}	k_{cat}/K_M
		(μM)	(s^{-1})	($\text{mM}^{-1} \text{s}^{-1}$)	(μM)	(s^{-1})	($\text{mM}^{-1} \text{s}^{-1}$)
A0	A100V, R104M, D133S	10 ± 1	1.41 ± 0.03	137	11 ± 1	3.49 ± 0.1	298
A1	A0 + F96D	na	na		na	na	
A2	A0 + W58E	na	na		na	na	
A3	A0 + E196L	na	na		na	na	
A4	A0 + F96Y	489 ± 40	0.62 ± 0.03	1.2	1121 ± 146	1.70 ± 0.14	1.5
A5	A0 + W58V	598 ± 89	0.55 ± 0.06	0.9	1254 ± 347	0.84 ± 0.19	0.7
A6	A0 + E196A	34 ± 4	3.01 ± 0.13	87	53 ± 7.5	3.14 ± 0.16	59

Table 2.4. Summary of Kinetic Data for dCK Library A

Similar in size to the previous set of variants, Library B was made up of nine members. Notably, the new library used ssTK3 instead of ssTK1A as parental sequence. The switch in template was based on selectivity data obtained from Library A that indicated neutrality in regard

to enantioselectivity for ssTK3 but showed an undesirable 2-fold preference for the D- over L-isomer for ssTK1A. The new template in combination with the refined machine-learned score function was used to identify amino acid substitutions that either destabilized interactions with D-Thy or resulted in structural changes favoring bound L-Thy in the phosphoryl-acceptor site. Besides ssTK3 (B0), eight variants (B1–B8) carrying single amino acid changes in seven positions (V55E/F, M85Y, P89F, L102Y, V130T, A138I, L191A) were prepared. None of these amino acid residues were in immediate contact with the substrates (primary shell residues) but instead were located in second and third shell positions. The steady-state kinetic parameters of the eight variants for D- and L-Thy clearly showed functional improvements over Library A. Only two candidates had no detectable activity. The remaining six variants (B3–B8) exhibited either neutral enantioselectivity (B3–B5) or favored the L-isomer by up to 2.3-fold (B6–B8).

Variant	Mutations	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)
B0	R104M, D133N	18 ± 4	1.61 ± 0.11	87	18 ± 3.1	1.49 ± 0.06	81
B1	B0 + P89F	na	na		na	na	
B2	B0 + A138I	na	na		na	na	
B3	B0 + L102Y	20 ± 2	2.69 ± 0.11	129	20 ± 2.6	2.59 ± 0.12	127
B4	B0 + M85Y	53 ± 5	1.73 ± 0.05	32	47 ± 22.2	1.34 ± 0.41	28
B5	B0 + V55E	1145 ± 153	1.29 ± 0.12	1.1	655 ± 96	0.67 ± 0.04	1
B6	B0 + L191A	48 ± 6	1.92 ± 0.08	39	30 ± 5.2	0.69 ± 0.04	23
B7	B0 + V130T	36 ± 6	1.30 ± 0.08	36	55 ± 15.3	0.95 ± 0.11	17
B8	B0 + V55F	188 ± 13	2.65 ± 0.06	14	391 ± 21	2.56 ± 0.06	6
B6-II	B0 + V130T, L191A	99 ± 4	6.21 ± 0.09	63	158 ± 13	2.82 ± 0.09	18
B8-II	B0 + V55F, V130T	91 ± 12	1.34 ± 0.07	15	210 ± 31	0.91 ± 0.07	4
B6-III	B0 + V55F, V130T, L191A	252 ± 34	2.40 ± 0.10	9.5	632 ± 31	2.10 ± 0.10	3

Table 2.5. Summary of Kinetic Data for dCK Library B

A closer review of the kinetic parameters for active variants in Library B highlights the complexity and subtlety of the enzyme's functional changes in response to these targeted amino acid substitutions. In variants B3–B5, the functional changes balance out each other, hence

conserving the parental D/L-selectivity. The mutation at position 102 in variant B3 does not affect the apparent binding affinity for D- and L-Thy yet causes roughly a 1.7-fold increase in catalytic activity for both substrates. In contrast, the substitution at position 85 in variant B4 results in approximately 2.5-fold higher K_M values for the two substrates yet does not significantly affect catalytic rates. Finally, replacing Val with Glu at position 55 (B5) is overall equally detrimental to both substrates, resulting in 80-fold drops in specific activity. However, the particular functional decline for D-Thy originates from a combined 2.5-fold drop in k_{cat} and 30-fold increase in K_M , while for L-Thy it comes from a slightly (1.2-fold) lower k_{cat} but a 60-fold increase in K_M . Interestingly, substitution of a Phe in the same position (B8) is also detrimental to K_M yet reverses the previously observed trend by causing a 20-fold and 10-fold drop in the apparent binding constant for D-Thy and L-Thy, respectively. Furthermore, the V55F substitution boosts catalytic activity for both substrates by 1.5-fold, creating a variant that exhibits a 2.3-fold net gain in enantioselectivity for the L-nucleoside. Similar improvements in enantioselectivity could also be detected upon amino acid changes at positions 191 (variant B6) and 130 (variant B7). Substitutions in these positions resulted in distinct but less dramatic functional changes, causing 2- to 3-fold increases in the K_M values for both substrates. However, the unfavorable binding effect for the L-isomer in B6 is more than compensated for by a raise in catalytic activity for L-Thy relative to D-Thy. In variant B7, the desired L-selectivity arises from a more favorable Michaelis–Menten constant for the L-isomer compared to its D-form analogue. In summary, analysis of Library B identified three positions (residues 55, 130, and 191) that individually resulted in a 1.7- to 2.3-fold shift in enantioselectivity in favor of the unnatural nucleoside L-Thy. In comparison to the previously reported L-selective engineered human TK2,(25) the kinetic parameters for our three variants (B6–B8) already match the catalytic performance for L-Thy while showing superior discrimination of D-Thy.

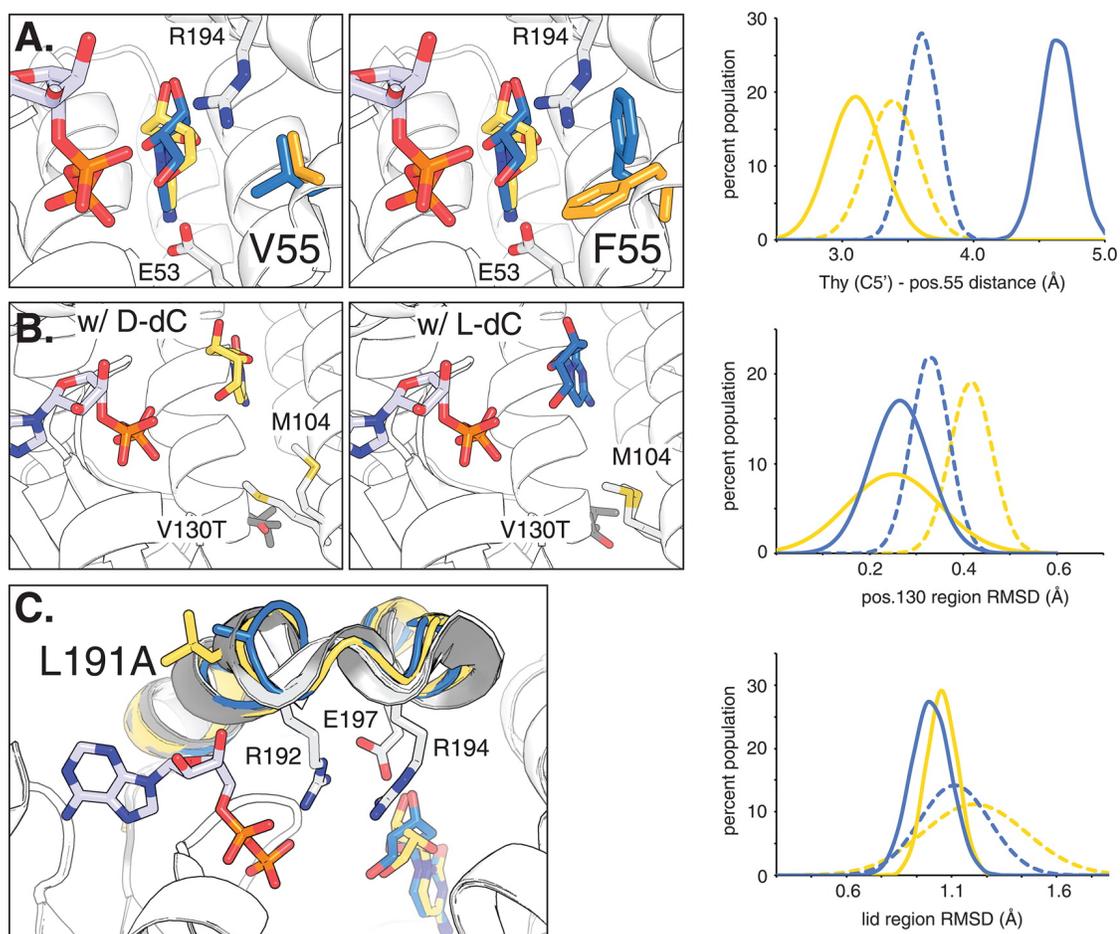


Figure 2.7. Molecular Dynamics Simulation of dCK Variants. For reference, ADP is bound in the phosphoryl donor site. The overlaid structures of D- and L-dC in the phosphoryl acceptor binding pocket are highlighted in yellow and blue, respectively. (A) Comparison of parental ssTK3 (V55) and variant B8 (F55) with their dominant side chain orientations at position 55 highlighted in yellow (for D-isomer) and blue (for L-isomer). Neighboring residues E53 and R194 are marked. The graph shows the overall distribution of preferred side chain conformations as a measure of the distance between the substrate C5' position and nearest side chain atom for ssTK3 (dotted lines) and B8 (solid lines). (B) Effect of V130T substitution on M104 side chain formation in the presence of D-dC and L-dC, respectively. The graph shows the lower conformational flexibility of residues near T130 (variant B7) versus V130 (B0) via the region's root square mean deviation (RMSD). (C) Impact of L191A substitution on the enzyme's lid region (highlighted in gray) and its three catalytic residues (R192, R194, and E197). The reduced conformational flexibility of variant B6 (A191) compared to parent ssTK3 (L191) is reflected in the lower, more narrow RMSD distribution.

To rationalize the observed functional contributions of amino acid changes in these three locations (positions 55, 130, and 191), we performed molecular dynamics simulations to sample conformational differences in protein structure in the presence of D- or L-Thy. The V55F substitution appears to influence enantioselectivity through sterics as the model predicts the bulky Phe side chain to adopt different rotamer conformations in the presence of the two substrates. Upon binding of the L-isomer to parent ssTK3 and variant B8, the F55 side chain orients itself parallel to the substrate, moving it away from the substrate by approximately 1 Å relative to V55. Although too far for direct interactions with substrate, the conformational change seems to affect the position of two neighboring catalytic residues, R194 that interacts with the triphosphate and E53 that serves as general base for nucleophilic activation of the substrate's 5'-hydroxyl group. In contrast, the model in the presence of the D-nucleoside indicates for F88 to assume an alternate orientation, rotating the side chain by $\sim 120^\circ$. The extended conformation protrudes into the substrate binding pocket, causing steric clashes with bound D-Thy and interfering with substrate binding as observed in our experiments. In contrast, the substitution of V130 with the more hydrophilic Thr in variant B7 results in only minor conformational differences. The analysis of the MD trajectories suggests a stabilizing effect, reflected in a smaller regional RMSD that indicates a decline in protein backbone flexibility of residues surrounding T130. In addition to these changes in dynamics, a notable structural change is observed for the adjacent side chain of M104. This position was originally mutated to enable phosphorylation of thymine nucleosides by dCK.(22) Upon substitution of V130 with Thr, energy minimization favors an alternate M104 rotamer, repositioning the side chain closer to the substrate, which could in part explain the measured effects on catalysis. Finally, the mutation at position 191 in variant B6 is located in the kinase's lid, an extended loop region that undergoes a conformational change upon substrate binding to establish multiple critical binding interactions with phosphoryl donor and acceptor involving residues R192, R194, and E197. While the location of the residue on the protein surface

is unlikely to directly influence enzyme performance, MD simulations indicate a change in the conformational flexibility of the region upon introduction of the Ala substitution, eliminating side chain torsion constrains. The RMSD of the loop region drops from a broad distribution averaging 1.1–1.2 Å for ssTK3 to a more narrow, less flexible conformational state in B6. Given the increased catalytic activity of B6, these results are consistent with a more defined structural arrangement of the loop in favor of a catalytically competent conformation.

The promising results from our single-site mutagenesis of ssTK3 also raised the question whether combination of these individual amino acid changes would result in additive or even synergistic functional effects. We therefore built two double mutants, combining V130T with either L191A (B6-II) or V55F (B8-II), as well as the triple mutant (B6-III). The subsequent kinetic analysis indicated improvement in enantioselectivity for all three variants. While the effects in B6-II and B8-II were roughly additive at 3.5- to 3.8-fold L-selectivity, the preference for the unnatural stereoisomer in B6-III declined to 3.2-fold. As seen for variants B3–B8, the overall gains resulted from a combination of K_M and k_{cat} effects. In B6-II, the increase in K_M values appeared to be additive, while k_{cat} values increased quite dramatically. These findings were contrasted by the kinetic parameters for B8-II, which suggests compensatory effects in K_M , averaging the double mutant's apparent binding affinity for both substrates relative to the single-site variants. At the same time, the catalytic rates in B8-II largely reflect the lower activity seen in variants B7 and B8. For our triple mutant (B6-III), the change in K_M values appeared to be mostly additive, raising the apparent binding constant for D- and L-Thy to 632 and 252 μM , respectively. In regards to k_{cat} , the three substitutions were compensatory. Overall, B6-III showed the desired preference for the L-isomer, yet the high K_M values make it an unlikely candidate for practical applications.

Substrate		B0	B6-II	B8-II
L-Thymidine	K_M (μM)	18 ± 3	158 ± 13	210 ± 31
	k_{cat} (s^{-1})	1.49 ± 0.06	2.82 ± 0.09	0.91 ± 0.07
	k_{cat}/K_M ($\text{s}^{-1} \text{mM}^{-1}$)	81	18	4
D-Thymidine	K_M (μM)	10 ± 1	33 ± 2	128 ± 4
	k_{cat} (s^{-1})	3.03 ± 0.2	4.49 ± 0.09	2.75 ± 0.04
	k_{cat}/K_M ($\text{s}^{-1} \text{mM}^{-1}$)	294	132	21
L-FMAU	K_M (μM)	18 ± 1	18 ± 1	49 ± 4
	k_{cat} (s^{-1})	0.46 ± 0.01	3.96 ± 0.04	2.75 ± 0.05
	k_{cat}/K_M ($\text{s}^{-1} \text{mM}^{-1}$)	25	220	55
L-FMAU/D-Thy (fold-change)		0.3	12	14
L-FMAU/D-dC (fold-change)		0.1	1.7	2.6

Table 2.6. Kinetic Parameters for dCK Variants with Native Substrates and L-FMAU

With B6-II and B8-II as lead candidates, we expanded our evaluation of kinetic properties to two additional substrates; L-FMAU as reference PET reporter and D-dC as the preferred substrate of our original dCK templates. The nucleoside analogue L-FMAU, while a moderately good substrate for ssTK3, has become an excellent substrate for B6-II and B8-II. In fact, L-FMAU is preferred over D-Thy and D-dC by a significant margin based on specificity constants (k_{cat}/K_M) and apparent binding constants (K_M). Compared to D-dC and D-Thy, B6-II has a 2- and 9-fold lower K_M value for L-FMAU. In respect to catalytic performance, the nucleoside analogue is turned over faster than D-dC and D-Thy by 1.7- and 12-fold, respectively. Similarly, B8-II favors binding of L-FMAU by 2.5- and 4-fold relative to D-dC and D-Thy, respectively, and shows a preference in specificity constant of 2.6- and 14-fold for the nucleoside analogue over the two natural substrates. The actual kinetic parameters of the two candidates for L-FMAU are equally promising. At K_M values of 18 and 50 μM and specific activities that match activity

levels of native enzymes with their natural substrates, B6-II and B8-II are effective activators for the PET reporter under physiologically relevant conditions. Finally, the catalytic performance of our two leads surpasses previously reported PET reporter kinases for L-FMAU.(13, 25) On the basis of in vitro experiments, B6-II and B8-II outperform engineered variants of HSV-tk1 and human TK2 with respect to specific activity and selectivity.

In summary, two rounds of semi-rational protein engineering of promiscuous dCK variants have yielded two effective orthogonal kinases with potential use in bioimaging systems in combination with L-FMAU as PET reporter. While benefiting from an extensive collection of structural data for nucleoside kinases, the creation of a predictive framework for enantioselectivity in combination with small libraries to probe the model and refine its accuracy through machine-learning algorithms has enable us to identify lead enzyme variants in a very time- and cost-effective manner. Next, the application of these novel reporter kinases in vivo will need to be tested for validating the predictive power of our bench experiments. At the same time, our findings support a broader, more systematic search for additional amino acid substitutions that might further enhance the functional performance of these kinases as potential PET reporter enzymes.

2.3 Methods

2.3.1 Materials

Reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless indicated otherwise. Restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). Pfu DNA polymerase (Stratagene, La Jolla, CA) was used for the PCR. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit, and PCR products were purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

2.3.2 Computer Models

Structural models for dCK bound to D- and L-Thy were based on PDB codes 3KFX and 3HP1, respectively.(23, 28) Ensembles of mutant structures were assembled in the Rosetta macromolecular modeling program, using side chain replacement and subsequent energy minimization.(29) The predicted structures were used to calculate a feature set of statistically correlated (Rank Sum test) components for machine learning composed of six terms (36-39) Our initial efforts utilized a standard linear score function to predict mutant efficiency, while our subsequent attempts used a non-parametric score function. Several non-parametric functions were evaluated, the k-nearest neighbor algorithm showed the best agreement to known data at various stages of design using leave-one-out validation as the evaluation criteria. Implementation of machine learning to identify potential mutation was performed in MATLAB (MathWorks, Natick, MA).

MD simulations were performed to approximate binding affinities for both D- and L-nucleosides, as well as nucleoside analogues bound to selected dCK variants. Atomic models were assembled using side chain replacement and molecular superposition based on the previously mentioned crystallographic data.(23) The structures were evaluated using the CHARMM22 all atom force field in reference to an implicit solvation model (SASA) using the CHARMM package.(40) The structures were minimized using cycles of steepest descent and conjugate gradient, heated to 300 K, equilibrated for 200 ps, and trajectories were recording for a 1 ns simulation time.

2.3.3 Site Directed Mutagenesis

Mutations in dck (NCBI code: BT019942) were created by primer overlap extension. The resulting PCR products were cloned into pET-14b vector (Novagen) via NdeI and SpeI

endonuclease restriction sites. Individual plasmid constructs were transformed into electrocompetent *E. coli* strain DH5 α and grown on LB-agar in the presence of ampicillin (100 $\mu\text{g}/\text{mL}$). Correct gene constructs were confirmed by DNA sequence analysis.

2.3.4 Protein Expression and Purification

Individual plasmids were transformed into *E. coli* strain BL21(DE3)pLysS and cultured in 250 mL 2-YT media containing ampicillin (100 $\mu\text{g}/\text{mL}$) and chloramphenicol (34 $\mu\text{g}/\text{mL}$). Cell cultures were grown to an OD₆₀₀ of ~ 0.6 at 37 °C, followed by induction with 0.3 mM IPTG for 2 h at 30 °C. Next, cell cultures were centrifuged (4000g, 4 °C, 30 min), and pellets were resuspended in 10 mL of lysis buffer (50 mM Tris-HCl (pH 8), 300 mM NaCl, 10 mM imidazole), supplemented with 50 μL of protease inhibitor cocktail (Sigma), 5 μL of benzonase (Novagen), and 0.5 mg of lysozyme (Sigma). After incubation on an orbital shaker at 4 °C for 20 min, cells were sonicated (8 \times 10 s pulses with 20 s pauses).

