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.

_____   _____

Pinar Iyidogan   Date

Engineering The Substrate Specificity Of Human Deoxycytidine Kinase

By

Pinar Iyidogan
Doctor of Philosophy

Chemistry

_____
Stefan Lutz, Ph.D.
Advisor

_____
Dennis C. Liotta, Ph.D.
Committee Member

_____
Vincent P. Conticello, Ph.D.
Committee Member

Accepted:

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

_____
Date

Engineering The Substrate Specificity Of Human Deoxycytidine Kinase

By

Pinar Iyidogan
B.S., Hacettepe University, 1999
M.S., Hacettepe University, 2002

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

2008

Abstract

Engineering The Substrate Specificity Of Human Deoxycytidine Kinase

By Pinar Iyidogan

Human deoxycytidine kinase (dCK) is responsible for the phosphorylation of a number of clinically important nucleoside analog (NA) prodrugs in addition to its natural substrates, 2′-deoxycytidine, 2′-deoxyguanosine, and 2′-deoxyadenosine. To improve the catalytic activity and tailor the substrate specificity of dCK, libraries of mutant enzymes were constructed and tested for thymidine kinase activity. Random and site-saturation mutagenesis were employed to probe for residue positions with an impact on substrate specificity, identifying positions Arg104 and Asp133 in the active site as key residues for substrate specificity. The results illuminate the key contributions of these two amino acids to enzyme function by demonstrating their ability to moderate substrate specificity. Clinically relevant NAs mostly differ from their endogenous counterparts in respect to various modifications at the ribose moiety. To determine the molecular aspects of modified ribose binding, site-saturation mutagenesis was utilized at six amino acid positions in the active site of dCK. All the library members were evaluated by an *in vivo* screening to find a variant with improved turnover rate for AZT. Kinetic analysis of the selected mutants indicates that the enhanced AZT activity could not be fulfilled with the chosen methodology. Therefore, novel screening methods with positive selection should be developed for further engineering with improved NA phosphorylation. Besides the active site residues, there are other structural elements in the dCK structure that

determines the substrate specificity. To test this hypothesis, chimeragenesis was utilized using dCK and the C-terminal region of *Dm*-dNK to probe the role of phosphoryl donor binding loop in regards to phosphoryl donor and acceptor specificity in dCK. Kinetic analysis of the chimeras indicates that the phosphoryl donor preference was reversed from UTP to ATP as a result of swapped subdomains from *Dm*-dNK. In order to further investigate the function of donor base-sensing loop, two specific amino acid positions Asp241 and Phe242 were selected for alanine scanning mutagenesis in dCK. The results elucidate the sequence dependency of phosphoryl donor preference and the structure-function relationship in this small phosphoryl donor binding loop region.

Engineering The Substrate Specificity Of Human Deoxycytidine Kinase

By

Pinar Iyidogan
B.S., Hacettepe University, 1999
M.S., Hacettepe University, 2002

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

2008

ACKNOWLEDGEMENTS

to the most important and inspirational people in my life: My mom and dad.

In the past months and years I have had both wonderful and tragic experiences in my life,

and all of you made those experiences very memorable by always being there for me.

Sincerely,

Pinar Iyidogan

TABLE OF CONTENTS Page

LIST OF FIGURES

LIST OF TABLES

Abbreviations

AraC   1-β-D-arabinosylcytosine

ATP   Adenosine triphosphate

AZT   3′-azido-3′deoxythymidine

CD   Circular dichroism

dA   2′-deoxyadenosine

dC   2′-deoxycytidine

dCK   Deoxycytidine kinase

dFdC   2′,2′-difluorodeoxycytidine

dG   2′-deoxyguanosine

Dm-dNK   *Drosophila melanogaster* deoxyribonucleoside kinase

dN   2′-deoxyribonucleosides

DNA   Deoxyribonucleic acid

dNK   Deoxyribonucleoside kinases

dNTP   2′-deoxyribonucleoside triphosphates

DTT   Dithiothreitol

HSV1-TK   Herpes simplex type 1 thymidine kinase

hTK2   Human thymidine kinase 2

IPTG   Isopropyl-β-D-thiogalactopyranoside

MgCl$_2$ Magnesium (II) chloride

NA   Nucleoside analog

NaCl   Sodium chloride

NTP	Ribonucleoside triphosphates
PCR	Polymerase chain reaction
SDS	Sodium dodecyl sulfate
T	Thymidine
tk	Thymidine kinase
UTP	Uridine triphosphate

Chapter 1

General Introduction

1. Introduction

1.1 Protein engineering

Protein engineering constitutes a powerful tool to generate novel proteins that serve for the practical applications in the fields of medicine, enzymology, food and chemical industry [1]. The desired properties of the target protein with altered catalytic activity, substrate specificity and protein stability are attained by using rational and random methods [2].

Rational approaches involve replacing amino acids in an enzyme that are selected based upon sequence and structural knowledge with the use of site-directed mutagenesis. Despite of the limited database on detailed protein structures and difficulties in predicting the specific substitutions for the preferred outcome, rational design still remains a potent technique for the dissection of the individual contributions of amino acid residues and to test the proposed hypothesis.

Various procedures have been developed for random methods that could be divided into two main groups as random mutagenesis and directed evolution. Irrespective of the prior three-dimensional structural knowledge, randomization of the entire gene of interest by error-prone PCR and/or recombination of one or more parent genes for diversity is followed by iterative rounds of screening and selection to identify enzyme variants with novel properties and/or function. Indeed, the most promising profiles are generally the result of multiple amino acid substitutions and replacements far from the active site that cause subtle changes in the structure can have a large and unpredictable impact on the

desired function. The major challenge for these approaches is to develop effective screening and selection methods for large libraries.

Rational, random and evolution-based methods have all been employed successfully, yet combination of different approaches is the best tactic for rapid progress. In this dissertation, we employed both random and rational methods to engineer the human deoxycytidine kinase with altered substrate specificity and improved catalytic activity towards therapeutical substrates.

1.2 Synthesis of endogenous DNA precursors

In all organisms, 2′-deoxyribonucleoside triphosphates (dNTPs) are building blocks of DNA and required for the DNA replication and repair. Mammalian cells utilize two routes to synthesize dNTPs, the *de novo* pathway and the salvage pathway.

De novo versus salvage pathway

The major dNTP supply is provided through the *de novo* pathway. All four ribonucleotides are synthesized from small molecules (amino acids, ribose-5′-phosphate, CO_2 and NH_3) to ribonucleoside monophosphates (NMPs) and subsequently phosphorylated to ribonucleoside diphosphates (NDPs). At this stage, ribonucleotide reductase (RNR) converts NDPs to deoxyribonucleoside diphosphates (dNDPs) via reducing the 2′-hydroxyl group of the ribose to the corresponding dNDP. The (deoxy)ribonucleoside diphosphates get phosphorylated to triphosphates (NTPs and dNTPs) by nonspecific nucleoside diphosphate kinases. Unlike *de novo* pathway, the salvage pathway recycles the free 2′-deoxyribonucleosides (dNs) released from DNA

degradation (intracellular) or food digestion (extracellular) back to deoxyribonucleotides. Nucleoside carrier proteins can transport dNs into the cell by facilitated diffusion or active transport. Deoxyribonucleoside kinases (dNKs) then catalyze the initial phosphorylation of a deoxyribonucleoside to a deoxyribonucleoside monophosphate (dNMP), which is considered as the key rate-limiting step since the phosphorylated nucleotides are trapped in the cell due to their negative charge [3]. Deoxyribonucleoside monophosphate kinases (dNMPK) and nucleoside diphosphate kinases (NDPK) convert the synthesized dNMPs into diphosphates (dNDPs) and triphosphates (dNTPs), respectively. The resulted dNTPs from salvage pathway are substrates for nuclear DNA repair and mitochondrial DNA synthesis.

Figure 1.1   Simplified overview of salvage and *de novo* pathway.
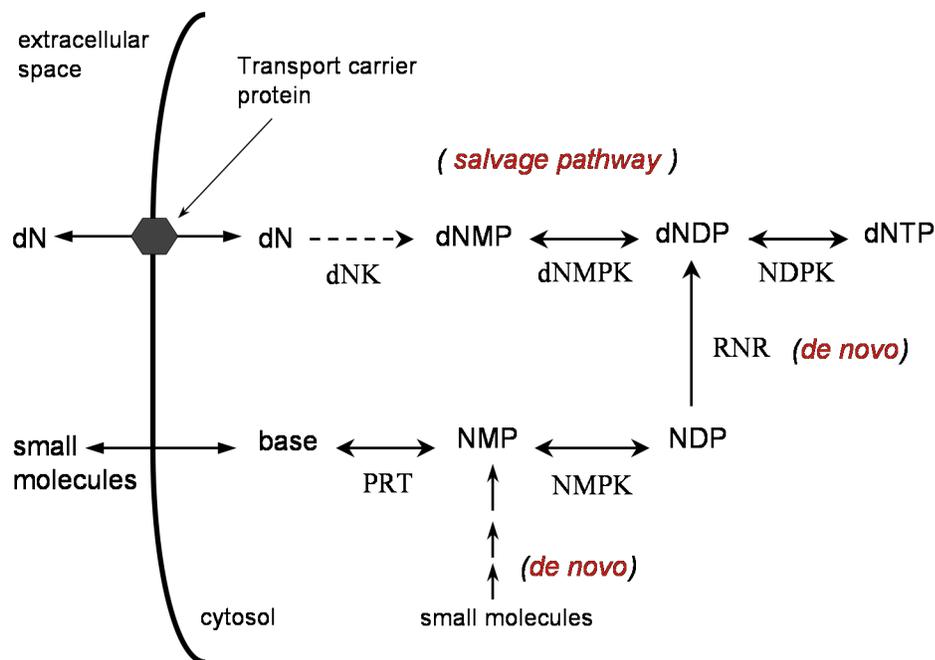
After the uptake of dNs into the cell by nucleoside carrier proteins, dNKs catalyze the initial phosphorylation step. (d)NMPKs and NDPKs catalyze the subsequent phosphorylation steps of both the *de novo* and salvage pathways (PRT indicates phosphoribosyltransferases).

1.3 Deoxyribonucleoside kinases

The deoxyribonucleoside kinases (dNKs) are often the rate-determining enzymes in the salvage pathway and performing the transfer of the phosphoryl group from a nucleoside triphosphate (generally ATP as donor) to dNs. During this process, a highly conserved glutamic acid (Glu) is suggested to be involved in the catalysis. The Glu residue acts as a general base and deprotonates the 5′-OH of the dN, activating the oxygen to attack nucleophilically the γ-phosphate of the phosphate donor to give dNMPs and ADP [4]. Several positively charged residues (e.g. Arginine) and magnesium ion (Mg^{2+}) are required for this reaction to stabilize the transition state. The mechanism for this initial phosphorylation step is shown in Figure 1.2.

Figure 1.2   General mechanism for deoxyribonucleoside kinases.
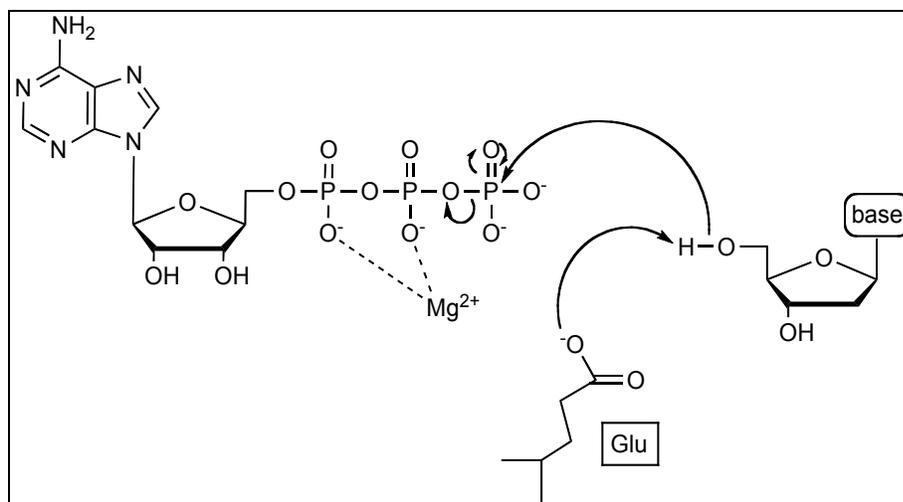
The figure has been modified from Eriksson *et al.*, 2002.

Deoxyribonucleoside kinases are divided into two subfamilies derived from the similarities in sequences and crystal structures; type 1 and type 2 kinases [5]. Type 1 family includes kinases from different origins such as human, insect, bacterial and viral

kinases. Type 1 dNKs are homodimeric proteins with broad substrate specificities and share distinct structural motifs. Several members of the type 1 family are described in detail in the following section. In contrast, type 2 kinases are homotetramers with narrower substrate specificity (e.g. Human TK1 and thymidine kinases from bacteria and viruses).

