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.

Signature:

Lingfeng Liu Date

Directed Evolution and Computational Design of Nucleoside Analog Kinases

By

Lingfeng Liu
Doctor of Philosophy

Chemistry

Dr. Stefan Lutz
Advisor

Dr. David G. Lynn
Committee Member

Dr. Fredric M. Menger
Committee Member

Accepted:

Lisa A. Tedesco, Ph. D.
Dean of the James T. Laney School of Graduate Studies

Date

Directed Evolution and Computational Design of Nucleoside Analog Kinases

By

Lingfeng Liu

M. Sc., Hunan University, China, 2003

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

2009

Abstract

Directed Evolution and Computational Design of Nucleoside Analog Kinases

By Lingfeng Liu

Nucleoside analog (NA) prodrugs represent a promising group of viral and cancer therapeutics. Despite the relative ease of synthesizing diverse NAs, < 30 NAs have been FDA-approved. Difficulties arise since the prodrugs need to be phosphorylated by cellular nucleoside/nucleotide kinases to become activated triphophates. The NA-triphosphates then function as DNA replication terminators. Nevertheless, the human deoxynucleoside kinases generally catalyze NAs with poor efficiency, causing low drug potency and failure of many NAs in vivo.

As a potential solution, previous research found the co-administration of an exogenous kinase can accelerate the NA phosphorylation in vivo and increase drug potency. Kinase engineering by directed evolution provides an efficient strategy to evolve specific and efficient NA kinases from parental nucleoside kinases. Nevertheless, limited success has been achieved because of the lack of efficient selection/screening protocols that directly monitor NA phosphorylation. In this dissertation, I developed a

fluorescence-activated cell sorting (FACS)-based screening that combines fluorescent nucleobases and modified 2'-deoxyriboses from NA prodrugs, and the phosphorylation was followed by monitoring the entrapment of the fluorescent NAs through FACS. Using this screening, an orthogonal ddT kinase was evolved with 20,000-fold higher specificity and 6-fold higher catalytic efficiency for ddT starting from *Drosophila melanogaster* deoxynucleoside kinase (*Dm*dNK), a broad substrate-specificity kinase. To enhance the capability of searching through the sequence space for NA kinase activity, we used Rosetta program to switch the substrate specificity of *Dm*dNK to ddT. Combining Rosetta design with a mutation learned from directed evolution also led to a specific ddT kinase with improved thermostability. Each method developed orthogonal kinases which we propose will be more effective at phosphorylating NAs while minimizing perturbation to cellular nucleoside metabolism.

We tested these ideas in both *E. coli* bacteria and three different human cancer cell lines, and the results support our hypothesis that specific NA kinases increase NA potency more efficiently than promiscuous kinases, suggesting an alternative perspective for clinical applications of evolving orthogonal enzymes with minimized perturbation to cellular environment.

Directed Evolution and Computational Design of Nucleoside Analog Kinases

By

Lingfeng Liu

M. Sc., Hunan University, China, 2003

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

2009

Acknowledgements

My advisor, Dr. Stefan Lutz is one of the most important reasons for me to be able to proudly present here my dissertation research. I failed to find any words that can accurately describe my deepest gratitude to him for his continuous support, inspiration, challenge, encouragement, patience, helpful advices and friendly jokes throughout my graduate school life. Five years ago, it was him who got me excited about directed evolution and the greater scientific field beyond. I've never lost my passion about scientific research ever since, which benefited my entire graduate research experience. From him, I learned that I should always believe in myself; work hard to pursue my dream; stand up to frustrations; be thoughtful but always speak out when I have an opinion; be critical but constructive to other people's work; and many many more. It's been a truly honor for me to work in his lab and enjoy the wonderful journey as a graduate student.

I'd like to thank my two committee members, Dr. David Lynn and Dr. Fred Menger. Dr. Lynn is one of the most inspirational scientists I've ever met. In his mind, there are scientific questions, but there are no boundaries in order to search for answers. I appreciate his advices, criticisms that helped me to become better in scientific thinking and research. I'm grateful for all the encouragements and suggestions Dr. Menger has given to me during the past five years. I also like to extend my thanks to Dr. Vince

Conticello for his funny spirit that made me laugh out loud many times. I appreciate all the help I received from Ann Dasher, Patti Barnett, Sarah Keller, Steve Krebs and the entire Chemistry staff.

Thanks to our collaborators, Dr. Yongfeng Li and Dr. Dennis Liotta from Emory, as well as Dr. Paul Murphy and Dr. David Baker from University of Washington. My research wouldn't have been carried on without their great help.

The former and current Lutz lab members have been wonderful colleagues to me. Some became my close friends: Ying Yu, Zhen Qian, Joseph Lichter, Monica Gerth, Yichen Liu. They made our lab a fun place to work at and I wish them all the best in their career as well as their life.

I had a wonderful class full of energetic people who are passionate about science. The first year we started at Emory, often after midnight, we were working in the Chemistry building together to support and help each other. Outside of class, we played hard and shared a lot of happy time together. Those are sweet memories and I thank them dearly for making my graduate student life full of laughs. I was also very fortunate to make a lot of good friends outside of my lab and my class. Anil Mehta and Andrew Palmer have been wonderful friends to talk to about science. I've enjoyed every conversation with them so much and I'm always even more proud of what I do after

talking to them.

I came from a tiny town in the central part of China. My parents were not able to enjoy good education while they were young because of economic and historical reasons. But they did everything they could to fully support my brother's and my education. One of the most valuable things I learned from them is their optimism to face either good or hard time, which has essentially shaped my perspective about my life and the world around me. I'd like to thank them and also my younger brother and sister-in-law for the sweet family I'm so lucky to have. Also, many thanks to Ted and Pat Childers for loving me and supporting me ever since we met several years ago.

In the end, to my fiancé, Seth Childers, thank him for understanding and appreciating who I am, trusting me more than I trust myself, influencing me to make the best out of myself, and bringing lots of joy to my life.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Current HIV/AIDS and cancer therapeutic strategies

According to the World Health Organization (WHO), cancer accounted for 7.9 million deaths (around 13% of all deaths) in 2007, and deaths from cancer worldwide are projected to rise, with an estimated 12 million deaths in 2030. Two million people died from AIDS and 2.7 million new infections were reported in 2007. By the end of 2007, 33 million are living with HIV infection and 2 million among them are children under 15 years of age. In order to treat these diseases, more therapeutics development is required.

1.1.1 HIV/AIDS therapeutic strategies

One major group of drugs for treatment of both viral infection and cancer development are nucleoside anlogs (NAs) (Galmarini *et al.*, 2003; Schinazi *et al.*, 2006; Parker, 2009). More than 4 million HIV patients in low and middle income countries received antiretroviral therapy (ART) at the close of 2008, representing a thirty-six percent increase in one year and a ten-fold increase over five years. Antiretroviral drugs are broadly classified by the targeted phase of the virus life-cycle: 1) Nucleoside reverse transcriptase inhibitors (NRTI) are NAs that inhibit reverse transcription by terminating viral DNA elongation upon incorporation; 2) Non-nucleoside reverse transcriptase inhibitors (NNRTI) directly bind to the enzyme and interfere with its function; 3) Protease inhibitors (PI) inhibit the HIV protease activity responsible for cleaving nascent

proteins from final assembly of new viral particles; 4) Integrase inhibitors block HIV integrase activity, so that the viral DNA can not integrate into the infected cells DNA; and 5) Entry or fusion inhibitors interfere with binding, fusion and entry of the virus into the host cells. Among the 32 FDA-approved ART drugs, the 13 nucleoside analog drugs (e.g. 5-fluoro-1-(2*R*,5*S*)-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine (FTC); 3'-azido-3'-dexoythymidine (AZT), *etc*.) serving as NRTIs represent the largest group (Figure 1-1).

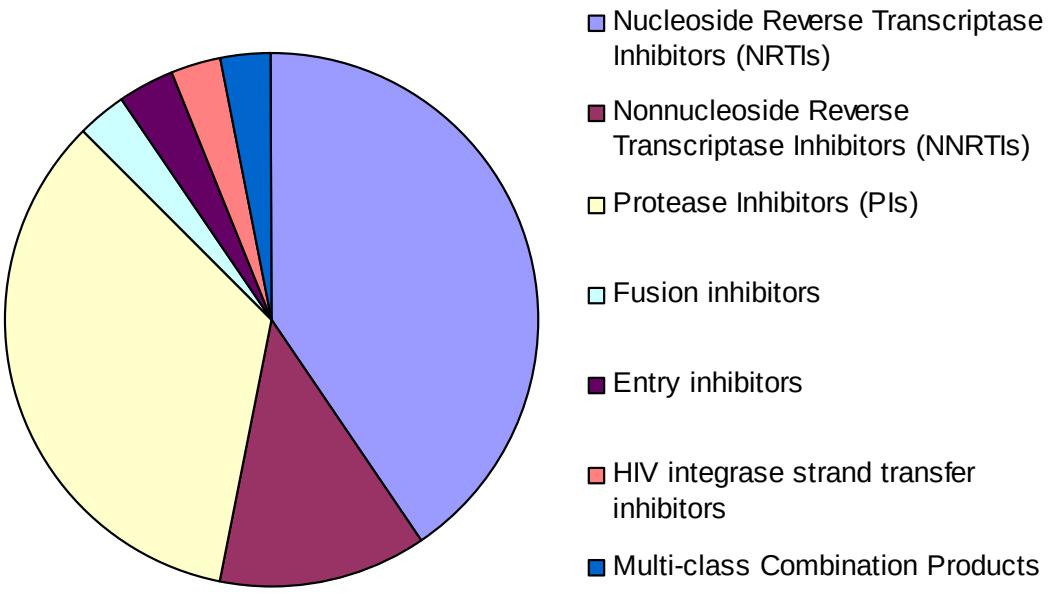


Figure 1-1: Distribution of antiretroviral drugs for the treatment of HIV infection. Among the 32 total FDA-approved drugs, thirteen are Nucleoside Reverse Transcriptase Inhibitors (NRTIs), four are Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs), eleven are Protease Inhibitors (PIs), and then the rest are one drug for each type: fusion inhibitor, entry inhibitor, integrase strand transfer inhibitor and multi-class combination products.
(http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm)

1.1.2 Nucleoside analogs in cancer treatment

The other major NA prodrug target is cancer. According to National Cancer Institute (NCI)'s SEER Cancer Statistics Review covering 1975-2006, the age-adjusted cancer incidence rate was 463 per 100,000 per year. Furthermore, the age-adjusted annual death rate was 187 per 100,000 men and women. Alarmingly, based on rates from 2004-2006, it is projected that 41% of men and women born today will be diagnosed with cancer at some time during their lifetime. On January 1, 2006, in the United States there were

approximately 11.4 million people alive who had a history of cancer. And it is estimated that ~1.5 million individuals will be diagnosed with cancer and approximately 560,000 people will die of cancer in 2009 (Horner MJ *et al.*, 2009). To battle against the rising threat from cancer, NAs also represent a group of promising drugs. For instance, Cytosine arabinoside (ara-C) is used for the treatment of hematological malignancies such as acute myeloid leukemia and non-Hodgkin lymphoma (Wang *et al.*, 1997), 2'-deoxy-2',2'-difluorocytidine (Gemcitabine) treats various carcinomas including non-small cell lung cancer (Tiseo *et al.*, 2009), pancreatic cancer (Neri *et al.*, 2009) and breast cancer (Dent *et al.*, 2008). 1-(β-L-2-hydroxymethyl-1,3-dioxolane-4-yl)-cytosine (Troxacitabine) is used for the treatment of acute myelogenous leukemia (Swords *et al.*, 2007). Whereas, 2-Chloro-9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-9H-purin-6-amine (Clofarabine) is approved for treatment of paediatric leukaemia (Bonate *et al.*, 2006), and 9-β-D-arabinofuranosylguanine (araG) is used for treatment of patients with T-cell acute lymphoblastic leukemia and lymphoblastic lymphoma (Curbo *et al.*, 2006), each of these NAs act in a similar manner to slow down the growth of cancer cells.

1.2 Nucleosides and nucleoside analog design

1.2.1 Structure and mechanism of NAs

Four 2’-deoxynucleosides (Thymidine (T), deoxycytidine (dC), deoxyadenosine (dA) and deoxyguanosine (dG), Table 1-1) exist in nature. The 2’-deoxynucleoside triphosphates (dNTPs) work as building blocks for DNA which are connected one after

another through phosphate bond backbone (Figure 1-2). Each nucleoside is composed of two moieties: a purine or purimidine nucleobase connected to 2'-deoxyribose through a N-glycosidic bond. Analogs of these nucleosides can impede DNA's natural function and be exploited for treatment of cancer and viral infections (Galmarini *et al.*, 2003; Schinazi *et al.*, 2006; Parker, 2009). Several NAs' structures are compared with natural 2'-deoxyribonucleosides in Table 1-1. Most nucleoside analog prodrugs have altered sugar moieties, and some have modifications on the nucleobase moieties. Each of these modifications can substantially impact gene extension/replication upon incorporation into DNA. The major ways in which NAs work as chain terminators include: a). preventing phosphate bond formation after being added to the end of the DNA chain; b). inhibiting DNA/RNA polymerases, or c). inhibiting replication by preventing base pairing upon incorporating into DNA.

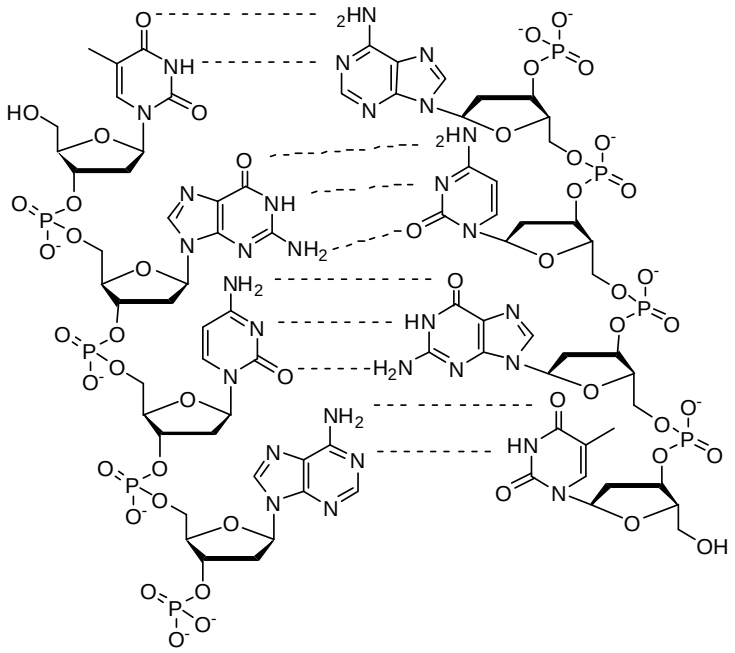
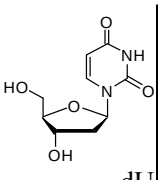
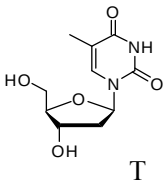
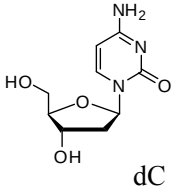
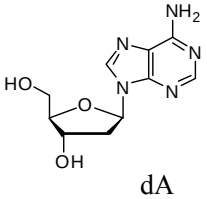
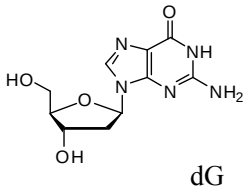
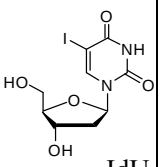
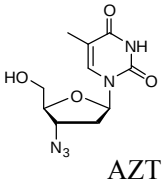
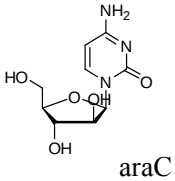
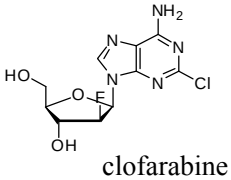
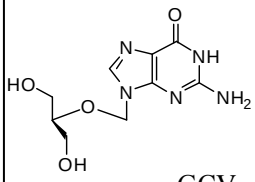
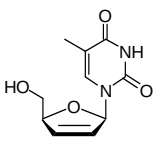
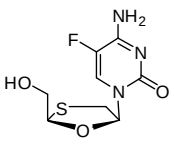
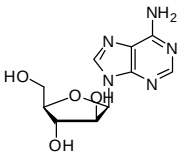


Figure 1-2: DNA double helix structure.

Table 1-1: Deoxyribonucleosides and NA prodrugs

Deoxy-nucleosides	 dU	 T	 dC	 dA	 dG
Nucleoside analog prodrugs	 IdU	 AZT	 araC	 clofarabine	 GCV
		 d4T	 FTC	 araA	

1.2.2 Challenges in NA drug development

AZT is a thymidine analog in which the 3'-hydroxyl group is replaced by an azido group. After being phosphorylated by cellular kinases into active triphosphate, AZT-triphosphate inhibits the HIV reverse transcriptase by competing with TTP (Yarchoan *et al.*, 1989). AZT has function as a chain terminator upon incorporating into DNA because the 3'-azido group can not form a phosphate bond with the subsequent 2'-deoxynucleoside. Another example is an analog of 2'-deoxyuridine, 2'-deoxy-5-iodo-uridine (IdU) that is used as an anti-herpes virus drug (Seth *et al.*, 2005). The IdU triphosphate can be incorporated into viral DNA and blocks further base pairing. Overall, it is clear that NAs provide a powerful therapeutics strategy. However, ironically the number of FDA-approved NA drugs has been limited. A major reason for the limited success has been the inability of NA prodrugs to become activated as NA triphosphates, a task carried on by cellular 2'-deoxynucleoside kinases (dNKs) and 2'-deoxynucleotide kinases which generally lack the efficiency for phosphoryl transfer to NAs. Strategies to overcome this obstacle will be necessary to advance the use of NAs as effective cancer and viral medications. One strategy I will discuss in this thesis is the possibility to develop NA kinases as gene therapy agents to accelerate NA phosphorylation in vivo.

1.3 The salvage pathway and phosphorylation problems for NAs

To function as gene replication terminator, NA prodrugs need to be sequentially transformed into the corresponding NA-triphosphates through three steps of phosphorylation. The phosphorylation is catalyzed by cellular nucleoside/nucleotide kinases, which are the core enzymes involved in the salvage pathway of cellular dNTP synthesis (Figure 1-3). These enzymes include 2'-deoxynucleoside kinases (dNK, e.g. thymidine kinase), 2'-deoxynucleoside monophosphate kinases (dNMPK, e.g. thymidylate kinase), and nucleoside diphosphate kinase (NDPK). The three enzymes work in a cascade manner to recycle scavenged 2'-deoxynucleosides and to activate intermediates of the *de novo* nucleotide biosynthetic pathway (Figure 1-1). Following cellular uptake via membrane nucleoside transporters, the prodrugs become activated by these kinases to reach their triphosphate state, the NA anabolites then turn into competitive substrates for low-fidelity polymerases and reverse transcriptases found in cancer cells and viruses, respectively. The incorporation of NA-triphosphates causes termination of the DNA replication, preventing further disease proliferation. On the contrary, the replication machinery of normal, healthy cells has higher fidelity, protecting the host from the lethal effects of these suicide substrates.

Unfortunately, human nucleoside/nucleotide kinases are typically inefficient at turning over NAs. For most of the nucleoside analogs studied, the first step of phosphorylation catalyzed by dNKs to form monophosphates is the rate-limiting step for

the entire process of triphosphates synthesis (Eriksson *et al.*, 2002). In general, the human dNKs phosphorylate NAs with noticeably lower efficiency than natural 2'-deoxyribonucleosides, which reduces the potency of existing prodrugs and results in the accumulation of cytotoxic reaction intermediates (Gentry, 1992; Arnér *et al.*, 1995; Eriksson *et al.*, 2002). Moreover, inability to phosphorylate the NAs leads to the failure of many NA prodrug candidates *in vivo* (Shi *et al.*, 1999), preventing clinically applications.

Previously, it was found that NA phosphorylation could be improved by co-administration of prodrugs with exogenous, broad-specificity kinases (Moolten, 1986; Culver *et al.*, 1992; Zhu *et al.*, 1998; Jordheim *et al.*, 2006; Solaroli *et al.*, 2007). While biochemical and pre-clinical experiments have demonstrated the effectiveness of the strategy in principle, the studies also uncovered limitations caused by dNKs' significantly lower activity for NAs compared with native substrates (Eriksson *et al.*, 2002). Moreover, the broad substrate specificity of exogenous kinases raised concerns as it interferes with the tightly regulated 2'-deoxynucleoside metabolism (Song *et al.*, 2005; Mathews *et al.*, 2007). Identifying orthogonal NA kinases through enzyme engineering hence provides promising solutions for the selective and efficient activation of NA prodrugs.

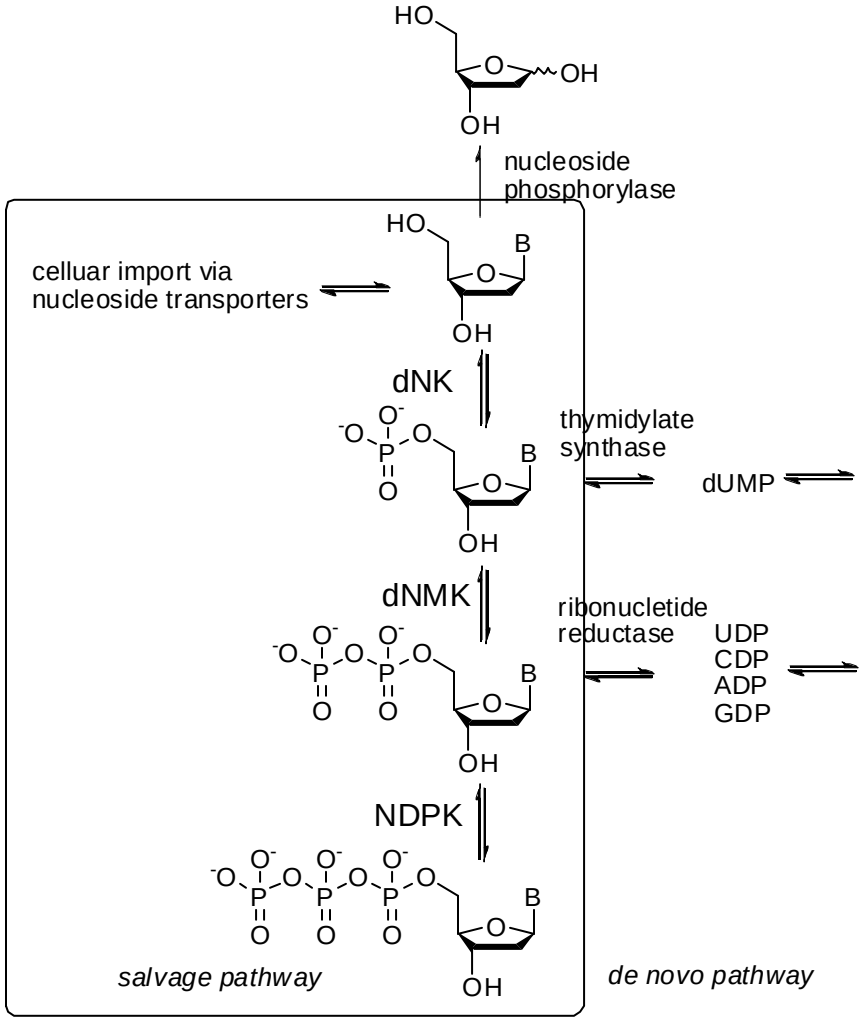


Figure 1-3: Nucleoside salvage pathway for the production of 2'-deoxynucleoside triphosphates (dNTPs). Scavanged 2'-deoxynucleosides are imported by membrane-bound nucleoside transporter proteins, and then serve as substrates for dNKs. Some nucleosides undergo cleavage by phosphorylase before going into other pyrimidine and purine metabolism pathways. Separately, thymidylate synthase and ribonucleotide reductase catalyze the final reactions of the nucleotide *de novo* pathways, producing TMP and dNDPs which are converted to their triphosphates by dNMPK and NDPK. dNK = 2'-deoxynucleoside kinase, dNMPK = 2'-deoxynucleoside monophosphate kinase, NDPK = nucleoside diphosphate kinase (adapted from (Lutz *et al.*, 2009)).