Cellular debris was separated via centrifugation (16,000g, 4 °C, 30 min), and the supernatant was equilibrated with 1 mL of Ni-NTA agarose resin (Qiagen) for 90 min at 4 °C. The resin was loaded on a prep-column (BioRad) and washed with 10 mL of lysis buffer, followed by 10 mL of wash buffer (50 mM Tris-HCl (pH 8), 300 mM NaCl, 50 mM imidazole). Finally, target protein was eluted with 2 mL of elution buffer (50 mM Tris-HCl (pH 8), 300 mM NaCl, 250 mM imidazole). The protein was exchanged into storage buffer (50 mM Tris-HCl (pH 8), 500 mM NaCl, 5 mM MgCl₂, 2 mM DTT) and concentrated using Amicon ultracentrifugation tubes (MWCO 10 kDa, Millipore). Aliquots were flash frozen in liquid nitrogen and stored at -80 °C. Typical yields for purified protein were 10 mg/L with $>95\%$ purity based on SDS-PAGE analysis. Individual kinase variants were evaluated by thermodenaturation experiments in the CD spectrophotometer and showed no significant changes in stability compared to the parental enzymes (data not shown).

2.3.5 Enzyme Kinetics

The catalytic parameters of individual kinase variants were measured using a spectrophotometric coupled-enzyme assay.⁽²⁰⁾ Variants were tested with two substrates; D- or L-Thy was added to reaction stock solution (50 mM Tris-HCl (pH 8), 0.1 M KCl, 5 mM MgCl₂, 1 mM DTT, 1 mM ATP, 0.21 mM phosphoenolpyruvate, 0.18 mM NADH, and 2 units/mL pyruvate kinase and 2 units/mL lactate dehydrogenase). Assays were performed in triplicate at 37 °C, and the absorbance change at 340 nm was measured in the presence of 10–100 nM enzyme per reaction with 1–1000 μM substrate. Steady-state kinetic parameters were calculated using nonlinear least regression analysis to the Michaelis–Menten equation in MATLAB (MathWorks, MA).

2.4 References

1. Brader, P., Serganova, I., and Blasberg, R. G. (2013) Noninvasive molecular imaging using reporter genes *J. Nucl. Med.* 54, 167– 172
2. Serganova, I., Ponomarev, V., and Blasberg, R. (2007) Human reporter genes: potential use in clinical studies *Nucl. Med. Biol.* 34, 791– 807
3. Yamamoto, M. and Curiel, D. T. (2010) Current issues and future directions of oncolytic adenoviruses *Mol. Ther.* 18, 243– 250
4. Yaghoubi, S. S., Campbell, D. O., Radu, C. G., and Czernin, J. (2012) Positron emission tomography reporter genes and reporter probes: gene and cell therapy applications *Theranostics* 2, 374– 391
5. Tjuvajev, J. G., Stockhammer, G., Desai, R., Uehara, H., Watanabe, K., Gansbacher, B., and Blasberg, R. G. (1995) Imaging the expression of transfected genes in vivo *Cancer Res.* 55, 6126– 6132
6. Jacobs, A., Tjuvajev, J. G., Dubrovin, M., Akhurst, T., Balatoni, J., Beattie, B., Joshi, R., Finn, R., Larson, S. M., Herrlinger, U., Pechan, P. A., Chiocca, E. A., Breakefield, X. O., and Blasberg, R. G. (2001) Positron emission tomography-based imaging of transgene expression mediated by replication-conditional, oncolytic herpes simplex virus type 1 mutant vectors in vivo *Cancer Res.* 61, 2983– 2995
7. Jacobs, A., Voges, J., Reszka, R., Lercher, M., Gossmann, A., Kracht, L., Kaestle, C., Wagner, R., Wienhard, K., and Heiss, W. D. (2001) Positron-emission tomography of vector-mediated gene expression in gene therapy for gliomas *Lancet* 358, 727– 729
8. Dempsey, M. F., Wyper, D., Owens, J., Pimlott, S., Papanastassiou, V., Patterson, J., Hadley, D. M., Nicol, A., Rampling, R., and Brown, S. M. (2006) Assessment of 123I-FIAU imaging of herpes simplex viral gene expression in the treatment of glioma *Nucl. Med. Commun.* 27, 611– 617

9. Penuelas, I., Mazzolini, G., Boan, J. F., Sangro, B., Marti-Climent, J., Ruiz, M., Ruiz, J., Satyamurthy, N., Qian, C., Barrio, J. R., Phelps, M. E., Richter, J. A., Gambhir, S. S., and Prieto, J. (2005) Positron emission tomography imaging of adenoviral-mediated transgene expression in liver cancer patients *Gastroenterology* 128, 1787–1795
10. Yaghoubi, S. S., Jensen, M. C., Satyamurthy, N., Budhiraja, S., Paik, D., Czernin, J., and Gambhir, S. S. (2009) Noninvasive detection of therapeutic cytolytic T cells with ¹⁸F-FHBG PET in a patient with glioma *Oncology* 6, 53–58
11. Barton, K. N., Stricker, H., Brown, S. L., Elshaikh, M., Aref, I., Lu, M., Pegg, J., Zhang, Y., Karvelis, K. C., Siddiqui, F., Kim, J. H., Freytag, S. O., and Movsas, B. (2008) Phase I study of noninvasive imaging of adenovirus-mediated gene expression in the human prostate *Mol. Ther.* 16, 1761–1769
12. McCracken, M. N., Gschweng, E. H., Nair-Gill, E., McLaughlin, J., Cooper, A. R., Riedinger, M., Cheng, D., Nosala, C., Kohn, D. B., and Witte, O. N. (2013) Long-term in vivo monitoring of mouse and human hematopoietic stem cell engraftment with a human positron emission tomography reporter gene *Proc. Nat. Acad. Sci. U.S.A.* 110, 1857–1862
13. Gambhir, S. S., Bauer, E., Black, M. E., Liang, Q., Kokoris, M. S., Barrio, J. R., Iyer, M., Namavari, M., Phelps, M. E., and Herschman, H. R. (2000) A mutant herpes simplex virus type 1 thymidine kinase reporter gene shows improved sensitivity for imaging reporter gene expression with positron emission tomography *Proc. Nat. Acad. Sci. U.S.A.* 97, 2785–2790
14. Likar, Y., Dobrenkov, K., Olszewska, M., Shenker, L., Cai, S., Hricak, H., and Ponomarev, V. (2009) PET imaging of HSV1-tk mutants with acquired specificity toward pyrimidine- and acycloguanosine-based radiotracers *Eur. J. Nucl. Med. Mol. Imaging* 36, 1273–1282

15. Song, S., Pursell, Z. F., Copeland, W. C., Longley, M. J., Kunkel, T. A., and Mathews, C. K. (2005) DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity, *P Proc. Nat. Acad. Sci. U.S.A.* 102, 4990– 4995
16. Gon, S., Napolitano, R., Rocha, W., Coulon, S., and Fuchs, R. P. (2011) Increase in dNTP pool size during the DNA damage response plays a key role in spontaneous and induced-mutagenesis in *Escherichia coli* *Proc. Nat. Acad. Sci. U.S.A.* 108, 19311– 19316
17. Niu, C., Bao, H., Tolstykh, T., Micolochick Steuer, H. M., Murakami, E., Korba, B., and Furman, P. A. (2010) Evaluation of the in vitro anti-HBV activity of clevudine in combination with other nucleoside/nucleotide inhibitors *Antiviral Ther.* 15, 401– 412
18. Riddell, S. R., Elliott, M., Lewinsohn, D. A., Gilbert, M. J., Wilson, L., Manley, S. A., Lupton, S. D., Overell, R. W., Reynolds, T. C., Corey, L., and Greenberg, P. D. (1996) T-cell mediated rejection of gene-modified HIV-specific cytotoxic T lymphocytes in HIV-infected patients *Nat. Med.* 2, 216– 223
19. Sabini, E., Ort, S., Monnerjahn, C., Konrad, M., and Lavie, A. (2003) Structure of human dCK suggests strategies to improve anticancer and antiviral therapy *Nat. Struct. Biol.* 10, 513– 519
20. Gerth, M. L. and Lutz, S. (2007) Mutagenesis of non-conserved active site residues improves the activity and narrows the specificity of human thymidine kinase 2 *Biochem. Biophys. Res. Commun.* 354, 802– 807
21. Ponomarev, V., Doubrovin, M., Shavrin, A., Serganova, I., Beresten, T., Ageyeva, L., Cai, C., Balatoni, J., Alauddin, M., and Gelovani, J. (2007) A human-derived reporter gene for noninvasive imaging in humans: mitochondrial thymidine kinase type 2 *J. Nucl. Med.* 48, 819– 826

22. Iyidogan, P. and Lutz, S. (2008) Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity *Biochemistry* 47, 4711– 4720
23. Hazra, S., Sabini, E., Ort, S., Konrad, M., and Lavie, A. (2009) Extending thymidine kinase activity to the catalytic repertoire of human deoxycytidine kinase *Biochemistry* 48, 1256– 1263
24. Liu, L., Murphy, P., Baker, D., and Lutz, S. (2010) Computational design of orthogonal nucleoside kinases *Chem. Commun.* 46, 8803– 8805
25. Campbell, D. O., Yaghoubi, S. S., Su, Y., Lee, J. T., Auerbach, M. S., Herschman, H., Satyamurthy, N., Czernin, J., Lavie, A., and Radu, C. G. (2012) Structure-guided engineering of human thymidine kinase 2 as a positron emission tomography reporter gene for enhanced phosphorylation of non-natural thymidine analog reporter *probe J. Biol. Chem.* 287, 446– 454
26. Gil, J. S., Machado, H. B., Campbell, D. O., McCracken, M., Radu, C., Witte, O. N., and Herschman, H. R. (2013) Application of a rapid, simple, and accurate adenovirus-based method to compare PET reporter gene/PET reporter probe systems *Mol. Imaging Biol.* 15, 273– 281
27. Arner, E. S. and Eriksson, S. (1995) Mammalian deoxyribonucleoside kinases *Pharmacol. Ther.* 67, 155– 186
28. Hazra, S., Ort, S., Konrad, M., and Lavie, A. (2010) Structural and kinetic characterization of human deoxycytidine kinase variants able to phosphorylate 5-substituted deoxycytidine and thymidine analogues *Biochemistry* 49, 6784– 6790
29. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J. J., Karanicolas, J., Das, R.,

- Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules *Methods Enzymol.* 487, 545– 574
30. Sabini, E., Hazra, S., Konrad, M., and Lavie, A. (2007) Nonenantioselectivity property of human deoxycytidine kinase explained by structures of the enzyme in complex with L- and D-nucleosides *J. Med. Chem.* 50, 3004– 3014
31. Sabini, E., Hazra, S., Konrad, M., Burley, S. K., and Lavie, A. (2007) Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase *Nucleic Acids Res.* 35, 186– 192
32. Friedman, J. H., Bentley, J. L., and Finkel, R. A. (1977) An algorithm for finding best matches in logarithmic expected time *ACM Trans. Math. Software* 3, 209– 226
33. Breiman, L., Friedman, J. H., Stone, C. J., and Olshen, R. A. (1984) Classification and Regression trees, *CRC Press*, New York.
34. Amit, Y. and Geman, D. (1997) Shape quantization and recognition with randomized trees *Neural Comput.* 9, 1545– 1588
35. Scholkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001) Estimating the support of a high-dimensional distribution *Neural Comput.* 13, 1443– 1471
36. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes *J. Mol. Biol.* 326, 1239– 1259
37. Dunbrack, R. L., Jr. and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences *Protein Sci.* 6, 1661– 1681
38. Lazaridis, T. and Karplus, M. (1999) Effective energy function for proteins in solution *Proteins* 35, 133– 152

39. Kuhlman, B. and Baker, D. (2000) Native protein sequences are close to optimal for their structures *Proc. Natl. Acad. Sci. U.S.A.* 97, 10383– 10388
40. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoseck, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program *J. Comput. Chem.* 30, 1545– 1614

Chapter 3

Design-of-Experiments for Deoxycytidine Kinase

Abstract

Positron emission tomography (PET) based reporter systems enable the non-invasive assessment of gene and cell based therapies, concurrently monitoring transgene expression and migration. Recent systems co-localize a deoxynucleoside-based radiolabel with an ectopically expressed deoxynucleoside kinase through chemical modification. Unfortunately, wild type kinases suffer from poor activity for PET probes and retain high activity for native deoxynucleosides. Exogenous expression of these native kinases incurs rampant phosphorylation, threatening cellular metabolism. Recent advances in protein engineering enable the alteration of substrate specificity to evolve bio-orthogonal kinases. This approach used statistically optimized gene libraries to iteratively evolve highly selective reporter genes based on human deoxycytidine kinase (dCK). Each round library incorporated systematic recombinations of amino acid substitutions, and evaluated individual dCK variants for both desired substrate, L-ribose PET probe 2'-fluoro-2'-deoxy-1- β -L-arabinofuranosyl-5-methyluracil (L-FMAU), and undesired native substrate, D-deoxycytidine (D-dC). Emulating mechanisms of natural selection, each successive round library artificially evolved kinases preferring unnatural L-ribose deoxynucleoside analogs. The ultimate variants harbored up to 11 amino acid substitutions and displayed significantly improved activity for L-FMAU, and reduced native pyrimidine activity to

physiological irrelevance. The evolutionary landscape of candidate reporter genes was explored through exhaustive deconvolution, elucidating mechanisms of substrate specificity conferred by the introduced amino acid interactions. This approach used smaller, functionally enriched libraries to evaluate ~200 distinct sequences and identified several highly evolved dCK variants, with a concomitant 10-fold increase in L-FMAU specific activity and 100-fold preference over its native D-dC.

3.1 Introduction

Reporter genes are ectopically expressed proteins that enable researchers and clinicians to monitor gene and cell based therapies within living subjects. Such a reporter system would fulfill a clinically unmet need for routine and sensitive methods to establish a relationship between the pharmacokinetics at target cells and therapeutic effect. Among imaging technologies, radionuclide detection by positron emission tomography (PET) is a promising application for noninvasive visualization of cellular processes. (1-7) These two-component systems contain a primary encoded protein expressed within target tissue and a secondary exogenously introduced radionuclide probe (typically [^{18}F]). PET imaging computationally reconstructs whole-body distributions of radionuclide, identifying spatially enriched populations of co-localized reporter gene and probe.

A recent adaptation introduces an exogenous reporter gene, having alternate substrate preferences to the host, and chemically modifies natively inert radionuclide analogs. (8) Ectopic expression confines the innocuous probe within transgenic cells, enabling non-invasive monitoring of gene expression and migration. The earliest and most common PET reporter gene are based on herpes simplex virus type 1 thymidine kinase (HSV-TK). (9-11) Endogenous deoxynucleoside kinases are highly specific and cannot phosphorylate certain [^{18}F]-labeled analogs *in vivo*, such as L-2'-fluoro-5-methyluracil-arabinofuranosyl (L-FMAU). Interestingly, these deoxynucleoside analogs are substrates for HSV-TK, and membrane-bound transporters no longer export the nascent [^{18}F]-labeled monophosphate, accumulating radionuclides within target cells. (12,13) The combination of HSV-TK and PET-tracers have wide-spread application in preclinical studies to evaluate cellular dynamics, imaging engineered cells within living organisms. To date, HSV-TK has been the only PET reporter gene used to image human patients. (9-11,14)

Unfortunately, HSV-TK has several limitations, severely reducing its clinical value as a reporter gene. Perhaps most significant, immunocompetent hosts eliminate cells expressing HSV-TK, making the viral kinase not well suited for the repetitive, long-term applications. (15) Additionally, the high native activity of the viral kinase for native deoxynucleosides, coupled with elevated gene reporter expression, threatens nucleotide metabolism through unregulated phosphorylation. (16,17) The increased and imbalanced dNTP pools compromise DNA replication/repair, increasing cellular genotoxicity. (18) In an effort to circumvent HSV-TK limitations, researchers have recently focused on the human deoxynucleoside kinases as reporter genes. (19)

The inherently reduced immunogenic risks of human-derived kinases have led to a number of protein engineering studies on human deoxycytidine kinase (dCK) and thymidine kinase 2 (TK2). (20,21) Largely guided by crystallographic data, engineered dCK and TK2 kinases have broadened the substrate specificities of these enzymes to encompass deoxynucleoside-based PET probes. (22) More importantly, the functional gains from *in vitro* characterization have correlated to *in vivo* cell and animal PET-based models. (23) Unfortunately, these early-engineered kinases suffer poor activity for PET probes, and retain relatively high activity for natural pyrimidines, limiting their versatility as clinical reporter genes. (15,22)

As a quantitative benchmark, the following design criteria for PET reporter genes is proposed. These *in vitro* conditions describe an engineered kinase with sufficient activity for PET imaging, and (de)evolved recognition for native pyrimidines below physiological activity. The design criteria is explicitly defined to satisfy two conditions. First, the engineered kinase must possess a specific activity $>100 \text{ mM}^{-1}\text{s}^{-1}$ for L-FMAU (k_{cat}/K_M values). Second, the engineered kinase must preferentially bind L-FMAU (K_M value $<10 \text{ }\mu\text{M}$) over native substrates (K_M values of $>50 \text{ }\mu\text{M}$).

To date, no native or engineered PET kinase reporters meet the criteria and most fail both conditions.

An emerging strategy within protein engineering describes the use a design-of-experiments methodologies and employ smaller, functionally enriched libraries to optimize amino acid sequences for an arbitrary function. These approaches algorithmically derive sequence-to-function relationships from experimental data, and identify superior enzyme variants from systematic recombinations of amino acid substitutions. With an explicit focus on efficiency and speed, these methods have considerably improved the scope of protein engineering in recent years. (24) Technologies, such as ProSAR and ProteinGPS, have previously evolved enzyme variants (harboring up to 30 amino acid substitutions) to meet a variety of design criteria, substantially increasing reaction rates at higher temperatures (typically k_{cat} focused) for a desired conversion. (25-30)

This work bridges the clinically unmet need for highly selective PET reporter genes with recent advances in protein engineering. The work implements design-of-experiments methodology to engineer selective PET reporter genes, however the major distinction from previous literature is the multi-constraint design and explicit focus on substrate recognition (K_M). This work describes the methodology to design computationally optimized gene libraries and evolve unnatural substrate specificity. This approach ultimately identified several reporter gene candidates, reducing substrate recognition for native pyrimidines below physiological levels ($K_M > 100 \mu\text{M}$), while significantly increasing L-FMAU activity an order of magnitude over the starting sequence.

3.2 Results and Discussion

3.2.1 Round 1

In this initial design, a library was synthesized to survey enantioselective features without prejudice from previous experiments. The library consisted of 66 distinct dCK variants, each based on the double substituted R0 (R104M, D133N) template, exhibiting broad substrate specificity for a variety of nucleoside analogs and neutral enantioselectivity. (22) Amino acid substitutions were then systematically introduced, selected from a set of 51 amino acid substitutions.

The first 40 substitutions are naturally observed variations among the deoxynucleoside kinase family. Briefly, the dCK was protein sequence compared against Genbank to retrieve homologous sequences, and identified alternate amino acids through global sequence alignment. The resulting thousands of individual substitutions were rank ordered within a scoring matrix (i.e. amino acid frequency, genetic mutations, etc.) as previously described. (26-28) The top 40 scoring amino acid substitutions were arbitrarily selected, distributed evenly both in primary and tertiary structure. The remaining 11 amino acid substitutions were rationally selected, focused around the active site.