1.3.1 Human deoxyribonucleoside kinases

There are four deoxyribonucleoside-specific kinases in human cells; deoxycytidine kinase (dCK), and thymidine kinase 1 (TK1) are located in cytosol and the remaining two are mitochondrial enzymes, deoxyguanosine kinase (dGK), and thymidine kinase 2 (TK2) [3]. These enzymes have complementary substrate specificities with ability to phosphorylate all of the natural substrates in their respective cellular compartments. The basic properties of human dNKs are summarized in Table 1.1.

Table 1.1 Summary of the properties of human dNKs.

Property	*Type-1*			*Type-2*
	dCK	**dGK**	**TK2**	**TK1**
Substrate specificity	dC, dA, dG	dG, dA	T, dC, dU	T, dU
Subcellular localization	cytosol	mitochondria	mitochondria	cytosol
Molecular mass (kDa)	30	30	29	25
Oligomerization state	homodimer	homodimer	homodimer	homotetramer
Cell cycle regulation	constitutive expression	constitutive expression	constitutive expression	S-phase regulated

Basic properties of all four human dNKs are summarized related to their physiological environment [5].

Cytosolic deoxycytidine kinase, **dCK**, (EC 2.7.1.74) is constitutively expressed and present in high levels in lymphoid tissues [6, 7]. Human dCK consists of 260 residues and forms a homodimer of two 30 kDa subunits [8, 9]. The substrate specificity of dCK is broad including 2′-deoxycytidine (dC), 2′-deoxyadenosine (dA) and 2′-deoxyguanosine (dG) [8]. dCK phosphorylates its natural substrates with different specificities; the lowest K_M for dC, and much higher K_M values for both purines with catalytic efficiencies in the range of 2 to 6×10^4 $M^{-1}s^{-1}$ [5]. Any endogenous cellular nucleoside triphosphate, except dCTP, which is a feedback inhibitor, could function as a phosphate donor for dCK [10]. Additionally, a recent study reports that even inorganic triphosphate (PPP_i) is able to serve as a phosphate donor for dCK [11]. Previous investigations indicated that UTP is the preferred phosphate donor and more efficient than ATP [12, 13].

Cytosolic thymidine kinase, **TK1**, (EC 2.7.1.21) is strictly cell cycle regulated and expressed in all tissues except non-proliferating cells [3]. TK1 consists of 234 residues with a calculated molecular size of 25.5 kDa. The molecular size of native TK1 as found by Sherley *et al.* was 96 kDa, indicating a homotetramer [14]. The substrate specificity of TK1 is more restricted than the other salvage deoxyribonucleoside kinases. TK1 accepts thymidine (T) and 2′-deoxyuridine (dU) as substrates with the highest catalytic efficiency constant for its primary substrate, T, which is equal to 8×10^6 $M^{-1}s^{-1}$, about 12 times higher than that of dCK with dC. The preferred phosphate donors are ATP and dATP, while dTTP serves as the feedback inhibitor [15].

Mitochondrial thymidine kinase, **TK2**, is present in most tissues, in correlation with mitochondrial distribution in the tissue and expressed constitutively [3]. Although there is no crystal structure available the size of the cloned TK2 polypeptide is about 29 kDa and shown to be a homodimer in native conditions. TK2 has broader substrate specificity than TK1 by phosphorylating all pyrimidine dNs efficiently (T, dU, and dC) with the catalytic activities in the range of 10^5 $M^{-1}s^{-1}$ [5, 16]. ATP and CTP can be used as phosphate donors and dTTP serves as the feedback inhibitor as well as dCTP [17].

Mitochondrial deoxyguanosine kinase, **dGK**, (EC 2.7.1.113) is found in brain, muscle, liver, and resting and mitogen-stimulated lymphocytes [18]. dGK was cloned from a human brain cDNA library (30 kDa in size) including a N-terminal mitochondrial targeting signal peptide for mitochondrial import [19]. This enzyme accepts the naturally occurring purine dNs as substrates with low K_M values of 8 μM and 60 μM for dG and dA, respectively [3]. The k_{cat}/K_M value is 6 x 10^4 $M^{-1}s^{-1}$ for dG and 4 x 10^3 $M^{-1}s^{-1}$ for dA [5]. Similarly to dCK, the recombinant dGK can use both ATP and UTP, as well as the PPP_i as phosphate donors [11]. UTP improved the phosphorylation efficiency in comparison with ATP for all dGK substrates investigated [20]. The end products dGTP and dATP both serve as feedback inhibitors [21].

1.3.2 Multisubstrate insect deoxyribonucleoside kinases

Insects have only one dNK with ability to phosphorylate all natural dNs. The first multisubstrate kinase to be discovered and cloned was from fruit fly, *Drosophila melanogaster* (*Dm*-dNK; EC 2.7.1.145) [22, 23]. *Dm*-dNK is one the most efficient

kinase among all the known dNKs and phosphorylates pyrimidine and purine dNs. The end product of the best substrate for *Dm*-dNK has been shown to be the most potent feedback inhibitor, which is TTP [24]. Other insect kinases have also identified, such as *Anopheles gambiae* and *Bombyx mori* dNK, which utilize all dNs [25, 26].

Table 1.2   Summary of k_{cat}/K_M ($M^{-1}s^{-1}$) values for several dNKs.

		Pyrimidines		Purines	
		dC	T	dA	dG
Type 1	dCK	2×10^5	-	8×10^4	6×10^4
	dGK	-	-	4×10^3	6×10^4
	TK2	3×10^4	9×10^5	-	-
	Dm-dNK	1×10^7	2×10^7	9×10^4	2×10^4
Type 2	TK1	-	8×10^6	-	-

The catalytic activities are taken from a review by Eriksson *et al*. [5].

1.3.3   Viral thymidine kinases

Herpes simplex virus type 1 thymidine kinase (HSV1-TK) is a homodimer with a monomer size of 40 kDa [27]. HSV1-TK has a broad specificity like other type 1 kinases and phosphorylates T, dU and dC. Moreover, thymidine monophosphate (TMP) serves also as a substrate for HSV1-TK with thymidylate activity [27]. Other recently solved structure belongs to this family is the Varicella zoster virus TK (VZV-TK) [28]. VZV-TK is a deoxypyrimidine kinase with similar sequence patterns to HSV1-TK [29, 30]. Vaccinia virus TK (VV-TK) is a homotetramer with a molecular weight of 80 kDa and belongs to the type 2 family [31]. VV-TK has a narrower substrate specificity that is similar to TK1 with a preference towards T [32].

1.3.4 Bacterial deoxyribonucleoside kinases

Thymidine kinase from *Escherichia coli* was the first bacterial dNK to be discovered and characterized [33, 34]. Various Gram-positive bacterial TKs from *Ureaplasma urealyticum* (*Uu*-TK) and *Bacillus anthracis* (*Ba*-TK) were characterized which belong to the type 2 family [35, 36]. Several bacterial dNKs were discovered from *Bacillus subtilis*, one dAK/dCK with dA and dC specificity and the other with dGK activity [37].

A deoxyadenosine kinase from *Mycoplasma mycoides* subsp. *Mycoides* (*Mm*-dAK) has been kinetically characterized with dA and dG activity and its crystal structure represents similar properties to the type 1 kinase family [38].

1.4   Type 1 deoxyribonucleoside kinases

1.4.1   Basic structural properties of type 1 dNKs

The focus of this dissertation is primarily on dNKs that belong to the type 1 family. Recently, crystal structures of several type 1 dNKs have been solved and share a similar topology; a Rossmann fold as revealed in Figure 1.3. The architecture of each subunit is an α/β/α sandwich with a parallel multi-stranded β-sheet core surrounded by α-helices. Two helices from each monomer form a four-helix bundle to create a dimer interface.

The highly conserved three sequence motifs are important for the enzyme function; P-loop, LID region and ERS motif (Figure 1.3; PDB code: 1P5Z [39]). The consensus sequence of GXXXXGKS/T (X can be any residue), with threonine substituting serine in some cases, is known as P-loop (phosphate binding loop) or Walker motif A [40]. The P-loop starts after the β1 strand and followed by α1 helix, which binds and positions the α

and β-phosphoryl groups of the phosphoryl donor via backbone interactions involving the amide hydrogens and phosphate oxygens. In addition, the oxygen atom of the hydroxyl group from the conserved serine or threonine residue establishes the coordination of Mg^{2+} ion in combination with several water molecules [39].

Figure 1.3 General fold and structural motifs of Type 1 dNKs.
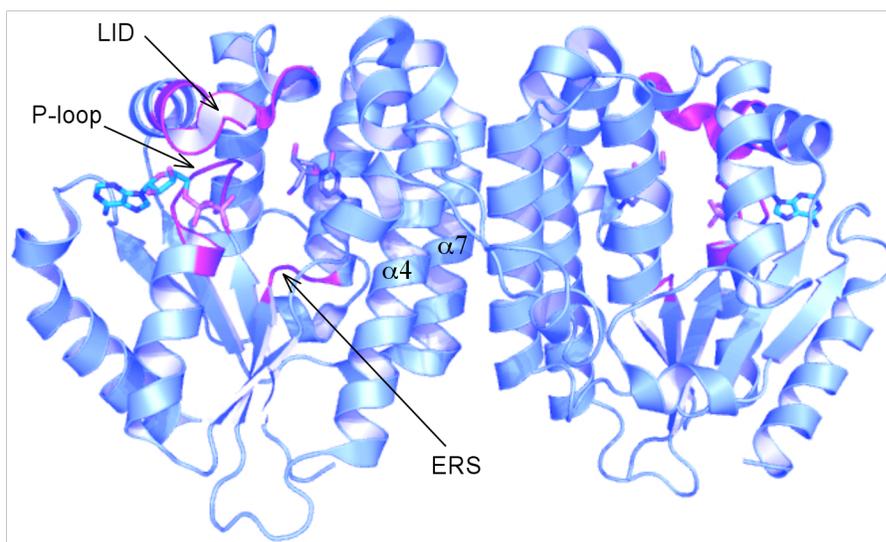
Overall fold of dCK dimer in the presence of dC and ADP. Three sequence motifs are highlighted in red; P-loop, ERS and LID region. Helices α4 and α7 form the dCK dimer interface.

The second conserved motif known as LID region (RXXXRXR) is located between two α-helices (α8 and α9 in dCK). This flexible loop closes on the phosphoryl donor facilitated by arginine interactions when it binds and also participates in catalysis. The last motif, ERS, is involved in ligand binding by forming hydrogen bond from the arginine residue to the 5′-OH group of the substrate. Furthermore, the arginine forms an extensive hydrogen-bonding network to preposition itself for substrate binding.

1.4.2 Substrate specificity and functional versatility link to the structure of type 1 dNKs

While the type 1 dNKs share similar structural architecture despite the low sequence identities, each kinase has distinct but overlapping substrate specificity as summarized in Table 1.2. The monomeric subunits of several kinases originated from different species bound to their respective substrates are presented in Figure 1.4.

Several conserved residues between the structures have been identified from their primary sequences:

(a) a glutamine (Gln) residue that forms hydrogen-bonding interaction with the nucleobase portion of the 2′-deoxyribonucleoside,

(b) a glutamic acid-tyrosine (Glu-Tyr) pair, which is hydrogen-bonded to the 3′-hydroxyl group of the deoxyribose moiety and

(c) a glutamic acid-arginine (Glu-Arg) couple that is at hydrogen-bonding distance to the 5′-hydroxyl group of the substrate and the glutamic acid presumably acts as the general base during the catalysis [4, 39, 41, 42].

The molecular nature of the determinants of substrate specificities has become available since the crystal structures of these kinases (dCK, dGK and *Dm*-dNK) were determined [39, 41].

Figure 1.4   3D structures of several type 1 dNKs.