1.4 Cellular dNKs, exogenous kinases and substrate specificity

In order to tackle the cellular phosphylation problem for NAs, enzyme engineering of kinases in the salvage pathway has largely focused on the first phosphorylation step catalyzed by dNKs because it is often the rate limiting step (Figure 1-3). dNKs exhibit generally high specificity towards natural 2'-deoxyribonucleosides and possess distinct yet complementary preferences for the individual nucleobases (Figure 1-4). In humans, two of the dNKs (2'-deoxycytidine kinase (dCK) & thymidine kinase 1 (TK1)) are cytoplasmic kinases while thymidine kinase 2 (TK2) and 2'-deoxyguanosine kinase (dGK) are expressed in the mitochondria. Although all four enzymes have distinct substrate preferences, the two enzyme pairs complement each other to ensure phosphorylation of the four natural DNA building blocks. These enzymes are categorized into two subfamilies based on their crystal structures (Figure 1-4): type-I dNKs have a parallel β-sheet in the core, flanked by multiple helices on both sides and a short LID region over the active site; type-II dNKs possess a central parallel β-sheet, flanked by three helices. Additionally, an extended lasso region, which serves as lid over the active site, is held in place by a ligated zinc atom.

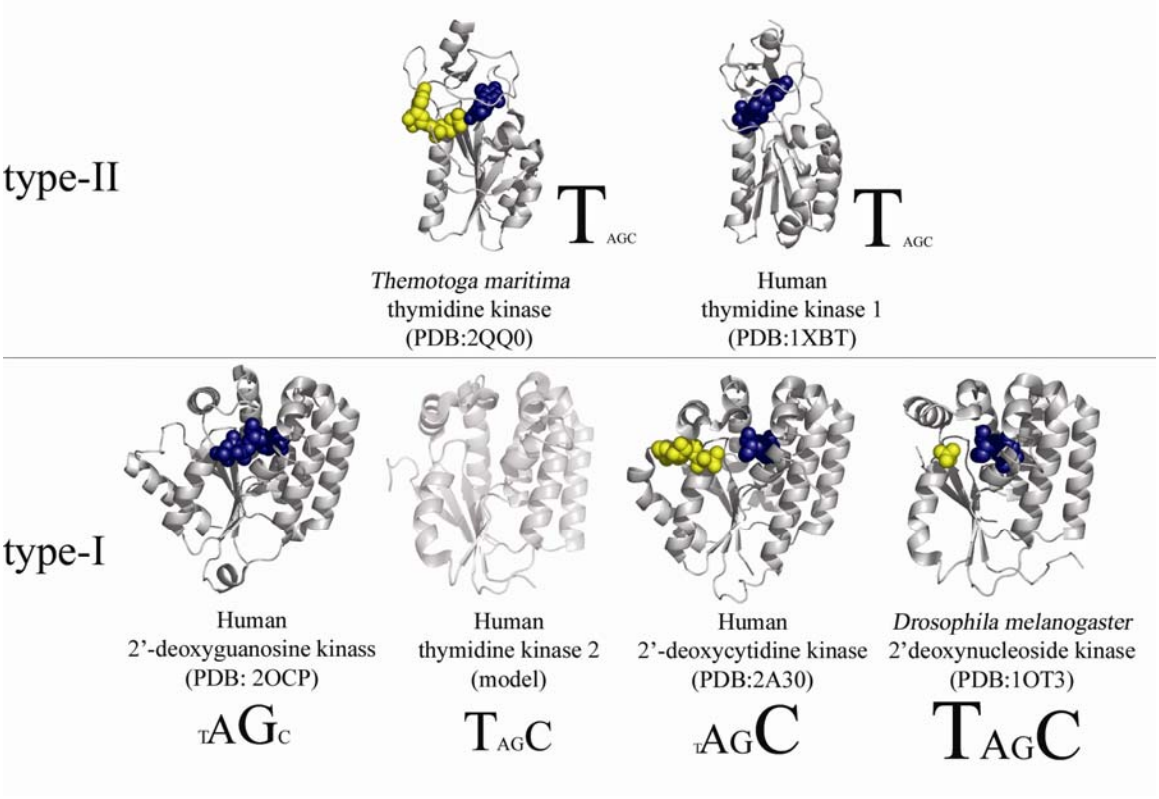


Figure 1-4: Summary of structure and substrate specificity of the 2'-deoxynucleoside kinase family. Members are divided into type-I and type-II subfamilies based on their distinct structural features. Representative protein structures, shown as grey ribbons, are based on crystal structure coordinates expect for the homology-based model of human TK2. Where available, substrate occupying the phosphoryl-acceptor binding site is shown as blue spheres while substrate binding in the phosphoryl-donor site is marked by yellow spheres. Substrate profiles are displayed beneath the structures, reflecting each enzyme's catalytic efficiency for the four natural 2'-deoxynucleosides (T, dA, dG, dC) by their font sizes. (adapted from (Lutz *et al.*, 2009))

In regards to NA activation, type-II dNKs play an important role in phosphorylating the two anti-HIV prodrugs 3'-azido-3'-deoxythymidine (AZT) and 2',3'-dideoxy-2',3'-didehydrothymidine (d4T), as well as various thymidine derivatives used for boron neutron-capture therapy (Al-Madhoun *et al.*, 2004; Welin *et al.*, 2004).

Nevertheless, the subfamily's strict specificity for thymine and uracil-nucleosides, as well as the substrate's direct interactions with the protein backbone, has complicated enzyme engineering and limited its potential therapeutic applications. Members of the type-I dNK subfamily on the contrary show great versatility in regards to substrate specificity, even though they all share a common protein scaffold (Arnér *et al.*, 1995) (Figure 1-2). For example, human dGK exclusively recognizes purine nucleosides as phosphoryl acceptors while TK2 phosphorylates only pyrimidine nucleosides. Even broader specificity can be found in dCK which activates 2'-deoxyribose carrying cytosine, adenine and guanine moieties while 2'-deoxynucleoside kinase from *Drosophila melanogaster* (*Dm*dNK) effectively phosphorylates all four natural 2'-deoxynucleosides, although with a preference for pyrimidines (Munch-Petersen *et al.*, 2000). The substrate adaptability of type-I dNKs has made members of this subfamily the target of many enzyme engineering studies.

1.5 Kinase engineering by directed evolution and the existing challenges

1.5.1 Directed evolution for enzyme engineering

Since natural nucleoside kinases generally prefer the natural 2'-deoxyribonycleosides, enzyme engineering by rational design and directed evolution presents a promising solution to seek orthogonal NA kinases. Directed evolution in particular has provided scientists and engineers with a powerful tool to understand how proteins and enzymes work, as well as to tailor substrate specificity of biocatalysts to novel substrates and

function (Bernhardt *et al.*, 2009; Gerlt *et al.*, 2009; Tao *et al.*, 2009; Tracewell *et al.*, 2009). The process is a lab-based procedure that mimics Darwinian evolution, applying iterative cycles of diversification and selection to stepwise modify an existing protein towards progeny with the desired properties. In practice, one or more "parental" genes that encode the target proteins are forced to diversify through saturation and/or random mutagenesis, as well as in vitro recombination methods, creating gene libraries of typically 10^3-10^{10} members (Lutz *et al.*, 2004). Following overexpression in a suitable host organism or *in vitro*, the corresponding protein libraries are evaluated for the desired property (e.g. substrate specificity, thermostability *etc.*) by selection or screening assays (Lutz *et al.*, 2004). One of the most common selection strategies is genetic complementation of auxotrophic host strains which can process libraries of 10^6- 10^7 members. Many other screening protocols are based on LB-agar or micro-titer plate assays and can typically be used to analyze up to 10^4 library members. Recently, fluorescence-activated cell sorting (FACS) and *in vitro* compartmentalization were used in library screening for up to 10^8 members (Farinas, 2006; Griffiths *et al.*, 2006). Candidates that excel under the selection/screening conditions are isolated and can either serve as "parents" for the subsequent generation of library evolution or undergo detailed protein characterization.

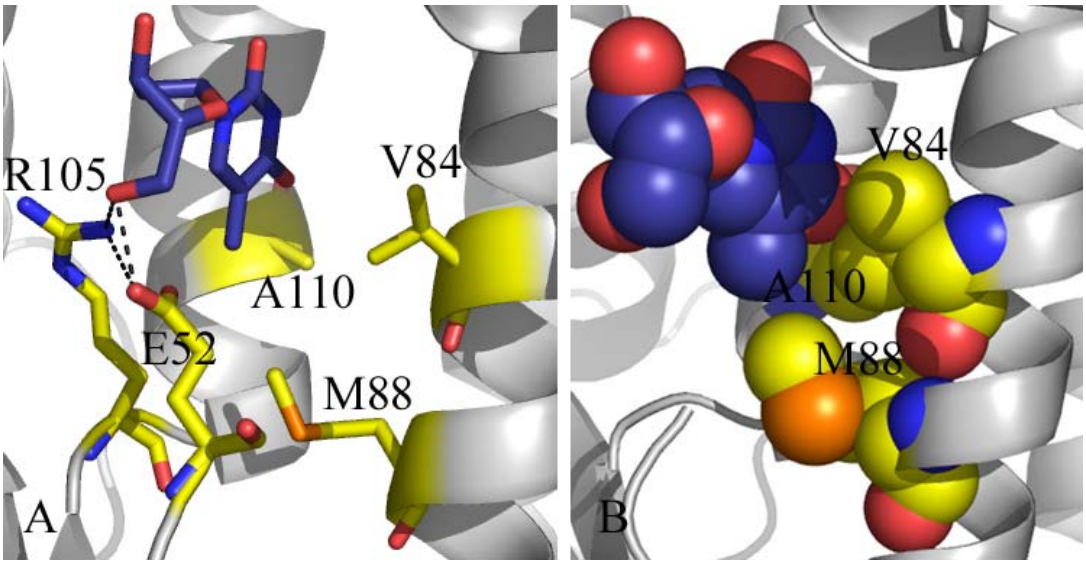


Figure 1-5: Structure of thymidine-bound 2'-deoxynucleoside kinase from *Drosophila melanogaster* (PDB code: 1OT3 (Galmarini *et al.*, 2003; Mikkelsen *et al.*, 2003; Parker, 2009)): A) Arg105 and Glu52 are highly conserved residues across the type-I kinase family that form critical hydrogen bonding interactions with the 5'-hydroxyl group of the substrate thymidine. B) Val84, Met88 and Ala110 form a hydrophobic binding pocket for the nucleobase.

1.5.2 Progress toward NA kinase engineering

Various nucleoside kinases have been engineered to improve NA substrate specificity and catalytic efficiency. The diverse substrate profiles of type-I kinase family members raised the question of whether nucleobase recognition could be tailored by altering the active site binding pocket. Earlier random mutagenesis experiments with *Dm*dNK, as well as crystallographic studies of hdCK had identified residues in similar positions, namely Val84/Met88/Ala110 (Ala100/Arg104/Asp133 in hdCK) as critical for nucleobase specificity (Knecht *et al.*, 2002; Sabini *et al.*, 2003)(Figure 1-5 and 1-6). The three residues Val84/Met88/Ala110 in *Dm*dNK form a large, generic hydrophobic

binding pocket, capable of accommodating both pyrimidine and purine bases. Indeed, mutagenesis at these three positions could alter the enzyme's specificity between purines and pyrimidines (Knecht *et al.*, 2002). In contrast, Arg104/Asp133 in hdCK are involved in an elaborate hydrogen-bonding network in the active site, allowing for favorable binding interaction between the exocyclic amino group of the substrate and Asp133, a contact that would be lost upon binding of thymidine. Furthermore, the network also preorients the Arg104 side chain in a conformation which creates a steric clash with the 5-methyl group of thymine. hdCK variants with thymidine activity were obtained by rationally varying these two residues (Iyidogan *et al.*, 2008). Interestingly, a separate study using random mutagenesis found similar variants with key mutations at the two residues. Two most frequent mutants Arg104Met/Asp133Ser and Arg104Met/Asp133Thr revealed a ~4000-fold enhanced thymidine kinase activity, and the substrate specificity profile became more *Dm*dNK-like.

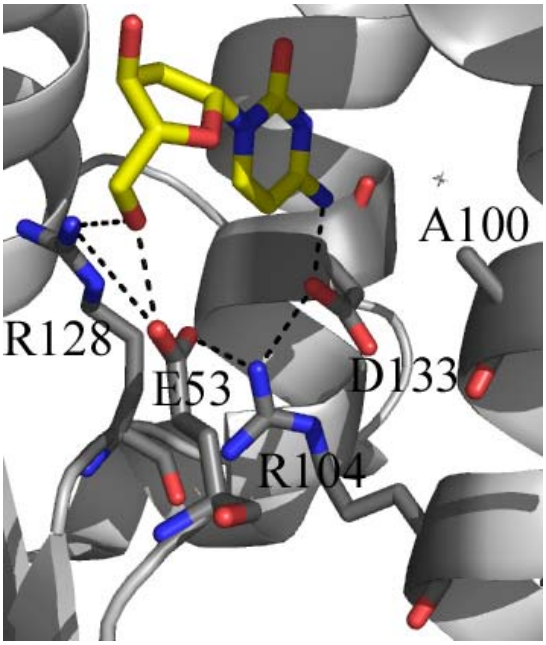


Figure 1-6: Active site of human 2'-deoxycytidine kinase (PDB code: 2A30) (Godsey *et al.*, 2006). The enzyme carrys 2'-deoxycytidine in the phosphoryl acceptor binding site. Similar to *Dm*dNK, Arg128 and Glu53 form critical hydrogen bonding interactions with the 5'-hydroxyl group of the substrate. The hydrogen bonding network is also extended to include Arg104 and Asp133 which in turn interacts with the exocyclic amino group on the substrate's nucleobase.

1.5.3 Library generation strategies used for kinase engineering

To make potent diverse libraries, random mutagenesis, DNA shuffling and non-homologous recombination strategy (SCRATCHY) all have proven useful. Random mutagenesis by error-prone PCR is most commonly used to generate diversities at random positions throughout the gene sequence. For example, variants of HSV1-TK have been found with decreased K_M values (up to 126-fold) for ganciclovir and acyclovir from a random mutagenesis library (Kokoris *et al.*, 2002). DNA shuffling can conveniently shuffle gene sequences together based on sequence homology. A library composed of HSV-1/HSV-2 TK chimeras was created by DNA shuffling, and chimera proteins were identified with 6 to 44-fold higher specificity toward AZT compared to parent enzymes (Christians *et al.*, 1999). Furthermore, the non-homologous recombination technique SCRATCHY was used to create a chimera library based on *Dm*dNK and hTK2 and variants were found with higher activity for ddT, ddC and d4T than the parent enzymes (Gerth *et al.*, 2007).

1.5.4 Biases in auxotrophic complementation selection

One reoccurring issue in dNK engineering has been the restrictions posed by using the thymidine kinase-deficient *E. coli* strain KY895. The auxotrophy strain is deficient in natural thymidine kinase activity and can be used to perform auxotrophic complementation experiments on kinase libraries as a selection for thymidine phosphorylation (Igarashi *et al.*, 1967). When *E. coli* KY895 is grown in the presence of

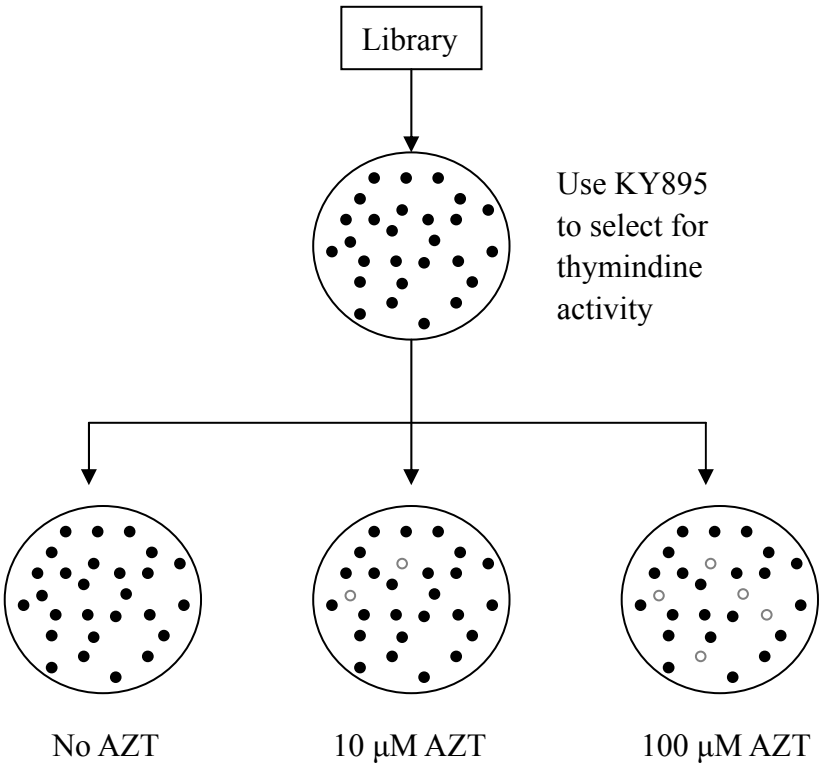


Figure 1-7: Selection protocol for AZT activity.

5-fluoro-2'-deoxyuridine (FdU) and uridine, the phosphorylated FdU inhibits thymidylate synthase, shutting down the *de novo* biosynthetic pathway for thymidine triphosphate (Figure 1-3). In parallel, uridine will inhibit thymidine phosphatase, preventing recycling of thymidine metabolites. Hence, the *E. coli* is solely dependent on the nucleoside salvage pathway, relying on intake of thymidine and its subsequent phosphorylation by an exogenous thymidine kinase (provided on the plasmid DNA), together with its own dNMPKs and NDPKs to facilitate growth and replication. In directed protein evolution, the golden rule has been "you get what you select for" (You *et al.*, 1996; Zhao *et al.*, 1997). Library selection based on *E. coli* auxotroph KY895 would enrich library members with thymidine kinase activity, excluding possibility to select orthogonal NA

kinase activity. An alternative is *in vivo* screening on replica plates that counts on cytotoxicity of NAs (Christians *et al.*, 1999). A previous selection strategy used for AZT phosphorylation is described in Figure 1-7. The library of interest was transformed into *E. coli* auxotroph KY895 (TK^-) to select for thymidine activity. Surviving cells were then grown on selection plates with increasing concentrations of AZT. Finally, phosphorylation of AZT would result in accumulation of toxic analogs, leading to cell death. Therefore, the cells which disappear on AZT selection plates potentially carry active AZT kinases that retain thymidine activity. There are two major disadvantages with this selection protocol: a) replica plating, while directly evaluating library members for NA activation, is very laborious and ultimately depends on the toxicity of the phosphorylated NA to the host, a criterion that in our tests in *E. coli* could not be extended beyond AZT; b) pre-selection for thymidine kinase activity leads to variants with broadened substrate activity and excludes orthogonal NA kinase variants. This is a fatal flaw of the technique as it directly excludes orthogonal kinases with high specificity for NAs. Orthogonal nucleoside analog kinases are ideal for gene therapy since they eliminate competition in vivo against natural nucleosides and avoid interfering with nucleoside metabolism pathways. In light of these biases and functional deficiencies of existing assays, we sought the design and implementation of a selection or screening technique for orthogonal NA kinases.

1.6 Fluorescent nucleoside analogs

However, the challenge exists in monitoring and designing screening protocols for phosphorylation reactions, which only adds a phosphate group to the 5'-OH of the NA. One result of phosphorylation is that the negatively charged monophosphates of NAs are trapped inside the cells while the non-phosphoryated NAs can be transported through cell membrane with high efficiency via promiscuous membrane-bound nucleoside transporters. How could the entrapment be monitored?

Previously, fluorescent nucleoside analogs have been developed for analyzing the interaction of DNA/RNA with corresponding binding proteins (Jean *et al.*, 2001; Liu *et al.*, 2002; Hawkins, 2007; Sandin *et al.*, 2007; Yang *et al.*, 2007). Perhaps these fluorescent NAs may provide a spectroscopic signature for following the entrapment of phosphorylated NAs. Chemical structures of selected fluorescent NAs are shown in Figure 1-8. 6-Methyl-3-(*β*-D-2-deoxyribofuranosyl)furano[2,3-*d*]pyrimidin-2-one (Furano-T, **6**) is a fluorescent thymidine analog with a maximum excitation/emission wavelengths of 331nm 431nm, respectively, and a quantum yield of ~0.3. While 6-methyl-3-(2-deoxy-β-D-ribofuranosyl)-3H-pyrrolo[2,3-*d*]pyrimidin-2-one) (Pyrrolo-dC, **7**) is a fluorescent deoxycytidine analog with maximum excitation/emission wavelength at 345nm and 470nm with quantum yield around 0.2 (Berry *et al.*, 2004). Additionally, pteridines are highly fluorescent compounds that were first isolated from butterfly wings in 1889. Various pteridines have been effectively used to replace purine bases on nucleosides to make purine-like fluorescent probes (Hawkins *et al.*, 1997; Hawkins,

2007). For example, the deoxyguanosine analog 6-methyl isoxanthopterin (6MI, **10**) has excitation/emission maxima at 340nm/431nm with relative quantum yield around 0.70. In comparison, the relative quantum yields for 4-amino-*6*-methyl-8-(2-deoxy-β-D-ribofuranosyl)-7(8H)-pteridone (6MAP, **9**), a deoxyadenosine analog, is 0.39 with excitation maxima of 330 nm respectively and emission maxima at 430 nm (Hawkins *et al.*, 1997). If these fluorescent bases are used on NAs, fluorescence can be used as the signal to follow entrapment of phosphorylated NAs, enabling development of a fluorescence-based screening method for evolving NA kinases. Furthermore, relatively small nucleobase modifications can avoid cellular background fluorescence by red-shift absorbance, while minimizing the perturbation on phosphorylation by nucleoside kinases.

The hypothesis in this thesis is these fluorescent bases can be combined with critical modifications of the 2'-deoxyribose (Figure 1-8) that lead to chain termination, resulting in fluorescent NAs that can be used in directed evolution to directly observe and select for the cellular entrapment of fluorescent NA. The degree of fluorescence can then be directly correlated with NA kinase activity.

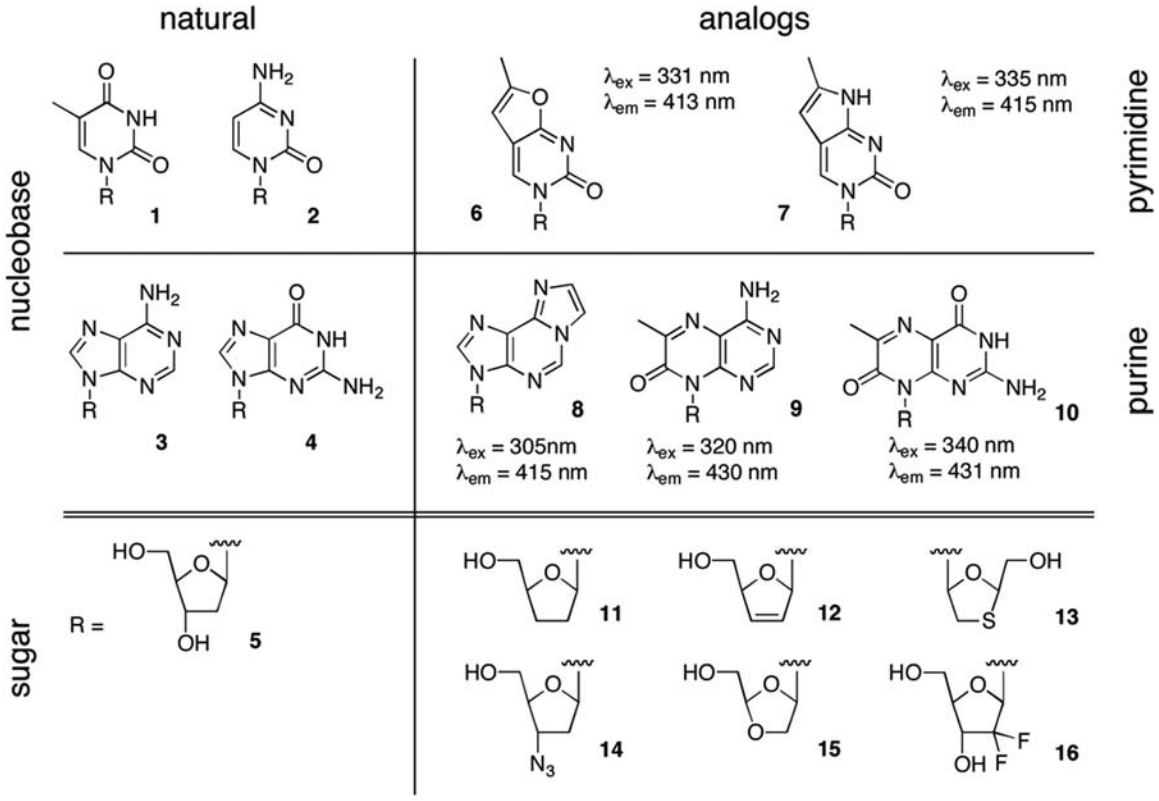


Figure 1-8: Overview of modular nucleoside design containing the base moiety and the sugar moiety. Many NAs are composed of a natural pyrimidine (**1,2**) or purine (**3,4**) moiety docked to a sugar analog (**11–16**) instead of the native 2'-deoxyribosyl portion (**5**). For the creation of fluorescent NAs, an extension of the heterocycles, as exemplified in furano and pyrrolo-pyrimidines (**6,7**), as well as etheno-adenine (**8**) and pterines (**9,10**), red-shifts the excitation/emission spectrum of the nucleobase and increases the quantum yield. When coupled to sugar analogs (**11–16**), these fluorescent versions of the NA become useful molecular probes for studying their cellular uptake and metabolism (Zhang *et al.*, 2006; Parker, 2009) and could serve as substrates to facilitate screening for phosphorylation activity in directed evolution of NA kinases. (Adapted from (Liu *et al.*, 2009))

1.7 FACS-based screening for directed evolution

Fluorescence-activated cell sorting (FACS) will be used to screen for cells with accumulation of fluorescent NAs. FACS is a specialized type of flow cytometry which can sort mixtures of biological cells or other particles (microbeads, emulsions) into two or more groups based on fluorescence intensity. Sample cells enter a FACS sorter one cell at time and are subsequently illuminated by a focused laser beam. The measuring optics quantitatively record specific light scattering and fluorescent properties of each individual cell. A charge can be applied to cells of interest; this deflects them into a collection tube through the application of an electrostatic field. It provides fast (Modern FACS equipment can routinely sort $>10^7$ events per hour.) and accurate separation of cells with desired characteristics (Givan, 1992; Shapiro, 2004). FACS has quickly become an important tool for protein engineering due to its high-throughput nature and capability of separating cells carrying different fluorophores.