The 48 total substitutions were incorporated into the library using a design-of-experiments methodology. The distribution of amino acid substitutions is visualized in the figure below, with each variant (row) containing three amino acid substitutions and each substitution (column) repeated approximately 5 times. Overall, the figure describes a sparse landscape to conservatively probe for substitutions favoring L-FMAU over D-dC.

Round 1 genes were synthesized and transformed into bacterial expression hosts. The clones were then subcultured and purified the recombinant proteins in parallel. The design criteria required a

detailed screen to obtain Michaelis-Menten parameters for each variant, specifically to evolve substrate recognition based on K_M values. Deoxynucleoside kinases have been traditionally characterized by steady state kinetics, measuring reaction rates at multiple concentrations for each substrate. Unfortunately, this labor-intensive approach was not amenable for kinetic characterization of hundreds of dCK variants in an efficient manner. While not as accurate, numerical integration of a single time-resolved reaction has been used to experimentally derive Michaelis-Menten parameters, providing estimates of substrate recognition (K_M) and overall activity (k_{cat}/K_M). Assay conditions were selected to screen kinetic parameters with physiological relevance, sensitive to K_M values within 10-50 μM and catalytic efficiencies within 50-500 $\text{mM}^{-1}\text{s}^{-1}$. This focused activity screen enables the robust evaluation of several rounds of kinase libraries, with sufficient kinetic detail to artificially evolve substrate specificity.

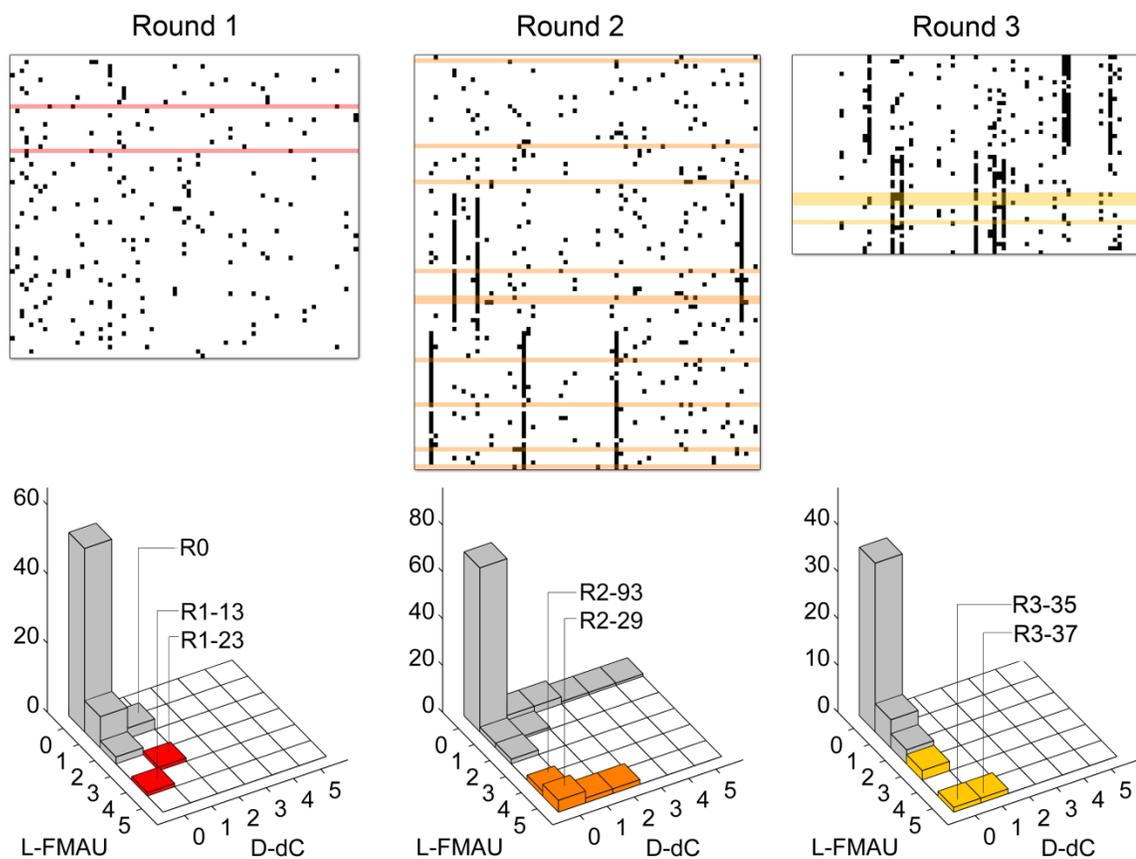


Figure 3.8. Substitution Matrices and Activity Histograms for Rounds 1-3. The top images describe the substitution matrix from each round, with each row representing an enzyme variant, and each column representing a modeled amino acid substitution. The presence/absence of a coordinated square is representative of amino acid changes from the starting sequence. The corresponding activity is shown below, describing fold activity change over the starting sequence (R0). The initial leftmost library describes a sparse landscape, merely probing for regions of benefit. Progressing rightward to more advanced libraries describe sequence convergence and the evolution of highly selective kinase populations

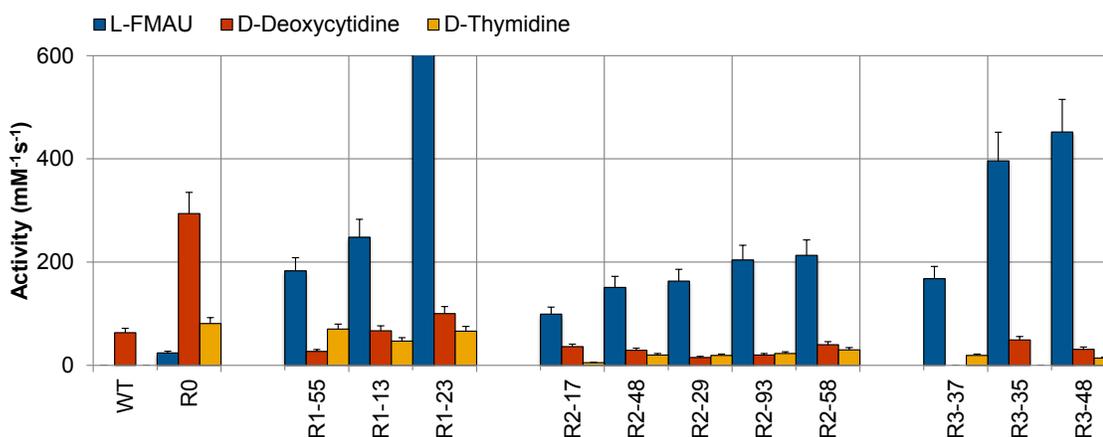


Figure 3.9. Catalytic Efficiencies of Lead Variants. The wild type dCK displays activity for thymidine, and by extension the target substrate L-FMAU. For this reason, this work used a double substituted starting sequence, denoted as R0. After screening the first library the lead candidates displayed preference for L-FMAU, however retain activity for the native substrates. The second round generally maintains activity for L-FMAU, but is able to reduce native activity to physiological irrelevance. The final third round results in variants with further improved selectivity, however display a diminishing return on catalytic improvement.

The variant R1-13 (N39R, V55I, I233V) boasts an order of magnitude improvement to L-FMAU activity ($k_{cat}/K_M > 250 \text{ mM}^{-1}\text{s}^{-1}$) and favorable binding interactions ($K_M < 1 \text{ }\mu\text{M}$), while slightly reducing activity for D-dC 5-fold. The V55I substitution, found in the proximity of the 2'-fluorine of L-FMAU, introduces a second side chain torsion (X_2) within the active site. While V55I is too far for a direct interaction, the larger isoleucine may adopt differential conformations in the presence of either substrate. The conformational change may influence neighboring residues to favor the polarized C-F bond, and not inherently favor an L-ribose substrate. This assertion is supported by the activities of native D-pyrimidines, which are indistinguishable from L-enantiomers. R1-13 is an encouraging initial lead, demonstrating a selective improvement to L-FMAU activity.

Variant	Substitutions	L-FMAU				D-Deoxycytidine			D-Thymidine		
		K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	0	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	63	K_M (μM)	k_{cat} (s^{-1})
WT	(none)	>1000	0.00	0	<1	0.06 ± 0.00	63	>1000	0.00	0	
R0	R104M, D133N	18 ± 2	0.45 ± 0.01	24	10 ± 2	3.03 ± 0.20	294	18 ± 3	1.49 ± 0.09	81	
R1-13	R104M, D133N, N39R, V55I, I233V	<1	0.24 ± 0.00	248	17 ± 8	0.41 ± 0.07	67	8 ± 2	0.42 ± 0.02	47	
R1-23	R104M, D133N, E90S, C146H, I26V	2 ± 0	2.00 ± 0.01	791	11 ± 7	1.20 ± 0.02	100	22 ± 3	1.50 ± 0.20	66	
R2-29	R104M, D133N, N60K, N195D, I200V, N224D	14 ± 6	2.31 ± 0.32	163	81 ± 21	1.24 ± 0.17	15	77 ± 10	1.47 ± 0.010	19	
R2-93	R104M, D133N, E90S, C146H, E87Q, L141M, T153A	8 ± 2	1.69 ± 0.07	204	82 ± 10	1.69 ± 0.07	20	64 ± 7	1.49 ± 0.07	23	
R3-35	R104M, D133N, E90S, C146H, E87Q, L141M, T153A K88D, I200V, R219K, N224D	8 ± 1	3.23 ± 0.94	396	104 ± 18	5.13 ± 0.36	49	>500	>3.00	<10	
R3-37	R104M, D133N, E90S, C146H, E87Q, T153A N60K, W161F, R219K	26 ± 4	4.42 ± 0.21	168	>500	>3.00	<10	263 ± 90	1.38 ± 1.40	19	

Table 3.7. Steady State Kinetics of Lead Variants for Selected Substrates

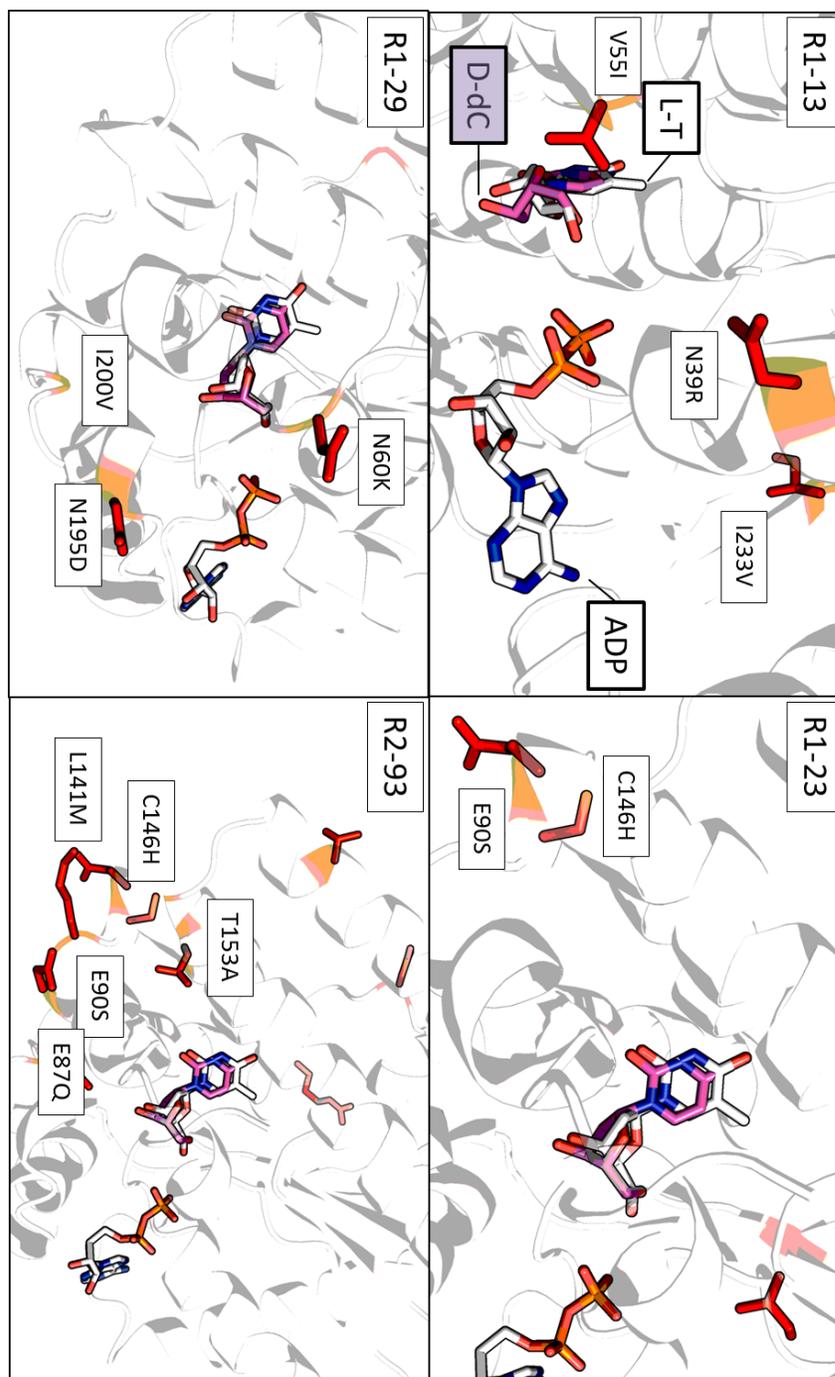


Figure 3.10. Positions of dCK Variant Substitution. R1-13 may gain catalytic benefit from the V55I substitution, selectively interacting with the 2'-fluoro position in L-FMAU. R2-29 contains several substitutions to the 'lid' region, which may confer improved selective. R1-23 is the evolutionary predecessor to R2-93, the latter containing several flanking substitutions to the nearby active site loop.

The variant R1-23 (E90S, C146H, I26V) is a significant improvement over the starting sequence, having the high L-FMAU specific activity ($791\text{mM}^{-1}\text{s}^{-1}$). More interestingly, the variant demonstrates enantioselective properties, consistently preferring the L-enantiomers over the native D-deoxynucleosides by 8-fold. Both the E90S and C146H substitutions are in close proximity to the 3'-hydroxyl of ribose, and may synergistically interact to improve L-ribose binding. While this Round 1 variant surpasses any known reporter gene in activity, subsequent rounds may improve clinical value by lowering D-deoxynucleoside activity even further.

3.2.2 Round 2

The Round 2 library incorporated the previous round's experimental data to enrich the evolution of dCK variants with improved preference for L-FMAU. Perhaps most explicit, Round 2 variants have an equal chance to have the Round 1 leads R1-13 or R1-23 as a template, instead of the original R0 sequence. The 96 Round 2 variants again used design-of-experiments to combine 4-5 additional substitutions from a set of 48. However, these amino acid substitutions were biased to the experimental data from Round 1.

Herein lies the advantage of the statistical methods. This approach can not only identify favorable substitutions; enabling enriched sampling in subsequent rounds, but more importantly detect deleterious mutations, and reduce the likelihood of unproductive conformations. The sequence-to-function model described each of the 66 Round 1 variants as a linear combination of the 48 individual substitutions, described previously. (26-28) Each substitution was repeated approximately five times, providing multiple microenvironments with corresponding experimental activities. The Round 1 activity was used as a dataset for machine learning algorithms to deconvolute and estimate the individual contribution of each substitution.

The resulting ranked list of the amino acid substitutions was used to design future libraries, based on predicted changes to L-FMAU/D-dC activity. Of the 48 Round 2 substitutions, 12 were

the top substitutions from the algorithmic deconvolution described above. A second set of 12 were rationally selected, at positions which neighbored the top substitutions. The final set of 24 was phylogenetically derived, an unbiased continuation from the previous round. This can also be visualized in the bivariate histograms shown previously, with dense substitution focused in regions biased by previous experimental data.

Round 2 demonstrated significant improvements over the previous round, identifying at least 8 variants with a L-FMAU preference greater than the Round 1 leads. The algorithmic deconvolution biased the library to avoid deleterious mutations, and Round 2 exhibited dramatically improved stability. Over 95% of variants had detectable protein yields (compared to 67% from Round 1).

Interestingly, both Round 1 and 2 have a similar 25% active variants (defined as activity at least half R0), however the latter begins to evolve highly specialized variants. This population drift is best illustrated in previously described histograms, as a nearly equal number of variants evolved toward two orthogonal trajectories. Among the 23 active variants, seven variants significantly improve preference for the native D-dC (>10-fold over R0), while another eight prefer unnatural L-FMAU. Of particular interest to the design criteria, is the small population of highly L-FMAU selective variants. Interestingly, variants in this region share common substitutions (E90S, E87Q, C146H) and suggest sequence convergence on enantioselective features. Among these L-FMAU selective variants R2-29 and R2-93 emerged as strong PET reporter candidates and were selected for further investigation.

The round 2 variant R2-93 (E90S, C146H, E87Q, L141M, T153A) is a pedigreed kinase – a direct evolution of the Round 1 lead R1-23. In addition to the common substitutions (E90S, C146H), the improved R2-93 contains flanking substitutions in the same region (E87Q, L141M, and T153A), resulting in a synergistic refinement of substrate specificity. The variant retains the

order of magnitude activity improvement for L-FMAU (From R1-23), while reducing activity for D-dC even further to 10-fold. Similar to its evolutionary predecessor, substrate specificity is generalizable to L-ribose, as kinetic comparisons of deoxynucleoside enantiomers distinctly favor L-ribose.

R2-29 (N60K, N195D, I200V, N224D) spontaneously emerged from screening and shares no common substitutions with previously identified kinases. Two neighboring substitutions (N195K and I200V) are located on a catalytic 'lid' region, implicated in the transfer of the gamma phosphate. (31) Mutational perturbations in this region may selectively influence catalytic orientation, as dCK-bound enantiomers have distinct phosphoryl transfer geometries. (32) The role of surface exposed residues N60K and N224D is unclear, but may improve protein stability/solubility to improve general kinase activity and/or mutational tolerance. Curiously, this variant also possesses enantioselective properties, favoring the L-enantiomers of native pyrimidines, albeit to a lesser extent than R2-93.

While both R2-93 and R2-29 are sequentially distinct, both variants have similar kinetic properties and exhibit a 5-fold increase in L-FMAU activity and a 500-fold improvement to selectivity compared to the starting template. These Round 2 leads fulfill the original design criteria, describing high activity for the L-ribose PET probe, and a marked reduction in native D-ribose metabolism. These variants, containing up to seven amino acid substitutions from wild type, are strong PET reporter gene candidates and merit preclinical investigation.

3.2.3 Round 3

While the previous round's lead variants had impressive catalytic profiles, L-FMAU selectivity can be evolved even further with an aggressive substitution strategy. The library consisted of 48 variants, composed of the 24 top amino acid substitutions using the previously described algorithmic deconvolution. Each variant had 4-5 additional substitutions using either R2-29 or

R2-93 as a template. Round 3 was a focused attempt for sequence optimization, and the substitution distribution is matrices described earlier, describing dense clusters of mutagenesis within narrow regions in primary sequence, a stark contrast to the sparse nature of the initial round.

Despite the relatively aggressive mutagenesis, Round 3 had continued protein stability, as only a single variant was not purified. While the final library maintained 25% active variants (similar to previous rounds) every active variant improved L-FMAU activity compared to the starting template. This trend when comparing Round 2, Round 3 removes the D-dC selective population entirely through genetic selection.

The Round 3 population is clustered about some apparent functional minima, as the increasingly homogenous sequences exhibit diminishing returns. If you recall, approximately 8% of Round 2 variants were catalytically superior to the Round 1 leads. Similarly, 6% of Round 2 variants have a kinetic advantage over Round 2. Interestingly, optimized sequences for R2-93, were observed suggesting dCK variants have varying tolerances for future design. The best two variants, R3-37 and R3-35 were selected for further study.