Type 1 deoxyribonucleoside kinases		
A dCK	**B** dGK	**C** *Dm*-dNK
2QRN; (dCMP, UDP & Mg^{+2})	2OCP ; (dATP)	1OT3 ; (T)

D HSV-1 TK	**E** *Mm*-dAK
1VTK ; (TMP; ADP)	2JAS ; (dATP & Mg^{+2})
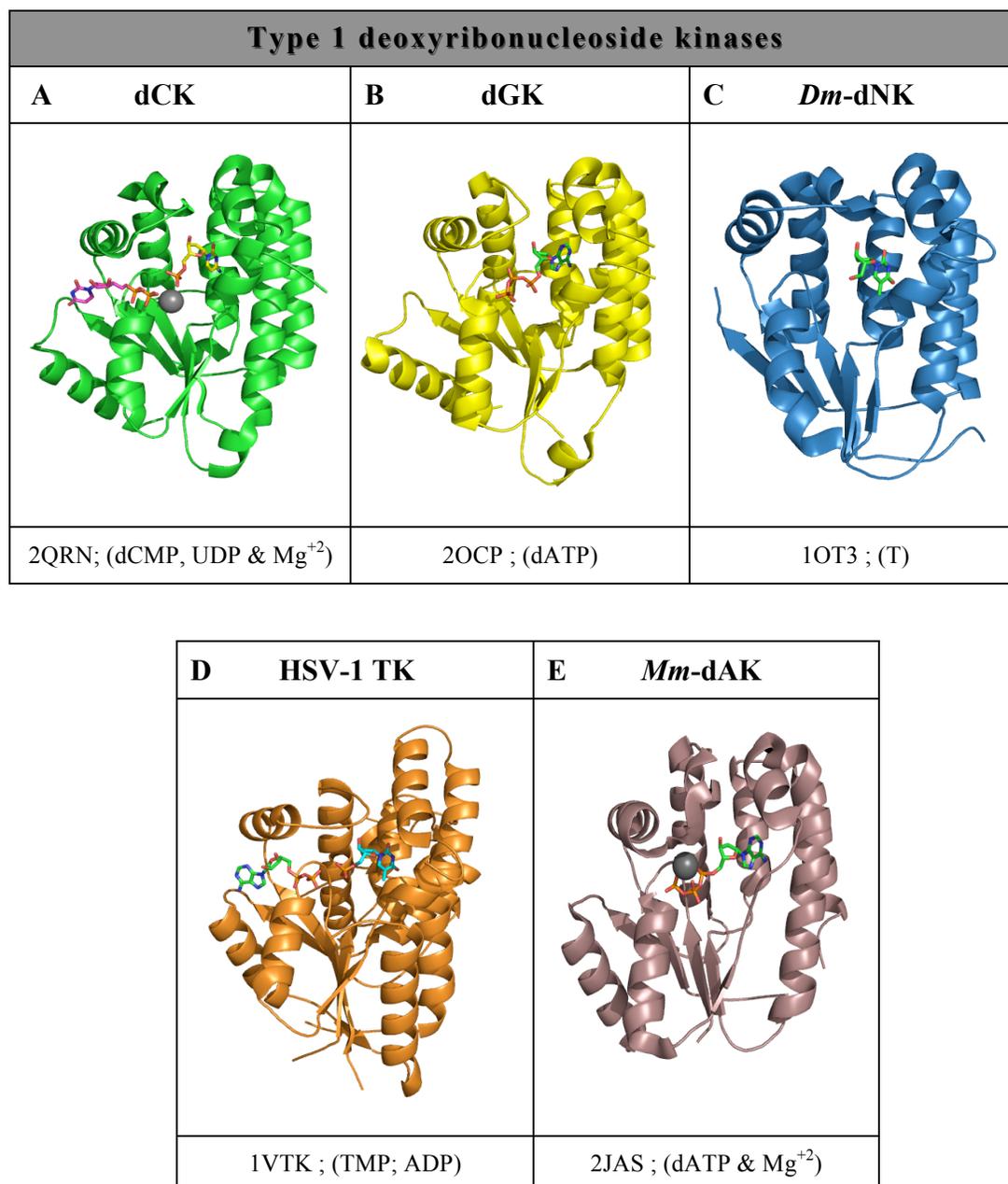
The five kinases have similar folds with a parallel five-stranded β-sheet core. **A;** The subunit structure of dCK (green) with bound dCMP, UDP and Mg^{+2} [43]. **B;** The subunit structure of dGK (yellow) with bound dATP [41]. **C;** The subunit structure of *Dm*-dNK (blue) with bound T [41]. **D;** The subunit structure of HSV-1 TK (orange) with bound TMP and ADP [4]. **E;** The subunit structure of *Mm*-dAK (purple) with bound dATP and Mg^{+2} [44]. All the substrates are shown in stick representation. The figures were made using PyMol.

The substrate specificity of dGK is partially overlapping with dCK, but dGK is more restrictive as indicated from the previous research. The reason for the restrictive substrate specificity of dGK was suggested to be the result of unfavorable interaction between the exocyclic amino group of dC and the extended conformation of Arg118 (Arg104 in dCK) in the active site. Arg118 provides hydrogen bonds to the N7 atom and 6-carbonyl of the guanine ring, while the amino group at 6-position of adenine makes the binding of this base less favorable. Asp147 (Arg133 in dCK) holds Arg118 tightly in place and the presence of Arg-Asp pair discriminates against T binding via steric clash between Arg and the methyl group of thymine. Furthermore, Asp causes unfavorable hydrogen-bonding interactions with a thymine base in this position as observed for dCK as well.

The structure of *Dm*-dNK has revealed a wider and non-specific substrate-binding site shaped by hydrophobic residues and thus provides an explanation for the ability of *Dm*-dNK to phosphorylate all four natural substrates [41]. The aforementioned Arg-Asp pair in dCK and dGK is substituted with Met88-Ala110 pair in *Dm*-dNK, which excludes the steric effect and the specific hydrogen bonding network, respectively.

This hypothesis was investigated by mutating the nonconserved residues Val84, Met88 and Ala110 in the *Dm*-dNK active site to the corresponding residues in dCK and dGK [45]. The resulted mutants converted the broad substrate specificity of wild type *Dm*-dNK from a predominantly pyrimidine specific enzyme to a purine specific enzyme with reduced pyrimidine phosphorylation. In an independent experiment using dCK with the complementary substitutions to mimic the active site of *Dm*-dNK, a triple mutant of dCK, Ala100Val/Arg104Met/Asp133Ala, showed the ability to phosphorylate T [39]. Moreover, *Dm*-dNK and the closely related TK2 share the same amino acids at the three

positions, yet both kinases differ in their substrate specificity. Given the fact that TK2 is a strict pyrimidine kinase and no detected activities for both purine deoxyribonucleosides, additional residues or mechanisms still remain to be uncovered for modulation of substrate specificity. Besides confirming the importance of these nonconserved residues for the substrate selectivity, more insight into the evolutionary preferences could be unraveled by the structure-function relationships derived from analogous mutational studies within the type 1 family members.

1.5   The role of type 1 dNKs in therapy of diseases

1.5.1   Nucleoside analogs in chemotherapy and antiviral therapy

Cellular kinases are responsible for the activation of nucleoside analog (NA) prodrugs that are not functional until phosphorylated to their triphosphate forms. The NAs are transported into cell using the same nucleoside transporter system as natural dNs [46]. Once the NAs are phosphorylated, the charged entities accumulate inside the cell and undergo subsequent sequential phosphorylation steps. The first activation step is carried out by dNKs, which phosphorylate the administered NA into a NA-monophosphate (NAMP). The NAMPs are then phosphorylated in two steps to NA-triphosphates by dNMPK and NDPK [47].

The triphosphate anabolites of a NA can act as a chain terminator for low-fidelity viral polymerases or reverse transcriptase and interfere with viral proliferation by means of halting viral DNA synthesis [48]. In addition, these NA-triphosphates hinder the DNA elongation by blocking or interrupting cellular DNA polymerases thus induce apoptosis [49]. Furthermore, some NAs can inhibit other nucleotide biosynthesis pathway enzymes

simultaneously for instance effective anticancer agents gemcitabine, also known as dFdC (2′,2′-difluorodeoxycytidine) and clofarabine, also known as CAFdA (2-chloro-9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-9*H*-purin-6-amine) inactivate RNR in its diphosphate and triphosphate forms, respectively [50, 51]

The first major group of antiretroviral drugs was NAs acting as competitive substrate inhibitors for the treatment of retroviral infections, primarily HIV infection. NAs possess unique modifications located at the nucleobase portion and/or at the 2′ and 3′ positions of the sugar moiety. The alterations of several NAs are highlighted in red in Figure 1.5.

One example of a 3′-carbon modification is AZT (3′-azido-3′-deoxythymidine), an antiviral agent used against HIV. Like most other 3′-sugar modifications, AZT is a strong DNA polymerase inhibitor because it lacks the 3′-hydroxyl group essential for polymerase chemistry and DNA chain elongation [52].

2′-ribose modified NAs also act as DNA polymerase inhibitors, but their mode of DNA chain termination is not direct. It should be noted that although many of these nucleoside analogs have additional mechanisms of cell replication inhibition such as inhibition of RNR and depletion of dNTP pools. Examples of 2′-modified NAs include potent anticancer agents, AraC (1-β-D-arabinosylcytosine) and gemcitabine [53, 54].

Figure 1.5 Structures of natural dNs and NA derivatives.
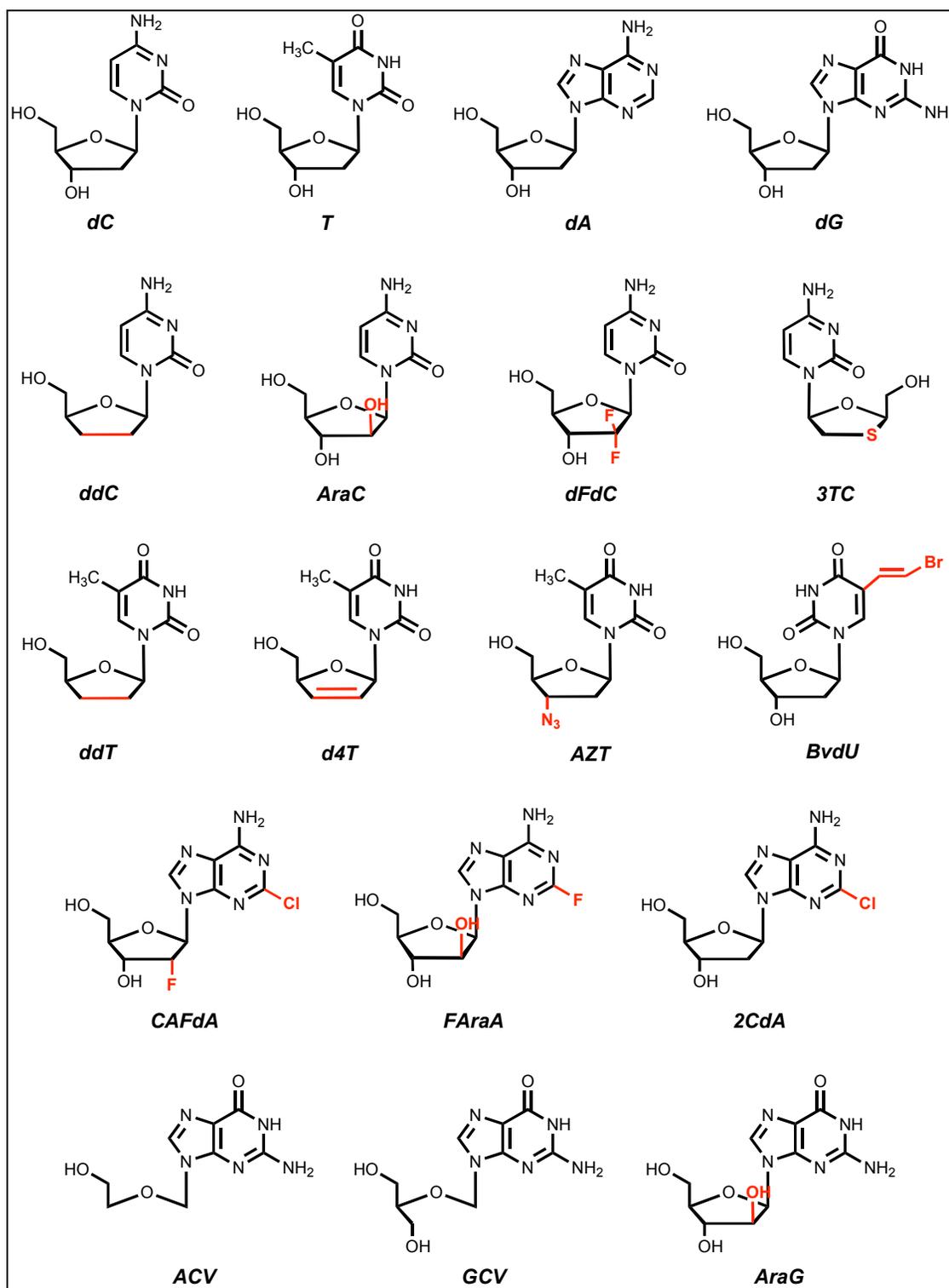
All the NA structures, mentioned within the text, were prepared in ChemBioDraw11.

1.5.2 Suicide gene/chemotherapy

In the salvage pathway, dNKs are the first and mostly regarded as the rate-limiting enzymes in the activation of NAs via phosphorylation [55]. Therefore, these therapeutically critical kinases with high catalytic efficiencies towards administered NA prodrugs are important for the therapy regime and their potential application in suicide gene therapy.

The fundamental of suicide gene/chemotherapy is to transduce a target cell line with a gene that encodes an exogenous dNK and following the administration of NAs, cells that carry the transduced kinase convert the prodrug to a cytotoxic agent and become sensitive to NA-induced apoptosis [56]. This system is implemented for HSV1-TK and ganciclovir, GCV, (9-(1,3-dihydroxy-2-propoxymethyl)-guanine) through first delivering the gene of interest into the cancer cells via a vector, e.g. adenoviruses [57]. Subsequent expression of the HSV1-TK gene inside the cancer cells eventually initiates the phosphorylation of administered GCV prodrug (not a substrate for any human dNKs) hence the triphosphorylated NA executes cancer cell death.