For directed evolution, it is critical to connect phenotype with genotype so the identified variants could be traced back to their gene sequences. There are four ways of coupling the genotype and phenotype together for FACS (Yang *et al.*, 2009): 1) cell surface display of active enzymes, 2) reporting of enzyme activity with GFP or its variants, 3) entrapment of the product within the cells, 4) in vitro compartmentalization. Aharoni *et al.* reported the high-throughput screening of glycosyltransferase mutants using FACS (Aharoni *et al.*, 2006). The principle of screening is built on the fact that

sialic acid is the only charged sugar under physiological conditions. While a neutral glycan acceptor with a fluorophore label can diffuse in and out of cells rather freely, the acceptor with negatively charged sialic acid transferred to it, will accumulate inside the cell since charged species can not easily pass through cell membrane. Therefore, cells with fluorescent product accumulation could be sorted by FACS, allows screening of glycosyltransferases activity. In a separate study, Olsen *et al.* used synthetic FRET pair labeled peptide for evolving proteases with desired substrate specificity (Olsen *et al.*, 2000). A fluorescent dye was labeled on positively charged peptide sequence to form a FRET pair. Cleavage of the scissile bond would release the FRET pair quencher group, and the fluorescent labeled peptide would remain on the bacteria surface, which could be sorted by FACS. Using desired the FACS-based strategy, researchers were able to successfully improve serine protease OmpT and change the enzyme's substrate preference.

In comparison, a similar strategy can exploit a convenient nucleoside feature that kinase catalysis would lead to phosphorylated product entrapment in the cellular compartment while the unphosphorylated substrates can freely transport in and out of cell membrane via high efficiency membrane-bound nucleoside transporters (Zhang *et al.*, 2006). Using fNAs as substrates for directed evolution of NA kinases, diverse libraries could be analyzed by FACS to directly monitor NA phosphorylation in individual cells

based on accumulation of the fluorophore inside host cells. The high fluorescence cells can be isolated and kinase candidates may then be characterized for NA catalysis.

1.8 Combining computational design with directed evolution for enzyme engineering

Among various ways for enzyme engineering (e.g. rational design, directed evolution and computational design), each method alone shows its own advantages and shortcomings. Rational design is most direct and requires the least effort but is highly dependent on scarce sequence-structure-function relationship knowledge of the target protein. On the other hand, directed evolution is complementary to rational design since it selects for desired function from a library of proteins without the need for complete understanding of the parent enzymes. However, directed evolution can be relatively time-consuming and is also limited by current library construction technology as well as screening/selection strategies. Computational design provides the digital version of directed evolution through which creation of diversity and selecting promising candidates are all done in computational program in much shorter time period. However, current computational programs which could be used for protein catalyst design are still on the way of improving modeling accuracy and reliability. Since each strategy alone has limitations, the combination of different strategies might offer opportunities for more efficient and productive enzyme engineering.

Rosetta is one of the leading computational tools for protein structure prediction

which has proved its powerful capability in designing new catalytic activity. Recently, the program was successfully used in the de novo design of protein catalysts for retro-aldol reaction on non-natural substrates (Jiang *et al.*, 2008; Rothlisberger *et al.*, 2008). Furthermore, Rosetta was combined with directed evolution by setting up a sound starting point for further modification by directed evolution (Rothlisberger *et al.*, 2008). The computational program first designed eight enzymes that use two different catalytic motifs to catalyze the Kemp elimination, then in vitro evolution was applied to enhance the computational designs, achieving a >200-fold increase in catalytical efficiency. Recently, Rosetta was also successfully used in altering substrate specificity of a human guanine deaminase (Murphy *et al.*, 2009). It raises the question that whether the computational approach could provide an alternative route to design orthogonal NA kinases. The *in silico* approach enables a more thorough search of sequence space, potentially identifying novel solutions missed due to experimental limitations in library size. In addition, it provides a more quantitative and predictive framework for protein engineers to explore questions of biocatalyst stability and substrate specificity. Furthermore, combining Rosetta design with the FACS-based screening could minimize limitations and opens up possibilities for faster NA kinase engineering with high specificity and efficiency.

Chapter 2. Development of FACS-based screening for kinase phosphorylation

(published, Liu *et al*., *Nucl. Acids Res.* 2009, 37, 4472-4481)

2.1 Introduction

Nucleoside analogs (NAs) form a major group of drugs used for therapy of existing and newly emerging viral infections and cancer (Crumpacker, 1996; Parker, 2009). Upon transporting NAs through cell membrane by relatively promiscuous nucleoside transporters, NAs undergo intracellular activation by nucleoside and nucleotide kinases into triphosphate forms. Upon activation, drugs can function as DNA chain extenstion/replication terminator. However, endogenous deoxynucleoside kinases often lack the capability of phosphorylating NAs with high efficiency. This inefficiency not only limits the potency of existing prodrugs, but also is a main reason for many new NAs' failure to work in vivo. Previously, it was found that introducing an exogenous 2'-deoxyribonucleoside kinase (dNK) can significantly enhance the potency of a NA prodrug (Moolten, 1986; Culver *et al.*, 1992; Jordheim *et al.*, 2006; Solaroli *et al.*, 2007). Therefore, development of dNKs with adequate activity for NAs is a major strategy to improve the drug efficiency for gene therapy. Unfortunately, most of the natural dNKs show significantly lower activity for NAs compared to natural substrates. Many groups have attempted directed evolution of natural nucleoside kinases using site saturation, and random mutagenesis, as well as homologous recombination and chimeragenesis to evolve

kinase variants with the desired NA activity (Christians *et al.*, 1999; Gerth *et al.*, 2007; Iyidogan *et al.*, 2008). However, these efforts have been hampered by the inability to directly screen or select library members for NA phosphorylation.

During the past two decades, the two techniques commonly used for dNK library analysis have been: a) in vivo selection, using genetic complementation of the auxotrophic *E. coli* strain KY895 (Igarashi *et al.*, 1967), or b) in vivo screening on replica plates, testing for cytotoxicity of NAs (Christians *et al.*, 1999). The auxotrophic selection protocol evolves variants with NA kinase function while maintaining thymidine kinase activity. However, if used in gene therapy, introduction of this residual kinase activity would likely interference with cellular dNTP metabolism (Mathews *et al.*, 2007). In contrast, orthogonal NA kinases would minimize these effects and be the ideal candidates for anti-viral/cancer therapy. Unfortunately, the selection based on thymidine kinase deficient *E. coli* strain excludes these desired library variants. Alternatively, replica plating is used for directly selections of NA phosphorylation activity. The technique is not only time-consuming and laborious, but also depends on the cytotoxicity of the phosphorylated NAs, a criterion that in our tests is limited to 3'-azido-3'-deoxythymidine (AZT). Therefore, development of selection/screening methods that directly evaluates library members for NA phosphorylation function will advance the therapeutic potential of a wide range of NAs.

Interestingly, while nucleosides and most NAs can be efficiently transported in and out the cellular environment with the help of promiscuous membrane nucleoside transporters (Zhang *et al.*, 2006), while the phosphorylated products remain entrapped in the host cells because the added negative charge prevents the export of the monophosphates. A screening protocol to follow phosphorylation could be built on the entrapment effect if detectable signal from nucleosides can be monitored. Even though natural nucleobases show intrinsic fluorescence at physiological conditions, direct measurements are impractical due to the compounds' low quantum yields and overlapping absorption with aromatic amino acids and small-molecule metabolites (e.g. ATP and NADH). Nucleosides with small synthetic modifications on the pyrimidine or purine moiety can red-shift fluorescent properties and increase fluorophore quantum yield, allowing detection of entrapped phosphorylated NAs, making suitable substrate candidates for screening NA kinase activity.

A number of fluorescent 2'-deoxyribonucleosides, used for DNA structure studies, have been reported including the *etheno*-derivatives of adenine (Barrio *et al.*, 1972; Secrist III *et al.*, 1972; Boryski *et al.*, 1988) and cytosine, tricyclic guanine, as well as furano and pyrrolo-pyrimidines and pterines (Bergstrom *et al.*, 1982; Hawkins *et al.*, 1997; Berry *et al.*, 2004) (Figure 1-8). In this chapter, the furano-pyrimidine (fT, Figure 2-1) is substituted for thymine base to evaluate the concept of using fluorescent NAs to follow nucleoside kinase-catalyzed phosphorylation.

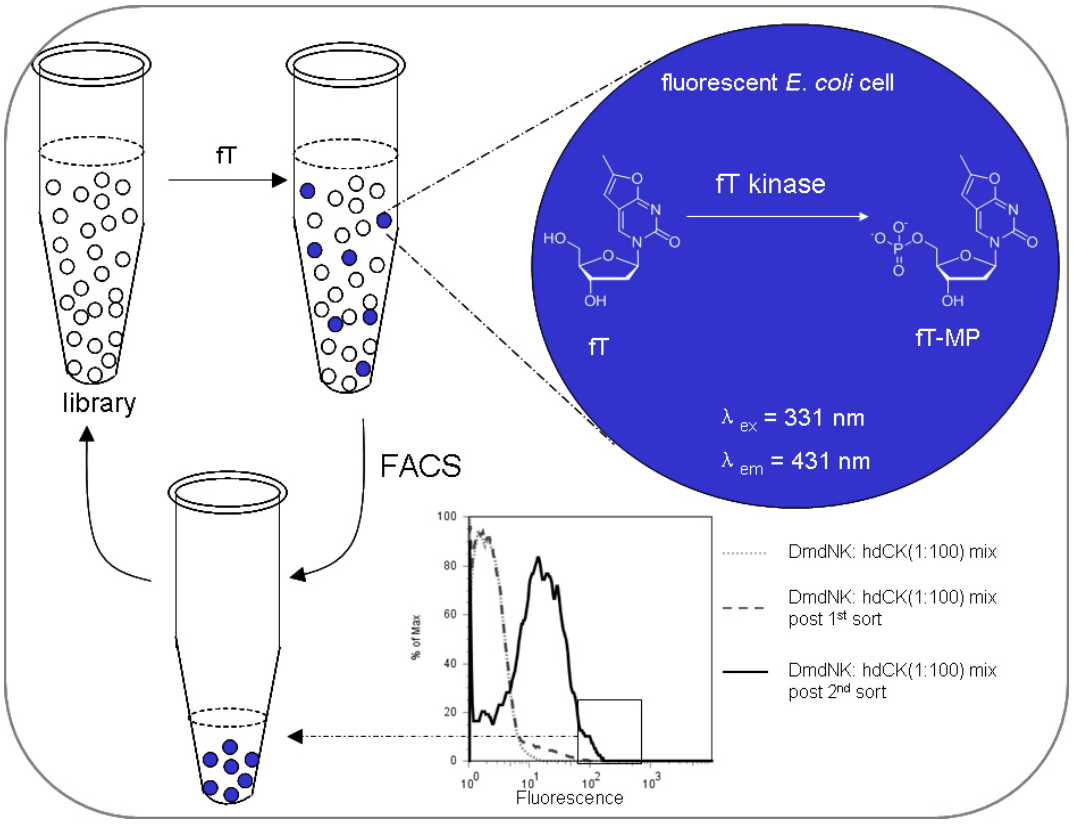


Figure 2-1: Scheme of enriching cells expressing *Dm*dNK from the mixture of cells expression *Dm*dNK (an enzyme with high thymidine/fT activity) and hdCK (an enzyme with little thymidine/fT activity) using fluorescence-activated cell sorting (FACS). Blue circles represent *E. coli* cells with accumulation of fT-monophosphate. Open circles represent cells with few fT-monophosphate entrapped, thus show low fluorescence.

The concept of using fluorescence-activated cell sorting (FACS) to screen for fT phosphorylation is explained in Figure 2-1. When bacteria, expressing thymidine kinase, are incubated with fT, broad-specificity nucleoside transporters facilitate the efficient translocation of the neutral fT compound across the membrane. In the presence of a functional thymidine kinase, the product of the enzymatic reaction, fT-monophosphate,

will accumulate inside the cell as its negative charge prevents exportation. FACS can subsequently be employed to isolate hosts with high fluorescence intensity.

2.2 Results and discussion

Focusing on thymidine analogs, we determined the furano-pyrimidine derivatives to be advantageous for detection and development of NA kinases. The relatively small modification of the nucleobase minimizes potential of selecting variants that might be ineffective against NA containing natural bases. More importantly, the sugar moiety is unchanged, as this is the critical functional portion of most NAs. The fluorescent analog of thymidine fT is commercially available. Initial experiments to implement and validate the concept of cellular fNA entrapment by dNKs were performed with fT and wild type 2'-deoxynucleoside kinase from *Drosophila melanogaster* (*Dm*dNK) which possesses high thymidine kinase activity. As a negative control, we used human 2'-deoxycytidine kinase (hdCK) which shows no detectable thymidine kinase activity. Both enzymes can easily be overexpressed in *E. coli*.

2.2.1 Fluorescence property of fT

Fluorescence excitation/emission spectra of fT were obtained (Figure 2-2). The maximum excitation/emission wavelengths are 331nm/431nm, respectively, and the quantum yield of the fluorophore is ~0.3, making it suitable for flow cytometry analysis.

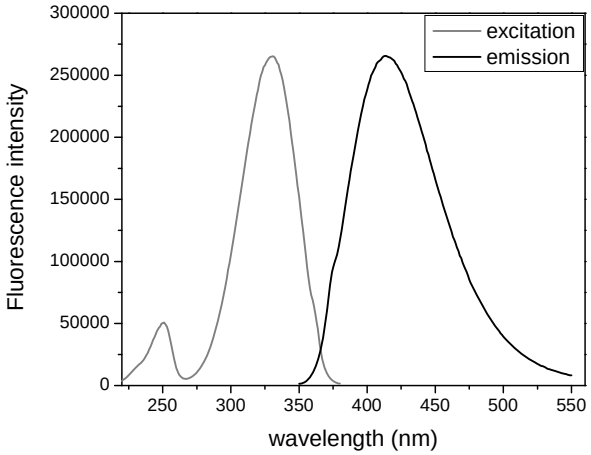


Figure 2-2: Fluorescence spectra of fT. The excitation/emission maximum are 331/431 nm.

2.2.2 Effect of furano-modification on thymine base on fT as substrates for *Dm*dNK

It is critical to learn if fluorescent nucleosides act as similar substrates for nucleoside kinases as their parent nucleosides. To evaluate the functional similarity, catalytic performance of *Dm*dNK for native substrates versus furano-modified nucleosides was determined. As shown in Table 2-1, *Dm*dNK's catalytic efficiency for fT is only 2-fold lower than T. The lower k_{cat}/K_M is mostly due to lower k_{cat} value, while no significant K_M change was caused by the base modification in fT. Since the difference in kinetics is not significant, the fluorescent nucleosides behave similarly to their non-fluorescent parent nucleosides as substrates for *Dm*dNK, and they could be used as suitable substrates for NA kinase evolution.

Table 2-1: Kinetic data of *Dm*dNK for T/fT at 25°C.

substrates	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($10^3 s^{-1} \cdot M^{-1}$)
T	3.2 ± 0.1	2.0 ± 0.1	1600 ± 130
fT	1.8 ± 0.1	2.0 ± 0.1	895 ± 70

2.2.3 Toxicity study for fT

If fT is lethal to the bacteria host, as more efficient kinases accumulate higher amounts of NA in the cytoplasm, would result in host death and loss of the most efficient candidates over the course of directed evolution. With this concern, toxicity of fT was determined on *E. coli* expressing *Dm*dNK by measuring the survival rate as a function of fT concentration. The result shows that up to 100 μM fT, no toxicity is observed, and when 200 μM fT is present, 50% of the cells die. Therefore, fT concentrations between 5-80 μM were used in directed evolution experiments.

2.2.4 Evaluation of different promoters for kinase expression in *E. coli* in the presence of fluorescent nucleosides

E. coli bacteria was chosen as the biological host for the screening since its broad-specificity nucleoside transporters can facilitate efficient trans-membrane equilibration of fNAs while absence of an endogenous type-I dNK minimizes the risk of false positives. Additionally, most of dNKs are easily expressed in *E. coli* with satisfying

yield. Therefore, *E. coli* was used as the host for this fluorescence nucleoside analog-based screening protocol.

An ideal expression system for the screening should have a tunable promoter so protein expression levels can be adjusted for applying appropriate selection pressure for directed evolution. Equally important, the host cells should have homogeneous expression of kinases, to avoid variability in cellular kinase level that could result in false negatives and positives. Hence, we evaluated a number of DNA plasmid and promoters, including pDIM vector with lac promoter, pET vector with T7 promoter and pBAD vector with P_{BAD} promoter. In Figure 2-3, the accumulation of fT-monophosphate upon overexpression of *Dm*dNK (black line; positive control) and hdCK (grey line; negative control) was analyzed by flow cytometry and compared for the three systems.

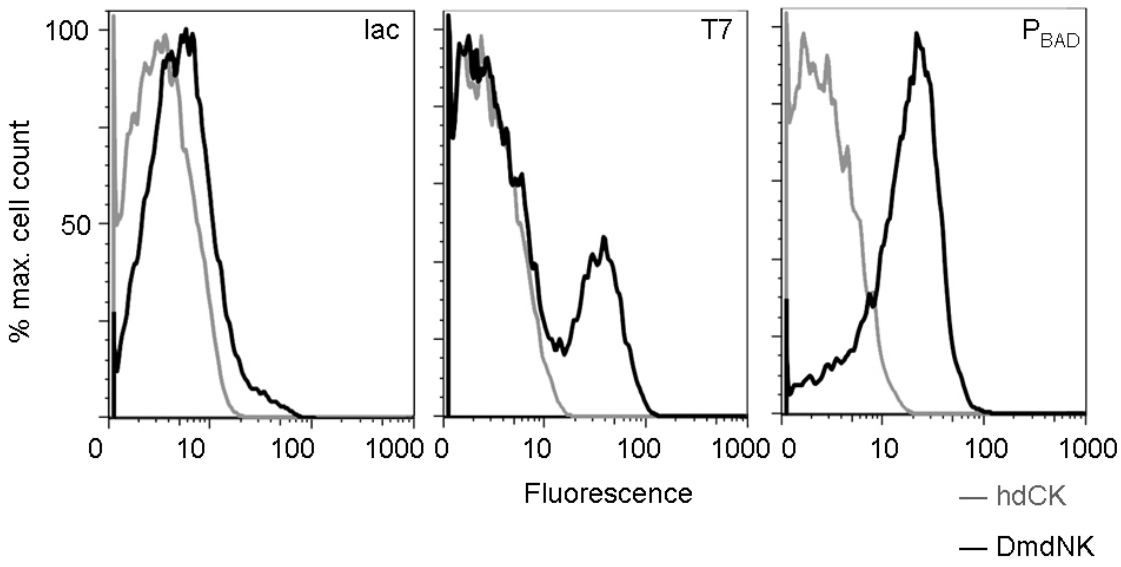


Figure 2-3: Comparison of three different vector/promoter systems in expressing nucleoside kinases and trapping the compound fT. (Adapted from (Liu *et al.*, 2009))

After 4 h of 1 mM IPTG induction, the *lac* promoter (pDIM (3) in *E. coli* KY895 (4)) shows only a small difference in signal, suggesting low expression levels of the target kinase. Under the same conditions, the split fluorescence intensity profile with the T7 promoter (pET-14b in *E. coli* BL21(DE3)) indicates heterogeneous (all-or-none) kinase production. All-or-none expression refers to the heterogeneous population of cells in which some are fully induced and others are poorly induced. The "on" phenotype is a result of inducer importers being turned on when a cell is exposed to the inducer, resulting in increased uptake of the inducer (Novick *et al.*, 1957). In contrast, protein expression under control of the P_{BAD} promoter (4 h induction with 0.2% arabinose; pBAD-HisA in *E. coli* TOP10) results in distinct populations with a 20 to 30-fold difference in fluorescence intensity for the positive and negative controls. The

homogeneous expression is achieved by deletion of the genes for arabinose transport (*araE* and *araFGH*) and arabinose degradation (*araBAD*) from the *E. coli* genome, resulting in tight repression in the absence of inducer and a 1000-fold induction range (Khlebnikov *et al.*, 2001; Khlebnikov *et al.*, 2002). Due to the homogeneous kinase expression observed with pBAD in combination with tunable induction by arabinose, all experiments for screening of the directed evolution libraries were performed using this expression system.

2.2.5 Method evaluation: enrichment experiments on fT phosphorylation

To evaluate the reliability of this FACS-based screening protocol, a simulated library made of known ratio of two mixed population: cells expressing *Dm*dNK and cells expressing hdCK was sorted using fT as the substrate (Figure 2-1). Since the catalytic efficiency of *Dm*dNK for thymidine is ~10^5 higher than activity of hdCK (Table 2-2), the *Dm*dNK gene should be enriched by sorting for the high fluorescent cell populations if the screening protocol is accurate and reliable.

Table 2-2: Kinetic data of hdCK and *Dm*dNK using Thymidine as substrate.

	hdCK	*Dm*dNK
k_{cat} / K_M ($s^{-1}M^{-1}$)	2×10^2	1.2×10^7

To evaluate FACS sorting capabilities, *E. coli* strain TOP10 transformed with pBAD vectors carrying hdCK and *Dm*dNK genes were prepared for protein expression and then mixed at three different ratios (*Dm*dNK : hdCK = 1:10, 1:100, 1:1000) before being exposed to fT, as described in the Methods. The mixed cells were sequentially sorted three times to enrich the high fluorescence cell population (Figure 2-4a), then the final sorted cells were recovered on LB plates. Finally, colony PCR was used to determine the post-sort ratio of the two genes based on the size difference of the two PCR products (hdCK: 783 bp; *Dm*dNK: 675 bp). To our satisfaction, FACS sorting of the pre-mixed *Dm*dNK + hdCK pools enriched *Dm*dNK gene by >100-fold (Figure 2-4b), indicating the FACS-based screening is suitable for enriching NA kinases by directly monitoring cellular NA phosphorylation. The newly developed screening strategy was named FLUPS for FLUorescent nucleoside analog Phosphorylation Screen.