The variant R3-35 (E87Q, K88D, E90S, L141M, C146H, T153A, I200V, R219K, N224D) is a marked improvement over its parent (R2-93), having increased activity for the PET probe L-FMAU. While the exact mechanism of these additional amino acids substitutions is unclear, R3-35 shares two substitutions (I200V, N224D) with another Round 2 lead R2-29 implicated in selective enzyme turnover improvements. Elements of this catalytic improvement can be observed, as R3-35 has a selective 2-fold increase in k_{cat} for L-FMAU, while K_M values are largely unchanged for each substrate. As such, R3-35 preserves enantioselectivity, preferring L-ribose to D-ribose substrates, and boosts performance for the PET probe. This final variant has a

nearly 20-fold improvement for L-FMAU over the starting template, while maintaining similar ratios of specific activity to the parent.

The variant R3-37 (E90S, C146H, E87Q, T153A, N60K, W161F, R219K) is also an improvement over the R2-93 parent, with increased discrimination against native D-deoxynucleosides. The additional substitutions are all surface exposed residues and their roles are again largely unclear. The W161F substitution was identified early among Round 1 variants to selectively benefit L-FMAU. N60K was identified in several Round 2 variants, including the lead R2-29. R219K is common to top Round 3 sequences, including both R3-35 and R3-37. These additional substitutions disproportionately affect substrate recognition for native metabolites. Amazingly, the evolved deoxycytidine kinase variant, no longer recognizes its native cytosine base, and has no detectible activity for either D-dC or L-dC enantiomers. The variant retains minimal activity for D-thymidine, however prefers the L-ribose enantiomer nearly an order of magnitude in both k_{cat}/K_M and K_M . The highly evolved R3-37 has 9 amino acid substitutions from wild type and exhibits at least 100-fold preference for L-FMAU over D-dC and a 40-fold discrimination over D-thymidine based on ratios of specific activity.

After screening three rounds of dCK variants, several kinases surpassing the original design criteria were discovered, having both high activity for the PET probe (k_{cat}/K_M L-FMAU > 100 $\text{mM}^{-1}\text{s}^{-1}$), and preferential substrate recognition (K_M L-FMAU < 10 μM , K_M D-dC > 50 μM). The method enabled the accelerated evolution of kinases for unnatural substrate specificities. The Round 1 library identified a handful of selective reporter genes, one of which, R1-23, has the highest reported *in vitro* activity for L-FMAU. Round 2 refined this sequence with R2-93, retaining the high activity for L-FMAU, while reducing activity for native substrates below physiological relevance. The final Round 3 variants pushed this selectivity even further, as the evolved dCK variant no longer even recognizes its natural substrate D-dC. These variants, among

others, are highly evolved kinases, and remain strong candidates for bio-orthogonal PET reporter genes and warrant pre-clinical testing.

3.2.4 Experimental Deconvolution

While the computational model estimate activity changes from individual amino acid substitutions, this section describes the efforts to experimentally investigate the combinatorial interactions and elucidate mechanisms of substrate selectivity. Exploring this evolutionary landscape provides further insight into protein engineering, as contemporary design strategies are based on additivity/cooperativity. (33,34) In this regard, this approach is unique, as does not assume an explicit order to introduce amino acid substitutions to evolve desired function. For this discussion, this work focused on exhaustive deconvolution of the Round 2 lead variants, and characterized each of the 16 combinations of the four-substituted R2-29 and the 32 combinations for the five-substituted R2-93.

The starting template R0 favors the natural metabolite by an order of magnitude over L-FMAU. The starting variant's substrate preference for D-dC over L-FMAU is based on a 2-fold lower K_M , and a 5-fold k_{cat} deficiency for the unnatural analog. The variants R2-29 and R2-93 not only overcome the disadvantage to L-FMAU, restoring the 10-fold activity loss, but simultaneously reduce native deoxynucleoside activity to physiological irrelevance. Despite the similar catalytic performance for desired L-FMAU and undesired D-dC, the lead variants share no common substitutions, suggesting distinct mechanisms of selectivity.

The variant R2-29 (N60K, N195D, I200V, N224D) spontaneously emerged from Round 2 screening with no discernible predecessor, as none of the four substitutions are found in Round 1. The substitutions N195K and I200V, located on a catalytic 'lid' region, are individually tolerated, having little consequence to activity. In contrast, the surface exposed N22D and N60K significantly increase enzyme turnover (k_{cat}) by greater than 10-fold for L-FMAU, but slightly

reduce substrate recognition (K_M) by about 2-fold. The exact mechanism of functional improvement is unclear, however may be related to a general increase in protein stability/solubility.

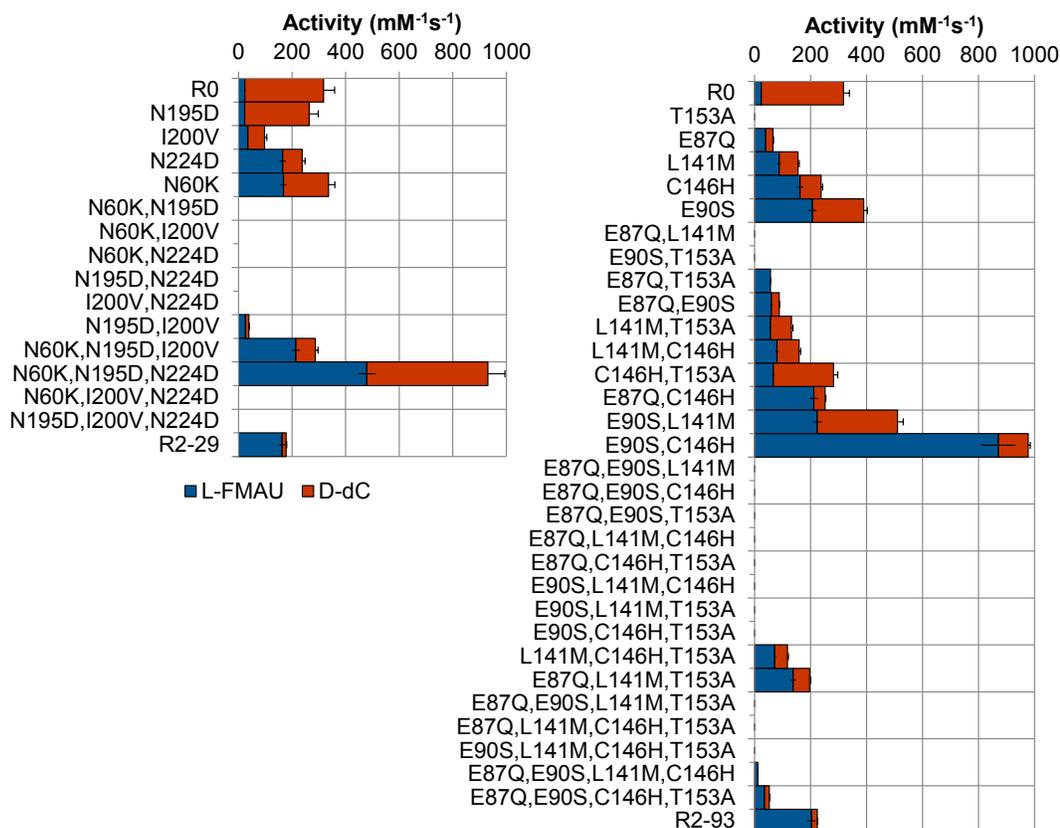


Figure 3.11. Catalytic Efficiencies of Deconvoluted Variants. The lead variants from Round 2 were experimentally deconvoluted, exhaustively sampling combinations of amino acid substitutions. Individual substitutions did appear to have benefit, however had limited additivity in combination. The requisite selectivity was observed in combinations of 3-4 substitutions, but were catalytically inferior to the ultimate lead variants.

Interestingly, these four substitutions are not immediately additive, as only a single double combination (N195D/I200V) had detectable activity. This active variant, N195D/I200V, combines the two 'lid' substitutions, resulting in a 2-fold preference for L-FMAU, albeit suffering from poor substrate recognition for both substrates ($K_M > 100 \mu\text{M}$). The triple

combinations begin to recover activity, specifically with N60K/N195D/N224D variant, inheriting the 10-fold increase turnover from the surface substitutions (N60K, N224D) with no discernible loss in substrate recognition. Another triple combination N60K/N195D/I200V significantly improves over the N195D/I200V double combination with the additional surface substitution, having a 5-fold binding preference for L-FMAU over D-dC, however does not improve turnover.

The final R2-29 quadruple variant inherits favorable kinetic properties of the triple combinations, maintaining the 5-fold binding preference for L-FMAU and improving L-FMAU turnover another 5-fold. The final variant has evolved a 10-fold preference for L-FMAU over D-dC, based on ratios of specific activity. Interestingly, this variant would likely not have been identified using an additive model, as all double combinations are explicitly worse than individual substitutions.

Unlike the previous variant, R2-93 (E90S, C146H, E87Q, L141M, T153A) is directly evolved from Round 1. Three substitutions are from Round 1 (E90S, E87Q, C146H) and the remaining two (L141M, T153A) were rationally selected based on spatial proximity. All five substitutions are located near the substrate 2'/3' ribose position, and likely play a direct role in substrate recognition. In fact, all substitutions (except T153A) increase specific activity for L-FMAU, with little/no changes to K_M values. The most impressive of these individual substitutions are E60S and C146H, selectively increasing L-FMAU k_{cat} by 6-10 fold.

In general, the individual substitutions are highly cooperative, with the double substituted E90S/C146H having a 40-fold improvement in L-FMAU, coupled with a 3-fold loss in D-dC activity. This selectivity is largely based on a 10-fold preference in K_M , suggesting a fundamental binding preference for the unnatural L-ribose substrate. Other double combinations (E87Q/E90S, E87S/C146H, and E90S/C146H), have similar catalytic preferences, and remain highly specific and active kinases.

Interestingly, the triple and quadruple combinations are largely inactive, suggesting higher order combinations become increasingly difficult to design, similar to R2-29. Among the few active variants, substrate affinity is significantly reduced ($K_M > 100 \mu\text{M}$), however preference for L-FMAU is still observed. Curiously, the intermediate triple/quadruple variants are catalytically inferior to both their predecessors (single/double) and the ultimate quintuple combination successor. The variant R2-93 is a serendipitous combination of amino acid substitutions, and simultaneously achieves a 10-fold increase in desired L-FMAU activity, and 10-fold decrease in native deoxynucleoside activity.

For both R2-29 and R2-93, individual substitutions demonstrate limited additive behavior, and describe the challenges in incorporating higher combinations of substitutions without detriment. These highly evolved variants with impressive catalytic performances simply could not have been identified with the cumulative addition of amino acid substitutions. Deconvolution of both variants describes evolutionary trajectories requiring spontaneous combinations of two or three substitutions to perceive any benefit, highlighting the limited understanding of the protein sequence-to-function relationship. While the exact amino acid substitution combinations may be somewhat fortuitous, their successful identification highlights the advantage of the design strategy to aggressively sample higher order combinations.

3.3 Conclusion

These results demonstrate a systematic exploration of amino acid substitutions to evolve unnatural L-selective deoxynucleoside kinases, for clinical applications in gene/cell based therapies. This iterative approach represents a paradigm shift toward smaller, functionally rich libraries, indicative of modern protein engineering. Each successive library demonstrated greater fitness, concomitantly evolving improved stability/solubility with improved catalytic preference over native metabolites. By the final third round, every active variant demonstrated an

improvement over the starting template sequence, enabling the successful identification of several PET reporter gene candidates. These kinases simultaneously increased desired L-FMAU above current engineered PET reporter genes, while reducing native D-deoxynucleoside activity to physiological irrelevance. Future efforts will confirm *in vitro* activity within an *in vivo* environment and establish new techniques for non-invasive gene therapies.

3.4 Material and Methods

4.4.1 Library Design

The Round 1 library consisted of 66 variants each using R0 as a template. Each variant contained three additional substitutions; selected with equal chance from a subset of 48 amino acid substitutions. Of the Round 1 substitutions, 36 were selected without bias from sequence homology and 12 were rationally selected from active site proximity.

The Round 2 library consisted of 96 variants, each having an equal chance to use R0, R1-13 or R1-23 as a template. Each variant contained an additional 4-5 amino acid substitutions, selected from a subset of 48. Of the Round 2 substitutions, 12 were top candidates from the previous round, another 12 were rationally selected based on proximity to these top candidates, and 24 were selected without bias from homology. Among the 96 variants, the 12 top substitutions were biased to occur twice as often, compared to the remaining 36.

The final Round 3 library consisted of 48 variants, each having an equal chance to use R2-29 or R2-93 as a template. Each variant contained an additional 4-5 amino acid substitutions, selected with an equal chance from a subset of the 24 top substitutions from the previous round.

3.4.2 Library Construction and Evaluation

Reagents and chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless indicated otherwise.

The dCK variants with an N-terminal 6xHis tag were synthesized by DNA2.0 and incorporated into the pJ401 vector, containing ampicillin resistance and an IPTG inducible T5 promoter. The genes were transformed into *E. coli* strain BL21(DE3), and single colonies were used to inoculate small volume cultures, and stored at -80°C.

Frozen stabs were used to inoculate 5 mL of LB media and grown for 7 hours at 37°C to an approximate OD₆₀₀ of 0.5. The grown cells were induced to a final concentration of 0.4 mM IPTG, and expressed for 2 hours at 30°C. Cells were harvested by centrifugation and stored at -80°C until purification.

For activity screening, cell pellets were thawed on ice, incubated with 120 µL of BugBuster (Millipore) at 4°C for 30 minutes. The cellular debris was separated through centrifugation, and the lysate was incubated at 4°C with 50 µL of HisMag beads (Invitrogen). The protein-bound beads were washed 1x200 µL of binding buffer (300 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM Imidazole), 3x200 µL of wash buffer (300 mM NaCl, 50 mM Tris-HCl pH 8.0, 50 mM Imidazole), and eluted with 100 µL of elution buffer (300 mM NaCl, 50 mM Tris-HCl pH 8.0, 250 mM Imidazole). Protein yields were generally uniform, at 10 µg per 5 mL culture, determined by absorbance ($\epsilon_{280} = 55,190 \text{ M}^{-1}\text{cm}^{-1}$), and with appropriate purities (>90%), based on SDS-PAGE analysis.

For measuring steady state parameters, the expression/purification protocols were scaled-up. Purified plasmid was transformed into BL-21(DE3), with clones isolated on LB agar, and subcultured in 50 mL LB culture. The crude lysate was purified using Ni-NTA agarose resin column (Qiagen), and exchanged in storage buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM MgCl₂) and stored at -80°C until use. Typical yields were 250 µg per 50 mL culture, and isolated at >95% purity.

3.4.3 Enzyme Characterization

Kinase activity was evaluated using an established coupled spectrophotometric assay. (20,21) Briefly, the reaction was monitored at 37°C in a 0.5 mL reaction, containing 50 mM Tris-HCl pH 8.0, 100 mM KCl, 5 mM MgCl₂, 1 mM ATP, 0.2 mM PEP, 0.2 mM NADH, 5 U of pyruvate kinase (Roche) and lactate dehydrogenase (Roche). Substrate phosphorylation was monitored by the coupled oxidation of NADH ($\epsilon_{280} = 6,220 \text{ M}^{-1}\text{cm}^{-1}$) with a Cary 50 UV-Vis Spectrophotometer (Agilent).

For initial screening kinetics, two time-resolved experiments using 2 μL or 5 μL of eluted protein (approximately 0.08 and 0.20 μM) were conducted. Substrate concentrations of L-FMAU (Carbosynth) and D-dC were determined by UV absorbance ($\epsilon_{260} = 8,840 \text{ M}^{-1}\text{cm}^{-1}$ and $\epsilon_{260} = 6,650 \text{ M}^{-1}\text{cm}^{-1}$ respectively). Initial concentrations were fixed to 100 μM and followed for 30 minutes at 37°C. The change in concentrations over time was fit to the Michaelis-Menten equation, and K_M and k_{cat} parameters were estimated using a numerical solution from Runge-Kutta approximation.

For steady state validation kinetics, approximately 0.05 μM of purified kinase is added to 0-512 μM substrate (concentrations determined by UV absorbance). The initial reaction rates were determined in triplicate 10 minute assay at 37°C, and fit to the Michaelis-Menten equation using nonlinear regression. Estimation of kinetic parameters was implemented in MATLAB (Mathworks, Natick, MA)

3.5 References

1. Min, J. J., & Gambhir, S. S. (2008). Molecular imaging of PET reporter gene expression. *Handb Exp Pharmacol*(185 Pt 2), 277-303.
2. Che, J., Doubrovin, M., Serganova, I., Ageyeva, L., Zanzonico, P., & Blasberg, R. (2005). hNIS-IRES-eGFP dual reporter gene imaging. *Mol Imaging*, 4(2), 128-136.
3. Moroz, M. A., Serganova, I., Zanzonico, P., Ageyeva, L., Beresten, T., Dyomina, E., Burnazi, E., Finn, R. D., Doubrovin, M., & Blasberg, R. G. (2007). Imaging hNET reporter gene expression with 124I-MIBG. *J Nucl Med*, 48(5), 827-836.
4. Reubi, J. C., Kvols, L., Krenning, E., & Lamberts, S. W. (1990). Distribution of somatostatin receptors in normal and tumor tissue. *Metabolism*, 39(9 Suppl 2), 78-81.
5. MacLaren, D. C., Gambhir, S. S., Satyamurthy, N., Barrio, J. R., Sharfstein, S., Toyokuni, T., Wu, L., Berk, A. J., Cherry, S. R., Phelps, M. E., & Herschman, H. R. (1999). Repetitive, non-invasive imaging of the dopamine D2 receptor as a reporter gene in living animals. *Gene Ther*, 6(5), 785-791.
6. Genove, G., DeMarco, U., Xu, H., Goins, W. F., & Ahrens, E. T. (2005). A new transgene reporter for in vivo magnetic resonance imaging. *Nat Med*, 11(4), 450-454.
7. Brader, P., Serganova, I., & Blasberg, R. G. (2013). Noninvasive molecular imaging using reporter genes. *J Nucl Med*, 54(2), 167-172.
8. Yaghoubi, S. S., Campbell, D. O., Radu, C. G., & Czernin, J. (2012). Positron emission tomography reporter genes and reporter probes: gene and cell therapy applications. *Theranostics*, 2(4), 374-391.
9. Haberkorn, U., Altmann, A., Morr, I., Knopf, K. W., Germann, C., Haeckel, R., Oberdorfer, F., & van Kaick, G. (1997). Monitoring gene therapy with herpes simplex virus thymidine kinase in hepatoma cells: uptake of specific substrates. *J Nucl Med*, 38(2), 287-294.