The phenomenon known as bystander effect, by which the introduced gene is responsible for eradication of neighboring cells in where it is not present. This effect could be a result of shuttling the toxic metabolites of NAs by cell-cell contacts [58]. Another strategy that has been used to overcome the bottleneck of GCV activation from the monophosphate to diphosphate form is by fusing the genes of HSV1-TK and guanylate kinase (GMPK) together. Compared to the HSV1-TK alone, a decrease in IC50 was observed [59]. HSV1-TK has been successfully exploited in various *in vitro* and *in vivo* cancer models as a suicide gene for cancer ablation with GCV and/or acyclovir (ACV) and for antiviral

activity of AZT against HIV [60-65]. The multisubstrate dNK from *Drosophila melanogaster*, has been suggested as an another candidate for suicide gene chemotherapy after some of the specifically chosen human cancer cell lines expressing *Dm*-dNK showed sensitization against several nucleoside analogs [66-68].

Protein engineering has been utilized to generate mutants of HSV1-TK with improved prodrug phosphorylation for GCV and ACV based on in vivo screening for prodrug sensitivity [69-71]. In addition, directed evolution of *Dm*-dNK resulted in mutants with increased specificity for several antiviral and anticancer NAs and some mutants were tested in various human cancer cell lines as potential suicide genes [72]. However, some progress has been made in this area, engineering dNKs with promising custom-made properties (enhanced specificity and activity) towards individual NAs would be more advantageous for the current suicide gene applications.

1.6   Previous engineering studies on dCK

Developments in protein engineering techniques over the past ten years have enabled dNKs to be evolved *in vitro* for the desired properties that facilitate the structure-function investigations at the molecular level and also to search for enzyme variants with enhanced NA prodrug activation. The ultimate goal of this field is to engineer novel kinases that would serve increased substrate affinity and catalytic rates for NAs, and low affinity for the naturally occurring competitors, dNs, in the cellular environment. The present studies described in this dissertation focuses on engineering these properties into dCK, as it is an invaluable human enzyme that is responsible for the phosphorylation of numerous NA prodrugs. Three of the eight FDA approved nucleoside analog reverse

transcriptase inhibitors (NRTIs) including ddC (2′,3′-dideoxycytidine), 3TC (2′-deoxy-3′-thiacytidine) and FTC (5-fluoro-2′-deoxy-3′-thiacytidine), as well as chemotherapeutic agents such as gemcitabine, AraC, clofarabine and cladribine, also known as 2CdA (2-chloro-2′-deoxyadenosine) are mainly activated by dCK [7, 73-78]. Moreover, drug resistance to several NAs has been linked to the loss of dCK activity in the cell [79-81]. Sabini and coworkers initiated the first mutational study of dCK via implementing a rational redesign of the active site residues that would mimic the *Dm*-dNK's by creating a triple dCK variant, Ala100Val/Arg104Met/Asp133Ala, according to the structurally equivalent residues [39]. Introducing these mutations into the dCK frame generated a variant with the ability to phosphorylate T and faster catalytic rates than wild-type dCK towards dC and gemcitabine. However, this particular study is not only lack in providing any insight into the detailed kinetic properties for the rest of dNs but also fails to explore other residues in the dCK structure that could be responsible for T phosphorylation. We employed the mutational studies explained in chapter 2 on dCK with the intention of finding some explanations to these statements.

Recent work by Godsey *et al*. is suggested that a flexible loop region also called "insert" plays a role in increasing the affinity for purine deoxyribonucleosides in dCK [82]. This flexible insert consists of 15 amino acids (between residue positions 65 and 79) located between α2 and α3 helices (Figure 1.6). The molecular rationale behind this interpretation is supported by other type 1 kinases. While dGK contains such an insert and phosphorylates both purines, human TK2 lacks the insert region and does not phosphorylate dA or dG. Besides TK2, *Dm*-dNK also lacks such an insert and the apparent binding constants for purines are much higher than pyrimidines.

Figure 1.6   Ribbon diagram of dCK monomer in the presence of dC and ADP.
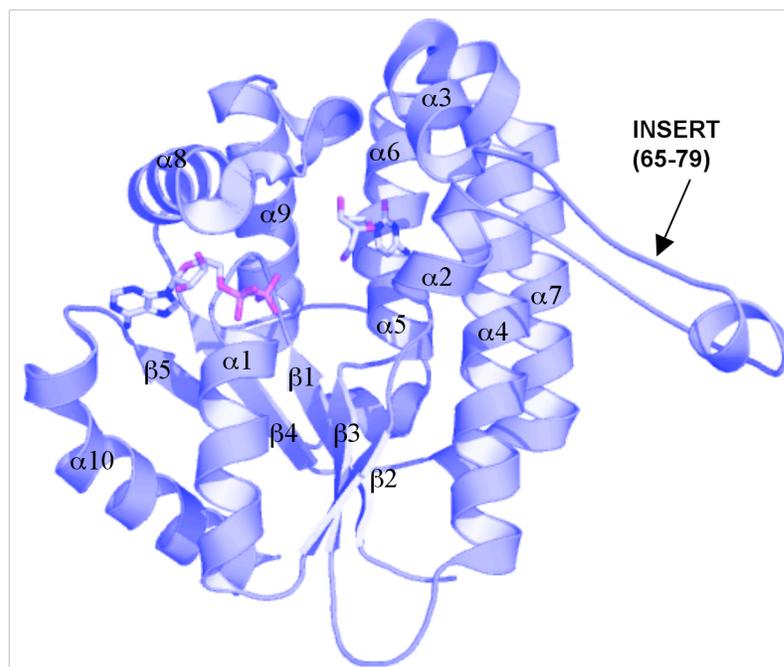
The deleted 15-residue "Insert" (residue 65-79) loop region between helix α2 and α3 is pointed by arrow.

A dCK variant was constructed by deleting this loop (dCKΔI) and the kinetic characterization of this variant using the natural substrates of dCK was compared to the wild-type enzyme. Deletion of the loop showed unchanged K_M for dC but caused 4-fold and 2-fold increase in the K_M values for dA and dG, respectively [82]. The mechanism by which the loop controls the specificity for purines versus pyrimidines is unclear; however, the authors speculate that the insert may affect the conformational flexibility of its flanking α-helices (α2 and α3), which contain residues that contribute to the binding of dNs.

Recently, Smal *et al.* suggested that phosphorylation of Ser74 residue, which is located in the insert region, enhances the intracellular activity of dCK [83]. To mimic the phosphorylation state at the Ser74 site, Ser74Glu mutation was introduced and the

following kinetic analysis of this dCK variant showed a 10-fold increase in its catalytic activity [84]. The authors suggest that phosphorylation or dephosphorylation of dCK is an intrinsic mechanism to regulate enzyme's intracellular activity.

1.7   Aims and scope of the presented dissertation

In this dissertation, I am trying to address some of the molecular aspects regarding the substrate recognition in human dCK and identify the elements that contribute to higher catalytic efficiencies towards NA prodrugs by applying protein engineering techniques.

I have utilized rational design and combinatorial mutagenesis to probe the influence of active site residues on the substrate specificity and catalytic performance of dCK. The results described in Chapter 2 demonstrated that two amino acid positions in the dCK active site play a critical role in altering the substrate specificity towards a generalist kinase or a specialist kinase by either excepting all natural and prodrug substrates or reversing the substrate specificity towards thymidine and relevant NA prodrugs.

In chapter 3, site-saturation mutagenesis was employed at the putative sugar-binding active site residues to generate dCK variants with improved specificities and efficient catalytic activities towards NAs modified at the sugar portion. I exploited an *in vivo* screening system to identify the dCK mutants with increased AZT sensitivity. The results suggest that this particular screening system might not be compatible with dCK properties for the desired engineering outcome.

In chapter 4, loop/subdomain swapping was utilized using dCK as the central protein scaffold and grafting C-terminal loops and subdomains from *Dm*-dNK to investigate the roles of phosphoryl donor-binding site regions on ATP versus UTP preference as

phosphoryl donors. Furthermore, I analyzed the effects of two specific amino acid residues on the donor specificity of dCK by site-directed mutagenesis. The results suggest that the engineered loop/domain regions of mutant proteins could alter the kinetic properties observed with ATP to those with UTP.

References

1.   Brannigan JA, Wilkinson AJ: **Protein engineering 20 years on.** *Nature* 2002, **3:**964-970.

2.   Chen R: **Enzyme engineering: rational redesign versus directed evolution.** *Trends Biotechnol* 2001, **19:**13-14.

3.   Arner ESJ, Eriksson S: **Mammalian deoxyribonucleoside kinases.** *Pharmacology & Therapeutics* 1995, **67:**155-186.

4.   Wild K, Bohner T, Folkers G, Schulz GE: **The structures of thymidine kinase from Herpes simplex virus type 1 in complex with substrates and a substrate analogue.** *Protein Science* 1997, **6:**2097-2106.

5.   Eriksson S, Munch-Petersen B, Johansson K, Eklund H: **Structure and function of cellular deoxyribonucleoside kinases.** *Cellular and Molecular Life Sciences* 2002, **59:**1327-1346.

6.   Arner ES, Flygar M, Bohman C, Wallstrom B, Eriksson S: **Deoxycytidine kinase is constitutively expressed in human lymphocytes: consequences for compartmentation effects, unscheduled DNA synthesis, and viral replication in resting cells.** *Exp Cell Res* 1988, **178:**335-342.

7.   Spasokoukotskaja T, Arner ES, Brosjo O, Gunven P, Juliusson G, Liliemark J, Eriksson S: **Expression of deoxycytidine kinase and phosphorylation of 2-chlorodeoxyadenosine in human normal and tumour cells and tissues.** *Eur J Cancer* 1995, **31A:**202-208.

8.   Datta NS, Shewach DS, Mitchell BS, Fox IH: **Kinetic properties and inhibition of human T lymphoblast deoxycytidine kinase.** *J Biol Chem* 1989, **264:**9359-9364.

9.   Chottiner EG, Shewach DS, Datta NS, Ashcraft E, Gribbin D, Ginsburg D, Fox IH, Mitchell BS: **Cloning and expression of human deoxycytidine kinase cDNA.** *Proc Natl Acad Sci U S A* 1991, **88:**1531-1535.

10.   Shewach DS, Reynolds KK, Hertel L: **Nucleotide specificity of human deoxycytidine kinase.** *Mol Pharmacol* 1992, **42:**518-524.

11.   Krawiec K, Kierdaszuk B, Kalinichenko EN, Rubinova EB, Mikhailopulo IA, Eriksson S, Munch-Petersen B, Shugar D: **Striking ability of adenosine-2'(3')-deoxy-3'(2')-triphosphates and related analogues to replace ATP as phosphate donor for all four human, and the Drosophila melanogaster, deoxyribonucleoside kinases.** *Nucleosides Nucleotides Nucleic Acids* 2003, **22:**153-173.

12.   White JC, Capizzi RL: **A critical role for uridine nucleotides in the regulation of deoxycytidine kinase and the concentration dependence of 1-beta-D-arabinofuranosylcytosine phosphorylation in human leukemia cells.** *Cancer Res* 1991, **51:**2559-2565.

13.   Hughes TL, Hahn TM, Reynolds KK, Shewach DS: **Kinetic analysis of human deoxycytidine kinase with the true phosphate donor uridine triphosphate.** *Biochemistry* 1997, **36:**7540-7547.

14. Sherley JL, Kelly TJ: **Human cytosolic thymidine kinase. Purification and physical characterization of the enzyme from HeLa cells.** *J Biol Chem* 1988, **263:**375-382.

15. Munch-Petersen B, Tyrsted G, Cloos L, Beck RA, Eger K: **Different affinity of the two forms of human cytosolic thymidine kinase towards pyrimidine analogs.** *Biochim Biophys Acta* 1995, **1250:**158-162.

16. Munch-Petersen B, Cloos L, Tyrsted G, Eriksson S: **Diverging substrate specificity of pure human thymidine kinases 1 and 2 against antiviral dideoxynucleosides.** *J Biol Chem* 1991, **266:**9032-9038.

17. Wang L, Munch-Petersen B, Herrstrom Sjoberg A, Hellman U, Bergman T, Jornvall H, Eriksson S: **Human thymidine kinase 2: molecular cloning and characterisation of the enzyme activity with antiviral and cytostatic nucleoside substrates.** *FEBS Lett* 1999, **443:**170-174.

18. Wang L, Karlsson A, Arner ES, Eriksson S: **Substrate specificity of mitochondrial 2'-deoxyguanosine kinase. Efficient phosphorylation of 2-chlorodeoxyadenosine.** *J Biol Chem* 1993, **268:**22847-22852.

19. Johansson M, Karlsson A: **Cloning and expression of human deoxyguanosine kinase cDNA.** *Proc Natl Acad Sci U S A* 1996, **93:**7258-7262.

20. Zhu CY, Johansson M, Permert J, Karlsson A: **Enhanced cytotoxicity of nucleoside analogs by overexpression of mitochondrial deoxyguanosine kinase in cancer cell lines.** *The Journal of Biological Chemistry* 1998, **273:**14707-14711.