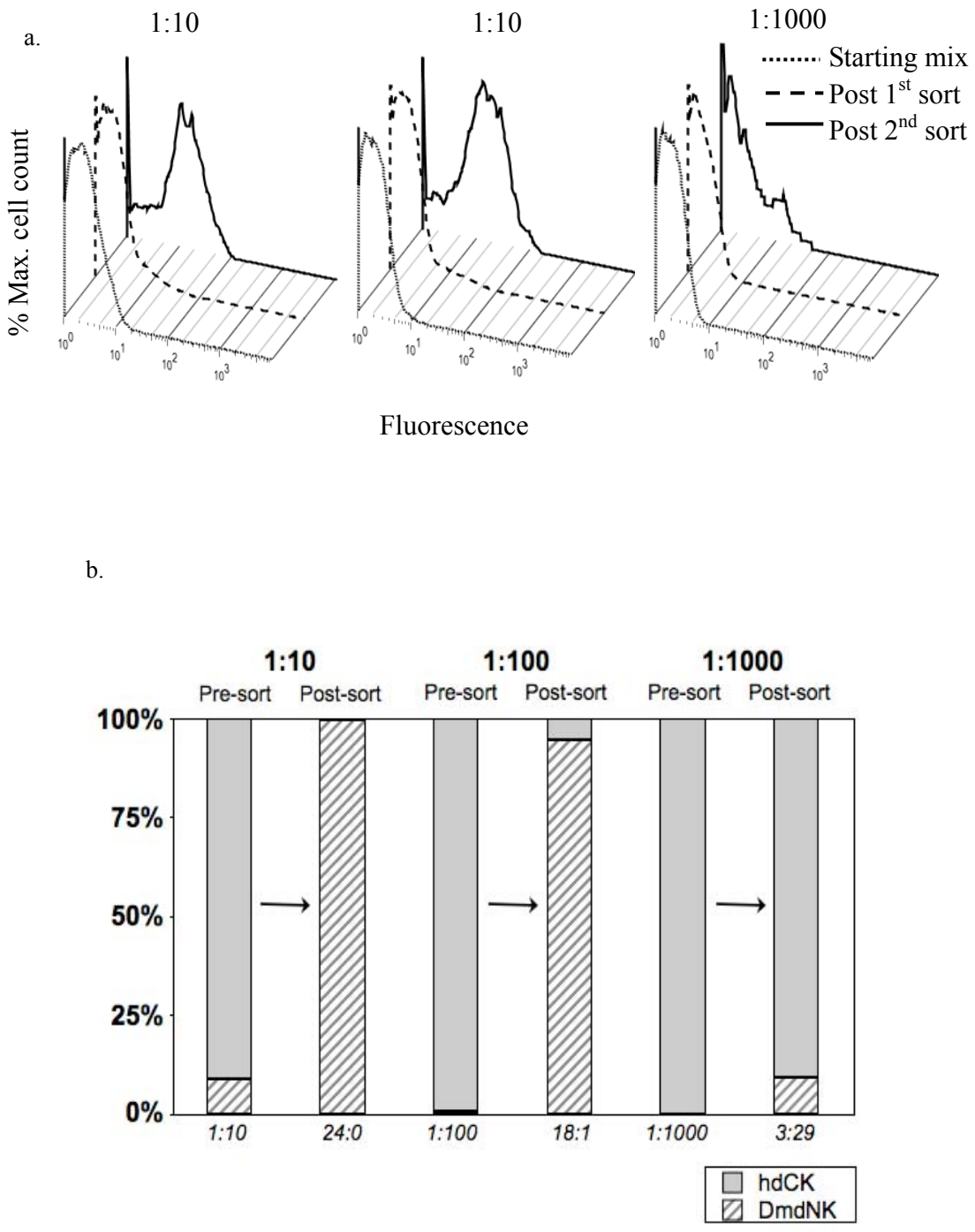


Figure 2-4: FACS sorting results of the enrichment experiment. a) FACS sorting scheme of the cell mixtures at three ratios; b) Enrichment of *Dm*dNK upon sorting. (Adapted from (Liu *et al.*, 2009))

2.3 Concluding remarks

Deoxynucleoside kinases are important protein engineering targets, due to their role in NA prodrug activation and potential application in suicide gene therapy. Ideally, to avoid unbalancing the cellular dNTP pool, an engineered kinase should exhibit high activity for the NA while effectively discriminating against phosphorylation of native substrates. Various directed evolution strategies have been attempted, yet lack of screening methods that identify library members with specific NA activity has resulted in limited success.

In this chapter, we proposed using fluorescent nucleoside analogs as substrates in directed evolution to establish a screening method that directly follows entrapped fNA-monophosphates. We validated the method by performing a simulated library screening with mixtures of *Dm*dNK and hdCK using fT as the target substrate. *Dm*dNK and hdCK were chosen for the simulated library screening due to the two enzymes' 10^5–fold difference in catalytic efficiency for thymidine. The FLUPS screening was able to enrich *Dm*dNK gene from the mixture by >100-fold, indicating capability of enriching nucleoside kinases with desired properties. Key parameters that could affect the final enrichment result include gate criteria selection in the sorting and number of rounds applied.

Fluorescent labeled NAs provide a good opportunity to develop a screening protocol

for directed evolution of nucleoside kinases by coupling with FACS sorting. However, addition of a fluorescent label often causes significantly diminished catalytic efficiency by nucleoside kinases (Hengstschlager *et al.*, 1993), due to potential bias the label can introduce to the process of evolving NA kinases. Here we discovered that furano-thymidine was a similar *Dm*dNK substrate as thymidine. This opens up the possibility of chemically synthesizing fluorescent NA as target substrates for NA kinase directed evolution.

Beside the fact that it is relatively straightforward to make modifications on the pyrimidine or purine bases to form fluorescent NAs, it is tricky for nucleoside kinases to catalyze both modified and natural nucleosides as the modifications might differentially bind to the enzymes. Therefore, to justify use of the fluorescent NAs to screen for NA kinases, it was critical to find fluorescent bases that have minimal catalytic impact. Therefore, it is very encouraging to discover that the furano-modifications on the thymidine base in fT and fddT does not change the substrates role for *Dm*dNK, a type I thymidine kinase well known for its broad substrate activity and high catalytic efficiency. Among the thymidine kinases family, type I kinases generally carry broader substrate activity than type I kinases benefiting from the more spacious and flexible active sites. *Dm*dNK specifically, was reported to be rather tolerant to substituents at 5-position of the pyrimidine ring owing to the extended cleft in the proximity (Johansson *et al.*, 2001), which could explain its high tolerance to fT, carrying the furano- addition to the

thymidine base between 5- and 4- positions.

The furano-thymine base provides a promising substitute to follow thymidine phosphorylation in *E. coli* cells. Its minor modification does not significantly alter it's ability to serve as a substrate for *Dm*dNK. Under identical assay conditions with fT incubation, the presence of an exogenous dNK with thymidine kinase activity results in a 20 to 30-fold increase in cellular fT fluorescence. This difference is sufficient to allow for ~100-fold enrichment of *E. coli* cells carrying a kinase with the desired substrate specificity through cell sorting in the flow cytometer. In summary, these results confirmed that *E. coli* cells expressing nucleoside kinases with desired catalytic activity can be reliably separated and enriched through FACS. In the next chapter, we will test whether our screening protocol is adaptable to evolution of specific kinases for a large category of NA prodrugs. Specifically, we will evaluate furano-thymine base combined with 2',3'-dideoxyribose to make a fluorescent version of ddT, a DNA chain replication terminator, for directed evolution of ddT kinases.

2.4 Materials and methods

2.4.1 Materials

Reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated. 6-Methyl-3-(β-D-2-deoxyribofuranosyl)furano-[2,3-*d*]pyrimidin-2-one (fT) was purchased from Berry & Associates (Dexter, MI). Fluorescence spectra were

measured with a FluoroMax-3 spectrofluorometer (HORIBA Jobin Yvon, Edison, NJ). *Pfu* Turbo DNA polymerase from Stratagene (La Jolla, CA) was used for all cloning unless otherwise indicated. Pyruvate kinase and lactate dehydrogenase were purchased from Roche Biochemicals (Indianapolis, IN) while restriction enzymes were acquired from New England Biolabs (Beverly, MA). Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). Plasmid DNA was isolated using the QiaPrep Spin MiniPrep kit and PCR products were purified with the QiaQuick PCR purification kit (Qiagen, Valencia, CA). All gene constructs were confirmed by DNA sequencing.

2.4.2 Steady-state fluorescence spectroscopy

Excitation spectra were collected from 230 nm to 380 nm while the emission wavelength set at 431 nm. Emission spectra were collected from 350 nm to 550 nm while the excitation wavelength set at 331 nm. Fluorescent compounds are 8.9 μM solutions in PBS buffer (pH 7.4).

2.4.3 Toxicity assay for fT

E. coli TOP 10 cells expressing *Dm*dNK from pBAD vector were grown until OD_{600} ~0.5, then protein expression was induced by 0.2% arabinose and after 4 h of expression, the cells would be incubated with fT (or thymidine as the control) for 2 hours before the cells were plated on LB plate with 50 μM ampicillin. The cell survival rate would be

compared between cells with fT and thymidine as an indicator of the cell toxicity.

2.4.4 FACS sorting of cell mixtures carrying *Dm*dNK and hdCK

The screening was performed by inoculating 2 ml LB media supplemented with ampicillin (50 μg/ml) with cells from the freezer stocks. After incubation at 37 °C for 13 h, cells (40 μl) were diluted 1/100 with LB media, containing ampicillin (50 μg/ml) and grown at 37 °C until the OD (600 nm) reached ~ 0.5. Protein expression was induced with arabinose (0.2 %) and incubation at 37 °C was continued for 4 h. Next, the cell culture was mixed with 22 μM fT. After continuing incubation at 37 °C for 2 h, cells were centrifuged and the pellet washed three times with PBS buffer (pH 7.4) before being resuspended in PBS buffer to ~ 1 x 10^8 cells/ml. Cell sorting was performed on a FACSVantage flow cytometer (Becton Dickinson, Franklin Lakes, NJ). The event detection was set to forward and side scattering. Sorting was performed at less than 2000 events/sec with excitation by a UV laser (351–364 nm) and emission detection through a band pass filter (424 ± 20 nm). Cells were collected into Eppendorf tubes containing SOC media, followed by incubated at 37 ºC for 2 h prior to plating on LB-agar plates supplemented with ampicillin (50 μg/ml). After overnight incubation at 37 °C, individual colonies were analyzed by PCR to determine the gene ratios of *Dm*dNK to hdCK.

2.4.5 Over-expression and purification of *Dm*dNK

*Dm*dNK gene was subcloned into pMAL-C2x (New England Biolabs) for expression as fusion protein with maltose binding protein (MBP) according to the manufacturer's protocol. Following induction of protein expression with IPTG at 37 °C for 4 h, cells were harvested by centrifugation (4000 *g*, 20 min, 4 ºC). Cell pellets were resuspended in buffer A (20 mM Tris-HCl pH 7.4, 300 mM NaCl, 1mM EDTA) and lysed by sonication on ice. Supernatant was collected after centrifugation (9,000 *g*, 30 min, 4 ºC) and loaded onto amylase resin (New England Biolabs). After two washes with five column volumes of buffer A, the target protein was eluted with three column volumes of buffer A supplemented with 10 mM maltose. Product fractions were concentrated in an Amicon-Ultra centrifugal filter unit (Amicon Bioseparations, Billercia, MA) and protein concentration quantified by measuring absorbance measurements at 280 nm (MBP-*Dm*dNK, ε = 106,230 $M^{-1}cm^{-1}$, calculated according to earlier method (PACE *et al.*, 1995)). Protein purity was verified by SDS-PAGE. Sample aliquots were flash-frozen and stored at -80 ºC.

2.4.6 Steady-state kinetic analysis by ITC

Isothermal titration calorimetry (ITC) was used to measure the kinetics of *Dm*dNK for thymidine and fT. Assays were performed at 25 °C in considerate of enzyme's stability during the test time period (~ 45 min). Reaction mixture contains 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2.5 mM $MgCl_2$ and 1 mM ATP. Measurements were made in triplicates on VP-ITC (MicroCal, Northampton, MA).

Chapter 3. Directed evolution of orthogonal ddT kinases using FLUPS (FLUorescent nucleoside analog Phosphorylation Screen)

(published, Liu *et al*., *Nucl. Acids Res.* 2009, 37, 4472-4481)

3.1 Introduction

Nucleoside analogs (NAs) have proved to be promising tools for antiviral and cancer therapies (Crumpacker, 1996; Galmarini *et al.*, 2003; Parker, 2009). However, many NAs are not efficiently transformed into tri-phosphates by natural deoxynucleoside kinases, which decreases the drug efficiency of current FDA-proved NA prodrugs and limits the development of new analogs (Arnér *et al.*, 1995; Shi *et al.*, 1999). Therefore the engineering of deoxynucleoside kinases into orthogonal, NA-specific kinases is of great interest for improving anti-viral & cancer therapies. Until now, success has been limited due to the lack of efficient screening or selection methods that directly monitor NA kinase activity (Kokoris *et al.*, 2002; Gerth *et al.*, 2007). The current selection strategies include *E. coli* auxotroph strain KY895 that selects for thymidine phosphorylation acitivity (Igarashi *et al.*, 1967), and live-death selection using replicating plates that depends on the cytotoxicity of the NA (Christians *et al.*, 1999). Unfortunately, the screening is only limited to AZT due to the cytotoxicity dependence. Additionally, the selection strategy would mislead the evolution towards kinases with broad substrate specificity while excluding variants with desired orthogonal NA activity. Exogenous kinases with broad substrate specificity might interfere with the tightly regulated

2’-deoxynucleoside metabolism. Therefore, the identification of orthogonal NA kinases, enzymes with changed rather than broadened substrate specificity, represents a formidable challenge with great potential benefits for therapeutics.

Inspired by enzyme library screening methods for novel proteases and glycosyltransferases based on fluorescent substrates (Varadarajan *et al.*, 2005; Aharoni *et al.*, 2006), in chapter 2, we developed a similar strategy for identifying enzymes that can effectively phosphorylate NAs. In concept, our idea is based on three observations: a) nucleosides can efficiently transport across the cell membrane while nucleoside phosphates are trapped inside the host cytoplasm, b) a majority of NA prodrugs carry functionally relevant substitutions in the molecule’s sugar moiety while leaving the nucleobase portion unchanged, and c) fluorescent nucleoside analogs (fNAs) with absorption maxima of >300 nm reduces signal interference from cellular auto-fluorescence, can serve as substrates for dNKs. Direct measurements of the natural nucleosides are impractical due to the compounds’ low quantum yields and overlapping absorption with aromatic amino acids and small-molecule metabolites (e.g. ATP and NADH). Fortunately, relatively small synthetic modifications of a nucleoside’s pyrimidine or purine moiety red-shift the absorption and increase quantum yield, enabling detection of these fNAs with high sensitivity in complex mixtures such as the cytoplasm. A number of fluorescent 2’-deoxyribonucleosides, used in DNA structure studies, have been reported including the *etheno*-derivatives of adenine and cytosine,

tricyclic guanine (Barrio *et al.*, 1972; Secrist III *et al.*, 1972; Boryski *et al.*, 1988), as well as furano and pyrrolo-pyrimidines and pterines (Bergstrom *et al.*, 1982; Hawkins *et al.*, 1997; Berry *et al.*, 2004) (Figure 1-8).

In Chapter 2, the FLUPS method was validated with the fluorescent version of thymidine (6-methyl-3-(β-D-2-deoxyribofuranosyl)furano-[2,3-*d*]pyrimidin-2-one; fT), using *E. coli*-based expression systems for human deoxycytidine kinase and *Drosophila malanogaster* deoxynucleoside kinase (*Dm*dNK). To apply the new screening method for NA kinase evolution, fluorescent nucleobases could be combined with NA prodrug sugar derivatives to generate suitable substrates for dNK library screening. The scheme of evolving NA kinases by using fluorescent 2', 3'-dideoxy-thymidine (fddT) in the FACS-based screening is shown in Figure 3-1. Initially bacterial broad-specificity nucleoside transporters facilitate the efficient translocation of fddT across the membrane. In the presence of a functional fddT kinase, the product of the enzymatic reaction (fddT –monophosphate) will accumulate inside the cell as its negative charge prevents exportation. Fluorescence-activated cell sorting (FACS) can subsequently be employed to isolate hosts with fluorescence intensity above background.

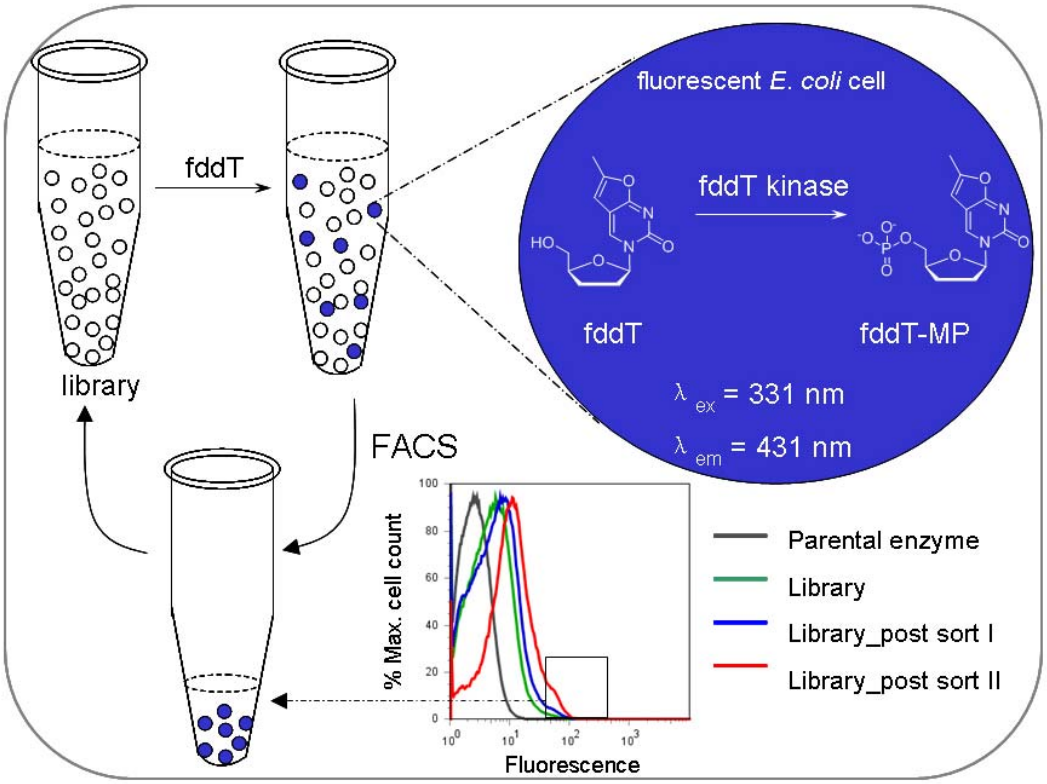


Figure 3-1: Scheme of screening for ddT/fddT kinase using FLUPS.

We chose 3'-deoxythymidine (ddT) as the first analog for directed evolution using FLUPS screening. Although chemically synthesized ddT-triphosphate can serve as DNA chain replication terminator (Fisher *et al.*, 1994), direct application of ddT to living cells shows little terminating DNA synthesis effect. The failure of ddT to exhibit biological activity is most likely caused by poor phosphorylation of ddT by cellular nucleoside and nucleotide kinases (Waqar *et al.*, 1984; Balzarini *et al.*, 1987). As such, ddT could be more efficient in chain termination in vivo if an exogenous kinase is provided with high

catalytic activity for ddT. To facilitate screening for ddT phosphorylation activity, a fluorescent version of ddT with spectral resolution for cellular fluorescence. We designed a synthetic route for fddT was established. Starting with uridine, a three-step synthesis eliminated the 2' and 3'-hydroxyl groups in the ribose moiety, followed by assembly of the furano-pyrimidine portion via palladium-catalyzed cross-coupling chemistry (Liu *et al.*, 2009). The ability to modify the sugar and base portions individually makes the approach applicable to a wide variety of pyrimidine nucleoside analogs and can potentially be expanded to purine analogs, using pterines as fluorescent substitutes for the nucleobase.

As a starting point for enzyme evolution, we chose *Dm*dNK, a member of the type-I dNK subfamily. (Johansson *et al.*, 2001; Eriksson *et al.*, 2002). Type-1 dNKs share a common fold but vary in their substrate specificities and overall catalytic efficiencies. Among type-1 dNKs, *Dm*dNK displays the broadest substrate specificity and also some of the highest activity for the four natural 2'-deoxyribonucleosides, as well as several NAs (Munch-Petersen *et al.*, 1998), making it a promising candidate for enzyme engineering.

3.2 Results and discussion

3.2.1 Effect of fddT's furano-modification as substrates for *Dm*dNK

In chapter 2, we have established that fT and thymidine are similar substrates for

*Dm*dNK, suggesting the furano-thymine base analog could be used in screening for kinases with improved catalytic efficiency toward NAs. Prior to directed evolution of a ddT kinase, *Dm*dNK's catalytic performance for ddT and fddT were further determined. As shown in Table 3-1, *Dm*dNK's catalytic efficiency for fddT is only 2-fold lower than ddT. The lower catalytic efficiency is mostly due to an effect of lower k_{cat} for both cases while no significant K_M changes was caused by the modification on the thymine base in fddT. Since 2-fold difference is a minimal difference, fddT could be used as substitute for ddT in FLUPS to evolve specific and efficient ddT kinases.

Table 3-1: Kinetic data of *Dm*dNK for T/fT/ddT/fddT at 25°C.

substrates	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($10^3 s^{-1} \cdot M^{-1}$)
T	3.2 ± 0.1	2.0 ± 0.1	1600 ± 130
fT	1.8 ± 0.1	2.0 ± 0.1	895 ± 70
ddT	0.22 ± 0.01	59 ± 1	3.7 ± 0.2
fddT	0.10 ± 0.01	84 ± 4	1.2 ± 0.1

3.2.2 Fluorescence property of fddT

Fluorescence excitation/emission spectra of fddT are obtained (Figure 3-2). The fluorescent properties of fddT and fT are very similar. The maximum excitation/emission wavelengths are 331nm, 431nm, respectively, and the quantum yield of the fluorophore is ~0.3, making it suitable for flow cytometry analysis

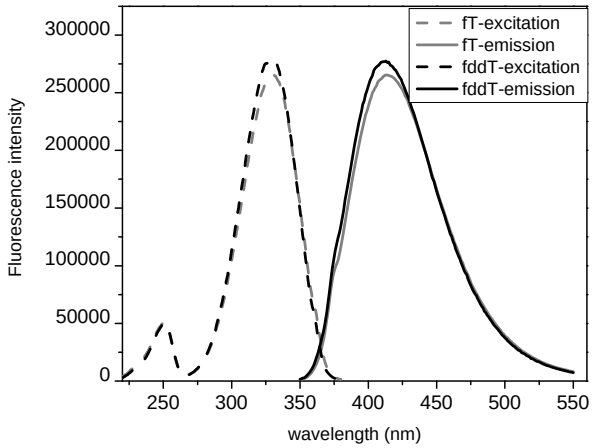


Figure 3-2: Fluorescence spectra of fT and fddT.

3.2.3 Evolution of an orthogonal ddT kinase

Using a combination of random mutagenesis by error-prone PCR and DNA shuffling, we created library diversity and screened for kinases with increased activity for fddT by FLUPS. After each round, a few sorted library members were picked for DNA sequence analysis while the remaining pool was used as templates to make the next generation library. The selection pressure for NA activity during FLUPS was gradually raised over the course of the engineering process. To favor library members that tightly bind fddT, the fNA concentration was lowered from initially 80 μM to 5 μM in the final rounds. To gain kinases with higher fddT turnover rate, the exposure time of cells to fluorescent substrate was shortened from 120 in the first round to 30 minutes in the later rounds. Finally, thymidine (up to 800 μM) was also added to compete with fddT binding to promote enzymes with reduced affinity for the native substrates. After three rounds of directed evolution, one representative variant (R3.V1) from the selected library was over-expressed for in vitro characterization. R3.V1 carries five mutations (K13N, T85M,

E172V, H193Y, S224C). While its activity for ddT seems largely unchanged, the mutant shows a drop in k_{cat}/K_M values for the natural 2'-deoxynucleosides by two to three orders of magnitude (Table 3-2). The decline in activity for the native substrates can be attributed to a combination of lower turnover numbers and higher *Michaelis-Menten* constants. To eliminate neutral and deleterious mutations, the *Dm*dNK variants selected from the 3rd generation were shuffled with the wild type enzyme by DNA shuffling, followed by FLUPS and DNA sequence analysis of selected candidates.