10. Koehne, G., Doubrovin, M., Doubrovina, E., Zanzonico, P., Gallardo, H. F., Ivanova, A., Balatoni, J., Teruya-Feldstein, J., Heller, G., May, C., Ponomarev, V., Ruan, S., Finn, R., Blasberg, R. G., Bornmann, W., Riviere, I., Sadelain, M., O'Reilly, R. J., Larson, S. M., & Tjuvajev, J. G. (2003). Serial in vivo imaging of the targeted migration of human HSV-TK-transduced antigen-specific lymphocytes. *Nat Biotechnol*, *21*(4), 405-413.
11. Dubey, P., Su, H., Adonai, N., Du, S., Rosato, A., Braun, J., Gambhir, S. S., & Witte, O. N. (2003). Quantitative imaging of the T cell antitumor response by positron-emission tomography. *Proc Natl Acad Sci U S A*, *100*(3), 1232-1237.
12. Niu, C., Bao, H., Tolstykh, T., Micolochick Steuer, H. M., Murakami, E., Korba, B., & Furman, P. A. (2010). Evaluation of the in vitro anti-HBV activity of clevudine in combination with other nucleoside/nucleotide inhibitors. *Antivir Ther*, *15*(3), 401-412.
13. Kokoris, M. S., & Black, M. E. (2002). Characterization of Herpes Simplex Virus type 1 thymidine kinase mutants engineered for improved ganciclovir or acyclovir activity. *Protein Sci*, *11*(9), 2267-2272.
14. Yaghoubi, S. S., Barrio, J. R., Namavari, M., Satyamurthy, N., Phelps, M. E., Herschman, H. R., & Gambhir, S. S. (2005). Imaging progress of herpes simplex virus type 1 thymidine kinase suicide gene therapy in living subjects with positron emission tomography. *Cancer Gene Ther*, *12*(3), 329-339.
15. Riddell, S. R., Elliott, M., Lewinsohn, D. A., Gilbert, M. J., Wilson, L., Manley, S. A., Lupton, S. D., Overell, R. W., Reynolds, T. C., Corey, L., & Greenberg, P. D. (1996). T-cell mediated rejection of gene-modified HIV-specific cytotoxic T lymphocytes in HIV-infected patients. *Nat Med*, *2*(2), 216-223.
16. Likar, Y., Dobrenkov, K., Olszewska, M., Shenker, L., Cai, S., Hricak, H., & Ponomarev, V. (2009). PET imaging of HSV-tk mutants with acquired specificity toward pyrimidine- and acycloguanosine-based radiotracers. *Eur J Nucl Med Mol Imaging*, *36*(8), 1273-1282.

17. Magrassi, L., Finocchiaro, G., Milanese, G., Benvenuto, F., Spadari, S., & Focher, F. (2003). Lack of enantioselectivity of herpes virus thymidine kinase allows safer imaging of gene delivery. *Gene Ther*, *10*(25), 2052-2058.
18. Kumar, D., Abdulovic, A. L., Viberg, J., Nilsson, A. K., Kunkel, T. A., & Chabes, A. (2011). Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Res*, *39*(4), 1360-1371.
19. Gerth, M. L., & Lutz, S. (2007). Mutagenesis of non-conserved active site residues improves the activity and narrows the specificity of human thymidine kinase 2. *Biochem Biophys Res Commun*, *354*(3), 802-807.
20. Likar, Y., Zurita, J., Dobrenkov, K., Shenker, L., Cai, S., Neschadim, A., Medin, J. A., Sadelain, M., Hricak, H., & Ponomarev, V. (2010). A new pyrimidine-specific reporter gene: a mutated human deoxycytidine kinase suitable for PET during treatment with acycloguanosine-based cytotoxic drugs. *J Nucl Med*, *51*(9), 1395-1403.
21. Campbell, D. O., Yaghoubi, S. S., Su, Y., Lee, J. T., Auerbach, M. S., Herschman, H., Satyamurthy, N., Czernin, J., Lavie, A., & Radu, C. G. (2012). Structure-guided engineering of human thymidine kinase 2 as a positron emission tomography reporter gene for enhanced phosphorylation of non-natural thymidine analog reporter probe. *J Biol Chem*, *287*(1), 446-454.
22. Iyidogan, P., & Lutz, S. (2008). Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity. *Biochemistry*, *47*(16), 4711-4720.
23. Gil, J. S., Machado, H. B., Campbell, D. O., McCracken, M., Radu, C., Witte, O. N., & Herschman, H. R. (2013). Application of a rapid, simple, and accurate adenovirus-based method to compare PET reporter gene/PET reporter probe systems. *Mol Imaging Biol*, *15*(3), 273-281.

24. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*, *485*(7397), 185-194.
25. Yoshikuni, Y., Ferrin, T. E., & Keasling, J. D. (2006). Designed divergent evolution of enzyme function. *Nature*, *440*(7087), 1078-1082.
26. Liao, J., Warmuth, M. K., Govindarajan, S., Ness, J. E., Wang, R. P., Gustafsson, C., & Minshull, J. (2007). Engineering proteinase K using machine learning and synthetic genes. *BMC Biotechnol*, *7*.
27. Ehren, J., Govindarajan, S., Moron, B., Minshull, J., & Khosla, C. (2008). Protein engineering of improved prolyl endopeptidases for celiac sprue therapy. *Protein Eng Des & Sel*, *21*(12), 699-707.
28. Midelfort, K. S., Kumar, R., Han, S., Karmilowicz, M. J., McConnell, K., Gehlhaar, D. K., Mistry, A., Chang, J. S., Anderson, M., Villalobos, A., Minshull, J., Govindarajan, S., & Wong, J. W. (2013). Redesigning and characterizing the substrate specificity and activity of *Vibrio fluvialis* aminotransferase for the synthesis of imagabalin. *Protein Eng Des & Sel*, *26*(1), 25-33.
29. 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. (2007). Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotechnol*, *25*(3), 338-344.
30. Chen, H. M., Borjesson, U., Engkvist, O., Kogej, T., Svensson, M. A., Blomberg, N., Weigelt, D., Burrows, J. N., & Lange, T. (2009). ProSAR: A New Methodology for Combinatorial Library Design. *J Chem Inf Model*, *49*(3), 603-614.
31. Lavie, A., Ostermann, N., Brundiers, R., Goody, R. S., Reinstein, J., Konrad, M., & Schlichting, I. (1998). Structural basis for efficient phosphorylation of 3'-azidothymidine

- monophosphate by *Escherichia coli* thymidylate kinase. *Proc Natl Acad Sci U S A*, 95(24), 14045-14050.
32. Hazra, S., Sabini, E., Ort, S., Konrad, M., & Lavie, A. (2009). Extending thymidine kinase activity to the catalytic repertoire of human deoxycytidine kinase. *Biochemistry*, 48(6), 1256-1263.
33. Kazlauskas, R. J., & Bornscheuer, U. T. (2009). Finding better protein engineering strategies. *Nat Chem Biol*, 5(8), 526-529.
34. Bommarius, A. S., Blum, J. K., & Abrahamson, M. J. (2011). Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr Opin Chem Biol*, 15(2), 194-200.

Chapter 4

Multi-Biosensor Detection of Nucleoside Analogs

Abstract

Therapeutic dosage monitoring (TDM) is the clinical practice of measuring a specific drug concentration within a patient's bloodstream to optimize individual dosing regimens. While TDM has demonstrated improved clinical outcomes in several therapies, routine monitoring is not advocated for most drugs due to a lack of established cost-effectiveness. As an alternative, this work describes a generalized low-cost detection methodology, leveraging interactions between pharmaceutical agents and their putative molecular targets. This model system focuses on TDM of nucleoside analogs, a potent drug class with activity against retroviruses, such as human immunodeficiency virus. Unlike most therapies, retroviral treatments routinely co-administer multiple nucleoside analogs as standard care, providing a unique challenge to detect several analytes within a sample. This study targeted eight compounds, consisting of the native DNA pyrimidines and six clinically relevant nucleoside analogs. These compounds were evaluated against three putative molecular targets displaying distinct substrate preferences: *T. maritima* thymidine kinase, *H. sapiens* deoxycytidine kinase, and *D. melanogaster* deoxynucleoside kinase. The chemometric model compares the observed time-course kinetics of an unknown sample to known enzyme-substrate interactions, simultaneously predicting both the identity and concentration of multiple analytes. Additionally, this methodology was capable of characterizing four analytes directly from human blood plasma, establishing a potential clinical application. This concept was expanded to a fluorometric assay to detect trace nucleoside analogs using a

cellphone camera, enabling the possibility of TDM in resource-limited and point-of-care environments. This modular detection strategy can readily translate to a variety of applications for the targeted identification of chemical agents.

4.1 Introduction

Nucleoside analogs represent a potent drug class for the treatment of retroviruses, such as human immunodeficiency virus (HIV) (1). These prodrugs are chemically modified compounds developed to mimic their physiological counterparts in order to exploit cellular metabolism. Within the cell, nucleoside analogs are subsequently phosphorylated by native kinases to mono-, di-, and triphosphorylated species, which then inhibit intracellular enzymes and/or destabilize genetic fidelity via direct incorporation (2, 3). Nucleoside analogs have unique dosage requirements, as combinations of multiple agents demonstrate superior efficacy over individual components and have since become standard of care for retroviral treatment (4). While dosage regimens can be complex, requiring daily administrations of multiple agents, these treatments are required to suppress viral loads and maintain quality of life (5).

Unfortunately, only three-fourths of retroviral patients undergoing treatment display suppressed viral loads, and a mere third are fully suppressed (6). Clinicians primarily attribute this observed treatment failure to poorly understood pharmacokinetics and lack of adherence (7-9). The blood plasma concentrations of nucleoside analogs have an established correlation to their clinical efficacy and ineffective dosing can result in toxicity at one extreme or evolved clinical resistance at the other, endangering present and future treatment options (10, 11). A proposed strategy to improve clinical outcomes is the routine monitoring of drug concentrations within a patient's blood, dynamically adapting dosage regimens to individual disease progression (10, 12-14).

This concept, therapeutic drug monitoring (TDM), is an established practice within clinical pharmacology, and is used to design treatments for optimized efficacy and reduced toxicity (15). In general, TDM is not considered cost-effective for most therapies, including nucleoside analogs, despite established improved clinical outcomes (16, 17). Unfortunately,

increasing numbers of high-risk infected populations, such as co-infection (i.e. HIV/Hepatitis C), cancer, and pregnant/pediatric patients, have disproportionate survival rates and potentially benefit from TDM (4, 13, 18). While physicians recommend clinical trials to assess the benefit of TDM for both general and specialty populations, the expense and instrument availability are presently prohibitive (15).

Traditional analytical methods to determine nucleoside analog concentrations are based on liquid chromatography coupled to mass spectrometry (19-21). Nucleoside analog therapies often co-administer multiple agents, requiring large volumes of blood (the requisite 5-25 mL is a specific concern for pediatric patients), labor-intensive sample preparation, and distinct column conditions to detect each analyte. Moreover, in resource-limited countries, having the greatest need for effective HIV treatment, instrumentation and personnel is often non-existent (22). For both domestic and global retroviral treatment, the clinical need for therapeutic detection requiring minimal training and deployment costs remains critically unmet.

A potential solution is to leverage the inherent selectivity of native enzymes to detect analytes directly from unmodified samples. Enzymes are routinely used as analytical reagents within biological samples; as the target species is selectively detected by changes during a monitored reaction. Relative to traditional analytical methods (i.e. chromatography), enzyme-based biosensors are simple, specific systems with rapid assay times and low cost (23, 24).

The enzymatic detection of nerve agents is often studied as a model system for biosensors. Neurotoxins have known activity against (aceto)cholinesterases and have an established molecular basis for inhibition (25). Early work assayed a single enzyme against representative insecticides and used a chemometric approach to identify compounds from observed kinetic rates (26). These studies demonstrated robust detection between two insecticides classes, organophosphates and carbamates, based on their respective irreversible and reversible

inhibition mechanisms. The concept was explored further using multiple cholinesterase homologs, establishing the ability to distinguish binary mixtures of insecticides from expected enzyme-substrate interactions (27). More recent strategies utilize similar arrays of cholinesterases, and analyze the kinetic output using machine-learning algorithms (28). These mathematical approaches significantly improve characterization of insecticide mixtures and further extended the limit of detection to nanomolar sensitivities (28). This technology has since enabled the detection of trace pesticides directly from soil, water, and food samples (29).

This work describes a similar strategy to quantify mixtures of natural and drug metabolites, however toward a clinical rather than an environmental application. This initial scope is limited to the quantification of eight compounds, including two native pyrimidine deoxynucleosides (thymidine, deoxycytidine) and six pyrimidine-based drugs of therapeutic relevance. These prodrugs are natively activated to their monophosphate forms by several native nucleoside kinases, having evolved distinct preferences for each substrate (30). The chemometric model uses an *in vitro* array of three recombinant nucleoside kinases to detect combinatorial mixtures of the target analytes. This modular approach is a potential platform for robust, low cost multi-biosensors for TDM and broader applications in other chemical disciplines.

4.2 Results and Discussion

4.2.1 Multi-biosensor Design

The present study was limited to the identification of pyrimidine-based analogs, a common scaffold for several therapeutic agents, and targeted eight representative compounds. Four of the investigated compounds were the native thymidine and representative derivatives. Zidovudine (AZT; 3'-azidothymidine) was the first HIV-therapy approved by the FDA and remains actively prescribed (3). L-FMAU (L-2'-fluoro-5-methyluracil-arabinofuranosyl) is an enantiomer of thymidine (resembling L-ribose), and has applications in positron emission

tomography (PET) imaging (31). Stavudine (D4T; 2',3'-didehydro-dideoxythymidine) is considered first-line therapy in resource limited countries, due to its low cost and availability (3). The remaining four substrates were the native deoxycytidine and its representative derivatives. Gemcitabine (GEM; 2',2'-difluorodeoxycytidine) is standard care in several carcinomas, as the analog is less debilitating than other chemotherapy agents (32). Zalcitabine (ddC; 2',3'-dideoxycytidine) is structurally similar to Deoxycytidine and Stavudine; its inclusion challenged the chemometric model with chemically similar species. Emtricitabine (FTC; 2',3'-dideoxy-5-fluoro-3'-thiacytidine) is often prescribed in combination with other antiretroviral agents and is considered standard of care for HIV treatment (33).

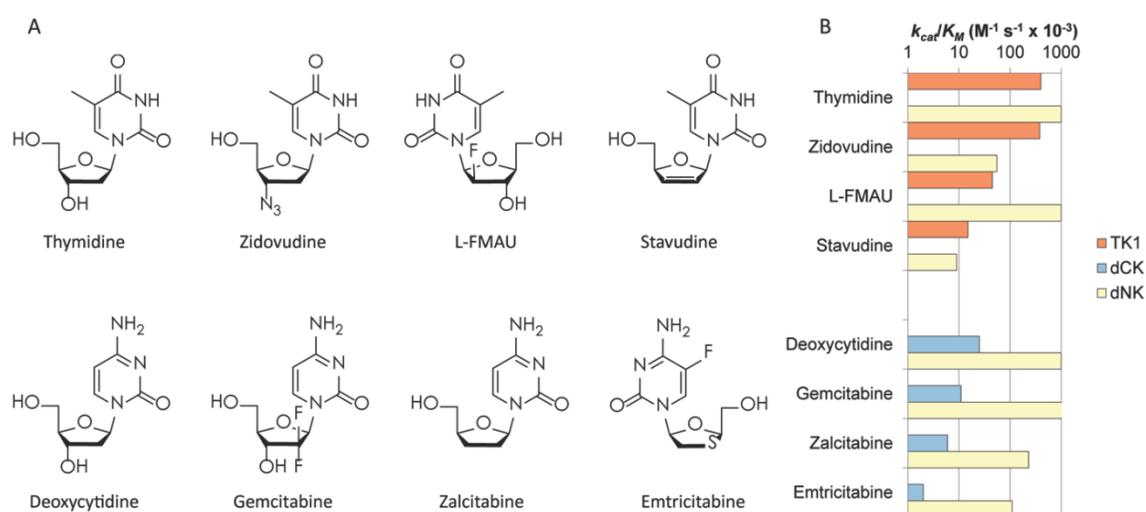


Figure 4.12. Investigated Substrates and Enzymes. (A) The top row depicts thymidine and its respective derivatives, while the bottom row depicts deoxycytidine and its respective derivatives. (B) Comparison of the catalytic efficiencies of each substrate interacting with the kinase array. TK1 displays selective activity for thymidine and its derivatives, while dCK displays exclusive activity for deoxycytidine and its derivatives. dNK exhibits activity for all investigated substrates.

An optimal multisensor contains a highly specific enzyme for each target analyte; unfortunately, native enzymes rarely fulfill this idealized role. In reality, natural selection provides a variety of homologous enzymes, evolving unique and diverse kinetic profiles (30).

Instead of a single enzymatic response, the experimental design contained an enzyme array with distinct patterns for each substrate. This modular approach combined time-resolved kinetics from multiple enzymes and identified an unknown compound from comparisons to known enzyme-substrate interactions. This investigation included three rationally selected nucleoside kinases exhibiting distinct preferences for the eight selected nucleoside analogs.

	TK1			dCK			dNK		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{mM}^{-1}\text{s}^{-1}$)
Thymidine	<1	0.40 ± 0.04	400		<i>not detected</i>	0	<1	14.9 ± 2.9	14 000
Zidovudine	<1	0.38 ± 0.03	380		<i>not detected</i>	0	8 ± 1	0.44 ± 0.08	55
L-FMAU	13 ± 2	0.59 ± 0.08	45		<i>not detected</i>	0	<1	6.10 ± 1.22	6 100
Stavudine	26 ± 0	0.38 ± 0.02	15		<i>not detected</i>	0	16 ± 4	0.14 ± 0.02	9
Deoxycytidine		<i>not detected</i>	0	<1	0.025 ± 0.005	25	<1	10.3 ± 2.0	10 000
Gemcitabine		<i>not detected</i>	0	17 ± 3	0.181 ± 0.033	11	<1	12.7 ± 2.5	12 000
Zalcitabine		<i>not detected</i>	0	4 ± 0	0.025 ± 0.002	6	4 ± 1	1.06 ± 0.21	230
Emtricitabine		<i>not detected</i>	0	10 ± 2	0.024 ± 0.001	2	7 ± 2	0.83 ± 0.16	110

Table 4.8. Steady State Michaelis-Menten Parameters for Enzymes and Substrates.

T. maritima thymidine kinase type 1 (TK1) showed exclusive activity for thymidine and its analogs (34). The thermophilic variant was selected for its high overexpression yield and ease of purification. *H. sapiens* deoxycytidine kinase (dCK) displayed selective activity for deoxycytidine and its analogs. The human enzyme is well studied among nucleoside kinases and is a known molecular target for several nucleoside-based drugs (35). *D. melanogaster* deoxynucleoside kinase (dNK) is known for its extremely high activity and promiscuity, displaying activity for all investigated substrates but showing sensitivity to substrate modifications at the 3'-ribose position (36).

4.2.2 Data Processing

The predictive model simulated reactions for established substrate-kinase interactions, and statistically compared the expected results to experimental traces of an unknown sample. The Bayesian model then returned the most likely substrate(s) and its respective concentration, based on interactions to all three kinases within the biosensor array.

Initial experiments focused on the blind characterization of a single component. The experimental trace (Q_k), described the time-dependent (t) substrate distribution assayed against enzyme k . Since each reaction was monitored to completion, the final absorbance change estimated initial substrate concentrations $[S_k]_0$ at a known enzyme concentration $[E_{total,k}]$. The expected time-course kinetics ($R_{i,k}$) for enzyme (k) interacting with substrate (i) was estimated using numerical integration of the Michaelis-Menten equation (Equation 1).