21. Herrstrom Sjoberg A, Wang L, Eriksson S: **Antiviral guanosine analogs as substrates for deoxyguanosine kinase: implications for chemotherapy.** *Antimicrob Agents Chemother* 2001, **45:**739-742.

22. Munch-Petersen B, Piskur J, Sondergaard L: **Four deoxynucleoside kinase activities from Drosophila melanogaster are contained within a single monomeric enzyme, a new multifunctional deoxynucleoside kinase.** *J Biol Chem* 1998, **273:**3926-3931.

23. Johansson M, Van Rompay AR, Degreve B, Balzarini J, Karlsson A: **Cloning and characterization of the multisubstrate deoxyribonucleoside kinase of Drosophila melanogaster.** *Journal of Biological Chemistry* 1999, **274:**23814-23819.

24. Mikkelsen NE, Johansson K, Karlsson A, Knecht W, Andersen G, Piskur J, Munch-Petersen B, Eklund H: **Structural basis for feedback inhibition of the deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase.** *Biochemistry* 2003, **42:**5706-5712.

25. Knecht W, Petersen GE, Munch-Petersen B, Piskur J: **Deoxyribonucleoside kinases belonging to the thymidine kinase 2 (TK2)-like group vary significantly in substrate specificity, kinetics and feed-back regulation.** *J Mol Biol* 2002, **315:**529-540.

26. Knecht W, Petersen GE, Sandrini MPB, Sondergaard L, Munch-Petersen B, Piskur J: **Mosquito has a single multisubstrate deoxyribonucleoside kinase characterized by unique substrate specificity.** *Nucleic Acids Research* 2003, **31:**1665-1672.

27. Chen MS, Prusoff WH: **Association of thymidylate kinase activity with pyrimidine deoxyribonucleoside kinase induced by herpes simplex virus.** *J Biol Chem* 1978, **253:**1325-1327.

28. Bird LE, Ren J, Wright A, Leslie KD, Degreve B, Balzarini J, Stammers DK: **Crystal structure of varicella zoster virus thymidine kinase.** *J Biol Chem* 2003, **278:**24680-24687.

29. Sawyer MH, Ostrove JM, Felser JM, Straus SE: **Mapping of the varicella zoster virus deoxypyrimidine kinase gene and preliminary identification of its transcript.** *Virology* 1986, **149:**1-9.

30. Sawyer MH, Inchauspe G, Biron KK, Waters DJ, Straus SE, Ostrove JM: **Molecular analysis of the pyrimidine deoxyribonucleoside kinase gene of wild-type and acyclovir-resistant strains of varicella-zoster virus.** *J Gen Virol* 1988, **69 (Pt 10):**2585-2593.

31. Black ME, Hruby DE: **Quaternary structure of vaccinia virus thymidine kinase.** *Biochem Biophys Res Commun* 1990, **169:**1080-1086.

32. Solaroli N, Johansson M, Balzarini J, Karlsson A: **Substrate specificity of three viral thymidine kinases (TK): vaccinia virus TK, feline herpesvirus TK, and canine herpesvirus TK.** *Nucleosides Nucleotides Nucleic Acids* 2006, **25:**1189-1192.

33. Okazaki R, Kornberg A: **Deoxythymidine Kinase of Escherichia Coli. I. Purification and Some Properties of the Enzyme.** *J Biol Chem* 1964, **239:**269-274.

34. Okazaki R, Kornberg A: **Deoxythymidine Kinase of Escherichia Coli. Ii. Kinetics and Feedback Control.** *J Biol Chem* 1964, **239:**275-284.

35. Carnrot C, Wehelie R, Eriksson S, Bolske G, Wang L: **Molecular characterization of thymidine kinase from Ureaplasma urealyticum: nucleoside analogues as potent inhibitors of mycoplasma growth.** *Mol Microbiol* 2003, **50:**771-780.

36. Carnrot C, Vogel SR, Byun Y, Wang L, Tjarks W, Eriksson S, Phipps AJ: **Evaluation of Bacillus anthracis thymidine kinase as a potential target for the development of antibacterial nucleoside analogs.** *Biol Chem* 2006, **387:**1575-1581.

37. Andersen RB, Neuhard J: **Deoxynucleoside kinases encoded by the yaaG and yaaF genes of Bacillus subtilis. Substrate specificity and kinetic analysis of deoxyguanosine kinase with UTP as the preferred phosphate donor.** *J Biol Chem* 2001, **276:**5518-5524.

38. Pol E, Wang L: **Kinetic mechanism of deoxyadenosine kinase from Mycoplasma determined by surface plasmon resonance technology.** *Biochemistry* 2006, **45:**513-522.

39. Sabini E, Ort S, Monnerjahn C, Konrad M, Lavie A: **Structure of human dCK suggests strategies to improve anticancer and antiviral therapy.** *Nat Struct Biol* 2003, **10:**513-519.

40. Ramakrishnan C, Dani VS, Ramasarma T: **A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins.** *Protein Eng* 2002, **15:**783-798.

41. Johansson K, Ramaswamy S, Ljungcrantz C, Knecht W, Piskur J, Munch-Petersen B, Eriksson S, Eklund H: **Structural basis for substrate specificities of cellular deoxyribonucleoside kinases.** *Nat Struct Biol* 2001, **8:**616-620.

42. Wild K, Bohner T, Aubry A, Folkers G, Schulz GE: **The three-dimensional structure of thymidine kinase from herpes simplex virus type 1.** *FEBS Lett* 1995, **368:**289-292.

43. Soriano EV, Clark VC, Ealick SE: **Structures of human deoxycytidine kinase product complexes.** *Acta Crystallogr D Biol Crystallogr* 2007, **63:**1201-1207.

44. Welin M, Wang L, Eriksson S, Eklund H: **Structure-function Analysis of a Bacterial Deoxyadenosine Kinase Reveals the Basis for Substrate Specificity.** *J Mol Biol* 2007, **366:**1615-1623.

45. Knecht W, Sandrini MP, Johansson K, Eklund H, Munch-Petersen B, Piskur J: **A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines.** *Embo J* 2002, **21:**1873-1880.

46. Plagemann PG, Wohlhueter RM, Woffendin C: **Nucleoside and nucleobase transport in animal cells.** *Biochim Biophys Acta* 1988, **947:**405-443.

47. Van Rompay AR, Johansson M, Karlsson A: **Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases.** *Pharmacology & Therapeutics* 2000, **87:**189-198.

48. Macchi B, Mastino A: **Pharmacological and biological aspects of basic research on nucleoside-based reverse transcriptase inhibitors.** *Pharmacol Res* 2002, **46:**473-482.

49.   Klopfer A, Hasenjager A, Belka C, Schulze-Osthoff K, Dorken B, Daniel PT: **Adenine deoxynucleotides fludarabine and cladribine induce apoptosis in a CD95/Fas receptor, FADD and caspase-8-independent manner by activation of the mitochondrial cell death pathway.** *Oncogene* 2004, **23:**9408-9418.

50.   Ostruszka LJ, Shewach DS: **The role of DNA synthesis inhibition in the cytotoxicity of 2',2'-difluoro-2'-deoxycytidine.** *Cancer Chemother Pharmacol* 2003, **52:**325-332.

51.   Xie KC, Plunkett W: **Deoxynucleotide pool depletion and sustained inhibition of ribonucleotide reductase and DNA synthesis after treatment of human lymphoblastoid cells with 2-chloro-9-(2-deoxy-2-fluoro-beta-D-arabinofuranosyl) adenine.** *Cancer Res* 1996, **56:**3030-3037.

52.   Balzarini J: **Metabolism and mechanism of antiretroviral action of purine and pyrimidine derivatives.** *Pharm World Sci* 1994, **16:**113-126.

53.   Kufe DW, Spriggs DR: **Biochemical and cellular pharmacology of cytosine arabinoside.** *Semin Oncol* 1985, **12:**34-48.

54.   Plunkett W, Huang P, Searcy CE, Gandhi V: **Gemcitabine: preclinical pharmacology and mechanisms of action.** *Semin Oncol* 1996, **23:**3-15.

55.   Van Rompay AR, Johansson M, Karlsson A: **Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases.** *Pharmacology & Therapeutics* 2003, **100:**119-139.

56.   Niculescu-Duvaz I, Springer CJ: **Introduction to the background, principles, and state of the art in suicide gene therapy.** *Mol Biotechnol* 2005, **30:**71-88.

57.   Moolten FL: **Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy.** *Cancer Res* 1986, **46:**5276-5281.

58.   Mesnil M, Yamasaki H: **Bystander effect in herpes simplex virus-thymidine kinase/ganciclovir cancer gene therapy: role of gap-junctional intercellular communication.** *Cancer Res* 2000, **60:**3989-3999.

59.   Willmon CL, Krabbenhoft E, Black ME: **A guanylate kinase/HSV-1 thymidine kinase fusion protein enhances prodrug-mediated cell killing.** *Gene Ther* 2006, **13:**1309-1312.

60.   Sterman DH, Recio A, Vachani A, Sun J, Cheung L, DeLong P, Amin KM, Litzky LA, Wilson JM, Kaiser LR, Albelda SM: **Long-term follow-up of patients with malignant pleural mesothelioma receiving high-dose adenovirus herpes simplex thymidine kinase/ganciclovir suicide gene therapy.** *Clin Cancer Res* 2005, **11:**7444-7453.

61.   Ural AU, Takebe N, Adhikari D, Ercikan-Abali E, Banerjee D, Barakat R, Bertino JR: **Gene therapy for endometrial carcinoma with the herpes simplex thymidine kinase gene.** *Gynecol Oncol* 2000, **76:**305-310.

62.   Marconi P, Tamura M, Moriuchi S, Krisky DM, Niranjan A, Goins WF, Cohen JB, Glorioso JC: **Connexin 43-enhanced suicide gene therapy using herpesviral vectors.** *Mol Ther* 2000, **1:**71-81.

63.   Niranjan A, Moriuchi S, Lunsford LD, Kondziolka D, Flickinger JC, Fellows W, Rajendiran S, Tamura M, Cohen JB, Glorioso JC: **Effective treatment of experimental glioblastoma by HSV vector-mediated TNF alpha and HSV-tk**

gene transfer in combination with radiosurgery and ganciclovir administration. *Mol Ther* 2000, **2:**114-120.

64. Culver KW, Ram Z, Wallbridge S, Ishii H, Oldfield EH, Blaese RM: **In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors.** *Science* 1992, **256:**1550-1552.

65. Guettari N, Loubiere L, Brisson E, Klatzmann D: **Use of herpes simplex virus thymidine kinase to improve the antiviral activity of zidovudine.** *Virology* 1997, **235:**398-405.

66. Zheng X, Johansson M, Karlsson A: **Retroviral transduction of cancer cell lines with the gene encoding Drosophila melanogaster multisubstrate deoxyribonucleoside kinase.** *J Biol Chem* 2000, **275:**39125-39129.

67. Zheng X, Johansson M, Karlsson A: **Nucleoside analog cytotoxicity and bystander cell killing of cancer cells expressing Drosophila melanogaster deoxyribonucleoside kinase in the nucleus or cytosol.** *Biochem Biophys Res Commun* 2001, **289:**229-233.

68. Zheng X, Johansson M, Karlsson A: **Bystander effects of cancer cell lines transduced with the multisubstrate deoxyribonucleoside kinase of Drosophila melanogaster and synergistic enhancement by hydroxyurea.** *Mol Pharmacol* 2001, **60:**262-266.

69. Black ME, Newcomb TG, Wilson HM, Loeb LA: **Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy.** *Proc Natl Acad Sci U S A* 1996, **93:**3525-3529.

70. Black ME, Kokoris MS, Sabo P: **Herpes simplex virus-1 thymidine kinase mutants created by semi-random sequence mutagenesis improve prodrug-mediated tumor cell killing.** *Cancer Res* 2001, **61:**3022-3026.

71. Kokoris MS, Sabo P, Adman ET, Black ME: **Enhancement of tumor ablation by a selected HSV-1 thymidine kinase mutant.** *Gene Ther* 1999, **6:**1415-1426.

72. Knecht W, Rozpedowska E, Le Breton C, Willer M, Gojkovic Z, Sandrini MP, Joergensen T, Hasholt L, Munch-Petersen B, Piskur J: **Drosophila deoxyribonucleoside kinase mutants with enhanced ability to phosphorylate purine analogs.** *Gene Ther* 2007.

73. Galmarini CM, Mackey JR, Dumontet C: **Nucleoside analogues and nucleobases in cancer treatment.** *Lancet Oncol* 2002, **3:**415-424.

74. Johansson M, Karlsson A: **Differences in kinetic properties of pure recombinant human and mouse deoxycytidine kinase.** *Biochem Pharmacol* 1995, **50:**163-168.

75. Shewach DS, Liotta DC, Schinazi RF: **Affinity of the antiviral enantiomers of oxathiolane cytosine nucleosides for human 2'-deoxycytidine kinase.** *Biochem Pharmacol* 1993, **45:**1540-1543.