Among the 4th-round mutants, R4.V3 represents a consensus candidate in regards to its mutations and thus was selected for detailed in vitro characterization. The enzyme variant carries four amino acid substitutions (T85M, E172V, Y179F, H193Y), newly introducing Y179F while eliminating K13N and S224C that were found in the previous round (Figure 3-3A/B). Separate experiments verified that the both K13N and S224C were neutral mutations.

In R4.V3, the substitutions in positions 172 and 179 are likely a direct consequence of the ribose modifications in fddT (Figure 3-3C). Both residues are part of the active site and, according to the crystal structure, form hydrogen bonding interactions directly (E172) or via a water molecule (Y179) to the 3'-hydroxyl group of the phosphoryl acceptor. Mutations in these two highly conserved positions have not previously been observed in kinase engineering, a fact that we attribute to biases resulting from the library analysis for

Table 3-2: Kinetic parameters for natural 2'-deoxyribonucleosides (T, dC, dA, dG) and nucleoside analog (ddT) of wild type enzyme (*Dm*dNK) and selected candidates from the directed evolution experiments after round 3 (R3.V1) and round 4 (R4.V3; R4.V1; R4.V2).

Enzymes		ddT	T	dC	dA	dG
*Dm*dNK	K_M (μM)	115 ± 22	2.7 ± 0.5	2.0 ± 0.4	98 ± 10	447 ± 54
	k_{cat} (s^{-1})	0.53 ± 0.03	12.9 ± 0.9	11.7 ± 0.7	15.8 ± 0.4	12.3 ± 0.4
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**4.6 ± 0.9**	**5000 ± 1300**	**6000 ± 1500**	**160 ± 20**	**28 ± 4**
R3.V1	K_M (μM)	123 ± 15	127 ± 17	103 ± 25	2639 ± 550	1299 ± 150
	k_{cat} (s^{-1})	0.46 ± 0.02	0.16 ± 0.01	0.19 ± 0.02	0.19 ± 0.02	0.059 ± 0.002
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**3.7 ± 0.5**	**1.3 ± 0.2**	**1.8 ± 0.6**	**0.07 ± 0.02**	**0.05 ± 0.01**
R4.V3	K_M (μM)	167 ± 9	319 ± 88	95 ± 19	> 3000	3603 ± 1371
	k_{cat} (s^{-1})	0.49 ± 0.01	0.031 ± 0.003	0.034 ± 0.002	< 0.01	0.036 ± 0.007
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**3.0 ± 0.2**	**0.10 ± 0.03**	**0.4 ± 0.1**	/	**0.010 ± 0.006**
R4.V1	K_M (μM)	102 ± 33	286 ± 27	156 ± 7	1561 ± 346	1182 ± 301
	k_{cat} (s^{-1})	0.27 ± 0.02	0.086 ± 0.002	0.094 ± 0.001	0.075 ± 0.006	0.040 ± 0.004
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**2.6 ± 1.0**	**0.30 ± 0.03**	**0.60 ± 0.03**	**0.05 ± 0.01**	**0.034 ± 0.012**
R4.V2	K_M (μM)	101 ± 21	111 ± 14	45.6 ± 12.7	1734 ± 619	1545 ± 154
	k_{cat} (s^{-1})	0.24 ± 0.01	0.079 ± 0.002	0.105 ± 0.009	0.16 ±0.03	0.191 ± 0.007
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**2.4 ± 0.5**	**0.7 ± 0.1**	**2.3 ± 0.8**	**0.09 ± 0.05**	**0.12 ± 0.02**

thymidine kinase activity in the *E. coli* auxotroph KY895. The two active site mutations E172V and Y179F not only eliminate the favorable hydrogen-bonding interactions but also actively discriminate against the native substrates by optimizing the binding pocket for the smaller, more hydrophobic 2',3'-dideoxyribose moiety of ddT. On the other hand, the mutations in residues 85 and 193 are difficult to explain their roles as these two positions are not in proximity of the active site of the enzyme. The kinetic analysis of R4.V3 shows sustained activity for ddT compared to R3.V1 and wild type *Dm*dNK, yet a complete loss in ability to phosphorylate the natural purine 2'-deoxynucleosides and a >10,000-fold reduction in the specific activity for thymidine and 2'-deoxycytidine (Table 3-2). All variations in substrate specificity for R4.V3 and R3.V1 resulted from a combination of lower k_{cat} and higher K_M values.

3.2.4 Mutation study by Reverse engineering

The functional contributions of the four mutations in R4.V3 were evaluated by kinetic characterization of enzymes reverse engineered for each individual amino acid substitution (Table 3-3). Reversing V172 back to the original glutamate resulted in the most dramatic changes. It raised the level of activity for the native 2'-deoxyribonucleosides by two to three orders of magnitude. R4.V3-E172 has catalytic performance very similar to wild type *Dm*dNK. The mutant's turnover rate is close to the parent enzyme, but the K_M values are 2 to 16-fold higher than *Dm*dNK, compromising its wild type-like overall performance. The F179Y reverse substitution in R4.V3-Y179 also

Table 3-3: Kinetic parameters of R4.V3 variants after reverse engineering to elucidate the functional contribution of individual mutations. The data for the natural 2'-deoxyribonucleosides (T, dC, dA, dG) and nucleoside analog (ddT) are reported.

Enzymes		ddT	T	dC	dA	dG
R4.V3-T85	K_M (μM)	49 ± 3	92 ± 15	334 ± 65	> 3000	> 3000
	k_{cat} (s^{-1})	1.36 ± 0.01	0.13 ± 0.01	0.13 ± 0.01	< 0.01	< 0.01
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**28 ± 2**	**1.4 ± 0.3**	**0.4 ± 0.1**	/	/
R4.V3-H193	K_M (μM)	154 ± 16	251 ± 36	105 ± 15	> 3000	1092 ± 243
	k_{cat} (s^{-1})	0.84 ± 0.02	0.077 ± 0.003	0.048 ± 0.002	< 0.01	0.051 ± 0.004
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**5.5 ± 0.7**	**0.3 ± 0.1**	**0.5 ± 0.1**	/	**0.05 ± 0.01**
R4.V3-E172	K_M (μM)	186 ± 28	43 ± 9	4 ± 1	218 ± 36	721 ± 22
	k_{cat} (s^{-1})	0.49 ± 0.03	13.9 ± 0.9	117 ± 0.5	12.4 ± 0.7	4.3 ± 0.1
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**2.6 ± 0.5**	**330 ± 90**	**3000 ± 750**	**60 ± 12**	**6 ± 2**
R4.V3-Y179	K_M (μM)	124 ± 15	188 ± 29	35 ± 3	3000 ± 500	1500 ± 200
	k_{cat} (s^{-1})	0.48 ± 0.01	0.22 ± 0.01	0.15 ± 0.01	0.25 ± 0.05	0.06 ± 0.02
	k_{cat}/K_M ($10^3 s^{-1}M^{-1}$)	**3.9 ± 0.5**	**1.2 ± 0.2**	**4.3 ± 0.6**	**0.08 ± 0.03**	**0.04 ± 0.01**

contributes towards the recovery of wild type activity, yielding a 10-fold higher specific activity for the native substrates as a result of improvements in k_{cat} and K_M values. Together, the two mutations are the major players responsible for R4.V3 discriminating against natural substrates. In contrast, swapping Y193 back to the original histidine in R4.V3-H193 leads to only minor changes in the catalytic performance, indicating that the mutation is neutral and not responsible for the observed changes in substrate specificity. At last, the reversion of T85M shows an unexpected effect on the enzyme's substrate specificity and overall catalytic efficiency. While leaving the specific activity for 2'-deoxycytidine and the purine substrates unaffected, R4.V3-T85 exhibits 10 and 14-fold higher activity for ddT and thymidine, respectively. For both substrates, the catalytic improvement benefited from a combination of higher turnover and lower apparent binding constant. A rational for this behavior seems less obvious as the side chain of threonine points away from the active site. Since the effect is thymine specific, it is hypothesized that T85M is a response to substitutions in C-4/5 of the nucleobase of fddT. The introduction of the furano moiety adds a large, sterically rigid substituent at C-4/5 of the pyrimidine. This extra bulk leads to unfavorable interactions and potential steric clashes with side chains of V84 and M88 in the active site binding pocket, an effect that the enzyme compensates for by mutating the neighboring T85 (Figure 3-3D). The mutation could also reposition the helix that carries all three residues, allowing for more favorable binding interactions.

Table 3-4: Kinetic data of kinases using fddT as the substrate.

Enzymes	k_{cat} (s^{-1})	K_M (μM)	$\boldsymbol{k_{cat}/K_M}$ **($10^3 s^{-1} \cdot M^{-1}$)**
*Dm*dNK	0.18 ± 0.01	139 ± 24	**1.3 ± 0.3**
R4.V3	0.019 ± 0.001	10 ± 1	**1.9 ± 0.2**
R4.V3-T85	0.032 ± 0.002	81 ± 17	**0.4 ± 0.1**
R4.V3-E172	0.10 ± 0.01	14 ± 3	**7.5 ± 2.0**
R4.V3-Y179	0.035 ± 0.001	14 ± 1	**2.5 ± 0.2**
R4.V3-H193	0.024 ± 0.001	10 ± 1	**2.5 ± 0.4**

Our hypothesis about the functional role of T85M mutation is supported by the change in the kinetic properties of R4.V3-T85 for fddT. The enzyme has 8-fold higher K_M(fddT) value compared to R4.V3, leading to 5-fold lower catalytic efficiency (Table 3-4). While considered together with the fact that R4.V3-T85 shows 10 to 14-fold higher k_{cat}/K_M for T and ddT, indicates that T85M substitution is involved in remodeling the substrate nucleobase binding pocket, representing a perfect example of "you get what you select for" rule in directed evolution. Mutation in position 85 have previously been observed in the engineering of *Dm*dNK (Knecht *et al.*, 2000). The authors also speculate that position 85 is important for binding site topology.

While efforts to further investigate this hypothesis via protein crystallography are still in progress, the reverse engineering experiments indicate that the effect of T85M is strictly additive, and by simply reversing this mutation, we achieved a truly orthogonal ddT kinase with improved catalytic activity. Furthermore, the kinetic data for R4.V3-T85

shows that the two active site mutations in position 172 and 179 are responsible for a moderate but noteworthy 6-fold increase in ddT phosphorylation catalytic efficiency over the parental *Dm*dNK.

In summary, four rounds of directed evolution in combination with FLUPS have yielded enzymes that no longer phosphorylate native substrates under physiological conditions. Two key substitutions of highly conserved residue E172 and Y179 in the enzyme active site result in K_M-values for the native substrates that are at least 10-fold above the typical cellular concentration of <10 μM. At the same time, our engineered kinase shows 6-fold greater catalytic performance for ddT, making it a truly orthogonal NA kinase.

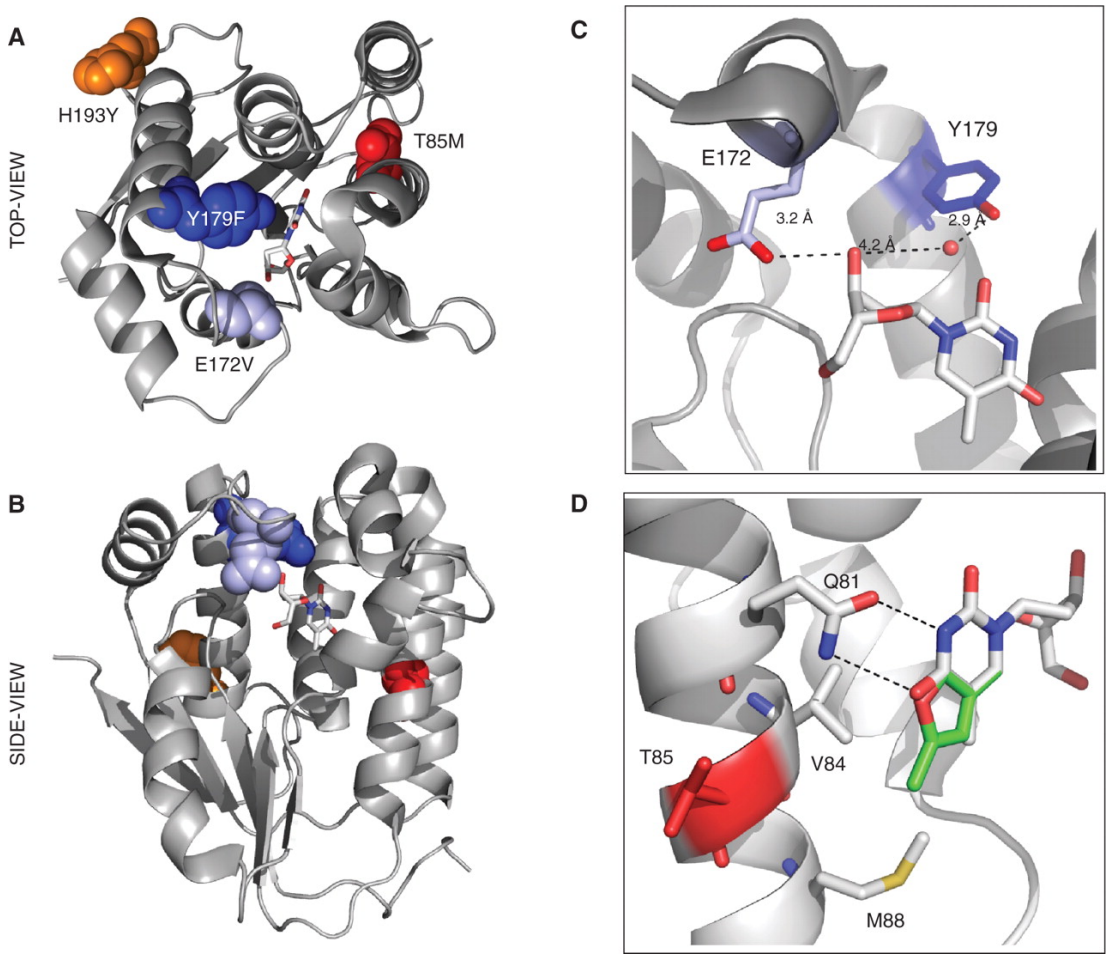


Figure 3-3: Structure model of R4.V3. The locations of the four mutations are highlighted in *Dm*dNK with thymidine, bound in the phosphoryl acceptor-binding site (PDB access#: 1OT3 (Mikkelsen *et al.*, 2003)). The top (A.) and side-view (B.) of *Dm*dNK show the close proximity of E172V (light blue) and Y179F (dark blue) to the substrate while H193Y (orange) and T85M (red) are located in more distant locations to the active site. C.) The close-up view of the active site shows the direct hydrogen-bonding interactions of E172 with the 3'-OH group of the substrate's ribose moiety. The OH group of Y179 contacts the same position on the substrate via a bridging water. D.) The overlay of thymidine and fT in the active site of *Dm*dNK indicates a potential steric clash of the fluorescent substrate's furano moiety with V84 and M88. The T85M mutation is believed to cause a slight conformational shift of the helical region, creating a more favorable binding pocket for substrates with modifications in the C4/C5-position of the nucleobase. (Adapted from (Liu *et al.*, 2009))

3.2.5 Competition experiments of orthogonal kinases in *E. coli*

When applied in a cellular environment, orthogonal NA kinases are predicted to show better performance than kinases with broad substrate specificity due to competition with natural nucleosides. To verify the hypothetical benefits of specific NA kinases *in vivo*, we compared the effects of R4.V3 to *Dm*dNK in *E. coli* cells when thymidine and fddT were present simutaneously. In Figure 3-4, *E. coli* cells expressing members of the fourth-round kinase library (black line) were able to accumulate more fddT than those expressing *Dm*dNK (grey line) when mixtures of thymidine and fddT were provided (1 to 100-fold molar excess of thymidine). At higher thymidine to fddT ratio, smaller difference was observed between performances of library members versus *Dm*dNK, resulting from the competition relationship between the natural substrate and fddT. Nevertheless, even at 8 mM thymidine (100:1 ratio to fddT), the evolved library variants still show more fddT entrapment than the parent kinase.

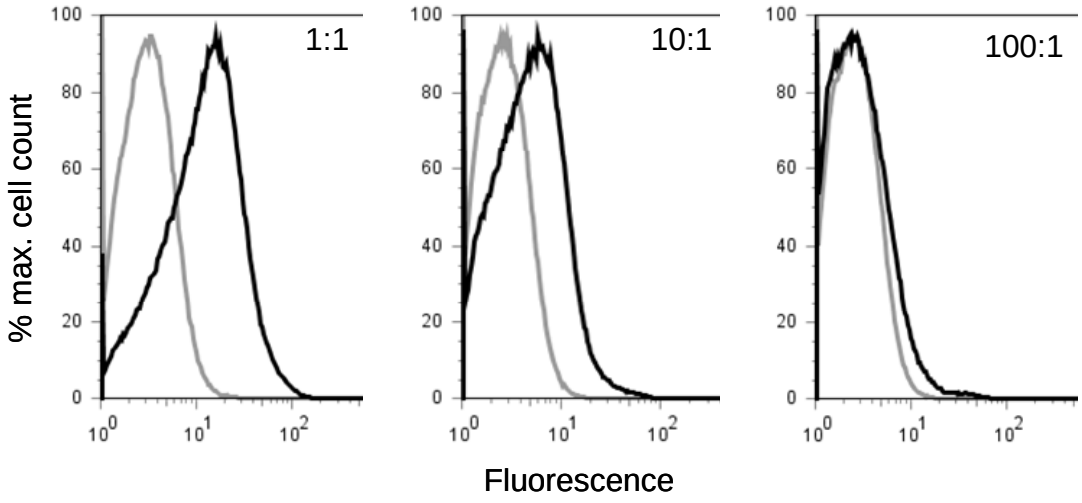


Figure 3-4: In vivo competition experiments in *E. coli*, studying the cellular accumulation of fluorescent ddT in the presence of increasing amounts of thymidine (1~100-fold molar excess of thymidine). Bacteria expressing members of the fourth-round kinase library (black line) consistently outperform cells expressing *Dm*dNK (gray line). At 80 μM fddT + 80 μM thymidine (1:1), the library shows approximately 10-fold higher fluorescence intensity over the wild-type enzyme. Although the favorable ratio decreases with increasing thymidine concentration, a direct consequence of the two substrates' competition for the active site, the evolved kinases slightly surpass *Dm*dNK even at 8 mM thymidine (100:1). (Adapted from (Liu *et al.*, 2009))

Next, R4.V3 was compared to *Dm*dNK for *in vivo* performance (Figure 3-5). Even though both enzymes show similar catalytic efficiency for fddT, *E. coli* cells expressing R4.V3 accumulated around 20-fold more fddT when incubated with a (10:1) mixture of thymidine and fddT, indicating R4.V3 has significantly higher apparent activity in an environment with natural nucleosides competing for active site of the enzyme. Rather than selecting a candidate with a greatly enhanced catalytic activity for fddT, the ability of R4.V3 to eliminate the competition of natural 2'-deoxynucleosides for occupancy of

the active site seems an equally successful strategy for increasing NA phosphorylation *in vivo*.

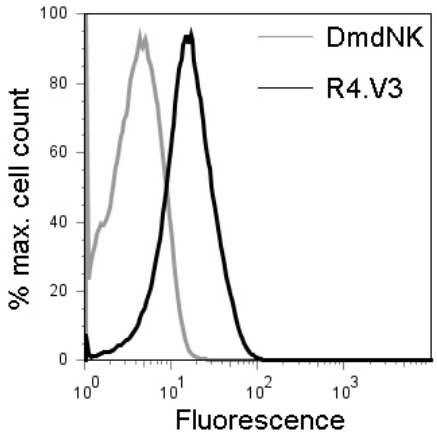


Figure 3-5: Cellular accumulation of fddT (10 μM) in bacteria expressing R4.V3 (black) or *Dm*dNK (gray) in the presence of a 10-fold molar excess of thymidine (100 μM).

3.2.6 Variants performance in cancer cell lines

To further verify the benefits of orthogonal kinases for NA activation in mammalian cells, we evaluated the cytotoxicity of ddT in three cancer cell lines upon co-expression of either *Dm*dNK, R4.V3, or no kinase. The genes of interest were transduced into human alveolar epithelial cells (A549) (Lieber *et al.*, 1976), breast adenocarcinoma cells (MCF-7) (Dickson *et al.*, 1986), and nasopharyngeal epidermal carcinoma cells (KB) (Eagle, 1955) via the pLXSN/retroviral expression system and cell viability was assessed after exposure to ddT for 96 hours (Figure 3-6). In the absence of an exogenous kinase (vector only control), A549 and MCF-7 cells show some sensitivity at the highest levels of ddT (500 μM), lowering the cell count by approximately 25-30%, while KB growth seems unaffected by the NA. In contrast, the presence of *Dm*dNK and R4.V3 impacts the viability of all three cell lines upon exposure to ddT, suggesting that elevated levels of

phosphorylated ddT raise the cytotoxicity of the NA. In A549 and MCF-7, the viability of cells expressing *Dm*dNK seems only slightly affected at 50 and 100 μM ddT but drops below 50% at 500 μM of NA. In the case of KB cells, co-expression of the fruitfly enzyme has a more muted effect, even at the highest concentration of ddT, only a moderate 20% drop was observed. Co-expression of our engineered kinase, R4.V3, shows a dramatic reduction in cell viability, falling by 30 to 50% at the lowest concentration (50 μM) of ddT in all three cell lines. In A549, cells are almost completely eradicated at NA concentrations of 100 μM and above while in MCF-7, viability drops steadily with increasing amounts of prodrug, reaching ~10% at 500 μM ddT. Beneficial effects in KB cells are also detectable. Although the overall response is weaker than in A549 and MCF-7, the cell counts at 50 μM ddT are significantly lower in comparison to the *Dm*dNK and vector-only experiments, reaching approximately 50% viability at 500 μM ddT. It is unclear why the cell count in KB drops rapidly at 50 μM ddT but then remains unchanged despite a 10-fold increase in prodrug. Since these data were reproducible in three independent experiments, the result might be linked to lower overall expression levels of our target enzyme and differences in NA metabolism in this particular cell line. The co-expression of a reporter such as GFP or western blots with kinase-specific antibodies will offer a better assessment of the protein levels in future experiments. Since control experiments with transfected cell lines in the absence of ddT show the same growth behavior than the original, uninfected cell lines over the time of

our experiments, the possibility of the exogenous kinase itself affecting cell proliferation are not a major concern in these experiments but could be investigated separately.

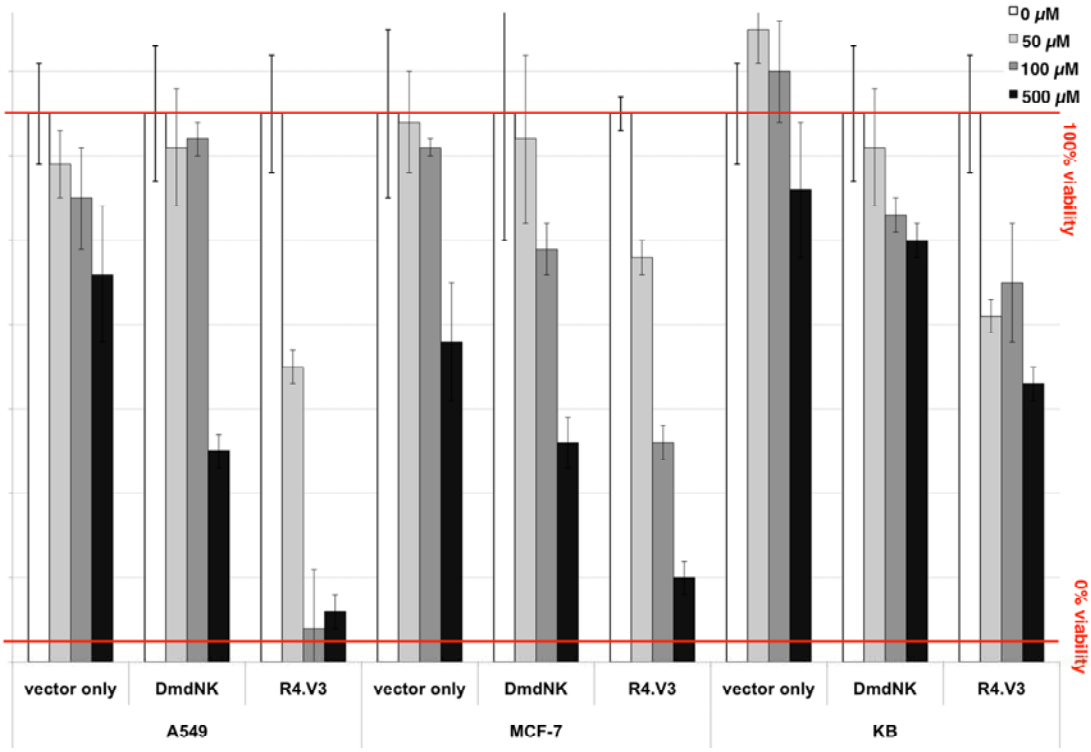


Figure 3-6: Viability of three cancer cell lines (A549, MCF-7, KB), following infection with adenovirus carrying R4.V3, *Dm*dNK, or vector-only, upon treatment with ddT at 0 μM (white bars), 50 μM (light grey), 100 μM (dark grey), and 500 μM (black) for 4 days. Normalized cell counts are shown.