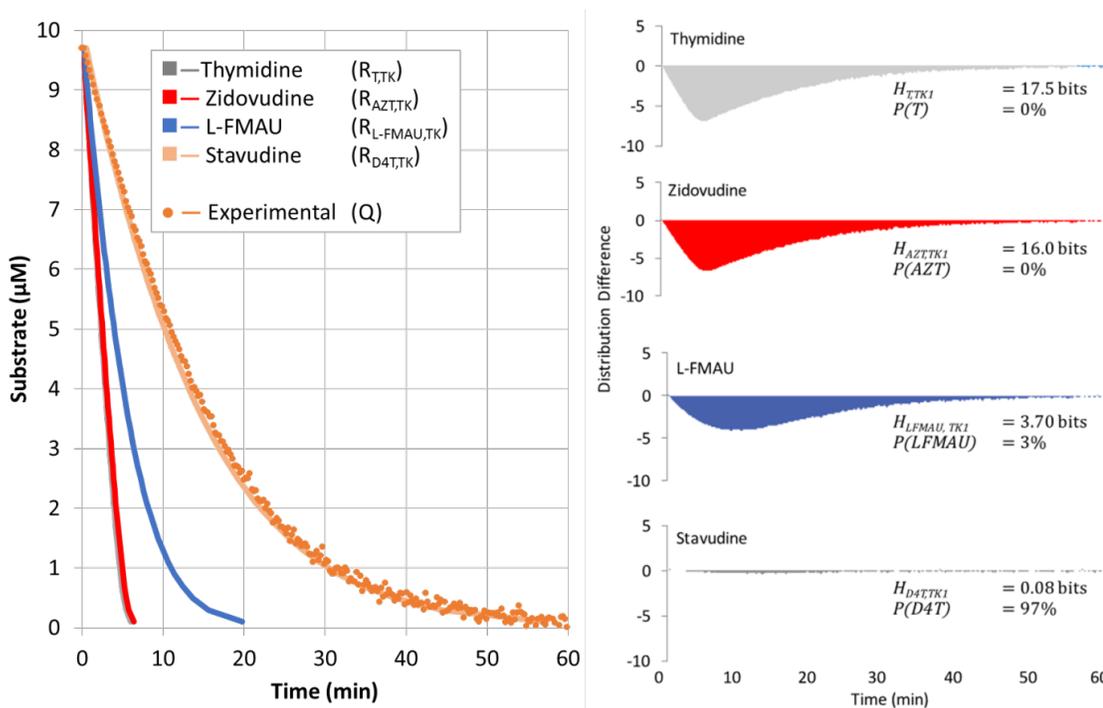


Figure 4.13. Data Analysis to Compare Observed Measurements to Expected Outputs. General method to compare analyte to unknown (Q) to expected outputs ($R_{i,k}$) based on Michaelis-Menten kinetics. The leftmost image depicts the experimental data of an unknown sample (Q) interacting with a known enzyme, in this case TK1. The solid lines are expected outputs from known Michaelis-Menten parameters for thymidine and its derivatives. The rightmost image visualizes the difference between the expected ($R_{i,k}$) and observed (Q) distributions, and quantified uncertainty using the Kullback-Leibler divergence ($H_{i,k}$). The identity of the unknown sample is estimated using a Boltzmann-like ensemble (P) for the probability distribution.

$$\frac{d[R_{i,k}]}{dt} = \frac{-d[S_i]}{dt} = \frac{K_{cat,i,k}[E_{total,k}][S_i]}{K_{M,i,k} + [S_i]} \quad (1)$$

The uncertainty between experiment and model was quantified with the Kullback-Leibler divergence ($H_{i,k}$), shown in Equation 2 (37). This uncertainty was translated into a Boltzmann-like ensemble (Equation 3) estimating the probability distribution of the unknown sample against known substrates (Y_i), from an experimental observation (X_i). In this work, $k=3$ (three kinases) and $j=8$ (eight possible substrates) for a substrate of interest (i).

$$H_{i,k} = \sum_t Q_k(t) \log \left(\frac{Q_k(t)}{R_{i,k}(t)} \right) \quad (2)$$

$$P(X_k|Y_i) = \frac{e^{-H_{i,k}}}{\sum_j e^{-H_{j,k}}} \quad (3)$$

The experimental information from each kinase was then combined using the joint probability the unknown sample matches a substrate of interest (Y_i), given the experimental kinetic traces from all three kinases (X_k), shown in Equation 4. The statistical model, derived from Bayes Theorem, assumed conditional independence, stating the observed activity was not correlated among kinases.

$$P(Y_i|X_1, \dots, X_k) = \frac{P(Y_k) \prod_i P(X_k|Y_i)}{\sum_j P(Y_j) \prod_i P(X_k|Y_j)} \quad (4)$$

The higher order mixtures had a similar mathematical framework; however, the expected substrate distributions were described as a linear sum of Michaelis-Menten reactions. A modification was initial concentrations of mixtures were estimated from nonlinear regression to the multiple reaction rates for each substrate. In most cases, two or more distinct kinetic rates were observed (i.e. ‘slow’ and ‘fast’ reactions), which was deconvoluted to estimate concentrations for the distinct substrates. If two substrate concentrations could not be determined

from the kinetic data, the algorithm assumed an equimolar mixture. The identities were attributed using a similar Bayesian framework to return the most likely combination of substrates that match the observed kinetics.

The blood plasma samples were modeled as a linear sum of four Michaelis-Menten reactions: Thymidine, Deoxycytidine, Zidovudine, and Emtricitabine. The initial concentrations of the four substrates were estimated using non-linear regression.

4.2.3 Single Component Evaluation

The initial set of experiments described a chemometric approach using Michaelis-Menten kinetics to characterize unknown solutions. Eight solutions (A1-A8) were prepared, each containing a single substrate. The individual models for each enzyme (TK1, dCK, dNK) are represented in the figure below, depicting the probability distribution across the eight possible substrates. The reactions were performed to completion and the final absorbance quantified the substrate concentration in each sample. The identity of each substrate was estimated from comparing the experimental time-course kinetics to expected outputs from known substrate interactions.

The model robustly distinguished thymidine and deoxycytidine analogs due to the TK1 and dCK kinases, displaying mutually exclusive activity. The individual enzymes identified the most likely substrate, and kinetic data from dNK was largely redundant, affirming the original prediction. These estimated probabilities were combined in the fusion model shown in the figure below, as kinetic information from all three kinases result in high confidence predictions for the identity of an unknown sample.

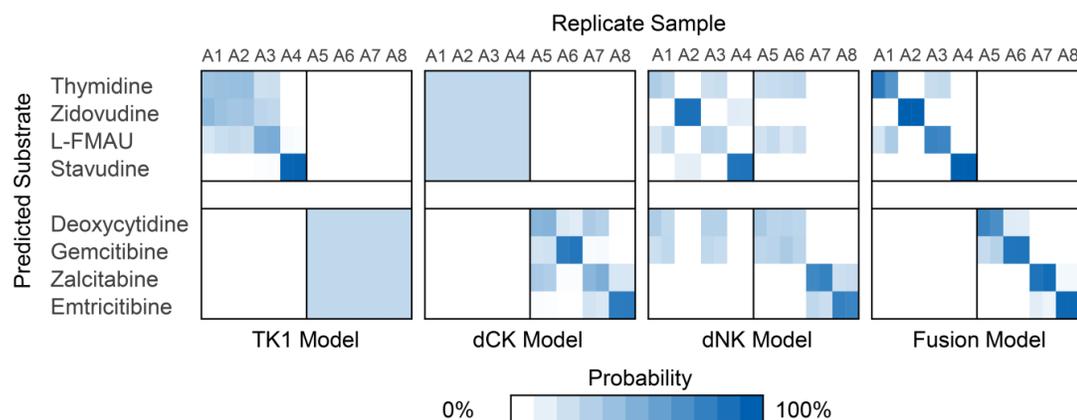


Figure 4.14. Probability Distribution of Single Component Samples. Each column represents a replicate sample of A1-A8 assayed against the indicated kinase. Each row represents a statistical comparison from the expected kinetic output of the indicated kinase with substrate to the experimental sample. The color intensity qualitatively describes the probability of each candidate model. The rightmost image describes the fusion model, combining kinetic information for each kinase, returning the most likely substrate present in each the duplicate sample.

	A1 (T)	A2 (AZT)	A3 (L-FMAU)	A4 (D4T)	A5 (dC)	A6 (Gem)	A7 (ddT)	A8 (FTC)
TK1	41%	40%	52%	97%	25%	25%	25%	25%
dCK	25%	25%	25%	25%	56%	78%	55%	78%
dNK	29%	89%	30%	89%	27%	27%	78%	76%
Fusion	78%	100%	79%	100%	81%	81%	93%	93%
[S] (μM)	9.7 ± 1.3	12.7 ± 2.3	10.5 ± 0.2	10.4 ± 0.4	11.2 ± 0.5	9.9 ± 0.4	12.4 ± 2.3	14.8 ± 1.0

Table 4.9. Analyte Probability for Single Component Analysis

In this initial dataset, the chemometric framework not only correctly identified the unknown substrate, but also accurately quantified the total substrate within each sample. While the identification of unknown metabolites has limited clinical significance, as patients/physicians are aware of their prescribed drugs, this dataset validated the predictive framework. Moreover, the robust and rapid identification of unknown compounds has direct applications in other disciplines, such as forensic chemistry.

4.2.4 Binary Mixture Evaluation

Next, the statistical framework was challenged to analyze mixtures of nucleosides/analogs in an attempt to deconvolute the identity and concentration of each

component. The 28 possible binary mixtures of the eight deoxynucleosides (B1-B28) were prepared, containing differing concentrations of each component. The systematic mixtures were assayed against the enzyme array and the results of the chemometric model are shown in the figure below, describing the probability distribution across all possible combinations.

Similar to the previous single component analysis, this methodology demonstrated the ability to identify and quantify the binary composition of an unknown sample. Mixtures containing thymidine (or analog) and deoxycytidine (or analog) were the easiest to characterize, as TK1 and dCK display mutually exclusive nitrogen base recognition. Among these samples, the predictive framework had nearly identical performance to the previous single component analysis. Similarly, mixtures of thymidine and its analogs also inherited high confidence, as TK1 and dNK display opposite preferences for the 3'-modified Zidovudine and enantiomeric L-FMAU. The disparate activity of these two kinases resulted in high probability predictions, and was necessary toward the experimental design of the enzyme array.

Deoxycytidine and its derivatives had the lowest confidence predictions due to the similar observed activity for 3' substituted substrates in both dCK and dNK. This highlights the importance of kinase selection and a potential limitation of enzyme-based characterization. Specifically, the mixtures Deoxycytidine/Zalcitabine (B24) and Deoxycytidine/Emtricitabine (B25) exhibited similar kinetic responses. While the first component (Deoxycytidine) was attributed with high confidence, the second substrate had an ambiguous characterization between the two species (Zalcitabine or Emtricitabine).

Sample	TK1	dCK	dNK	Fusion	Analytes (1/2)	Analyte 1	Analytes 2
B1	63%	16%	28%	89%	T/AZT	22.5 μ M	22.4 μ M
B2	33%	16%	23%	76%	T/L-FMAU	15.2 μ M	23.2 μ M
B3	49%	16%	25%	99%	T/D4T	18.8 μ M	20.3 μ M
B4	12%	16%	34%	51%	T/D-dC	23.4 μ M	11.9 μ M
B5	12%	23%	34%	82%	T/GEM	20.2 μ M	10.2 μ M
B6	12%	14%	32%	86%	T/ddC	23.3 μ M	14.0 μ M
B7	12%	12%	31%	84%	T/FTC	17.5 μ M	13.7 μ M
B8	30%	16%	99%	100%	AZT/L-FMAU	14.9 μ M	11.3 μ M
B9	47%	16%	45%	91%	AZT/D4T	23.8 μ M	18.9 μ M
B10	11%	16%	38%	75%	AZT/dC	18.2 μ M	13.0 μ M
B11	11%	23%	37%	98%	AZT/GEM	20.0 μ M	10.7 μ M
B12	11%	14%	40%	97%	AZT/ddC	19.3 μ M	12.5 μ M
B13	11%	12%	49%	96%	AZT/FTC	21.4 μ M	11.5 μ M
B14	30%	16%	32%	88%	L-FMAU/D4T	19.0 μ M	22.5 μ M
B15	9%	16%	28%	45%	L-FMAU/dC	19.3 μ M	10.9 μ M
B16	9%	23%	29%	85%	L-FMAU/GEM	23.8 μ M	8.2 μ M
B17	9%	14%	43%	88%	L-FMAU/ddC	19.3 μ M	13.8 μ M
B18	9%	12%	46%	86%	L-FMAU/FTC	21.4 μ M	15.4 μ M
B19	24%	16%	48%	75%	D4T/dC	23.9 μ M	11.0 μ M
B20	24%	23%	47%	98%	D4T/GEM	22.9 μ M	9.5 μ M
B21	24%	14%	43%	98%	D4T/ddC	26.4 μ M	12.1 μ M
B22	24%	12%	67%	97%	D4T/FTC	24.9 μ M	13.3 μ M
B23	16%	87%	24%	85%	dC/GEM	n.d.	n.d.
B24	16%	33%	26%	41%	dC/ddC	11.5 μ M	13.6 μ M
B25	16%	55%	31%	51%	dC/FTC	10.3 μ M	18.1 μ M
B26	16%	73%	25%	82%	GEM/ddC	9.1 μ M	10.1 μ M
B27	16%	43%	30%	75%	GEM/FTC	11.4 μ M	17.2 μ M
B28	16%	40%	59%	99%	ddC/FTC	n.d.	n.d.

Table 4.10. True Analyte Probability within Sample. (n.d. = not determined)

Similarly, the chemometric model had difficulty quantifying mixtures of similarly reacting species, such as Zalcitabine/Emtricitabine (B28) and could not algorithmically deconvolute two distinct species. These limitations are largely irrelevant, as this work serves as a model system. Assay conditions were chosen to best characterize all substrates, and optimized conditions can improve discrimination for species of interest. Alternatively, the inclusion of other deoxynucleoside kinases will increase confidence and can be modularly incorporated into the chemometric model.

This blind validation assessed the ability to deconvolute mixtures to not only distinguish chemically similar species, but also quantify their abundance. Overall, the data suggests this

chemometric approach can be applied toward a clinical application, and used to analyze complex mixtures of native/analog nucleosides directly from biological samples.

4.2.5 Blood Plasma Quantification

In these set of experiments, human plasma samples were spiked with Zidovudine and Emtricitabine, simulating the TDM of a patient undergoing HIV-treatment. The challenge was the simultaneous detection of the four analytes found within blood (two nucleoside analogs and the native pyrimidines). Plasma samples contained 1 to 10 μM of additional nucleoside analog, and the chemometric framework quantified each of the four deoxynucleosides.

As mentioned earlier, the selection of enzyme was critical toward the design of biosensors, and simplified algorithmic analysis in this study. The three kinases displayed distinct activity and can be robustly deconvoluted to obtain the concentrations of each metabolite. The thymidine specific TK1 phosphorylates Zidovudine and Thymidine, while dCK phosphorylates Deoxycytidine and Emtricitabine. Finally, dNK concentrations were intentionally low ($<0.1 \mu\text{g}$) to have detectible activity for the native pyrimidines only. The kinetic responses are shown in the figure below, and the approach was able to quantify diminishing substrate concentration in each sample.

The assay maintained sensitivity for therapeutically relevant concentrations, appropriate for TDM, however displayed reduced sensitivity at dilute drug concentrations. This detection limit was determined by the analytical technique, and is not inherent to the chemometric model. If needed, higher sensitivity detection methods (such fluorometric), can be integrated into the statistical framework and extend limit of detection. Altogether, this approach demonstrated the ability to identify nucleoside analogs directly from unmodified human plasma and the results suggest multi-biosensors are viable candidates for TDM to treat retroviral infection.

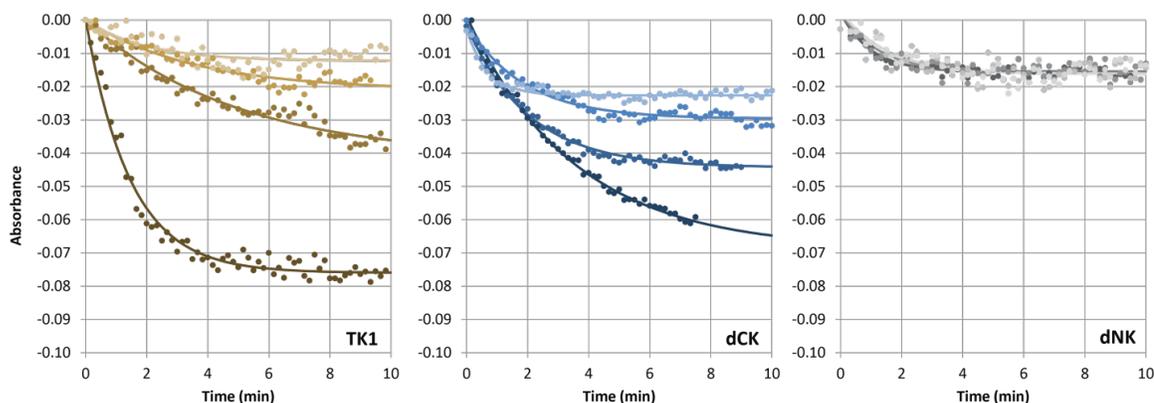


Figure 4.16. Time-course Kinetics of Plasma Samples. The color intensity corresponds to samples C1-C4, containing 10 μM (darkest), 5 μM , 2.5 μM , and 1 μM (lightest) of nucleoside analogs. TK1 displays activity for Thymidine and Zidovudine, dCK displays activity for Deoxycytidine and Emtricitabine, while dNK displays activity only for the native Thymidine and Deoxycytidine.

Analyte	C1 (μM)	C2 (μM)	C3 (μM)	C4 (μM)
Emtricitabine	10.2 ± 1.0	5.7 ± 0.6	3.2 ± 0.5	1.9 ± 0.5
Zidovudine	10.7 ± 1.1	5.5 ± 0.6	2.5 ± 0.4	1.1 ± 0.3
Deoxycytidine	1.0 ± 0.1	1.4 ± 0.1	1.6 ± 0.2	1.7 ± 0.5
Thymidine	1.5 ± 0.2	1.2 ± 0.1	0.9 ± 0.1	0.9 ± 0.2

Table 4.11. Estimated Composition of Spiked Plasma Samples

4.2.6 Point-of-Care Detection

While the previous assay has potential clinical utility, widespread adoption would still require on-site sampling and transport to off-site laboratories. To overcome this logistical limitation, this work adapted a previously described nanomolar sensitive fluorometric assay to detect nucleoside analogs, using a retail laser pointer and cellphone for point-of-care applications. The cellphone's camera retained detection at therapeutically relevant concentrations, obtaining time-resolved data without specialized equipment or personnel, and had similar performance to the established assay (38). With subsequent algorithmic analysis, this methodology enables low-cost, robust characterization of multiple analytes for therapeutic dosage monitoring. Moreover, the detected fluorophore is routinely used in several clinical assays (e.g. quantification of glucose, cholesterol, etc.). This fluorometric method can be re-optimized for focused detection at any desired concentration or utilize high-dynamic range for an expanded detection limit using

multiple exposures. This cell-phone based methodology is highly adaptable to a variety of point-of-care applications for immediate diagnostic results, even within resource-limited environments.