76. Kewn S, Veal GJ, Hoggard PG, Barry MG, Back DJ: **Lamivudine (3TC) phosphorylation and drug interactions in vitro.** *Biochem Pharmacol* 1997, **54:**589-595.

77. Kierdaszuk B, Krawiec K, Kazimierczuk Z, Jacobsson U, Johansson NG, Munch-Petersen B, Eriksson S, Shugar D: **Substrate/inhibitor specificities of human**

deoxycytidine kinase (dCK) and thymidine kinases (TK1 and TK2). *Adv Exp Med Biol* 1998, **431:**623-627.

78.   Lotfi K, Mansson E, Spasokoukotskaja T, Pettersson B, Liliemark J, Peterson C, Eriksson S, Albertioni F: **Biochemical pharmacology and resistance to 2-chloro-2'-arabino-fluoro-2'-deoxyadenosine, a novel analogue of cladribine in human leukemic cells.** *Clin Cancer Res* 1999, **5:**2438-2444.

79.   Stegmann AP, Honders MW, Hagemeijer A, Hoebee B, Willemze R, Landegent JE: **In vitro-induced resistance to the deoxycytidine analogues cytarabine (AraC) and 5-aza-2'-deoxycytidine (DAC) in a rat model for acute myeloid leukemia is mediated by mutations in the deoxycytidine kinase (dck) gene.** *Ann Hematol* 1995, **71:**41-47.

80.   Groschel B, Himmel N, Cinatl J, Perigaud C, Gosselin G, Imbach JL, Doerr HW, Cinatl J, Jr.: **ddC- and 3TC-bis(SATE) monophosphate prodrugs overcome cellular resistance mechanisms to HIV-1 associated with cytidine kinase deficiency.** *Nucleosides Nucleotides* 1999, **18:**921-926.

81.   Mansson E, Spasokoukotskaja T, Sallstrom J, Eriksson S, Albertioni F: **Molecular and biochemical mechanisms of fludarabine and cladribine resistance in a human promyelocytic cell line.** *Cancer Res* 1999, **59:**5956-5963.

82.   Godsey MH, Ort S, Sabini E, Konrad M, Lavie A: **Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: illuminating the role of main-chain reorganization.** *Biochemistry* 2006, **45:**452-461.

83.   Smal C, Vertommen D, Bertrand L, Ntamashimikiro S, Rider MH, Van Den Neste E, Bontemps F: **Identification of in vivo phosphorylation sites on human**

deoxycytidine kinase. Role of Ser-74 in the control of enzyme activity. *J Biol Chem* 2006, **281:**4887-4893.

84.   McSorley T, Ort S, Hazra S, Lavie A, Konrad M: **Mimicking phosphorylation of Ser-74 on human deoxycytidine kinase selectively increases catalytic activity for dC and dC analogues.** *FEBS Lett* 2008, **582:**720-724.

Chapter 2

Functional investigation of active site mutants responsible for modulating the human deoxycytidine kinase substrate specificity

2.1   Introduction

Human deoxycytidine kinase (dCK; EC 2.7.1.74) catalyzes the phosphorylation of 2′-deoxycytidine (dC), 2′-deoxyadenosine (dA) and 2′-deoxyguanosine (dG) to their corresponding monophosphates using nucleoside triphosphates as phosphoryl donors. This reaction is the first step of the deoxyribonucleoside salvage pathway, an alternative to *de novo* nucleotide biosynthesis, which, in combination with deoxyribonucleoside mono and diphosphate kinases, provides triphosphate anabolites for DNA replication and repair [1]. In addition to recycling natural 2′-deoxyribonucleosides, dCK catalyzes the initial, often rate-determining phosphorylation of several chemotherapeutic nucleoside analog (NA) prodrugs such as gemcitabine (2′,2′-difluorodeoxycytidine), AraC (1-β-D-arabinosylcytosine) and clofarabine [2-chloro-9-(2-deoxy-2-fluoro-β-D-arabinofuranosyl) -9*H*-purin-6-amine], as well as antiviral prodrugs including ddC (2′,3′-dideoxycytidine), 3TC (2′-deoxy-3′-thiacytidine), and FTC (5-fluoro-2′-deoxy-3′-thiacytidine) whose pharmacological activity depends on their triphosphate form [2-7]. Given the critical role of dCK in phosphorylating 2′-deoxyribonucleosides and NA, the enzyme has undergone extensive structural and functional characterization in recent years. Multiple sequence alignments and crystallographic studies have identified dCK as a member of the type 1 deoxynucleoside kinase (dNK) subfamily [8]. Type 1 dNKs show a wide range of distinct substrate profiles including the purine-specific human deoxyguanosine kinase (dGK) and the pyrimidine-specific human thymidine kinase 2 (TK2), as well as dNK from *Drosophila melanogaster* (*Dm*-dNK), which is a highly efficient phosphoryl transfer catalyst for all natural 2′-deoxyribonucleosides. Such functional versatility suggests that

these enzymes make an excellent framework for engineering novel substrate specificity, a thought that is further supported by their striking structural similarities. Recent crystallography studies of dCK and other family members show highly similar orientation of the protein backbone and substrate binding in the different kinases [9-13]. The fact that pyrimidine, purine, and nucleoside analog substrates bind in superimposable conformation in the active site has helped identifying a number of enzyme-substrate interactions, yet a detailed analysis of their contributions in individual enzymes is complicated by the low sequence identity among family members (typically 30-50%). As a consequence, it has inspired a series of random mutagenesis studies on the fruit fly enzyme. Early studies included mutagenesis over the entire length of *Dm*-dNK and led to the successful identification of multiple residues throughout the enzyme structure with impact on substrate specificity, yet the limited sample set and the distant mutations from the active site complicated a conclusive analysis [14]. Follow-up experiments focused on positions Val84, Met88, and Ala110 in the *Dm*-dNK active site (Figure 2.1), substituting these residues to Ala, Arg, and Asp, respectively [15]. These rational changes were inspired by the residues found in the same position in dGK and converted the native broad-specificity enzyme into a purine-specific kinase. Similar efforts in dCK were reported in the context of the first crystal structure, demonstrating the importance of the corresponding three active site residues in dCK, Ala100, Arg104, and Asp133, in regards to substrate specificity (Figure 2.1) [11]. Separately, rational design studies of the enzyme's lid region [16] and deletion of a distinct surface loop [17] have been reported but had no significant impact on the phosphoryl acceptor specificity of the enzyme.

Figure 2.1   Comparison of dCK and *Dm*-dNK active site residues with bound substrates.
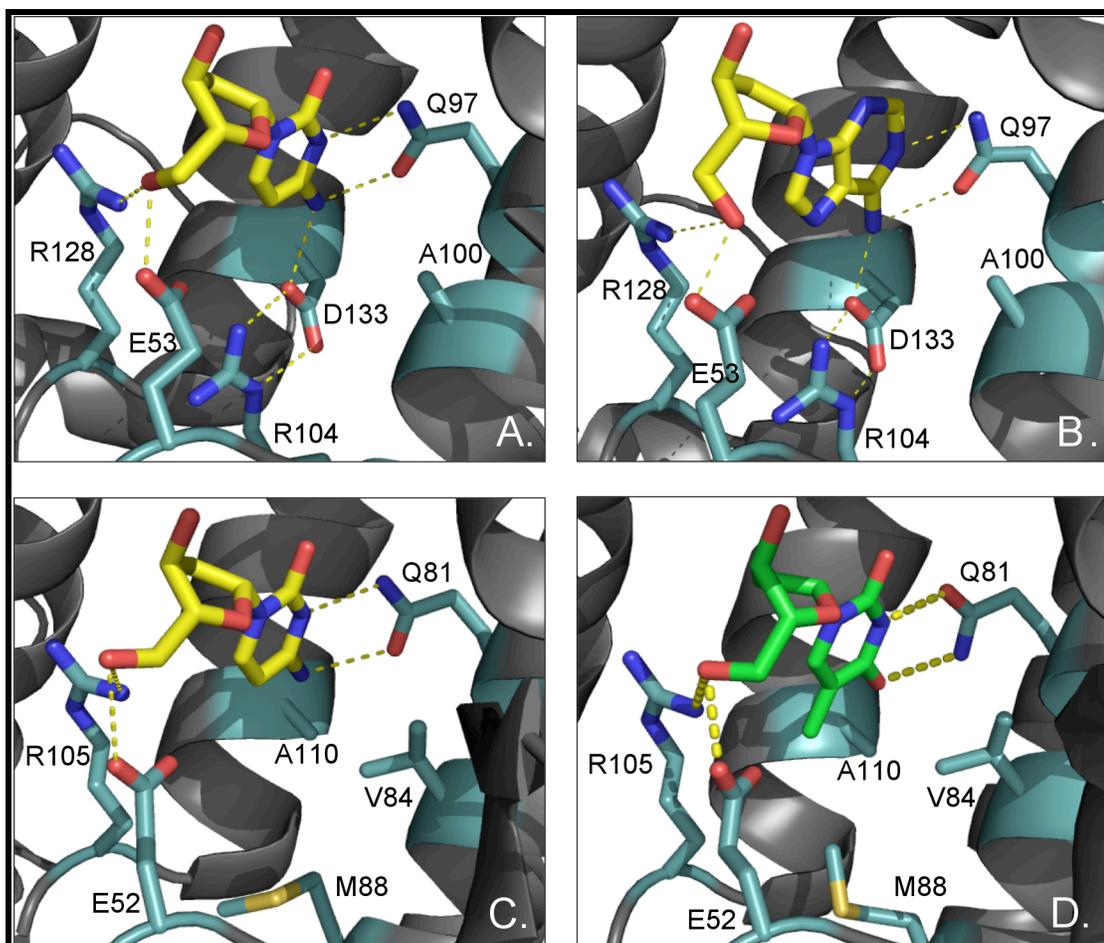
Detailed view of the active sites of dCK in the presence of dC (**A**; PDB code: 1P60 [11] and dA (**B**; PDB code: 2ZI3 [18], as well as *Dm*-dNK with dC (**C**; PDB code: 1J90 [9] and T (**D**; PDB code: 1OT3 [19]. Hydrogen-bonding interactions between the substrate's 5′-hydroxyl group, R128 and E53 (R105 and E52 in *Dm*-dNK) and nucleobase, Q97 (Q81 in *Dm*-dNK) are highly conserved. Position A100, R104 and D133 (V84, M88 and A110 in *Dm*-dNK) and the specific hydrogen bonds are highlighted in the substrate-binding pocket.

Enhancing the activity and tailoring the substrate specificity in dCK are desirable features for the creation of orthogonal NA kinases with potential future therapeutical applications. We have therefore initiated a more systematically exploration of the impact of active site

mutations on the catalytic performance of dCK using rational design and combinatorial protein engineering. Our experiments confirm the critical role of amino acid positions 104 and 133 in regards to substrate specificity. Selected mutations in these two positions created "generalist" catalysts with broad specificity for the natural 2′-deoxyribonucleosides. The simplicity of converting dCK into a generalist dNK supports the notion that the human enzyme is a close relative of the last common ancestor dNK [20, 21]. Furthermore, the generalist serves as an excellent template for subsequent enzyme engineering, best demonstrated by the complete reverse of dCK's substrate specificity to an exclusive thymidine kinase in one of the characterized enzymes. Besides enhancing the phosphorylation of the four natural substrates, alternate amino acid substitutions also introduce novel catalytic activity towards thymidine analogs such as 3′-azido-3′-deoxythymidine (AZT) and 2′,3′-dideoxythymidine (ddT), a first step towards the creation of orthogonal NA kinases.

2.2   Materials and Methods

2.2.1   Chemical reagents and bacterial strains

All reagents were purchased from Fisher (Pittsburg, PA) and Sigma & Aldrich (St. Louis, MO). Enzymes were purchased from New England Biolabs (Beverly, MA) unless indicated otherwise. *Pfu* Turbo DNA polymerase (Strategene, La Jolla, CA) was used for all cloning. PCR reactions for sequencing were performed using Taq DNA polymerase (NEB, Beverly, MA). Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). The QIAprep Spin Miniprep Kit, QIAquick Gel Extraction

Kit and QIAquick PCR Purification Kit were purchased for DNA samples (Qiagen, Valencia, CA). T4 DNA ligase was obtained from Promega (Madison, WI). Pyruvate kinase and lactate dehydrogenase were from Roche Biochemicals (Indianapolis, IN). All DNA manipulations were performed in *E. coli* DH5α-E (Invitrogen, Carlsbad, CA) using standard methodologies and *E. coli* strain BL21(DE3)pLysS (Novagen, Madison, WI) was used for protein expression.