Engineering dNKs with changed substrate specificity represent critical importance for future therapeutics applications. Enzymes with broader substrate specificity, often found in previous kinase engineering projects, have a clear disadvantage in situ as the NA must compete with the native 2'-deoxynucleoside substrates. Eliminating the competition with the natural substrates results in significant benefit for NA activation. As the kinetic and cell culture data for R4.V3 indicates, our ddT kinase significantly enhances the in

situ potency of ddT compared to the broad-specificity *Dm*dNK. We attribute this effect to R4.V3's capability to effectively discriminate against native substrates. Furthermore, orthogonal NA kinases reduce the risk of potential side effects due to fluctuations in the cellular dNTP pools. Although the long-term effects of auxiliary dNKs on cell cultures or an entire organism cannot be assessed easily, preventing the unregulated phosphorylation of natural 2'-deoxynucleosides should minimize their potential to interfere with other metabolic processes and DNA replication fidelity.

3.3 Concluding remarks

Nucleoside analogs (NAs) represent an important category of prodrugs for antiviral and cancer therapy, yet their biological potency is compromised due to inefficient activation by cellular deoxynucleoside kinases. We have designed and validated a new high-throughput screening method for the identification of engineered dNKs with NA activity, using a fluorescent version of the substrate. The approach was used to identified variants of *Dm*dNK whose substrate specificity changed by >10,000-fold in favor of our target substrate ddT. Besides evaluating the functional contributions of the introduced mutations by reverse engineering, we demonstrated significantly enhanced cytotoxicity of ddT in mammalian cells upon expression of our ddT kinase.

We believe that the concept of cellular entrapment of fNAs is a versatile strategy for screening directed evolution libraries of dNKs for variants with enhanced catalytic

efficiency and NA specificity. Structural variation in the prodrugs often concentrates on the ribose moiety, providing opportunities for modifying the nucleobase without compromising the enzyme's ability to phosphorylate the target substrate. The outlined synthetic methodology could be extended to a broad range of pyrimidine and purine nucleosides. Furthermore, directed evolution experiments using FLUPS are not limited to the dNK from fruitfly but should be functional for other members of the type-1 dNK family. Human dNKs are of particular interest for creating orthogonal NA kinases as they reduce the risk of undesirable immunogenic response in clinical applications. Last but not least, the screening of kinase libraries in combination with fluorescent substrates is not limited to *E. coli* but can be adapted to eukaryotic expression system such as *Saccharomyces cerevisiae* or mammalian cell lines. Both are well suited for FACS and could further optimize the search process for orthogonal NA kinases.

3.4 Materials and Methods

3.4.1 Directed Evolution Library Construction

Random mutagenesis libraries of wild type *Dm*dNK were created with the GeneMorph II kit (Stratagene, La Jolla, CA), using an average mutation frequency of 2-4 nucleotide changes per gene. DNA shuffling libraries were prepared by following the nucleotide exchange & excision technology (NExT) protocol (Kristian M. Müller, 2005). DNA fragments from the mutagenesis libraries and parental *dmdnk* were mixed at a 20:1 ratio for reassembly. Re-engineering of library members for function studies via

site-directed mutagenesis was performed by overlap extension PCR, following standard protocols (Gerth *et al.*, 2007). Final PCR products were cloned into pBAD-HisA (Invitrogen, Carlsbad, CA) via *Nco*I and *Hind*III restriction sites and electroporated into *E. coli* TOP10 (F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 nupG) (Invitrogen, Carlsbad, CA).

3.4.2 Library screening by fluorescence activated cell sorting

Library screening was performed by inoculating 2 ml LB media supplemented with ampicillin (50 μg/ml) at 37 °C until the OD (600nm) reached ~ 0.5. Protein expression was induced with arabinose (0.2 %) and incubation was continued for 4 h. The cell culture was mixed with 5 – 80 μM fNA. For competition experiments, the sample was complemented with an additional 50 – 800 μM thymidine. After incubation at 37°C for 30 - 120 min, cells were centrifuged and the pellet washed three times with PBS buffer (pH 7.4) before being resuspended in PBS buffer to ~ 1 x 10^8 cells/ml. Cell sorting was performed on a FACSVantage flow cytometer (Becton Dickinson, Franklin Lakes, NJ). The event detection was set to forward and side scattering. Sorting was performed at less than 2000 events/sec with excitation by a UV laser (351–364 nm) and emission detection through a band pass filter (424 ± 20 nm). Cells were collected into Eppendorf tubes, plated on LB-agar plates, and harvested for DNA sequence analysis and subsequent rounds of directed evolution.

3.4.3 Protein over-expression and purification

The genes of library candidates selected for in vitro studies were subcloned into pMALC2x for expression as fusion proteins with maltose binding protein (MBP) according to the manufacturer's protocol. Following induction of protein expression with IPTG at 37 °C for 4 h, cells were harvested by centrifugation (4000 *g*, 20 min, 4 ºC). Cell pellets were resuspended in buffer A (20 mM Tris-HCl pH 7.4, 300 mM NaCl, 1mM EDTA) and lysed by sonication on ice. Supernatant was collected after centrifugation (9,000 *g*, 30 min, 4 ºC) and loaded onto amylase resin (New England Biolabs, Beverly, MA). After two washes with five column volumes of buffer A, the target protein was eluted with three column volumes of buffer A supplemented with 10 mM maltose. Product fractions were concentrated in an Amicon-Ultra centrifugal filter unit (Amicon Bioseparations, Billercia, MA) and protein concentration quantified by measuring absorbance measurements at 280 nm (MBP-*Dm*dNK, ε = 106,230 $M^{-1}cm^{-1}$, calculated according to (PACE *et al.*, 1995)). Protein purity was verified by SDS-PAGE. Sample aliquots were flash-frozen and stored at -80 ºC.

3.4.4 Kinetic analysis

Selected candidates were subcloned into pMAL-C2x (New England Biolabs) for expression according to the manufacturer's protocol and purity of >95% was verified by SDS-PAGE. A spectrophotometric coupled-enzyme assay was used to measure the

phosphorylation activity for thymidine (Valentin-Hansen, 1978). Assays were carried on at 37 ºC, in a 500 µl reaction mixture containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2.5 mM $MgCl_2$, 0.18 mM NADH, 0.21 mM phosphoenolpyruvate, 1 mM ATP, 1 mM 1, 4-dithio-DL-threitol, 30 U / ml pyruvate kinase, 33 U / ml lactate dehydrogenase, and substrate with concentrations ranging from 1 µM to 7 mM. Measurements were made in triplicates and corrected for background. Data were fit to the Michaelis-Menten equation using Origin (OriginLab, Northhampton, MA) to get apparent K_m and V_{max}, and then k_{cat} is calculated from V_{max}.

3.4.5 Cell culture experiments

Wild type *dmdnk* and *r4.v3* were cloned into pLXSN (Clontech, Palo Alto, CA) using *EcoR*I and *Xho*I restriction sites. RetroPack PT67 packaging cells were cultured at 37 °C in Dulbecco's modified Eagle's medium (Mediatech, Manassas, VA) supplemented with 10% (V/V) fetal calf serum, 100 U/ml penicillin and 0.1 mg/ml streptomycin. The vectors were transfected into PT67 using Clonfectin (Clontech) following the manufacture's protocol. The cells were cultured continuously for 2 weeks in the presence of 0.6 mg/ml G418. Selected colonies were transferred to 6-well plates and retrovirus-containing media was collected after 48 h.

The three target cancer cell lines A549 (Lieber *et al.*, 1976), KB (Eagle, 1955) and MCF7 (Dickson *et al.*, 1986) were cultured in F-12K medium, Eagle's Minimum

Essential Medium and Eagle's Minimum Essential Medium supplemented with 0.01 mg/ml bovine insulin, respectively. All three media were further supplemented with 10% (V/V) fetal calf serum, 100 U/ml penicillin and 0.1 mg/ml streptomycin. Cells were grown at 37°C in a humidified incubator with a gas phase of 5% CO_2. The cell lines were transduced with retrovirus-containing media mixed with 6 μg/ml polybrene. The cells were incubated for 24 h and then cultured for 3 weeks in the presence of 0.6 mg/ml G418.

The viability of transfected cells in the presence of ddT was determined by MTT cell proliferation assay (Roche, Indianapolis, IN). Approximately 2000 cells per well were placed in 96-well plates. After 24 h, ddT was added to the medium at the indicated concentration. Medium supplemented with ddT was replaced with fresh solution every 24 h. Cell proliferation was assayed after 4 days as suggested by the manufacturer's protocol. All experiments were performed in triplicates.

Chapter 4. Computational Design of ddT kinase by Rosetta

(Manuscript submitted: Computational design of an orthogonal nucleoside analog kinase)

4.1 Introduction

Nucleoside kinases are key enzymes in the nucleoside salvage pathway which catalyze the first step of nucleoside phosphorylation (Johansson *et al.*, 2001; Eriksson *et al.*, 2002). Nucleoside analog (NA) prodrugs require the salvage pathway machineries to be effective against viral infection and cancer development. Indeed, phosphorylation of NAs to mono-phosphates is usually the rate-limiting step of prodrug activation. Specifically, human 2'-deoxyribonucleoside kinases (dNKs) are typically inefficient toward NAs, providing a major hurdle toward the development of new synthetic NA prodrugs (Eriksson *et al.*, 2002). For example, some NA tri-phosphates prove to have high in vitro gene replication termination activity, however failed to show equivalent anti viral/cancer activity in vivo where NAs were used (Waqar *et al.*, 1984; Shi *et al.*, 1999). One solution is to create more efficient NA kinases for coupled gene therapy (Moolten, 1986; Zheng *et al.*, 2000; Solaroli *et al.*, 2007). While nature provides a wide collection of kinases with potential NA activity, a common problem is that these native enzymes have higher activity for natural substrates relative to NAs. Therefore, significant effort has been devoted to engineering specific NA kinases.

In chapter 3, an orthogonal 2',3'-dideoxy thymidine (ddT) kinase was developed using directed evolution of *Drosophila melanogaster* deoxynucleosdie kinase (*Dm*dNK) in combination with FACS-based library screening (Liu *et al.*, 2009). The entire process took six months to reach variants with satisfying ddT specificity. Moreover, the variants carry a common mutation T85M that improves affinity towards the fluorescent analog base, but decreasing activity toward ddT containing the natural thymine base. An additional step was required to reverse the mutation to gain one order of magnitude higher ddT activity. An alternative approach was illustrated by recent enzyme design using computational methods (Jiang *et al.*, 2008; Rothlisberger *et al.*, 2008; Murphy *et al.*, 2009), raises the question of whether we could recapitulate our directed evolution results *in silico* with higher efficiency. Furthermore, computational approach enables a more thorough search of sequence space, potentially identifying unique solutions missed due to practical limitations in laboratory experiments. The goal is to design NA kinases that not only show significant shifts in substrate specificity (orthogonality), but simultaneously improve the catalytic efficiency for NA phosphorylation.

Rosetta is one of the leading computational tools for protein structure prediction which has proved its powerful capability in designing new catalytic activity by setting up a sound starting point for further directed evolution (Rothlisberger *et al.*, 2008). In this chapter, Rosetta was used to predict beneficial mutations in *Dm*dNK resulting in a ddT kinase that discriminates against natural nucleosides. After the Rosetta-proposed variants

were characterized, rational design strategies were used to combine mutations learned from directed evolution with Rosetta predicted changes to further improve ddT specificity. In the end, a *Dm*dNK variant was developed with comparable ddT activity and specificity to the orthogonal ddT kinase found in earlier directed evolution.

4.2 Results and discussion

4.2.1 Design ddT kinase from *Dm*dNK by Rosetta (accomplished by Murphy P.)

Using crystallographic information for *Dm*dNK in the presence of thymidine in the phosphoryl-acceptor binding site (PDB access code: 1OT3 (Mikkelsen *et al.*, 2003)), we applied an extension of the Rosetta suite of molecular modeling tools to redesigned the active site of the kinase (Das *et al.*, 2008). The computational method performed fixed-backbone design to optimize the specificity of *Dm*dNK for 3'-deoxythymidine (ddT) relative to Thy. Models of the transition state structures (TSS) were created based on the substrate structure identified in PDB structure 1OT3. The TSS for ddT was identical to that of Thy except for the substitution of the 3'-hydroxyl group with a hydrogen modeled in the geometrically optimal position for an aliphatic hydrogen.

Initial energy minimizations identified a set of three positions (L66, Y70, and V175) in the vicinity of the substrate binding pocket and designs were made with altered amino acid identities for these residues (Dahiyat *et al.*, 1997; Kuhlman *et al.*, 2000; Kortemme *et al.*, 2004). Given the small sample set (n=3), combinations of new amino acid

identities for these residues were sampled exhaustively and assignments for each of the 8000 (20^3) combinations were made. Residues within 10Å were allowed to repack and minimize, using the side chain degrees of freedom while the position of the substrate was fixed during this stage. Calculations were performed for TSS models with ddT and Thy independently. The predicted energy of interaction (ΔG) between the enzyme and ddT was used to estimate the catalyst's activity for the NA. Separately, the difference in energies of interaction (ΔΔG) was calculated by subtracting ΔG_{Thy} from ΔG_{ddT} to provide an approximation of relative specificity. Of the 8000 possibilities, designs with maximal ΔΔG and low ΔG_{ddT} values were identified to select designs with maximal ddT acitivity and specificity.

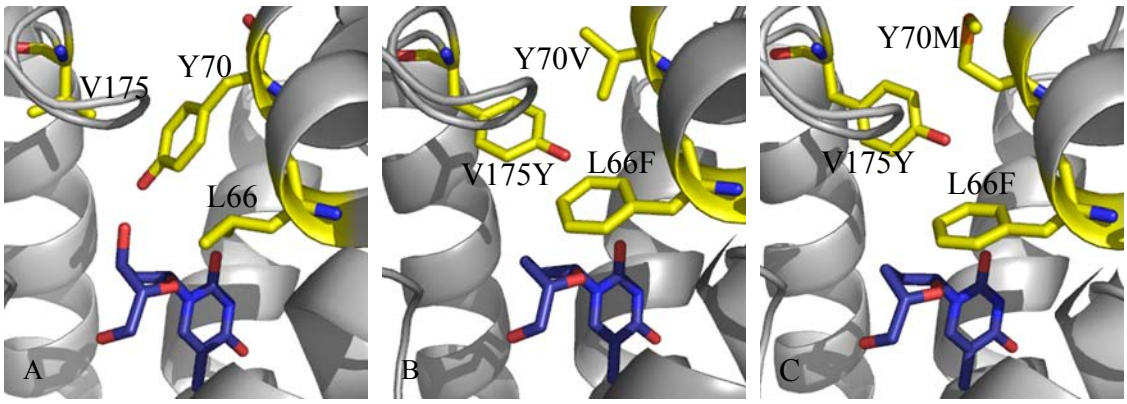


Figure 4-1: Rosetta structural models. A) Structure of thymidine-bound *Dm*dNK with L66, Y70 and V175 highlighted; B) Rosetta structure model of ddT-bound variant with three mutations: L66F, Y70V and V175Y; C) Rosetta structure model of ddT-bound variant with three mutations: L66F, Y70M and V175Y.

Among the top performers in the computational model, the predictions for position 66 clearly favored a benzyl side chain to sterically block the proper orientation of the native substrate's 3'-hydroxyl group. Then amino acids at Y70 and V175 need to be

replaced accordingly to accommodate the bulkier aromatic residue at position 66. The data was less conclusive about what exact substitutions to use in positions Y70 and V175. Position 175 seemed to favor large hydrophobic side chains (F, Y, W), while substitutions in position 70 were vague and seemed to be largely compensatory in nature, accommodating the new bulky neighboring groups in positions 66 and 175 (Figure 4-1).

To our surprise, none of the solutions proposed by Rosetta showed substitutions found in the directed evolution experiments. Amino acid substitutions at positions 172 and 179 could lead to specificity gain toward ddT (Liu *et al.*, 2009). Upon closer analysis, we attributed their absence in the computational design to the unfavorable effects of these mutations on overall protein stability (Table 4-3). On the other hand, our laboratory evolution experiments did not identify a single candidate with amino acid substitutions in positions 66, 70, or 175, even though Rosetta predicted these variants to alter substrate specificity and maintain stability. The absence of these solutions in directed evolution is most likely attributed to the fact that all three substitutions require at least two or three nucleotide changes per codon. For instance, the mutation at L66 is the most critical one in the Rosetta designs. Phe or Tyr should replace L66 to sterically block the 3'-hydroxyl group on the native 2'-deoxyribose. To replace Leu with Phe, the codon need to be mutated from "CTG" to "TTT" or "TTC", requiring two simultaneous nucleoside changes; while mutation to Tyr requires three nucleoside changes to "TAT" or "TAC". This is highly unlikely in a whole-gene random mutagenesis library with a total

of 2-4 nucleotide changes per 700-bases sequence (Wong *et al.*, 2006; Wong *et al.*, 2006). In addition, mutagenesis in one of the three positions likely requires compensatory changes of the neighboring amino acid(s) to preserve the enzyme's structural and functional integrity, further reducing the probability for such variants to emerge in our experimental libraries. Given the unique set of mutations, the Rosetta designs were constructed in the laboratory and tested for stability, and catalytic performance with native substrate (thymidine) versus the targeted nucleoside analog (ddT).

4.2.2 Kinetic analysis of Rosetta designed mutants

We started with Rosetta's most frequently observed substitution (L66F) and chose V175Y from among the suggested set of substitutions (F, Y, W). Within this framework, two variants were characterized carrying either Y70V (RosD3) or Y70M (RosD4) which, based on the model, fit well in the newly created cavity between F66 and Y175 (Fig. 1B, C). The two mutants were assembled by site-directed mutagenesis and expressed well in our *E. coli* expression system. Both proteins were purified to >90% purity (based on SDS-PAGE) and kinetic studies were performed to assess the two variants' catalytic performance with Thy and ddT. The k_{cat}/K_M values for thymidine declined by 20 and 58-fold for RosD3 and RosD4, respectively. In both enzymes, the decreases were largely due to higher K_M values for thymidine. At the same time, the catalytic efficiency for ddT phosphorylation was preserved in RosD3 and increased approximately 2.4-fold in RosD4 compared to *Dm*dNK, largely due to a drop in the *Michaelis-Menten* constant. Even

though the specificity for both mutants has been improved, the activity for natural substrates thymidine is still 8 or 50-fold higher than its activity for ddT. These variants need to be further improved to gain higher ddT specificity.

Table 4-1: Kinetics of Rosetta designed mutants in comparison to *Dm*dNK and R4.V3-T84, the variant from directed evolution.

Enzymes		ddT	T	Specificity*
*Dm*dNK	K_M (μM)	115 ± 22	2.7 ± 0.5	0.001
	k_{cat} (s^{-1})	0.53 ± 0.03	12.9 ± 0.9	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**4.6 ± 0.9**	**4,800 ± 1,300**	
R4.V3-T85	K_M (μM)	49 ± 3	92 ± 14	20
	k_{cat} (s^{-1})	1.4 ± 0.1	0.13 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**28 ± 2**	**1.4 ± 0.3**	
RosD3	K_M (μM)	79 ± 14	27 ± 3	0.016
	k_{cat} (s^{-1})	0.29 ± 0.01	6.3 ± 0.2	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**3.7 ± 0.7**	**234 ± 33**	
RosD4	K_M (μM)	36 ± 2	56 ± 2	0.13
	k_{cat} (s^{-1})	0.40 ± 0.01	4.6 ± 0.1	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**11.1 ± 0.9**	**83 ± 5**	

*Specificity=[k_{cat}/K_M (ddT)]/[k_{cat}/K_M (T)]

The thermostability was also evaluated for both Rosetta variants and compared with the wild type kinase. Our assay determined the reduction in enzyme activity after incubation at 37°C for 10 min. The results show *Dm*dNK retains 70% residual activity while RosD3 and RosD4 drop to 58% and 39%, respectively (Table 4-3).

4.2.3 Modifying Rosetta designed *Dm*dNK variants

To further improve specificity and efficiency of the Rosetta designed variants, we evaluated the possible complementary effects of mutations found by directed evolution with the Rosetta designed changes in *Dm*dNK. Previous experiments described in Chapter 3 had identified E172 as a key residue for substrates with modifications in the ribose moiety while Y179 was found to contribute albeit to a lesser degree (Liu *et al.*, 2009). Hence, we created a small site-directed mutagenesis library at position 172 covering eleven hydrophobic residues (V, P, H, I, L, Y, M, A, T, G and F). RosD4 was selected as a template for these experiments, due to its 2-fold higher ddT activity and 130-fold higher ddT specificity compared to the wild type kinase. Interestingly, the kinetic properties of all eleven second-generation variants show 2-fold or less variation in their k_{cat} and K_M values for ddT compared to RosD4 (Table 4-2). Similarly, the *Michaelis-Menten* constants for thymidine were largely unchanged while the turnover rates for the native substrate declined by 20 to 50-fold. The drop in k_{cat} for thymidine translates into substantial gains in relative substrate specificity. Among the tested variants, substitution of E172 to either V, T, L, and I showed the most significant functional improvements. The most substantial specificity improvement was observed in E172V, shifting the ratio by 50-fold in favor of ddT. Interestingly, the same mutation was found in our directed evolution library, possibly reflecting the significant counter selection pressure from high concentrations of thymidine applied during the library screening by FACS (Liu *et al.*, 2009). The remaining two variants exhibited 30 to 35-fold specificity gains. In contrast, E172I, despite slightly less specificity, retained higher

activity than its parental enzyme. A possible explanation for the differences in stability of these variants can be derived from computational models (Figure 4-2). All three substitutions at 172 (V, I and T), in conjunction with F66, remodel the enzyme active site to disfavor binding of substrates with 2'-deoxyribosyl moieties by eliminating the potential hydrogen-bonding partner and increasing steric constraint for the substrate's 3'-OH group. However, E172I shows noticeable tighter packing of the sec-butyl side chain compared to the isopropyl group of E172V and 2-hydroxyl-ethyl group of E172T, an observation consistent with the detected increases in protein thermostability. Upon 10 min incubation at 37°C, RosD4-E172V and RosD4-E172T retain only 28% and 29% of the activity, respectively. In contrast, RosD4-E172I still has 49% residual activity (Table 4-3). The E172I substitution contributes to the higher stability of the variant, which likely is the reason for its slightly better conservation of the enzyme's ddT activity.

Table 4-2: Kinetic analysis of variants based on RosD4 with substitutions at position E172.