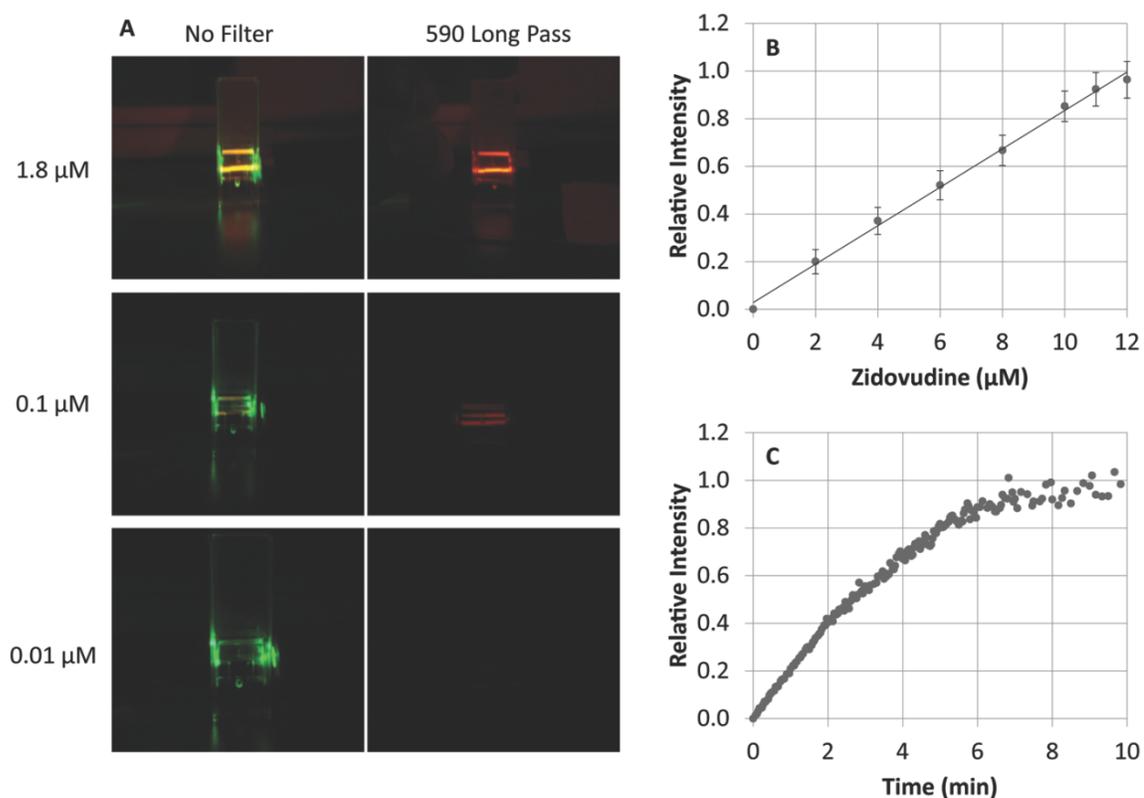


Figure 4.17. Cellphone Camera-based Nucleoside Analog Detection. (A) Captured images depicting experimental set up for laser pointer induced fluorescence of samples with known concentration (B) Calibration curve from triplicate Zidovudine samples, configured for optimal sensitivity at 10 μM . (C) Time-course kinetics of 10 μM Zidovudine interacting with TK1, using the same exposure settings.

4.3 Conclusion

This work outlines a methodology for the enzyme-based detection of multiple analytes using algorithmic analysis. In this model system, three nucleoside kinases were able to blindly identify and characterize unknown mixtures of eight nucleoside analogs at physiologically relevant concentrations. Additionally, this approach was able to quantify four analytes directly from unmodified human blood plasma, enabling potential clinical applications. Perhaps more interesting is the inherent versatility of a chemometric model. Additional substrates can be readily incorporated with known Michaelis-Menten parameters, and no experimental alterations. This framework not only supports monitoring for nucleoside analogs, but also any potential treatment regimen. An expanded roster of non-kinase enzymes allows for the detection of separate drug classes altogether, and future implementations enable the design of customized multisensors for individual patient treatment(s). Physicians can routinely monitor individual disease progression, dynamically adapting regimens to patient needs, and improve clinical outcomes. In a general case, multi-enzyme approaches are widely applicable for the detection of a broad range of chemical and biological agents. Further development may result in novel applications of these biosensors in other disciplines and increased overall use of enzymes within analytical chemistry.

4.4 Methods

4.4.1 Chemicals and Reagents

Unless otherwise specified, reagents were purchased from Sigma-Aldrich. The native deoxynucleosides (thymidine and deoxycytidine) were prepared as stock aqueous solutions of known concentrations determined by UV absorbance ($\epsilon_{260} = 9,000 \text{ M}^{-1} \text{ cm}^{-1}$; $8,600 \text{ M}^{-1} \text{ cm}^{-1}$ respectively) (39). The thymidine analogs: Zidovudine, L-FMAU (Carbosynth) and Stavudine ($\epsilon_{260} = 9,700 \text{ M}^{-1} \text{ cm}^{-1}$; $8,800 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon_{280} = 3,200 \text{ M}^{-1} \text{ cm}^{-1}$ respectively); in addition to the

deoxycytidine analogs Gemcitabine (Eli Lilly), Zalcitabine, and Emtricitabine were similarly prepared ($\epsilon_{260} = 9,400 \text{ M}^{-1} \text{ cm}^{-1}$; $9,100 \text{ M}^{-1} \text{ cm}^{-1}$, $9,600 \text{ M}^{-1} \text{ cm}^{-1}$) (40-42).

4.4.2 Protein Expression and Purification

The genes encoding for TK1, dCK and dNK were previously cloned into pET expression vectors (Invitrogen) with IPTG inducible promoters and an N-terminal 6xHis tag fusion (35, 43, 44). The expression vectors were transformed into BL21(DE3) cells, and selected on LB media with ampicillin as a resistance marker (100 $\mu\text{g}/\text{mL}$). Isolated clones were subcultured into LB media, grown to an OD_{600} of 0.5. Protein expression was induced with a final IPTG concentration of 0.4 mM at 20°C for 16 hours. The resulting cultures were pelleted and lysed via sonication.

For the purification of *H. sapiens* dCK and *D. melanogaster* dNK, the crude lysate was isolated using a Ni-NTA agarose resin column (Qiagen), according to manufacturer instruction. The resulting elutions were exchanged into storage buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM MgCl_2) and stored at -80°C until use. Typical yields were 5 mg/L titer at >90% purity, as determined by SDS-PAGE.

Purification of *T. maritima* thymidine kinase (TK1) followed similar protocol, with the modification that the crude lysate was heated to 80°C for 20 min. The impurities were precipitated through centrifugation at 4°C and the target protein, found in the supernatant, was exchanged into storage buffer. Typical yields were 20 mg/L titer at >90% purity, as determined by SDS-PAGE.

4.4.3 Enzyme Characterization

Kinase activity was monitored with a coupled spectrophotometric assay, as described previously (38). Briefly, a generic kinase phosphorylation event was monitored by the regeneration of consumed ATP (via pyruvate kinase), which was coupled to the molar equivalent

oxidation of NADH. Activity for each substrate was measured by UV absorbance (NADH: $\epsilon_{340}=6,220 \text{ cm}^{-1} \text{ M}^{-1}$) using a Cary Bio 100 Spectrophotometer (Agilent).

The reactions were performed at room temperature in 500 μL reactions, containing 50 mM Tris-HCl pH 8.0, 100 mM KCl, 5 mM MgCl_2 , 1 mM ATP, 0.2 mM phosphoenolpyruvate, 0.2 mM NADH, 5 U of pyruvate kinase (Roche) and 5 U lactate dehydrogenase (Roche). Unless otherwise specified, the total enzyme per reaction was constant (1.03 μg TK1, 1.52 μg dCK, 0.727 μg dNK), determined by UV absorbance ($\epsilon_{280}=10,680 \text{ cm}^{-1} \text{ M}^{-1}$; $56,380 \text{ cm}^{-1} \text{ M}^{-1}$; $39,880 \text{ cm}^{-1} \text{ M}^{-1}$ respectively) (45). Steady state Michaelis-Menten parameters were obtained from triplicate measurements using non-linear regression, and were in agreement with literature precedent (35, 43, 44).

4.4.4 Datasets

The experimental design contained a collection of three datasets, assayed against each of the three kinases within the array. The first two datasets (Dataset I and Dataset II) validated the chemometric model, while the final third dataset (Dataset III) simulated a potential clinical application. In a typical retroviral and cancer treatment regimen, drug blood plasma concentrations are 10-50 μM upon administration and decline according to their respective pharmacokinetic profile (46). As such, the scope was limited to the detection of 10 μM , comparable to mass spectrometry based methods (19-21).

The first dataset contained single component samples of the eight investigated substrates (A1-A8): Thymidine, Zidovudine, L-FMAU, Stavudine, Deoxycytidine, Gemcitabine, Zalcitabine, and Emtricitabine respectively. Each sample contained 10 μM of substrate, as determined by UV-absorbance. The time-course kinetics for each of the 24 reactions was collected in duplicate. Dataset I assessed the blind identification and quantification of a single component.

The second set contained all 28 possible binary combinations (B1-B28) of the eight investigated substrates. Each sample contained 20 μM of thymidine or respective derivative and 10 μM of deoxycytidine or respective derivative. Dataset II assessed the ability to characterize mixtures of compounds at differing concentrations.

The third set contained 50 μL of human blood plasma, spiked with Emtricitabine and Zidovudine, intended to simulate an HIV patient undergoing treatment. The concentrations ranged from 1 to 10 μM Zidovudine or Emtricitabine. The plasma samples were supplemented with a 10x buffer solution to a final concentration identical to the previously described assays. The solutions were allowed to equilibrate for 30 minutes, after which the diagnostic kinases were added, and the reaction was monitored until completion. Dataset III assessed the ability to quantify quaternary mixtures directly from human plasma, containing trace nucleoside analogs and the native pyrimidine deoxynucleosides as background.

4.4.5 Cell-Phone Based Detection

The fluorescence assay had been previously developed for the time-resolved detection of pyruvate, and uses pyruvate oxidase for analyte detection within biological samples (47, 48). The individual reactions contained 5 mM MgCl_2 , 0.2 mM phosphoenolpyruvate, 0.2 mM thiamine pyrophosphate, 10 μM FAD, 10 μM Amplex Red, supplemented with 0.2 U pyruvate kinase, 0.4 U pyruvate oxidase, and 0.2 U horseradish peroxidase. The resulting kinase activity is detected by the conversion of Amplex Red (Invitrogen) to resorufin having excitation/emission maxima at 570/583 nm. Pyruvate oxidase (originating from acidophilic bacteria) is inactive within blood plasma and reactions were performed in aqueous solutions.

The detection protocol was adapted to characterize deoxynucleosides using a cell-phone camera. A 5mW green laser pointer, transmitting a broad spectra at 532 nm, excites resorufin producing visible orange fluorescence. For quantification, a cell-phone camera was placed behind

a 590 nm long-pass filter and monitored nucleoside conversion through time-lapse image acquisition. An image was taken every 3 seconds, at optimal exposure settings for 10 μM of substrate (f-stop 4, shutter speed 500 ms, 200 ISO). The pixel intensities were adjusted for the Bayer filter and gamma correction to return a linear response to substrate conversion.

4.5. References

1. Jordheim, L. P., Durantel, D., Zoulim, F., and Dumontet, C. (2013) Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases, *Nat Rev Drug Discov* 12, 447-464.
2. Stein, D. S., and Moore, K. H. (2001) Phosphorylation of nucleoside analog antiretrovirals: a review for clinicians, *Pharmacotherapy* 21, 11-34.
3. Cihlar, T., and Ray, A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine, *Antiviral Res* 85, 39-58.
4. Dybul, M., Fauci, A. S., Bartlett, J. G., Kaplan, J. E., and Pau, A. K. (2002) Guidelines for using antiretroviral agents among HIV-infected adults and adolescents - The panel on clinical practices for treatment of HIV, *Ann Intern Med* 137, 381-433.
5. Gulick, R. M., Mellors, J. W., Havlir, D., Eron, J. J., Meibohm, A., Condra, J. H., Valentine, F. T., McMahon, D., Gonzalez, C., Jonas, L., Emini, E. A., Chodakewitz, J. A., Isaacs, R., and Richman, D. D. (2000) 3-year suppression of HIV viremia with indinavir, zidovudine, and lamivudine, *Ann Intern Med* 133, 35-39.
6. Cohen, S. M., Van Handel, M. M., Branson, B. M., Blair, J. M., Hall, H. I., Hu, X. H., Koenig, L. J., Skarbinski, J., Tracey, A., Mermin, J., and Valleroy, L. A. (2012) Vital Signs: HIV Prevention Through Care and Treatment-United States (Reprinted from MMWR, vol 60, pg 1618-1623, 2011), *Jama-J Am Med Assoc* 307, 247-250.
7. Simoni, J. M., Amico, K. R., Pearson, C. R., and Malow, R. (2008) Strategies for promoting adherence to antiretroviral therapy: a review of the literature, *Curr Infect Dis Rep* 10, 515-521.
8. Dixit, N. M., and Perelson, A. S. (2004) Complex patterns of viral load decay under antiretroviral therapy: influence of pharmacokinetics and intracellular delay, *J Theor Biol* 226, 95-109.

9. Pawelek, K. A., Liu, S., Pahlevani, F., and Rong, L. (2012) A model of HIV-1 infection with two time delays: mathematical analysis and comparison with patient data, *Math Biosci* 235, 98-109.
10. Back, D., Gatti, G., Fletcher, C., Garaffo, R., Haubrich, R., Hoetelmans, R., Kurowski, M., Lubner, A., Merry, C., and Perno, C. F. (2002) Therapeutic drug monitoring in HIV infection: current status and future directions, *AIDS 16 Suppl 1*, S5-37.
11. Winston, A., Jose, S., Gibbons, S., Back, D., Stohr, W., Post, F., Fisher, M., Gazzard, B., Nelson, M., Gilson, R., Orkin, C., Johnson, M., Palfreeman, A., Chadwick, D., Leen, C., Schwenk, A., Anderson, J., Gompels, M., Dunn, D., Khoo, S., Sabin, C., and Study, U. K. C. H. C. (2013) Effects of age on antiretroviral plasma drug concentration in HIV-infected subjects undergoing routine therapeutic drug monitoring, *J Antimicrob Chemother* 68, 1354-1359.
12. Peloquin, C. A. (2002) Therapeutic drug monitoring in the treatment of tuberculosis, *Drugs* 62, 2169-2183.
13. Burger, D. M. (2010) The Role of Therapeutic Drug Monitoring in Pediatric HIV/AIDS, *Ther Drug Monit* 32, 269-272.
14. Alsultan, A., and Peloquin, C. A. (2014) Therapeutic drug monitoring in the treatment of tuberculosis: an update, *Drugs* 74, 839-854.
15. Touw, D. J., Neef, C., Thomson, A. H., Vinks, A. A., Cost-Effectiveness of Therapeutic Drug Monitoring Committee of the International Association for Therapeutic Drug, M., and Clinical, T. (2005) Cost-effectiveness of therapeutic drug monitoring: a systematic review, *Ther Drug Monit* 27, 10-17.
16. Asboe, D., Aitken, C., Boffito, M., Booth, C., Cane, P., Fakoya, A., Geretti, A. M., Kelleher, P., Mackie, N., Muir, D., Murphy, G., Orkin, C., Post, F., Rooney, G., Sabin, C., Sherr, L., Smit, E., Tong, W., Ustianowski, A., Valappil, M., Walsh, J., Williams, M., Yirrell,

- D., and Subcomm, B. G. (2012) British HIV Association guidelines for the routine investigation and monitoring of adult HIV-1-infected individuals 2011, *Hiv Med* 13, 1-44.
17. Kredo, T., Van der Walt, J. S., Siegfried, N., and Cohen, K. (2009) Therapeutic drug monitoring of antiretrovirals for people with HIV, *Cochrane Db Syst Rev*.
18. Kirk, G. D., Merlo, C., O' Driscoll, P., Mehta, S. H., Galai, N., Vlahov, D., Samet, J., and Engels, E. A. (2007) HIV infection is associated with an increased risk for lung cancer, independent of smoking, *Clin Infect Dis* 45, 103-110.
19. Adaway, J. E., and Keevil, B. G. (2012) Therapeutic drug monitoring and LC-MS/MS, *J Chromatogr B Analyt Technol Biomed Life Sci* 883-884, 33-49.
20. Kumar, V. R., Reddy, B. P., Kumar, B. R., Sreekanth, K., and Babu, K. N. (2013) High throughput LC-MS/MS method for simultaneous determination of zidovudine, lamivudine and nevirapine in human plasma, *J Chromatogr B Analyt Technol Biomed Life Sci* 921-922, 9-14.
21. Novakova, L. (2013) Challenges in the development of bioanalytical liquid chromatography-mass spectrometry method with emphasis on fast analysis, *J Chromatogr A* 1292, 25-37.
22. Phillips, A. N., Pillay, D., Miners, A. H., Bennett, D. E., Gilks, C. F., and Lundgren, J. D. (2008) Outcomes from monitoring of patients on antiretroviral therapy in resource-limited settings with viral load, CD4 cell count, or clinical observation alone: a computer simulation model, *Lancet* 371, 1443-1451.
23. Martinez, A. W., Phillips, S. T., Carrilho, E., Thomas, S. W., 3rd, Sindi, H., and Whitesides, G. M. (2008) Simple telemedicine for developing regions: camera phones and paper-based microfluidic devices for real-time, off-site diagnosis, *Anal Chem* 80, 3699-3707.
24. Arora, A., Simone, G., Salieb-Beugelaar, G. B., Kim, J. T., and Manz, A. (2010) Latest developments in micro total analysis systems, *Anal Chem* 82, 4830-4847.

25. Fukuto, T. R. (1990) Mechanism of action of organophosphorus and carbamate insecticides, *Environ Health Perspect* 87, 245-254.
26. Danzer, T., and Schwedt, G. (1996) Multivariate evaluation of inhibition studies on an enzyme electrodes system with pesticides and mixtures of mercury and pesticides, *Anal Chim Acta* 325, 1-10.
27. Bachmann, T. T., and Schmid, R. D. (1999) A disposable multielectrode biosensor for rapid simultaneous detection of the insecticides paraoxon and carbofuran at high resolution, *Anal Chim Acta* 401, 95-103.
28. Bachmann, T. T., Leca, B., Vilatte, F., Marty, J.-L., Fournier, D., and Schmid, R. D. (2000) Improved multianalyte detection of organophosphates and carbamates with disposable multielectrode biosensors using recombinant mutants of *Drosophila* acetylcholinesterase and artificial neural networks, *Biosens Bioelectron* 15, 193-201.
29. Amine, A., Mohammadi, H., Bourais, I., and Palleschi, G. (2006) Enzyme inhibition-based biosensors for food safety and environmental monitoring, *Biosens Bioelectron* 21, 1405-1423.
30. Arner, E. S., and Eriksson, S. (1995) Mammalian deoxyribonucleoside kinases, *Pharmacol Ther* 67, 155-186.
31. Alauddin, M. M. (2012) Positron emission tomography (PET) imaging with ¹⁸F-based radiotracers, *Am J Nucl Med Mol Imaging* 2, 55.
32. Conroy, T., Desseigne, F., Ychou, M., Bouche, O., Guimbaud, R., Becouarn, Y., Adenis, A., Raoul, J. L., Gourgou-Bourgade, S., de la Fouchardiere, C., Bennouna, J., Bachet, J. B., Khemissa-Akouz, F., Pere-Verge, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., Ducreux, M., Unicanc, G. T. D., and Intergrp, P. (2011) FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer, *New Engl J Med* 364, 1817-1825.