2.2.2   Cloning of human deoxycytidine kinase

The gene for human dCK (dCK: NCBI access# P27707) was isolated with gene-specific primers G-1 and G-2 (Table 2.1) from a human thymus cDNA library (Clontech, Palo Alto, CA). Prior to subcloning of dCK, an internal *Nde*I restriction site was removed by introducing a silent mutation in Thr98 (ACA to ACG) via primer overlap extension [22], using the mutagenic primers G-3 and G-4 (Table 2.1). The corrected PCR product was then digested with *Nde*I and *Spe*I restriction enzymes and ligated into pDIM-PGX [23] for in vivo complementation and pET-14b (Novagen, Madison, WI) for protein overexpression. The *Nde*I and *Spe*I restriction sites were introduced to the dCK gene by the primers. All constructs were confirmed by DNA sequencing. Plasmid-specific primers (Table 2.1) P-1/P-2 for pDIM and P-3/P-4 primer pairs for pET-14b were used for the DNA sequence analysis.

Table 2.1 Primers used for plasmid construction and sequencing.

Primer	Sequence
G-1	5'-GCGCATATGGCCAGCTCTGAGGGGACCCGC-3'
G-2	5'-GCGACTAGTTCACAAAGTACTCAAAAACTCTTTG-3'
G-3	5'-GGTCTTTTACCTTCCAAACGTATGCCTGTCTCAGTCG-3'
G-4	5'-CGACTGAGACAGGCATACGTTTGGAAGGTAAAAGACC-3'
P-1	5'-CGCGCAATTAACCCTCACTAAAG-3'
P-2	5'-GAATAAGGGCGACACGGAAATG-3'
P-3	5'-GCGAAATTAATACGACTCACTATAGGG-3'
P-4	5'-GCTAGTTATTGCTCAGCGG-3'
R104NNS_f	5'-ATAAGAGCTCAGCTTGCCTCTCTGAATGG-3'
R104NNS_r	5'-AAGCTGCGCTCTTATSNNACTGAGACAGGCATACG-3'
D133NNS_f	5'-GGTATATTTTTGCATCTAATTTGTATGAATCTG-3'
D133NNS_r	5'-TAGATGCAAAAATATACCTSNNACTATACACAGATCG-3'
D133A_f	5'-GTGTATAGTGCCAGGTATATTTTTGCATCTAA-3'
D133A_r	5'-ATATACCTGGCACTATACACAGATCGTTCAA-3'
R104M_f	5'-CTGTCTCAGTATGATAAGAGCTCAGCTTGCCTC-3'
R104M_r	5'-GAGGCAAGCTGAGCTCTTATCATACTGAGACAGGC-3'
A100V_f	5'-CAAACGTATGTCTGTCTCAGTATGATAAG-3'
A100V_r	5'-ACTGAGACAGACATACGTTTGGAAGGTAA-3'
H202L_f	5'-GCAAGGCATTCCTCTTGAATATTTAGAGAAGCT-3'
H202L_r	5'-CTAAATATTCAAGAGGAATGCCTTGCTCTTC-3'
N133G_f	5'-TGTGTATAGTGGCAGGTATATTTTTGC-3'
N133G_r	5'-AATATACCTGCCACTATACACAGATCG-3'

2.2.3   Mutagenesis of dCK

2.2.3.1 Site-directed mutagenesis

The site-specific Asp133Ala, Arg104Met and Ala100Val amino acid mutations in dCK

were introduced by primer overlap extension PCR, using the corresponding D133A_f/r, R104M_f/r and A100V_f/r primer pairs (Table 2.1), respectively. The resulting gene products were subcloned into the pDIM and pET-14b vectors via the *Nde*I and *Spe*I restriction sites as described above. DNA sequencing confirmed the correct gene sequences.

2.2.3.2 Random mutagenesis

Random mutagenesis of dCK was performed by error-prone PCR using the GeneMorphII system (Stratagene, La Jolla, CA). Following the manufacturer's protocol, the mutation frequency was controlled via the template concentration in the PCR. Three libraries using 1 pg, 10 pg and 100 pg of pDIM-dCK in combination with the primer pair G-1/G-2 were generated in 50 μl reaction volume. The cycling conditions were: 95°C for 2 min; 29 cycles of 95°C for 30 sec, 53°C for 30 sec, 72°C for 75 sec; 1 cycle of 72°C for 10 min. Subsequently the amplification products were digested with *Nde*I and *Spe*I and ligated back into linearized pDIM vector, followed by transformation into competent thymidine kinase (tk)-deficient *E. coli* KY895 cells [24]. The cells were plated on LB-agar plates containing ampicillin (100 μg/ml) and incubated at 37°C overnight. Colonies were harvested by rinsing the agar plates twice with 10 ml 2YT medium [22], supplemented with glucose and glycerol to a final concentration of 2% and 15%, respectively. After flash freezing in liquid nitrogen, library aliquots were stored at -80°C. The mutation frequency in the three libraries was determined by DNA sequence analysis of twenty randomly picked colonies per library.

2.2.3.3 Site-saturation mutagenesis

Libraries with all nineteen amino acids in positions 104, 133, as well as 104 & 133 of dCK were created by primer overlap PCR, using primers R104NNS_f/r and D133NNS_f/r (Table 2.1). The degenerate codons (N = 1:1:1:1 mixture of dA, dG, dC, & T; S = 1:1 mixture of dG & dC) were introduced during automated primer DNA synthesis, using hand-mixing 2′-deoxyribonucleoside phosphoramidites to minimize library biases. The resulting PCR products were subcloned into the pDIM vector via the *Nde*I and *Spe*I restriction sites as described above and the gene sequences were confirmed by DNA sequencing. The libraries were transformed into *E. coli* KY895, plated on LB-agar plates containing ampicillin (100 μg/ml) and incubated overnight at 37°C. The resulting colonies were harvested and stored as outlined above.

2.2.4 *In vivo* complementation; functional selection for thymidine kinase activity

An aliquot of the frozen pDIM-dCK libraries was thawn and 20 μl cell suspension was used to inoculate 50 ml LB medium containing ampicillin (100 μg/ml) and incubated at 37°C. When the culture reached an OD_{600} of ~0.2, a 1-ml aliquot was diluted with 9 ml LB media and 1 ml of that solution (~2 × 10^6 cfu) was plated on selection plates (2% bacto-peptone; 0.5% NaCl; 1.5% noble agar; 0.2% glucose; 50 μg/ml carbenicillin; 60 μg/ml 5-fluoro-2′-deoxyuridine; 2 μg/ml T; 12.5 μg/ml uridine) [25, 26]. For selection experiments with the random mutagenesis libraries, the plates were incubated at ambient temperature for 36 h while site-saturation libraries were tested at 37°C for 16 h. Colonies that appeared over that time period were picked and restreaked on fresh selection plates to confirm the thymidine kinase activity, followed by DNA sequence analysis of mutants.

2.2.5 Protein expression and purification

For in vitro characterization, wild-type dCK and all mutant kinases were overexpressed as fusion proteins with a N-terminal hexa-His tag. Individual kinase genes, subcloned in pET-14b, were transformed into *E. coli* strain BL21(DE3)pLysS. Cell cultures were grown at 37°C in 200 ml 2YT media containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) to an $OD_{600} \sim 0.5$. Protein expression was induced with 0.1 mM IPTG for 4 h at 30°C except for wild-type dCK which was expressed at 37°C. The cells were centrifuged (4000*g*, 4°C, 20 min) and the cell pellet resuspended in 12 ml buffer A (50 mM Tris-HCl, pH 8; 0.3 M NaCl; 10 mM imidazole) and mixed with 60 μl protease inhibitor cocktail (Sigma), 6 μl Benzonase nuclease (Novagen) and 60 μl lysozyme (20 mg/ml, Sigma). After 15 min incubation on the orbital shaker at 4°C, cells were lysed by sonication on ice and the cell lysates centrifuged (10,000*g*, 4°C, 20 min). The clear supernatant was mixed with 1 ml Ni-NTA agarose resin (Qiagen), preequilibrated in buffer A, and incubated for 90 min at 4°C. The resin was loaded on a Prepcolumn (BioRad, Carlsbad, CA) and washed with 10 column volumes (CV) of buffer A. Unspecifically bound protein was then eluted with 4 CV buffer A containing 20 mM imidazole and 2 CV buffer A containing 50 mM imidazole, followed by protein elution in 3 CV of buffer A containing 250 mM imidazole. The combined product fractions were concentrated in an Amicon Ultra-4 ultrafiltration unit (Millipore, Bedford MA; MWCO: 10 kDa; 5000*g* at 4°C) and buffer-exchanged into storage buffer (50 mM Tris-HCl, pH 8; 0.2 M NaCl, 5 mM $MgCl_2$, 2 mM DTT). The protein yields were determined as >95% pure by SDS-PAGE. For stability reasons, the NaCl concentration in the storage buffer for mutant dCKs was increased to 0.5 M. Aliquots were stored at -80°C after flash

freezing in liquid nitrogen. The protein concentrations were determined by A_{280} measurements ($\varepsilon = 5.6 \times 10^4\,M^{-1}cm^{-1}$) [27, 28].

2.2.6 Enzyme activity assay

The kinase activity of recombinant enzymes was determined using a spectrophotometric coupled-enzyme assay [29, 30]. Briefly, 2′-deoxyribonucleosides and NAs at 1 to 5000 μM were prepared in reaction buffer containing 50 mM Tris-HCl (pH 7.5), 0.1 M KCl, 5 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 0.21 mM phosphoenolpyruvate, 0.18 mM NADH, and 2 units/ml pyruvate kinase and 2 units/ml lactate dehydrogenase. Assays were performed at 30°C or 37°C (as indicated), measuring the absorbance change at 340 nm in the presence of 0.3 – 4.3 μg enzyme per reaction. The enzyme amount was adjusted to limit NADH turnover to 10% over the time of the experiment. All experiments were performed in triplicate and kinetic data was determined by non-linear regression analysis using the Michaelis-Menten equation in Origin 7 (OriginLab, Northampton, MA).

2.2.7 Biophysical characterization

All far-UV circular dichroism (CD) experiments were performed on a Jasco J-810 spectropolarimeter (Jasco, Inc., Easton, MD) equipped with a Peltier unit for temperature control. Samples were prepared in 50 mM potassium phosphate buffer (pH 7.4) with 0.5 M KF, 5 mM $MgCl_2$ and 2 mM DTT. Spectra were collected at 20°C from 260-190 nm (0.5 nm increments) using a 0.02 cm path length quartz cell, a scan rate of 20 nm/min, bandwidth of 2 nm and a response time of 2 s. Protein concentrations ranged from 2.3 – 3.5 μM as determined by the absorbance at 280 nm. All spectra were corrected for

background buffer contribution. Each spectrum represents the average of 5 accumulation scans. Spectra were corrected for buffer absorbance and converted to mean residue ellipticity.

Thermal denaturation of the dCK constructs was monitored from 10°C to 85°C by the change in ellipticity at 222 nm with a temperature gradient of 50°C/hour. Samples were prepared in a 0.1 cm path length quartz cuvette. The data from each sample is the average of two independent experiments and for evaluation of the spectra; the buffer spectrum was subtracted from the protein spectrum. Using Origin 7, the melting point (T_M) was calculated by fitting the data to a Boltzmann distribution.

2.3   Results

2.3.1   Rational design of dCK, purification and activity measurements

Two rational design mutants of dCK with tk activity were prepared based on previous reports in the literature. Guided by the consensus sequence from multiple sequence alignments and crystal structure information, Lavie and coworkers mutated three positions in the active site, changing Ala100Val, Arg104Met, and Asp133Ala in dCK to expand the enzyme's substrate specificity (Figure 2.1) [11]. The substitutions in positions 104 and 133 are critical for T binding, as the Arg side chain will clash with the methyl group in thymine while Asp is unable to form hydrogen bonding interactions. In the absence of detailed kinetic data for the rational dCK mutants, we constructed the double mutant   Arg104Met/Asp133Ala   (rTK2)   and   triple   mutant Ala100Val/Arg104Met/Asp133Ala (rTK3) by site-directed mutagenesis and determined the catalytic properties of these two enzymes with natural dNs as substrates.

Figure 2.2   12% SDS-PAGE analysis of purified recombinant wild-type dCK.
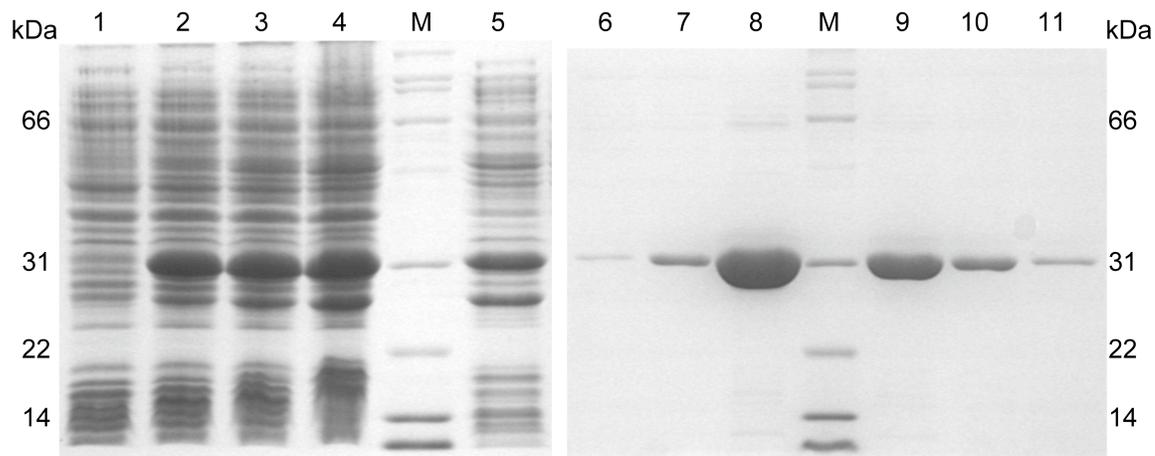
Proteins were visualized by Coomassie Blue staining. The molecular weight of standard markers (Lane M) are indicated. Lanes 1: *E. coli* culture before induction; Lanes 2, 3 and 4: Induced cultures of wild-type dCK after t= 2h, 3h and 4h expression, respectively. Lane 5: Soluble proteins after lysis; Lanes 6, 7, 8, 9, 10 and 11: Eluent fractions of purified 6-His tag dCK (31 kDa) from the Ni^{2+}- affinity column.