Enzymes		ddT	T	specificity
RosD4	K_M (μM)	36 ± 2	56 ± 2	0.13
	k_{cat} (s^{-1})	0.40 ± 0.01	4.6 ± 0.1	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**11.1 ± 0.9**	**83 ± 5**	
RosD4-E172I	K_M (μM)	35 ± 4	66 ± 7	3.7
	k_{cat} (s^{-1})	0.41 ± 0.01	0.21 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**11.7 ± 1.6**	**3.2 ± 0.5**	
RosD4-E172T	K_M (μM)	47 ± 3	91 ± 5	4.7
	k_{cat} (s^{-1})	0.44 ± 0.01	0.19 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**9.4 ± 0.9**	**2.0 ± 0.2**	
RosD4-E172A	K_M (μM)	50 ± 4	65 ± 5	2.5
	k_{cat} (s^{-1})	0.46 ± 0.01	0.24 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**9.3 ± 0.9**	**3.7 ± 0.4**	
RosD4-E172F	K_M (μM)	41 ± 3	75 ± 8	3.1
	k_{cat} (s^{-1})	0.35 ± 0.01	0.20 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**8.5 ± 0.8**	**2.7 ± 0.4**	
RosD4-E172L	K_M (μM)	58 ± 8	101 ± 21	3.5
	k_{cat} (s^{-1})	0.44 ± 0.02	0.22 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**7.6 ± 1.4**	**2.2 ± 0.5**	
RosD4-E172M	K_M (μM)	52 ± 6	66 ± 9	2.5
	k_{cat} (s^{-1})	0.35 ± 0.01	0.18 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**6.8 ± 1.0**	**2.7 ± 0.5**	
RosD4-E172P	K_M (μM)	42 ± 2	74 ± 12	3.0
	k_{cat} (s^{-1})	0.28 ± 0.01	0.16 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**6.7 ± 0.5**	**2.2 ± 0.5**	
RosD4-E172H	K_M (μM)	47 ± 2	77 ± 7	2.5
	k_{cat} (s^{-1})	0.29 ± 0.01	0.19 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**6.2 ± 0.5**	**2.5 ± 0.4**	
RosD4-E172Y	K_M (μM)	50 ± 6	86 ± 4	2.3
	k_{cat} (s^{-1})	0.28 ± 0.01	0.21 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**5.6 ± 0.9**	**2.4 ± 0.2**	
RosD4-E172V	K_M (μM)	35 ± 4	96 ± 15	6.4
	k_{cat} (s^{-1})	0.19 ± 0.01	0.08 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**5.4 ± 0.9**	**0.84 ± 0.18**	
RosD4-E172G	K_M (μM)	54 ± 7	92 ± 10	1.9
	k_{cat} (s^{-1})	0.23 ± 0.01	0.21 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**4.3 ± 0.7**	**2.3 ± 0.4**	

Table 4-3: Activity-based thermostability comparison using ddT as the substrate (data showing residual activity (V_{max}) after enzyme being incubated at 37°C for 10 min).

Enzyme	Residual activity
*Dm*dNK	70%
R4.V3-T85	34%
RosD3	58%
RosD4	39%
RosD4-E172V	28%
RosD4-E172I	49%
RosD4-E172T	29%
RosD4-(E172I V175W)	49%
RosD4-(E172I V175F)	23%

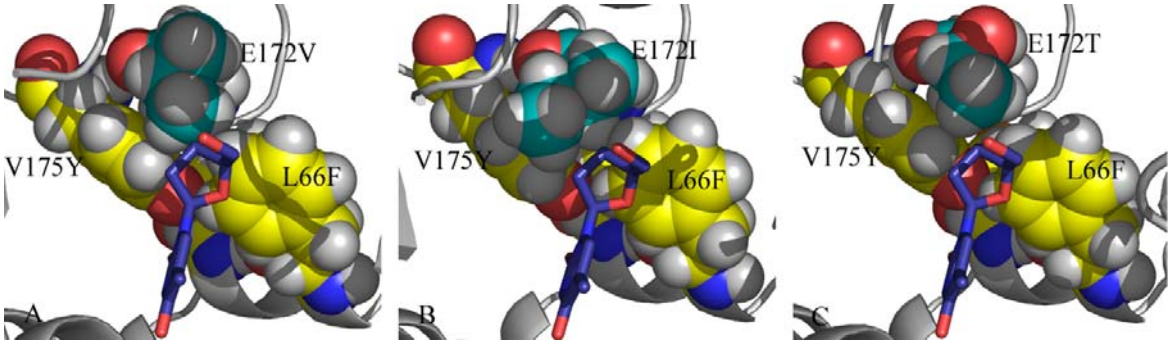


Figure 4-2: Rosetta structural models with ddT bound: A) RosD4-E172V; B) RosD4-E172I; C) RosD4-E172T. Ile at 172 shows tighter packing with F66 and Y175 than either Val or Thr.

In the Rosetta model, residues substituting V175 must accommodate various mutations at 66 and 70. Since V175 is in close proximity to 3'-C on the ribose moiety of ddT, an appropriate amino acid substitution could facilitate ddT binding and increase the

enzyme's ddT activity. Working with RosD4-E172I as the template, we prepared five mutations, replacing Y175 with I, L, M, F, or W (Table 4-4). Following protein expression and purification, kinetic properties with thymidine and ddT for these variants were characterized. Among the substitutions, only V175W and V175F showed improvements in their relative specificity. In RosD4-(E172I V175F), the gain in specificity results from a combination of lower turnover rates and higher K_M values, slightly more favorable to ddT than thymidine. More importantly, the elimination of the 4-hydroxyl group on the side chain caused a drop in protein's thermostability to 23% residual activity in our stability assay, the lowest value for any designed kinase in this study. In contrast, the replacement of the tyrosine side chain with an indole moiety in V175W translated into higher ddT catalytic efficiency by raising k_{cat} while lowering the catalyst's performance for the native substrates by increasing its apparent binding constant. In addition to these very favorable functional changes, the protein stability compared to RosD4-E172I remained at 49% residual activity. While the observed changes in catalytic function remain difficult to rationalize based on our current models, the computational structure predictions for these new variants can provide valuable guidelines. For RosD4-E172I V175W, the energy minimization in Rosetta causes the indole side chain of W175 to rotated 90˚ relative to Y175 and F175 (Figure 4-3). The conformational reorganization positions the aromatic side chain so that it can stack with the benzyl portion of the neighboring Y179, potentially contributing to stabilizing the enzyme scaffold. For RosD4-(E172I V175F), lacking the 4-hydroxyl group on tyrosine

resulted in the side chain of M70 pointing down into the space between 66 and 175 while it points up in the other two variants, forcing the aromatic ring of F66 to adopt a different angle. This side chain arrangement prevents ddT from binding to the active site thus decreases phosphorylation activity. The over-packed residues above the ddT binding pocket might also be the cause for decreased stability of the F175 variant. Overall, the amino acid at position 175 coordinates with the residues at 66 and 70 to shape the binding pocket of the ribose moiety of ddT, and contribute to high ddT activity and specificity.

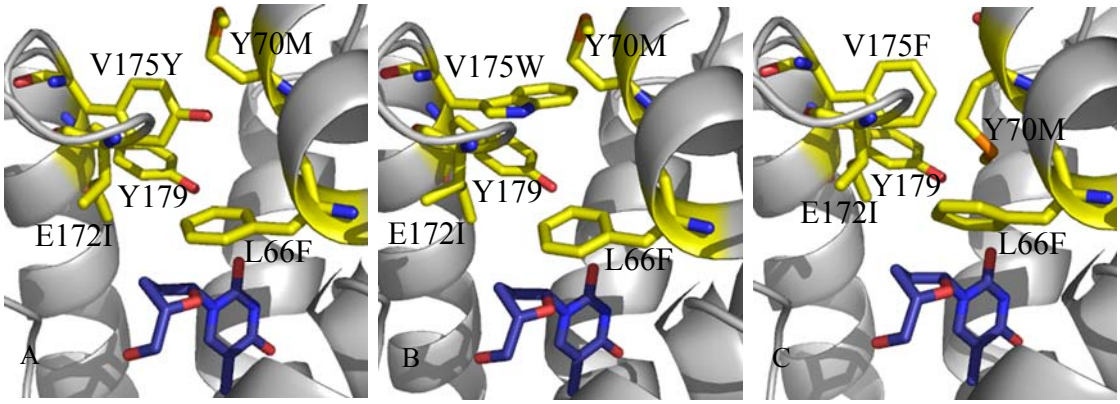


Figure 4-3: Rosetta structural models with ddT (shown in blue) bound: A) RosD4-E172I; B) RosD4-E172I V175W; C) RosD4-E172I V175F.

Table 4-4: Kinetics of variants based on RosD4-E172I with mutations at position 175.

Enzymes		ddT	T	specificity
RosD4-E172I	K_M (μM)	35 ± 4	66 ± 7	3.7
	k_{cat} (s^{-1})	0.41 ± 0.01	0.21 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**11.7 ± 1.6**	**3.2 ± 0.5**	
RosD4-(E172I V175W)	K_M (μM)	32 ± 4	173 ± 32	8.5
	k_{cat} (s^{-1})	0.65 ± 0.02	0.42 ± 0.02	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**20.5 ± 3.2**	**2.4 ± 0.9**	
RosD4-(E172I V175F)	K_M (μM)	120 ± 16	202 ± 79	4.7
	k_{cat} (s^{-1})	0.34 ± 0.01	0.12 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**2.8 ± 0.4**	**0.6 ± 0.3**	
RosD4-(E172I V175I)	K_M (μM)	800 ± 183	331 ± 131	1.6
	k_{cat} (s^{-1})	0.38 ± 0.04	0.10 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**0.48 ± 0.15**	**0.30 ± 0.15**	
RosD4-(E172I V175L)	K_M (μM)	302 ± 73	367 ± 75	3.0
	k_{cat} (s^{-1})	0.53 ± 0.04	0.22 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**1.8 ± 0.6**	**0.60 ± 0.15**	
RosD4-(E172I V175M)	K_M (μM)	353 ± 88	167 ± 25	1.3
	k_{cat} (s^{-1})	0.34 ± 0.03	0.13 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**1.0 ± 0.3**	**0.78 ± 0.15**	

In order to further improve enzyme specificity, amino acid variations at position 179 were also explored. In previous directed evolution experiments, mutating Y179 into Phe disabled the hydrogen bonding interaction between the hydroxyl group on tyrosine and 3'-OH of the natural 2'-deoxynucleosides via a water molecule, leading to a 10-fold gain in kinase specificity toward ddT. Among the five amino acids (L, I, V, A, F) used to substitute Y179, the first four all show improved ddT specificity (Table 4-5). However,

changes of Y179 to other hydrophobic residues consistently proved detrimental to catalytic performance for both ddT and thymidine. The specificity gain is a result of the substitutions being more deleterious for thymidine than ddT.

Table 4-5: Kinetics of variants based on RosD4-(E172I V175W) with mutations at position 179

Enzymes		ddT	T	specificity
RosD4-(E172I V175W)	K_M (μM)	32 ± 4	173 ± 32	8.5
	k_{cat} (s^{-1})	0.65 ± 0.02	0.42 ± 0.02	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**20.5 ± 3.2**	**2.4 ± 0.9**	
RosD4-(E172I V175W **Y179L**)	K_M (μM)	87 ± 18	511 ± 178	18.2
	k_{cat} (s^{-1})	0.35 ± 0.03	0.11 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**4.0 ± 0.3**	**0.22 ± 0.10**	
RosD4-(E172I V175W **Y179I**)	K_M (μM)	172 ± 42	818 ± 358	11.8
	k_{cat} (s^{-1})	0.22 ± 0.02	0.09 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**1.3 ± 0.4**	**0.11 ± 0.06**	
RosD4-(E172I V175W **Y179V**)	K_M (μM)	90 ± 21	480 ± 80	10.3
	k_{cat} (s^{-1})	0.32 ± 0.02	0.17 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**3.6 ± 1.0**	**0.35 ± 0.08**	
RosD4-(E172I V175W **Y179A**)	K_M (μM)	83 ± 24	548 ± 137	9.1
	k_{cat} (s^{-1})	0.25 ± 0.02	0.18 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**3.0 ± 1.0**	**0.33 ± 0.10**	
RosD4-(E172I V175W **Y179F**)	K_M (μM)	85 ± 13	318 ± 96	1.3
	k_{cat} (s^{-1})	0.24 ± 0.01	0.15 ± 0.01	
	k_{cat} / K_M ($10^3 s^{-1} M^{-1}$)	**2.8 ± 0.5**	**0.46 ± 0.17**	

4.3 Concluding remarks

In summary, computational redesign of *Dm*dNK by Rosetta in combination with site-directed mutagenesis in position E172 yielded a new designed kinase, RosD4-(E172I

V175W) whose catalytic performance closely resembles our previously evolved ddT kinase R4.V3-T85. Although the relative specificity of RosD4-(E172I V175W) for ddT over thymidine is approximately 2-fold lower than the laboratory-evolved variant, our new *in silico* design possess a lower K_M for ddT (32 μM versus 49 μM) and a significantly more favorable K_M ratio of ddT/thymidine of 5.4 compared to 1.9 for R4.V3-T85, features that have proven of critical importance *in situ* as they minimize interference with NA activation by the native 2'-deoxyribonucleosides (Chapter 3, 3.2.6). Furthermore, the designed kinase is significantly more stable than our previous ddT kinase.

Even though our final design RosD4-(E172I V175W) does integrate an additional mutation not identified by Rosetta, the computational design made major contributions towards the functional changes. In regards to catalytic efficiency (k_{cat}/K_M), the overall 2000-fold decline for thymidine can be attributed to an 80-fold drop upon substitution of positions 66, 70, and 175 and a 26-fold reduction upon introduction of E172I. For ddT, the 4.6-fold improvement in k_{cat}/K_M is almost exclusively the contribution of the three designed changes.

The computational method provides an alternative solution enabled rapid generation of a stable, orthogonal ddT kinase in a shorter time period. The beneficial mutations predicted by Rosetta was not found in previous directed evolution experiments, which

reflects the restrictions of directed evolution approach because of practical limitations in covering all possible variants at amino acid level. On the other hand, a distinct difference between directed evolution approaches and computational approaches is using ΔΔG criteria ensures a high level of stability, however it can overlook advantageous mutations for catalytical substrate specificity such as E172I because of the mutant's impact on stability. Future efforts to refine Rosetta to differentiate stability from activity will help advance the approach. As well directed evolution strategies that screen for stability will aid in the selection of desired variants with high stability. While future structural studies of selected engineered variants will be helpful to validate and refine our model predictions, the combination of computational design and directed evolution provides more efficient and thorough exploration of variability in protein sequence that could benefit our effort to develop orthogonal NA kinases.

4.4 Materials and methods

4.4.1Materials

All primers for cloning were from Integrated DNA Technologies (Coralville, IA). *Pfu* Turbo DNA polymerase (Strategene, La Jolla, CA) was used for all cloning. Restriction enzymes were purchased from New England Biolabs (Ipswich, MA) unless otherwise indicated. Pyruvate kinase and lactate dehydrogenase were purchased from Roche Biochemicals (Indianopolis, IN). Ampicillin and Chloramphenicol were purchased from Fisher Scientific (Fair Lawn, NJ). All other reagents were purchased from Sigma-Aldrich

(St. Louis, MO) unless otherwise indicated.

4.4.2 Protein expression and purification

The genes of interest are cloned into pMAL-C2x (New England Biolabs), then transformed into *E. coli* K12 TB1. Protein expression was carried on at 25ºC. A single colony was picked for 2 ml overnight culture of LB media with 100 μM Ampicillin present. Then 100 ml LB media was inoculated with 1 ml of the starting culture. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to final concentration 0.3 μM when OD_{600} reached ~0.5. The expression culture grew for another 12 h before cells were harvested by centrifugation.

Proteins were purified using amylose resin (New England Biolabs) following the manufacturer's guide. Protein purity was verified by SDS-PAGE to be >95%. Protein concentrations were quantified by measuring A280 (MBP-*Dm*dNK, ξ= 106,230 $M^{-1}cm^{-1}$, calculated according to (PACE *et al.*, 1995)). Protein aliquots were stored at -80ºC.

4.4.3 Steady-state kinetic assays

A spectrophotometric coupled-enzyme assay was used to measure the phosphorylation activity for nucleosides and ddT (Valentin-Hansen, 1978). Assays were carried on at 37 ºC, in a 500 μl reaction mixture containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2.5 mM MgCl2, 0.18 mM NADH, 0.21 mM phosphoenolpyruvate, 1 mM

ATP, 1 mM 1, 4-dithio-DL-threitol, 30 U / ml pyruvate kinase, 33 U/ml lactate dehydrogenase, and substrate with concentrations ranging from 1 μM to 7 mM. Measurements were made in triplicates. Data were fit to the *Michaelis-Menten* equation using Origin (OriginLab, Northhampton, MA) to get apparent K_M and V_{max}, and then k_{cat} is calculated from V_{max}.

4.4.4 Enzyme thermostability assay

Each protein sample was divided into two tubes and one was incubated at 37°C for 10 min while the other one was kept on ice, then V_{max} was detected for both samples with ddT (700 μM) and the residual activity upon incubation was calculated.

Chapter 5. Conclusions and Perspectives

5.1 Summary

Orthogonal nucleoside analog (NA) kinases are ideal enzymes to activate NA prodrugs in vivo for gene therapy approaches to cancer and viral infection treatment. Limited success has been achieved in engineering NA kinases largely due to the lack of selection/screening strategy that can directly monitor NA phosphorylation. In this dissertation, I have designed and implemented a new FACS-based screening strategy for directed evolution of efficient and specific NA kinases. The protocol was named FLUorescent nucleoside analog Phosphorylation Screen (FLUPS). The concept of the design is based on three aspects: (a) nucleosides and NAs can be efficiently transported across the cell membrane by membrane nucleoside transporters while their corresponding monophosphates, the products of the kinase-catalyzed phosphorylation would be trapped in the cellular compartment and (b) minor modification on the nucleobase could generate fluorescent NAs (fNAs) with an excitation maxima of >300 nm that serve as detectable substrates for kinase engineering. The method was validated using fluorescent version of thymidine (fT) in enrichment experiments on two kinases: the 2'-deoxynucleoside kinase from *Drosophila melanogaster* (*Dm*dNK) and human 2'-deoxycytidine kinase (hdCK). The former is one of the most active thymidine kinases and the latter has little activity for thymidine. In this preliminary study, *Dm*dNK was successfully enriched by >100-fold using the FACS-based screening. Next, a fluorescent version 2'-3'-deoxythymidine (fddT)

was used to identify deoxynucleoside kinase variants with favorable ddT substrate specificity. My experiment yielded an orthogonal ddT kinase (R4.V3-T85) with 6-fold higher catalytic efficiency and 20,000-fold specificity change for (R4.V3-T85) relative to *Dm*dNK. The results emphasize the benefit of our new screening strategy as substitutions in two key amino acid positions (172 and 179) have never been observed due to the previous auxotrophic selection system's bias towards maintaining thymidine kinase activity. Additionally, despite only moderate ddT activity gains, the true advantage of NA kinase orthogonality was observed in vivo. The enzyme's altered substrate profile minimized binding competition between ddT and the native 2'-deoxynucleosides, and was reflected in significantly elevated fluorophore accumulation and greater ddT cytotoxicity in bacteria and mammalian cell cultures. Finally, to explore and develop alternative strategies for NA kinase engineering, the computational program Rosetta (Das *et al.*, 2008) was applied to predict *Dm*dNK mutants with altered substrate specificity to make ddT kinases. Combining Rosetta prediction with beneficial mutation found in directed evolution generated a ddT kinase with comparable catalytic performance to R4.V3-T85 with the additional benefit of higher protein thermostability. Additionally, application of Rosetta accelerated the engineering process, leading to a desired ddT kinase in a shorter time period than using directed evolution alone.

In summary, directed evolution using the newly developed FACS-based screen and Rosetta's computational approach provided different solutions to alter *Dm*dNK into

orthogonal ddT kinases. In each case variants were not found in previous effort of engineering NA kinases. While each strategy shows its advantages and limitations, the combined approach including both directed evolution and Rosetta design would enable more efficient NA kinase engineering. The obtained orthogonal NA kinases present promising candidates for antiviral and cancer therapy. Furthermore, they helped to reveal the structure-function relationships within the family of deoxynucleoside kinases that should also provide new insights in nucleotide metabolism, as well as NA prodrug activation and design.

5.2 Comparison of our engineered kinases to previous evolution results

When compared to kinase variants from previous engineering experiments, the orthogonal ddT kinases found in my directed evolution experiments show clear advantage both in NA activity and substrate specificity. The top variants of HSV-thymidine kinase (Christians *et al.*, 1999; Kokoris *et al.*, 2002) and *Dm*dNK (Knecht *et al.*, 2000; Knecht *et al.*, 2002; Knecht *et al.*, 2007) typically show <2-fold gains in k_{cat}/K_M for the NA, while the best *Dm*dNK mutant from the FLUPS screening has 6-fold higher ddT activity than the parental enzyme. More importantly, mutants selected by genetic complementation using *E. coli* auxotroph strain KY895 or cytotoxicity screening generally retain high activity for the native substrates. The HSV-thymidine kinase variant with highest NA specificity show a preference for NA over thymidine (k_{cat}/K_M (NA) / k_{cat}/K_M (thymidine)) of 0.63 for AZT (Christians *et al.*,

1999) and 0.17 for ganciclovir (Kokoris *et al.*, 2002). For *Dm*dNK, the highest NA specificity reported is 0.14 for AZT versus thymidine (Knecht *et al.*, 2000) and 0.19 for ddC versus thymidine as well (Knecht *et al.*, 2007). In contrast, our final variant R4.V3-T85 shows 20-fold higher catalytic efficiency for ddT compared to thymidine, plus the K_M values for natural nucleosides are all >100 μM, well beyond the physiological concentration (<10 μM), making the mutant a truly orthogonal ddT kinase in vivo. This result further supports the benefit of applying FLUPS in directed evolution of specific and efficient NA kinases.

5.3 The advantages of using orthogonal enzymes for gene therapy

One of the most intriguing discoveries in this research project is the clear advantage of using orthogonal other than promiscuous enzymes in vivo. Even though R4.V3 showed similar level of ddT activity than *Dm*dNK, when *E. coli* bacteria expressing R4.V3 were exposed to fddT in presence of 10-fold excess thymidine, the cells selectively accumulated fddT-monophosphates. In contrast, cells expressing *Dm*dNK show only background fluorescence (section 3.2.5). Furthermore, R4.V3 showed higher apparent catalytic efficiency in vivo, benefiting from its orthogonality.

In mammalian cell lines, the expression of R4.V3 also increased the potency of ddT 2-10 fold versus control experiment with *Dm*dNK (section 3.3.6), further supporting the advantages of using orthogonal NA kinases for cellular NA phosphorylation. More

detailed research needs to be performed to better understand the mechanism of how broad substrate specificity kinase might interfere with cellular nucleoside/nucleotide metabolism, and how a high specificity NA kinase accomplishes more efficient NA activation in living cells. To investigate the potential problems with using broad substrate specificity dNKs for gene therapy, we can study the influence of exogenous kinase expression on the cellular nucleoside metabolites. Three cases should be compared to determine the difference: no exogenous enzyme expression as the background control, *Dm*dNK as the broad substrate specificity enzyme, orthogonal ddT kinase R4.V3 as the specific kinase. Upon protein expression in mammalian cell lines, the cellular nucleoside metabolites (2'-deoxynucleosides, 2'-deoxynucleoside monophosphates, diphosphates and triphosphates) could be quantified by HPLC-MS (Dipierro *et al.*, 1995).

5.4 Extension of FACS-based screening to other NAs

The FACS-based approach is a highly versatile tool for tailoring the substrate specificity in deoxynucleoside kinases. Fluorescent versions of all four nucleobases such as furano- and pyrrolo- pyrimidines, as well as pterines in combination with a sugar derivative of choice are synthetically accessible and can serve as target substrates for FACS-based screening (Figure 1-8). Moreover, the furano- moiety on fddT led to the T85M mutation in the enzyme to accommodate the extended base. NAs with even smaller modifying groups to enable fluorescent detection of the compounds could further minimize the interference of the additional moiety on directed evolution results. For

example, Lin and colleagues developed photoactivated 1,3-dipolar cycloaddition reaction between a photoactivatable diaryltetrazole and alkene-containing protein and successfully used the "photoclick chemistry" to label protein in *E. coli* cells (Song *et al.*, 2008; Song *et al.*, 2008; Wang *et al.*, 2008). Alternatively, if we replace the 5-methyl- group on thymidine with an allyl moiety, upon kinase-catalyzed phosphorylation, the photoclick chemistry could generate the corresponding fluorescence product from 5-allyl-deoxyuridine monophosphate trapped in the host cells. The fluorescent compound therefore can serve as a signal molecule in FACS-based screening for directed evolution of NA kinases (Figure 5-1). Another advantage of this strategy is less limitation with choice of fluorophores, so that we could potentially use fluoroscent compounds with more red-shifted spectral properties to minimize interferences from cellular fluorescence background, as well as higher quantum yield to increase the sensitivity of the screening protocol.