33. Wainberg, M. A., Brenner, B. G., and Turner, D. (2005) Changing patterns in the selection of viral mutations among patients receiving nucleoside and nucleotide drug combinations directed against human immunodeficiency virus type 1 reverse transcriptase, *Antimicro Agents Chemother* 49, 1671-1678.
34. Lutz, S., Lichter, J., and Liu, L. (2007) Exploiting temperature-dependent substrate promiscuity for nucleoside analogue activation by thymidine kinase from *Thermotoga maritima*, *J Am Chem Soc* 129, 8714-8715.
35. Iyidogan, P., and Lutz, S. (2008) Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity, *Biochemistry* 47, 4711-4720.
36. Munch-Petersen, B., Piskur, J., and Sondergaard, L. (1998) Four deoxynucleoside kinase activities from *Drosophila melanogaster* are contained within a single monomeric enzyme, a new multifunctional deoxynucleoside kinase, *J Biol Chem* 273, 3926-3931.
37. Lages, N. F., Cordeiro, C., Sousa Silva, M., Ponces Freire, A., and Ferreira, A. E. (2012) Optimization of time-course experiments for kinetic model discrimination, *PLoS One* 7, e32749.
38. Schelling, P., Folkers, G., and Scapozza, L. (2001) A spectrophotometric assay for quantitative determination of k_{cat} of herpes simplex virus type 1 thymidine kinase substrates, *Anal Biochem* 295, 82-87.
39. Cavaluzzi, M. J., and Borer, P. N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA, *Nucleic Acids Res* 32, e13.
40. Espey, M. G., Chen, P., Chalmers, B., Drisko, J., Sun, A. Y., Levine, M., and Chen, Q. (2011) Pharmacologic ascorbate synergizes with gemcitabine in preclinical models of pancreatic cancer, *Free Radic Biol Med* 50, 1610-1619.
41. Suhel, P., Baghel, U., Rajesh, P., Prabhakar, D., Engla, G., and Nagar, P. (2009) Spectrophotometric Method Development and Validation for Simultaneous Estimation of

- Tenofoir disoproxil fumarate and Emtricitabine in Bulk Drug and Tablet Dosage form, *J Pharm Clin Res* 1, 28-30.
42. Vinogradov, S. V., Poluektova, L. Y., Makarov, E., Gerson, T., and Senanayake, M. T. (2010) Nano-NRTIs: efficient inhibitors of HIV type-1 in macrophages with a reduced mitochondrial toxicity, *Antivir Chem Chemother* 21, 1-14.
43. Muthu, P., Chen, H. X., and Lutz, S. (2014) Redesigning human 2'-deoxycytidine kinase enantioselectivity for L-nucleoside analogues as reporters in positron emission tomography, *ACS Chem Biol* 9, 2326-2333.
44. Liu, L., Li, Y., Liotta, D., and Lutz, S. (2009) Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting, *Nucleic Acids Res* 37, 4472-4481.
45. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods Mol Biol* 112, 531-552.
46. Schulz, M., and Schmoltdt, A. (2003) Therapeutic and toxic blood concentrations of more than 800 drugs and other xenobiotics, *Int J Pharm* 58, 447-474.
47. Zhu, A., Romero, R., and Petty, H. R. (2010) Amplex UltraRed enhances the sensitivity of fluorimetric pyruvate detection, *Anal Biochem* 403, 123-125.
48. Zhu, A., Romero, R., and Petty, H. R. (2010) A sensitive fluorimetric assay for pyruvate, *Anal Biochem* 396, 146-151.

Chapter 5

Final Thoughts

Abstract

Inherent to its form, this dissertation focuses on a narrow set of applications, but does not exist in isolation. The fifth and final chapter draws connections within the original work presented and extends to contemporary literature, providing comment and perspective. The previous chapters describe two different enzyme engineering strategies to design reporter genes, enabling discussion into their strengths and weaknesses. These enzymes display favorable *in vitro* characterization and these candidates, along with others, may lead to a safe and stable reporter system. Once established, these systems will provide valuable clinical data for future treatments, such as gene therapy. A related focus of this dissertation is enzyme biosensors for reliable, affordable medical diagnostics. Investment into this and similar technologies is needed to provide analytical solutions within resource limited environments, and integration into the information age, perhaps having a transformative impact on healthcare. Finally, this chapter and dissertation closes with a general perspective on the applied sciences.

5.1 Computational Enzyme Engineering

Computer algorithms endeavor to solve the so-called “inverse folding problem”: the ability to design a sequence for an arbitrary structure/function. An early milestone was by Kuhlman et al. with the design of Top7, demonstrating an unnatural fold can be designed by computational methods (1). This framework eventually led to the development of *de novo* enzymes catalyzing unnatural reactions, such as in Siegel et al. which demonstrated enzyme-based catalysis of the traditional hallmarks of synthetic chemistry: Diels-Alder Kemp elimination, and Retro-aldol reactions: (2, 3). Despite the low activity of computationally derived enzymes, expectations were high at the time, and saw the routine use of computational chemistry to design biocatalysts.

Inspired by these and other achievements, initial work of this dissertation sought to computationally design therapeutic enzymes. The motivation was to evolve an orthogonal deoxycytidine kinase preferring unnatural L-nucleoside analogs over the native D-metabolites for applications as a non-invasive reporter gene. The computational framework was able to successfully identify individual amino acid substitutions to favor the unnatural enantiomer, however the majority of singly substituted variants displayed lower overall activity. Combinations of substitutions only moderately improve enantioselectivity, and higher substitution numbers significantly lowered activity. Herein lies the limitation of modern computational strategies: the design of entire enzyme sequences for activity is far from robust.

This work used established algorithms as tools for enzyme design, and discussion into theoretical methodologies and energy potentials therein is outside the scope of this work. For argument’s sake, suppose the current state of computational chemistry can accurately describe enzyme-substrate interactions. Within computational enzyme engineering, molecular dynamics (MD) simulation is considered most feasible to describe enzyme-substrate interactions, providing

detailed justification for the observed properties of engineered sequences (4, 5). Even work from previous lab members, Daugherty et al., confirm the benefit and predictive aspect of MD, however these simulations cannot be scaled and new algorithms are needed (6).

As an example, chapter 2 of this dissertation explicitly calculated energy potentials for 10,000 structures with similar approximations in molecular mechanics. To its credit, these algorithms did manage to find identified amino acid substitutions of benefit, however the limitation was within the combinatorial aspect of enzyme engineering. This work could not construct highly substituted variants. Expanding evaluation into double substituted variants would require 45 million structural calculations. Within a computational framework, a ‘brute-force’ approach is not feasible, even with high-performance computing. Undoubtedly, computational chemistry will develop improved algorithms exploiting increasingly powerful hardware, but current computational design is not reliable for applied use in enzyme engineering. Alternate algorithm-based approaches need to be investigated.

5.2 Design-of-Experiments

As described earlier, computational approaches reasonably identify amino acid substitutions of catalytic benefit, but demonstrate limited additivity with higher numbers of substitutions. This “hard limit” is not unique to computational approaches, and results from an assumption that beneficial substitutions are generally additive when combined (7). In fact, many protein design strategies assume additivity to construct stepwise superior catalysts (8, 9). Past efforts using these methods have demonstrated success utilizing large teams, screening hundreds of thousands of enzyme variants (10). While most combinations result in unexpected behavior, eventually a serendipitous variant will be found with a sufficient library (11, 12). In fact, most practitioners have little doubt any given approach has merit and will eventually improve an enzyme; rather, experts question which is most efficient.

Even the best enzyme engineering strategies sample an infinitesimal portion of sequences. Perhaps the most flawed aspect of these strategies is that only a fraction of information is retained. Significant investment is required to screen enzyme variants, but only details of the “best” variants are kept and the vast majority of information is simply discarded. Though not every sequence can be explored, one strategy is to minimize investigated variants by using smaller, functionally enriched libraries. Detailed characterization of these enriched libraries would not only provide information for beneficial substitutions, but also identify deleterious ones as well. A data-driven approach may result in the more efficient design of enzymes.

Chapter 3 investigated this concept to further develop dCK for improved L-enantioselectivity. Partnering with DNA2.0, we incorporated a design-of-experiments (DOE) methodology to iteratively design optimal dCK variant sequences. DOE is a generalized data-driven concept within applied statistics to algorithmically design experiments to gather the most information about an arbitrary system. The key distinction of applying DOE is that iterative rounds of enzyme variants had explicit knowledge of previous data. Subsequent rounds favored to mimic top ranking sequences, while biased against deleterious substitutions. Ultimately, this approach confirmed multiple variants demonstrating excellent *in vitro* discrimination as reporter gene candidates.

Data-driven approaches are rapidly emerging as the best strategy toward combining multiple substitutions for enzyme engineering. First described by O'Maille et al. using a Bayesian framework, the group successfully engineered a sesquiterpene synthetases for pharmaceutical development. This dissertation focused on a modern interpretation pioneered by the ProSAR and ProteinGPS algorithms developed by Codexis and DNA2.0 respectively (13, 14). These impressive catalytic achievements only mark an early stage of the implementation of data-driven approaches within enzyme engineering, as there is substantial room for improvement. Not to be

discounted, these approaches still require human intuition/artistry for design, as parameters used (e.g. number of substitutions per round) are largely empirical. With established guidelines, coupled with lowering costs of gene synthesis, the entirety of enzyme engineering may incorporate similar statistical models as routine practice.

5.3 Reporter Systems

A significant portion of this dissertation focused on the development of reporter systems to monitor of transgene expression and migration. These reporter systems are required for long-term implementation of gene and cell based therapies, specifically to provide clinical data from both animal and human models (15). Reporter systems are defined as a reporter gene coupled to a non-invasively detected probe. These genetic sequences typically encode enzymes that chemically modify and entrap an administered probe within target cells. The localized accumulation of exogenous probe is directly correlated to gene expression, providing a non-invasive method to monitor putative therapies. This section is not meant to exhaustively review reporter systems; but most experts consider human-derived reporter genes to have the most clinical merit due to their inherent non-immunogenicity. Among the six human reporters, it remains difficult to determine which system is of clear benefit, and standards of practice still need to be established (16-21).

This dissertation focused on one such system: human deoxycytidine kinase. The native enzyme is not suited as a reporter gene, having poor catalytic performances for non-invasive probes, and high activity for native metabolites. Previous chapters describe efforts to apply data-driven engineering and resulted in improved candidates which display exquisite *in vitro* performance, potentially overcoming the limitations of current reporter genes. This engineered system, in addition to work by others, are being investigated in cell and animal models, as part of the collaborative development of safe and stable reporter technology.

Hopefully, a reporter system will be approved for human testing; however, widespread adoption from academia and industry still has remaining challenges. Regulatory agencies will likely require established vector backbones and targeting sequences using reporter constructs for both *ex* and *in vivo* investigations (22). To date, an ideal vector has not been developed and current therapy is limited to somatic cell transferal.

Again this section is not meant to serve as a review of gene delivery vectors, but adeno-associated viruses (AAV) are considered by experts to be the most viable (23). AAV systems are attractive due to their low frequency of insertion, having stable chromosomal integration compared to alternative vectors (e.g. retrovirus, cationic particles) with non-negligible rates of random incorporation. To date, AAV have several limitations (i.e. immunogenicity, transfection efficiency, targeted delivery), however directed evolution efforts have mitigated some of these concerns and continued development may result in robust candidates (24, 25).

Once both reporter and vector systems have been established, conventional tissue models will soon be replaced by these validated, non-invasive assays to investigate disease progression with molecular precision. This combination of molecular imaging and reporter genes can monitor genetic and cellular processes in animal and eventually human patients. This new era of molecular medicine will allow physicians unprecedented access to evaluate treatment progress, a fundamental shift toward data-intensive patient care and individualized regimens, significantly improving therapeutic outcomes.

5.4 Enzyme Biosensors

Biosensors take advantage of the inherent specificity/sensitivity within biology to obtain complex analytical measurements with a simple, easy-to-use format. The uses for such devices are virtually limitless, ranging from medical diagnostics, process control to environmental/security applications. Enzyme sensors have firm historical precedent, first

described by Clark and Lyons, the progenitor to the modern ubiquitous use of glucose oxidase for diabetes management (26). Enzyme-based detection is a \$10 billion industry for routine at-home detection, yet enzymes have limited use within analytical chemistry, with the exception of niche circumstances (27). Much of the focus of analytical chemistry is toward the development of sophisticated procedures and instrumentation within controlled laboratory environments. This dissertation inverted this design constraint, instead developing portable, decentralized devices for point-of-care detection by non-specialists, a growing field of interest.

The developing world is often overlooked within scientific discourse, yet there is a desperate need for robust diagnostics that can be deployed within the field by professionals and volunteers alike. In this regard, sophisticated instrumentation is immaterial; and treatment is restricted by access to analytical technology. The most common diseases of poverty are HIV and tuberculosis, killing 3 million people every year, with the latter affecting a third of the world's population (28). The solution may be biosensors as selective, affordable and easy-to-use devices.

Chapter 4 of this dissertation continues discussion with the development of a 'proof-of-principle' detection of nucleoside analog drugs, which are standard-of-care for both HIV and tuberculosis treatment. The work described an array of deoxynucleoside kinases, each with distinct substrate preferences. The enzyme system directly assayed human plasma spiked with mixtures of drug to emulate patient blood samples. The resulting kinetic response was analyzed using a chemometric model to identify mixtures of native metabolites and administered drug. The distinguishing feature of this approach was the use of computational algorithms to deconvolute experimental measurements not only to detect but also quantify mixtures within complex biological solutions. In this information age, smartphone and tablet devices are uniquely positioned to profoundly change point-of-care diagnostics. To this end, Chapter 4 repurposed a cellphone camera as a diagnostic device with similar performance to traditional detection without

specialized equipment or personnel. While this work still requires development, future implementations can lead to the generalized detection of a broad range of chemical and biological agents.

The most important application for this and other biosensors is healthcare, symptomatic of a public health shift toward diagnostics and prevention, over explicit treatment. Current spending on healthcare is unsustainable, the United States alone spends 17% of its GDP on healthcare, having pervasive impact on all aspects of society (29). Information technology may provide a solution for data collection and analysis for individualized patient care within an existing infrastructure. This will stimulate a market for point-of-care and personal diagnostics that necessitates the need for new, inexpensive detection platforms. In this regard, smartphone enabled biosensors, like the one described in this dissertation, are ideal candidates, and can directly interface with technology. Having an established infrastructure, future medical technology will embrace distributed diagnostics, with phone and tablets repurposed to collect data, serving as conduits for analysis and discourse with professionals. While enzymes may not provide all analytical solutions, their potential benefits cannot be understated and will undoubtedly play a role in affordable healthcare.

5.5 General Perspective

A common theme of this dissertation is the exploration of data and its translation to actionable information. These concepts are not new to chemistry or even science, but particularly resonate in this information age. Across all scientific disciplines, the rate of data generation exceeds the ability to analyze, interpret and make decisions. Data has always been at the foundation of hypothesis-driven research; however the velocity to which data is collected and the volume of existing data provide unique opportunities within science. ‘Data-driven’ research

enables new types of projects, as featured in this dissertation, to leverage data as a resource to optimize experimental methods and make informed decisions.

Researchers within data-intensive environments will be required to develop data analytical skills in addition to their technical proficiency. This requires better interaction with data, integration from disparate sources, and reduction in complexity for interpretation. A growing market for computational expertise will be needed to not only construct analytical models but also visualize data to present findings for effective communication. At its core, data-driven research is not a revolutionary discipline, but rather the evolution of current methodologies. The ability to uncover meaningful relationships and patterns has been an integral to science—since its inception.

5.6 References

1. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy, *Science* 302, 1364-1368.
2. 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., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction, *Science* 329, 309-313.
3. 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., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes, *Science* 319, 1387-1391.
4. Kiss, G., Celebi-Olcum, N., Moretti, R., Baker, D., and Houk, K. N. (2013) Computational enzyme design, *Angew Chem Int Ed Engl* 52, 5700-5725.
5. Kiss, G., Rothlisberger, D., Baker, D., and Houk, K. N. (2010) Evaluation and ranking of enzyme designs, *Protein Sci* 19, 1760-1773.
6. Daugherty, A. B., Muthu, P., and Lutz, S. (2012) Novel protease inhibitors via computational redesign of subtilisin BPN' propeptide, *Biochemistry* 51, 8247-8255.
7. Wells, J. A. (1990) Additivity of mutational effects in proteins, *Biochemistry* 29, 8509-8517.
8. Reetz, M. T., Wang, L. W., and Bocola, M. (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space, *Angew Chem Int Ed Engl* 45, 1236-1241.
9. Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat Protoc* 2, 891-903.

10. Lutz, S. (2010) Beyond directed evolution--semi-rational protein engineering and design, *Curr Opin Biotechnol* 21, 734-743.
11. Weinreich, D. M., Delaney, N. F., DePristo, M. A., and Hartl, D. L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins, *science* 312, 111-114.
12. Reetz, M. T., and Sanchis, J. (2008) Constructing and analyzing the fitness landscape of an experimental evolutionary process, *Chembiochem* 9, 2260-2267.
13. O'Maille, P. E., Malone, A., Dellas, N., Andes Hess, B., Jr., Smentek, L., Sheehan, I., Greenhagen, B. T., Chappell, J., Manning, G., and Noel, J. P. (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases, *Nat Chem Biol* 4, 617-623.
14. Fox, R. J., Davis, S. C., Mundorff, E. C., Newman, L. M., Gavrilovic, V., Ma, S. K., Chung, L. M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J. C., Sheldon, R. A., and Huisman, G. W. (2007) Improving catalytic function by ProSAR-driven enzyme evolution, *Nat Biotechnol* 25, 338-344.
15. Alam, J., and Cook, J. L. (1990) Reporter genes: application to the study of mammalian gene transcription, *Anal Biochem* 188, 245-254.
16. Ponomarev, V., Doubrovin, M., Shavrin, A., Serganova, I., Beresten, T., Ageyeva, L., Cai, C., Balatoni, J., Alauddin, M., and Gelovani, J. (2007) A human-derived reporter gene for noninvasive imaging in humans: mitochondrial thymidine kinase type 2, *J Nucl Med* 48, 819-826.
17. Altmann, A., Kissel, M., Zitzmann, S., Kubler, W., Mahmut, M., Peschke, P., and Haberkorn, U. (2003) Increased MIBG uptake after transfer of the human norepinephrine transporter gene in rat hepatoma, *J Nucl Med* 44, 973-980.

18. Haberkorn, U. (2001) Gene therapy with sodium/iodide symporter in hepatocarcinoma, *Exp Clin Endocrinol Diabetes* 109, 60-62.
19. Rogers, B. E., Zinn, K. R., and Buchsbaum, D. J. (2000) Gene transfer strategies for improving radiolabeled peptide imaging and therapy, *Q J Nucl Med* 44, 208-223.
20. Olafsen, T., Sirk, S. J., Olma, S., Shen, C. K., and Wu, A. M. (2012) ImmunPET using engineered antibody fragments: fluorine-18 labeled diabodies for same-day imaging, *Tumour Biol* 33, 669-677.
21. Liang, Q., Satyamurthy, N., Barrio, J. R., Toyokuni, T., Phelps, M. P., Gambhir, S. S., and Herschman, H. R. (2001) Noninvasive, quantitative imaging in living animals of a mutant dopamine D2 receptor reporter gene in which ligand binding is uncoupled from signal transduction, *Gene Ther* 8, 1490-1498.
22. Brader, P., Serganova, I., and Blasberg, R. G. (2013) Noninvasive molecular imaging using reporter genes, *J Nucl Med* 54, 167-172.
23. Muzyczka, N. (1992) Use of adeno-associated virus as a general transduction vector for mammalian cells, In *Viral Expression Vectors*, pp 97-129, Springer.
24. Maheshri, N., Koerber, J. T., Kaspar, B. K., and Schaffer, D. V. (2006) Directed evolution of adeno-associated virus yields enhanced gene delivery vectors, *Nat Biotechnol* 24, 198-204.
25. Gray, S. J., Blake, B. L., Criswell, H. E., Nicolson, S. C., Samulski, R. J., and McCown, T. J. (2010) Directed evolution of a novel adeno-associated virus (AAV) vector that crosses the seizure-compromised blood-brain barrier (BBB), *Molecular Therapy* 18, 570-578.
26. Clark, L. C., Jr., and Lyons, C. (1962) Electrode systems for continuous monitoring in cardiovascular surgery, *Ann N Y Acad Sci* 102, 29-45.
27. Turner, A. P. (2013) Biosensors: sense and sensibility, *Chem Soc Rev* 42, 3184-3196.

28. Liu, L., Johnson, H. L., Cousens, S., Perin, J., Scott, S., Lawn, J. E., Rudan, I., Campbell, H., Cibulskis, R., and Li, M. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000, *The Lancet* 379, 2151-2161.
29. Berwick, D. M., and Hackbarth, A. D. (2012) Eliminating waste in US health care, *JAMA* 307, 1513-1516.