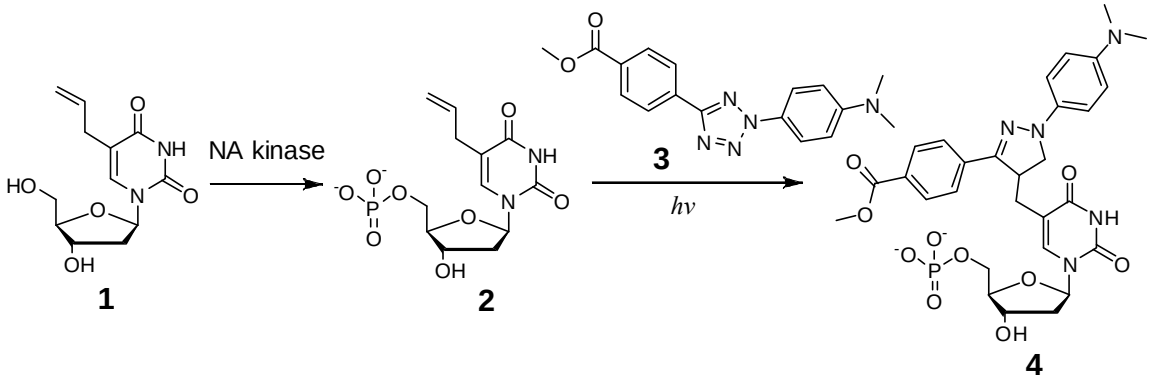


Figure 5-1: Scheme of using fluorescence tag added by photoclick chemistry to monitor NA phosphorylation catalyzed by corresponding kinases. First, 5-allyl-deoxyuridine (**1**) gets phosphorylated by an active kinase variant in the host cell, resulting in the entrapment of the monophosphate (**2**). Then 2-(4'-methoxyphenyl)-5-4''-dimethylamino-phenyl)tetrazole (**3**) is added to the cells and react with the alkene of the NA through photoactivated 1,3-dipolar cycloaddition. The entrapped fluorescent product (**4**) can be used as a quantitative signal to track phosphorylation of the NA.

On the other hand, NAs with modified sugar moiety other than 2'3'-dideoxyribose in ddT would make substrates with higher potential for directed evolution of NA kinases with more catalytic efficiency gain. We speculate that the reason for the moderate increase in k_{cat}/K_M for ddT in either directed evolution or computational design is the lack of a functional group at the 3'-position on the ribose moiety of ddT, providing no functional group for the enzyme to evolve tighter binding. An additional functional group might enlarge the opportunity for kinase variants to evolve and generate specific interactions with desired substrate, resulting in more efficient phosphoryl transfer. In regard of which chemical group to use at the 3'-position to substitute the native hydroxyl, thiols, amines and aldehydes are ruled out because of potential stability issue when used

in vivo. In contrast, methoxy, hydroxyl-methyl, methyl-keto provide good candidates due to chemical stability (Figure 5-2). In principal, natural dNKs would not accommodate these NAs as well because of the extra bulkiness of the substituents while kinase variants through directed evolution, might build up hydrogen bond interaction with the substituting functional group at 3'- position, becoming NA kinases.

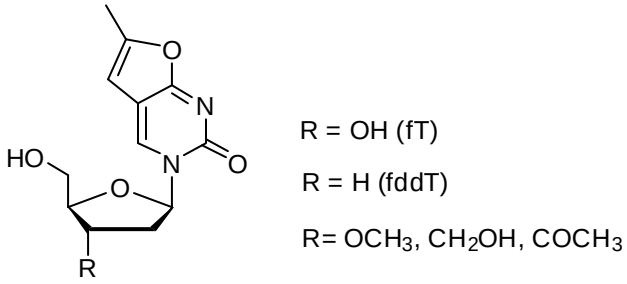


Figure 5-2: Alternative fNAs that could be used in FLUPS for directed evolution of NA kinases.

5.5 Engineer human kinases

Our first directed evolution experiment using the FACS-based screening was performed using *Dm*dNK as the starting enzyme to take advantage of its high catalytic activity and broad substrate specificity. Also, variants of *Dm*dNK have been previously tested in mammalian cell line and increased the potency of NA prodrugs (Zheng *et al.*, 2000; Knecht *et al.*, 2007; Solaroli *et al.*, 2007). However, a human enzyme would be more suitable for gene therapy as it lowers the risk of immunogenicity.

Two different approaches could be taken to engineer human nucleoside kinases. First, the FACS-based screening protocol could be used on libraries of human kinases for directed evolution of NA kinases. Second, since the human enzymes are generally less stable therefore pose challenges in directed evolution, mutations in *Dm*dNK dicovered in our evolution experiments may be introduced into human kinases (2'-deoxycytidine kinase (hdCK), 2'-deoxyguanosine kinase (hdGK) and thymidine kinase-2 (hTK2)) to produce NA specific kinases. These three human enzymes together with *Dm*dNK all belong to the type-I nucleoside kinase family based on their structural similarities (Munch-Petersen *et al.*, 2000; Eriksson *et al.*, 2002). Type-1 dNKs share the same fold, but vary widely in their substrate specificities and overall catalytic efficiencies. Despite their functional differences, the active site residues of these enzymes are highly conserved, especially the amino acid residue responsible for the ribose moiety recognition (Johansson *et al.*, 2001; Eriksson *et al.*, 2002; Gerth *et al.*, 2007; Iyidogan *et al.*, 2008). Therefore, the human enzymes could be transformed into NA kinases by introducing mutations found in *Dm*dNK that benefit NA specificity and activity while excluding natural nucleosides.

5.6 Combining computational design with directed evolution for kinase engineering

Applying Rosetta calculation to altering the substrate specificity of *Dm*dNK provided a distinct solution from our directed evolution experiments. While both approaches were able to generate NA kinases with desired specificity and catalytic efficiency, an approach

merging the two strategies could potentially present a powerful and efficient tool for evolving NA kinases. To facilitate FACS sorting, it is essential for the parental kinases to have some residual activity for the desired NA. Rosetta has proved its capability in setting up a sound starting point by de novo design for further directed evolution to improve enzyme catalytical efficiency (Rothlisberger *et al.*, 2008). Functional selection/screening on Rosetta design could lead the evolution toward desired catalytical activity. Overall, directed evolution discoveries could feedback to the computational algorithm for further improvement in design of enzyme activity and substrate specificity.

References

Aharoni, A., K. Thieme*, et al.* (2006). "High-throughput screening methodology for the directed evolution of glycosyltransferases." Nature Methods **3**(8): 609-614.

Al-Madhoun, A. S., J. Johnsamuel*, et al.* (2004). "Evaluation of Human Thymidine Kinase 1 Substrates as New Candidates for Boron Neutron Capture Therapy." Cancer Res **64**(17): 6280-6286.

Arnér, E. S. J. and S. Eriksson (1995). "Mammalian deoxyribonucleoside kinases." Pharmacology & Therapeutics **67**(2): 155-186.

Balzarini, J., G. Kang*, et al.* (1987). "The anti-HTLV-III (anti-HIV) and cytotoxic activity of 2',3'-didehydro- 2',3'-dideoxyribonucleosides: a comparison with their parental 2',3'-dideoxyribonucleosides." Mol Pharmacol **32**(1): 162-167.

Barrio, J. R., J. A. Secrist III*, et al.* (1972). "Fluorescent adenosine and cytidine derivatives." Biochemical and Biophysical Research Communications **46**(2): 597-604.

Bergstrom, D. E., H. Inoue*, et al.* (1982). "Pyrido[2,3-d]pyrimidine nucleosides. Synthesis via cyclization of C-5-substituted cytidines." The Journal of Organic Chemistry **47**(11): 2174-2178.

Bernhardt, P. and S. E. O'Connor (2009). "Opportunities for enzyme engineering in natural product biosynthesis." Current Opinion in Chemical Biology Biocatalysis and Biotransformation/Bioinorganic Chemistry **13**(1): 35-42.

Berry, D. A., K.-Y. Jung*, et al.* (2004). "Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2'-deoxycytidine for the study of oligonucleotides." Tetrahedron Letters **45**(11): 2457-2461.

Bonate, P. L., L. Arthaud*, et al.* (2006). "Discovery and development of clofarabine: a nucleoside analogue for treating cancer." Nature Reviews Drug Discovery **5**(10): 855-863.

Boryski, J., B. Golankiewicz*, et al.* (1988). "Synthesis and antiviral activity of novel N-substituted derivatives of acyclovir." Journal of Medicinal Chemistry **31**(7): 1351-1355.

Christians, F. C., L. Scapozza*, et al.* (1999). "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling." Nature Biotechnology **17**(3): 259-264.

Crumpacker, C. S. (1996). "Ganciclovir." N Engl J Med **335**(10): 721-729.

Culver, K., Z. Ram*, et al.* (1992). "In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors." Science **256**(5063): 1550-1552.

Curbo, S. and A. Karlsson (2006). "Nelarabine: a new purine analog in the treatment of hematologic malignancies." Rev Recent Clin Trials **1**(3): 185-92.

Dahiyat, B. I. and S. L. Mayo (1997). "De Novo Protein Design: Fully Automated Sequence Selection." Science **278**(5335): 82-87.

Das, R. and D. Baker (2008). "Macromolecular Modeling with Rosetta." Annual Review of Biochemistry **77**(1): 363-382.

Dent, S., H. Messersmith*, et al.* (2008). "Gemcitabine in the management of metastatic breast cancer: a systematic review." Breast Cancer Res Treat **108**(3): 319-31.

Dickson, R. B., S. E. Bates*, et al.* (1986). "Characterization of Estrogen Responsive Transforming Activity in Human-Breast Cancer Cell-Lines." Cancer Research **46**(4): 1707-1713.

Dipierro, D., B. Tavazzi*, et al.* (1995). "An Ion-Pairing High-Performance Liquid-Chromatographic Method for the Direct Simultaneous Determination of Nucleotides, Deoxynucleotides, Nicotinic Coenzymes, Oxypurines, Nucleosides, and Bases in Perchloric-Acid Cell-Extracts." Analytical Biochemistry **231**(2): 407-412.

Eagle, H. (1955). "Propagation in a Fluid Medium of a Human Epidermoid Carcinoma, Strain-Kb." Proceedings of the Society for Experimental Biology and Medicine **89**(3): 362-364.

Eriksson, S., B. Munch-Petersen*, et al.* (2002). "Structure and function of cellular deoxyribonucleoside kinases." Cell Mol Life Sci **59**(8): 1327-46.

Farinas, E. T. (2006). "Fluorescence Activated Cell Sorting for Enzymatic Activity." Combinatorial Chemistry; High Throughput Screening **9**: 321-328.

Fisher, M. A., P. N. S. Yadav*, et al.* (1994). "Identification of a pharmacophore for nucleoside analog inhibitors directed at HIV-1 reverse transcriptase." Journal of

Molecular Recognition **7**(3): 211-214.

Galmarini, C. M., L. Jordheim*, et al.* (2003). "Pyrimidine nucleoside analogs in cancer treatment." Expert Review of Anticancer Therapy **3**(5): 717-728.

Gentry, G. A. (1992). "Viral thymidine kinases and their relatives." Pharmacology & Therapeutics **54**(3): 319-355.

Gerlt, J. A. and P. C. Babbitt (2009). "Enzyme (re)design: lessons from natural evolution and computation." Current Opinion in Chemical Biology Biocatalysis and Biotransformation/Bioinorganic Chemistry **13**(1): 10-18.

Gerth, M. L. and S. Lutz (2007). "Non-homologous Recombination of Deoxyribonucleoside Kinases from Human and Drosophila melanogaster Yields Human-like Enzymes with Novel Activities." Journal of Molecular Biology **370**(4): 742-751.

Givan, A. L. (1992). Flow Cytometry: First Principles. New York, Wiley-Liss.

Godsey, M. H., S. Ort*, et al.* (2006). "Structural Basis for the Preference of UTP over ATP in Human Deoxycytidine Kinase: Illuminating the Role of Main-Chain Reorganization." Biochemistry **45**(2): 452-461.

Griffiths, A. D. and D. S. Tawfik (2006). "Miniaturising the laboratory in emulsion droplets." Trends in Biotechnology **24**(9): 395-402.

Hawkins, M. E. (2007). "Synthesis, purification and sample experiment for fluorescent

pteridine-containing DNA: tools for studying DNA interactive systems." Nature Protocols **2**(4): 1013-1021.

Hawkins, M. E., W. Pfleiderer*, et al.* (1997). "Fluorescence Properties of Pteridine Nucleoside Analogs as Monomers and Incorporated into Oligonucleotides." Analytical Biochemistry **244**(1): 86-95.

Hengstschlager, M. and E. Wawra (1993). "Cytofluorometric assay for the determination of thymidine uptake and phosphorylation in living cells." Cytometry **14**(1): 39-45.

Horner MJ, Ries LAG*, et al.* (2009). "SEER Cancer Statistics Review, 1975-2006." Surveillance, Epidemiology and End Results (SEER).

Igarashi, K., S. Hiraga*, et al.* (1967). "A Deoxythymindine Kinase Deficient Mutant of Escherichia coli. II. Mapping and Transduction Studies with Phage phi80." Genetics **57**(3): 643-654.

Iyidogan, P. and S. Lutz (2008). "Systematic Exploration of Active Site Mutations on Human Deoxycytidine Kinase Substrate Specificityâ€ " Biochemistry **47**(16): 4711-4720.

Jean, J. M. and K. B. Hall (2001). "2-Aminopurine fluorescence quenching and lifetimes: Role of base stacking." Proceedings of the National Academy of Sciences of the United States of America **98**(1): 37-41.

Jiang, L., E. A. Althoff*, et al.* (2008). "De Novo Computational Design of Retro-Aldol Enzymes." Science **319**(5868): 1387-1391.

Johansson, K., S. Ramaswamy*, et al.* (2001). "Structural basis for substrate specificities of cellular deoxyribonucleoside kinases." Nat Struct Biol **8**(7): 616-20.

Jordheim, L. P., C. M. Galmarini*, et al.* (2006). "Gemcitabine resistance due to deoxycytidine kinase deficiency can be reverted by fruitfly deoxynucleoside kinase, DmdNK, in human uterine sarcoma cells." Cancer Chemotherapy and Pharmacology **58**(4): 547-554.

Khlebnikov, A., K. A. Datsenko*, et al.* (2001). "Homogeneous expression of the P_{BAD} promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter." Microbiology **147**(12): 3241-3247.

Khlebnikov, A., T. Skaug*, et al.* (2002). "Modulation of gene expression from the arabinose-inducible araBAD promoter." Journal of Industrial Microbiology & Biotechnology **29**(1): 34-37.

Knecht, W., B. Munch-Petersen*, et al.* (2000). "Identification of residues involved in the specificity and regulation of the highly efficient multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster." Journal of Molecular Biology **301**(4): 827-837.

Knecht, W., E. Rozpedowska*, et al.* (2007). "Drosophila deoxyribonucleoside kinase mutants with enhanced ability to phosphorylate purine analogs." Gene Ther **14**(17): 1278-86.

Knecht, W., M. P. B. Sandrini*, et al.* (2002). "A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines." Embo Journal **21**(7): 1873-1880.

Kokoris, M. S. and M. E. Black (2002). "Characterization of Herpes Simplex Virus type 1 thymidine kinase mutants engineered for improved ganciclovir or acyclovir activity." Protein Science **11**(9): 2267-2272.

Kortemme, T., L. A. Joachimiak*, et al.* (2004). "Computational redesign of protein-protein interaction specificity." Nat Struct Biol **11**(4): 371-379.

Kristian M. Müller, S. C. S., Susanne Knall, Gregor Zipf, Hubert S. Bernauer, Katja M. Arndt (2005). "Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution." Nucleic Acids Res. **33**(13): e117.

Kuhlman, B. and D. Baker (2000). "Native protein sequences are close to optimal for their structures." Proceedings of the National Academy of Sciences of the United States of America **97**(19): 10383-10388.

Lieber, M., G. Todaro*, et al.* (1976). "A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells." International Journal of Cancer **17**(1): 62-70.

Liu, C. H. and C. T. Martin (2002). "Promoter clearance by T7 RNA polymerase - Initial bubble collapse and transcript dissociation monitored by base analog fluorescence." Journal of Biological Chemistry **277**(4): 2725-2731.

Liu, L., Y. Li*, et al.* (2009). "Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting." Nucl. Acids Res. **37**(13): 4472-4481.

Lutz, S., L. Liu*, et al.* (2009). "Engineering kinases to phosphorylate nucleoside analogs for antiviral and cancer therapy." CHIMIA International Journal for Chemistry: accepted.

Lutz, S. and W. M. Patrick (2004). "Novel methods for directed evolution of enzymes: quality, not quantity." Current Opinion in Biotechnology **15**(4): 291-297.

Mathews, C. K. and S. Song (2007). "Maintaining precursor pools for mitochondrial DNA replication." FASEB J. **21**(10): 2294-2303.

Mikkelsen, N. E., K. Johansson*, et al.* (2003). "Structural Basis for Feedback Inhibition of the Deoxyribonucleoside Salvage Pathway: Studies of the Drosophila Deoxyribonucleoside Kinase." Biochemistry **42**(19): 5706-5712.

Moolten, F. L. (1986). "Tumor Chemosensitivity Conferred by Inserted Herpes Thymidine Kinase Genes: Paradigm for a Prospective Cancer Control Strategy." Cancer Res **46**(10): 5276-5281.

Munch-Petersen, B., W. Knecht*, et al.* (2000). "Functional Expression of a Multisubstrate Deoxyribonucleoside Kinase from Drosophila melanogaster and Its C-terminal Deletion Mutants." Journal of Biological Chemistry **275**(9): 6673-6679.

Munch-Petersen, B., J. Piskur*, et al.* (1998). "Four Deoxynucleoside Kinase Activities from Drosophila melanogaster Are Contained within a Single Monomeric Enzyme, a New Multifunctional Deoxynucleoside Kinase." Journal of Biological Chemistry **273**(7): 3926-3931.

Murphy, P. M., J. M. Bolduc*, et al.* (2009). "Alteration of enzyme specificity by computational loop remodeling and design." Proceedings of the National Academy of Sciences **106**(23): 9215-9220.

Neri, B., G. Cipriani*, et al.* (2009). "Gemcitabine plus irinotecan as first-line weekly therapy in locally advanced and/or metastatic pancreatic cancer." Oncol Res **17**(11-12): 559-64.

Novick, A. and M. Weiner (1957). "Enzyme Induction as an All-or-None Phenomenon." Proceedings of the National Academy of Sciences of the United States of America **43**(7): 553-566.

Olsen, M. J., D. Stephens*, et al.* (2000). "Function-based isolation of novel enzymes from a large library." Nature Biotechnology **18**(10): 1071-1074.

PACE, C. N., F. VAJDOS*, et al.* (1995). "How to measure and predict the molar absorption coefficient of a protein." Protein Sci **4**(11): 2411-2423.

Parker, W. B. (2009). "Enzymology of Purine and Pyrimidine Antimetabolites Used in the Treatment of Cancer." Chemical Reviews **109**(7): 2880-2893.

Rothlisberger, D., O. Khersonsky*, et al.* (2008). "Kemp elimination catalysts by computational enzyme design." Nature **453**(7192): 190-195.

Sabini, E., S. Ort*, et al.* (2003). "Structure of human dCK suggests strategies to improve anticancer and antiviral therapy." Nature Structural Biology **10**(7): 513-519.

Sandin, P., P. Lincoln*, et al.* (2007). "Synthesis and oligonucleotide incorporation of fluorescent cytosine analogue tC: a promising nucleic acid probe." Nature Protocols **2**(3): 615-623.

Schinazi, R. F., B. I. Hernandez-Santiago*, et al.* (2006). "Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses." Antiviral Res **71**(2-3): 322-34.

Secrist III, J. A., J. R. Barrio*, et al.* (1972). "A Fluorescent Modification of Adenosine Triphosphate with Activity in Enzyme Systems: 1,*N6*-Ethenoadenosine Triphosphate." Science **175**(4022): 646-647.

Seth, A. K., A. Misra*, et al.* (2005). "Topical Liposomal Gel of Idoxuridine for the Treatment of Herpes Simplex: Pharmaceutical and Clinical Implications." Pharmaceutical Development and Technology **9**(3): 277-289.

Shapiro, H. M. (2004). Practical Flow Cytometry. New York, Wiley-Liss.

Shi, J., J. J. McAtee*, et al.* (1999). "Synthesis and Biological Evaluation of 2',3'-Didehydro-2',3'-dideoxy-5- fluorocytidine (D4FC) Analogues: Discovery of Carbocyclic Nucleoside Triphosphates with Potent Inhibitory Activity against HIV-1 Reverse Transcriptase1." Journal of Medicinal Chemistry **42**(5): 859-867.

Solaroli, N., M. Johansson*, et al.* (2007). "Enhanced toxicity of purine nucleoside analogs in cells expressing Drosophila melanogaster nucleoside kinase mutants." Gene Ther **14**(1): 86-92.

Song, S., Z. F. Pursell*, et al.* (2005). "DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity." Proceedings of the National Academy of Sciences of the United States of America **102**(14): 4990-4995.

Song, W., Y. Wang*, et al.* (2008). "Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells." J Am Chem Soc **130**(30): 9654-5.

Song, W., Y. Wang*, et al.* (2008). "A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins." Angew Chem Int Ed Engl **47**(15): 2832-5.

Swords, R. and F. Giles (2007). "Troxacitabine in acute leukemia." Hematology **12**: 219-227.

Tao, J. and J.-H. Xu (2009). "Biocatalysis in development of green pharmaceutical processes." Current Opinion in Chemical Biology Biocatalysis and Biotransformation/Bioinorganic Chemistry **13**(1): 43-50.

Tiseo, M., M. Bartolotti*, et al.* (2009). "First-line treatment in advanced non-small-cell lung cancer: the emerging role of the histologic subtype." Expert Review of Anticancer Therapy **9**(4): 425-435.

Tracewell, C. A. and F. H. Arnold (2009). "Directed enzyme evolution: climbing fitness peaks one amino acid at a time." Current Opinion in Chemical Biology Biocatalysis and Biotransformation/Bioinorganic Chemistry **13**(1): 3-9.

Valentin-Hansen, P. (1978). "Uridine-cytidine kinase from Escherichia coli." Methods Enzymol **51**: 308-14.

Varadarajan, N., J. Gam*, et al.* (2005). "Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity." Proceedings of the National Academy of Sciences of the United States of America **102**(19): 6855-6860.

Wang, W., C. Tzeng*, et al.* (1997). "High-dose cytarabine and mitoxantrone as salvage therapy for refractory non-Hodgkin's lymphoma." Jpn. J. Clin. Oncol. **27**(3): 154-157.

Wang, Y. Z., W. J. Hu*, et al.* (2008). "Discovery of long-wavelength photoactivatable diaryltetrazoles for bioorthogonal 1,3-dipolar cycloaddition reactions." Organic Letters **10**(17): 3725-3728.

Waqar, M. A., M. J. Evans*, et al.* (1984). "Effects of 2',3'-dideoxynucleosides on mammalian cells and viruses." Journal of Cellular Physiology **121**(2): 402-408.

Welin, M., U. Kosinska*, et al.* (2004). "Structures of thymidine kinase 1 of human and mycoplasmic origin." Proceedings of the National Academy of Sciences of the United States of America **101**(52): 17970-17975.

Wong, T. S., D. Roccatano*, et al.* (2006). "A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution." Journal of Molecular Biology **355**(4): 858-871.

Wong, T. S., D. Zhurina*, et al.* (2006). "The Diversity Challenge in Directed Protein

Evolution." Combinatorial Chemistry & High Throughput Screening **9**: 271-288.

Yang, G. and S. G. Withers (2009). "Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution." ChemBioChem **9999**(9999): NA.

Yang, K. S., S. Matsika*, et al.* (2007). "6MAP, a fluorescent adenine analogue, is a probe of base flipping by DNA photolyase." Journal of Physical Chemistry B **111**(35): 10615-10625.

Yarchoan, R., H. Mitsuya*, et al.* (1989). "Clinical-Pharmacology of 3'-Azido-2',3'-Dideoxythymidine (Zidovudine) and Related Dideoxynucleosides." New England Journal of Medicine **321**(11): 726-738.

You, L. and F. H. Arnold (1996). "Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide." Protein Eng. **9**(1): 77-83.

Zhang, J., X. Sun*, et al.* (2006). "Studies of Nucleoside Transporters Using Novel Autofluorescent Nucleoside Probes." Biochemistry **45**(4): 1087-1098.

Zhao, H. and F. H. Arnold (1997). "Combinatorial protein design: strategies for screening protein libraries." Current Opinion in Structural Biology **7**(4): 480-485.

Zheng, X., M. Johansson*, et al.* (2000). "Retroviral Transduction of Cancer Cell Lines with the Gene Encoding Drosophila melanogaster Multisubstrate Deoxyribonucleoside Kinase." Journal of Biological Chemistry **275**(50): 39125-39129.

Zhu, C., M. Johansson*, et al.* (1998). "Enhanced Cytotoxicity of Nucleoside Analogs by

Overexpression of Mitochondrial Deoxyguanosine Kinase in Cancer Cell Lines." Journal of Biological Chemistry **273**(24): 14707-14711.