Overall, the amino acid substitutions in rTK2 and rTK3 result in a significant enhanced pyrimidine activity with a simultaneous drop in purine specificity (Table 2.2). Most noticeable, the specificity constants (k_{obs}/K_M) for T increased 263-fold (rTK2) and 1105-fold (rTK3), respectively. The gains are largely due to lower K_M values with an additional 5 to 10-fold increase in T turnover rates in both mutants. Besides establishing catalytic activity for T in dCK, the mutations also significantly enhanced the enzymes' catalytic performance with dC. Although the binding constants for dC is slightly higher in the mutants compared to wild-type enzyme, the rate of turnover increases 10 to 20-fold, improving the specificity constant by up to an order of magnitude.

Table 2.2   Kinetic parameters of wild-type dCK and variants from rational design with natural 2′-deoxyribonucleosides.

substrate	enzyme	K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times\ 10^3$
dC	wild-type	1.0 ± 0.2	0.064 ± 0.002	63
	rTK2	2.0 ± 0.4	0.58 ± 0.03	287
	rTK3	1.54 ± 0.17	1.2 ± 0.03	777
T	wild-type	3485 ± 482	0.38 ± 0.03	0.095
	rTK2	74 ± 11	1.88 ± 0.09	25
	rTK3	31.4 ± 7.4	3.29 ± 0.29	105
dA	wild-type	81 ± 6	3.10 ± 0.06	38
	rTK2	412 ± 68	2.71 ± 0.19	7
	rTK3	598 ± 95	3.40 ± 0.27	6
dG	wild-type	154 ± 10	3.26 ± 0.07	21
	rTK2	1280 ± 63	1.23 ± 0.04	0.96
	rTK3	1364 ± 399	0.31 ± 0.06	0.23

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation.

As intended by Sabini *et al.* [11], the amino acid substitutions create a pyrimidine-specific human dNK by lowering the substrate specificity for purine nucleosides. In the case of dA, the steady-state kinetics indicates that mutagenesis leaves the turnover rates largely unaffected but causes an almost 5 to 7.5-fold reduction of the binding affinity from a moderate K_M of 81 µM for the wild-type enzyme to 412 µM (rTK2) and 598 µM (rTK3), respectively. For dG, the effects of the mutagenesis are more dramatic, showing a similar decline in the K_M values as seen for dA in addition to a 10-fold drop in the k_{obs} values. Overall, the catalytic performance of rTK2 and rTK3 with dG, compared with wild-type dCK, decreases 20-fold and 100-fold, respectively.

2.3.2 Random mutagenesis of dCK and functional selection

Random mutagenesis was performed by PCR amplification of the 747-bp dCK gene with low-fidelity polymerases (GeneMorphII system, Stratagene, TX). Ligation into pDIM generated a library of 4×10^5 members. We targeted a mutation frequency of ~ 2-3 amino acid changes per gene by adjusting the template concentration in the PCR reaction. DNA sequence analysis of 23 randomly picked library members (~18,000 bases) showed 2.67 ± 0.4 nucleotide substitutions per gene, translating into 2.1 ± 0.2 amino acid changes on the protein level. Given these colony counts, our library covers approximately 30% of the theoretical diversity (~1.2 million) [32].

Functional selection for tk activity among random mutagenesis library members was carried out in *E. coli* KY895 [24] and yielded two unique dCK mutants. For the genetic complementation experiment, approximately 2 million library members were plated on selection plates and incubated at either 37°C or ambient temperature. The lower temperature reduces the selection pressure for kinase activity as it slows bacterial growth, facilitating proper folding of mutant proteins and better accommodating enzymes with compromised overall stability. While no colonies were found at 37°C, 14 colonies grew at room temperature. Among the successful candidates, two sets of sequences with distinct mutations were identified (Table 2.3). The first mutant (epTK6) carried five amino acid substitutions (Asp47Glu/Arg104Gln/Asp133Gly/Asn163Ile/Phe242Leu). The second mutant (epTK16) encoded three amino acid changes (Arg104Gln/Asp133Asn/Leu202His). To further evaluate the catalytic performance of the two functional dCK mutants, we subcloned their corresponding genes into pET-14b and overexpressed the protein in *E. coli* BL21 as described in the experimental section.

Table 2.3   Summary of amino acid substitutions for all constructed dCK mutants.

mutant	amino acid residue						
	47	100	104	133	163	202	242
wild-type	D	A	R	D	N	L	F
rTK2			M	A			
rTK3		V	M	A			
epTK16			Q	N		H	
epTK16A			Q	N			
epTK6	E		Q	G	I		L
epTK6A			Q	G			
ssTK1			M	S			
ssTK1A		V	M	S			
ssTK2			M	T			
ssTK2A		V	M	T			
ssTK3			M	N			

2.3.3   Selected random dCK mutant purification and activity measurements

The overexpression of epTK6 and epTK16 yielded soluble protein at wild-type dCK levels for epTK6 but resulted in protein aggregates in the case of epTK16. Inspection of the three mutations in epTK16 by multiple sequence alignment against other type 1 dNKs and the dCK crystal structure suggested that mutagenesis of the highly conserved Leu202 could at least in part be responsible for inclusion body formation. To stabilize epTK16, we reversed Leu202His by site-directed mutagenesis, creating the double-mutant epTK16A (Arg104Gln/Asp133Asn). The new construct also complemented the *E. coli* auxotroph and, more importantly, produced soluble protein in good yields upon overexpression in the pET-system at 30°C. We reengineered epTK6 in a similar fashion, removing all but Arg104Gln/Asp133Gly to create epTK6A.

Table 2.4   Kinetic parameters of wild-type dCK and variants from random mutagenesis with natural 2′-deoxyribonucleosides.

substrate	enzyme	K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times\ 10^3$
dC	wild-type	1.0 ± 0.2	0.064 ± 0.002	63
	epTK16A	8.6 ± 1.5	1.30 ± 0.06	152
	epTK6	1.41 ± 0.36	0.21 ± 0.01	150
	epTK6A	1.04 ± 0.12	0.37 ± 0.01	356
T	wild-type	3485 ± 482	0.38 ± 0.03	0.095
	epTK16A	36 ± 7	1.51 ± 0.08	42
	epTK6	25 ± 5	0.68 ± 0.04	27
	epTK6A	25.7 ± 5.5	1.01 ± 0.05	39
dA	wild-type	81 ± 6	3.10 ± 0.06	38
	epTK16A	532 ± 99	2.00 ± 0.15	4
	epTK6	91 ± 21	1.08 ± 0.07	12
	epTK6A	231 ± 25	1.75 ± 0.07	8
dG	wild-type	155 ± 10	3.26 ± 0.07	21
	epTK16A	240 ± 23	1.92 ± 0.06	8
	epTK6	79 ± 10	1.18 ± 0.05	15
	epTK6A	81 ± 11	1.50 ± 0.07	19

The experiments were performed in triplets at 30°C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation.

The kinetic properties of epTK6, as well as the two redesigned mutants epTK6A and epTK16A were determined for all four natural dNs (Table 2.4). Similar to the rational design mutants, the specificity constants for dC and T increase significantly. For dC, all three mutants show 2 to 6-fold increases in k_{obs} / K_M. In the case of epTK6 and epTK6A, the apparent binding constants are unchanged and gains are largely due to higher turnover numbers. In contrast, epTK16A shows an almost 10-fold higher K_M value, possibly

caused by unfavorable steric and electrostatic interactions of the Asp133Asn mutation with the exocyclic amine on the substrate. The decline in apparent binding affinity is compensated by a 20-fold increase in k_{obs}, resulting in an overall 2-fold improvement in catalytic performance. For T, all three candidates show increases in their specificity constants by more than two orders of magnitude, a change that is mostly due to 100-fold lower K_M values. The similarity of the kinetic parameters in all three enzymes suggests that the observed improvements in T phosphorylation are not resulting from newly created beneficial interactions but more likely arise from eliminating unfavorable steric (Arg104) and electrostatic interactions (Asp133).

While rational design and random mutagenesis appear to affect pyrimidine substrate specificity similarly, the trends in the kinetic parameters for purines are noticeably different. In the case of dA, the lower catalytic activity is not caused by higher Michaelis-Menten constants alone as seen for the rational design mutants but results equally from a reduction in the k_{obs} and K_M values. For dG, all three random mutagenesis enzymes show wild type-like catalytic performance, generally experiencing less than two-fold changes in the kinetic parameters. These results are in sharp contrast to the 20 to 100-fold lower k_{obs}/K_M values measured for rTK2 and rTK3.

In summary, we have used random mutagenesis to search for alternative solutions to introducing tk activity in dCK. The tk selection in a large library with multiple random amino acid substitutions shows a clear preference for library members with mutations in position 104 and 133, suggesting that these two residues are major determinants for T specificity in dCK. Nevertheless, our results also reveal some limits to the chosen experimental design. The application of random mutagenesis did not yield either one of

the rational design variants rTK2 or rTK3. The result can be explained by the fact that an Arg104Gln change is possible by a single nucleotide substitution (CGA to CAA) while Arg104Met (CGA to ATG) requires simultaneous nucleotide changes in all three codon positions [33]. Confronted with the possibility that alternative amino acid substitutions in these two residues could produce enzymes with tk activity, we applied site-saturation mutagenesis to completely randomize both positions.

2.3.4 Site-saturation library construction and activity measurements

We generated three site-saturation libraries covering amino acid position 104 and 133 individually, as well as the double mutant 104 & 133. These experiments were motivated by our findings that rational design and random mutagenesis produced two distinct solutions for tk activity in dCK.

The individual site-saturation libraries were prepared with the help of degenerate NNS primers which offers a close to natural distributions for all amino acids while minimizing the occurrence of stop codons [34]. The size for the Arg104NNS and Asp133NNS single-site libraries (theoretical library size = 20) was 7000 cfu and 1500 cfu respectively. The Arg104NNS/Asp133NNS library consisted of 1.5×10^5 members, providing almost 400-fold coverage of the maximum diversity (theoretical library size is 400).

To evaluate the codon distribution in both positions in the naïve libraries, DNA sequences from 48 random samples of the Arg104NNS/Asp133NNS library were analyzed (Table 2.5). The nucleotide distribution in position 104 is relatively well balanced. We found 17 of the 20 natural amino acids in our small sample set. Consistent with the degeneracy of the genetic code, the most frequently found amino acids were Ser

and Arg, closely followed by Leu, Pro, Ala, Val, and Gly.

Table 2.5 Codon distributions in position 104 & 133 for site-saturation mutagenesis.

		dA	dG	dC	T
N (ideal)		0.25	0.25	0.25	0.25
S (ideal)		0	0.5	0.5	0
position 104	N	0.27	0.21	0.33	0.19
	N	0.21	0.26	0.32	0.21
	S	0.08	0.45	0.47	0
position 133	N	0.17	0.17	0.5	0.16
	N	0.19	0.08	0.35	0.38
	S	0	0.7	0.3	0

Furthermore, multiple codons for individual amino acids could be identified. The dA in the third nucleotide position maybe the result of contamination during oligonucleotide synthesis. The nucleotide distribution at Asp133 is less ideal, showing clear biases in all three codon positions. Despite of the skewed distribution, resulting in the overrepresentation of Pro and Leu, our sample set still contained 14 different amino acids, suggesting that all natural amino acids are represented in the current libraries.

The three libraries were subsequently tested for members with tk activity, using the previously described *E. coli* auxotroph. For both single-site mutagenesis libraries, the genetic complementation experiments did not yield any colonies. As both libraries are comprehensive and were sampled at 10-fold redundancy, we concluded that individual amino acid substitutions in either location are insufficient to establish thymidine kinase activity for in vivo complementation. In contrast, the double-mutant library yielded 16 cfu when incubated at 37°C. DNA sequence analysis shows that Met in position 104

dominates among the functional variants (13 out of 16) while Gln is a less frequent alternative (3/16). Position 133 tolerates a broader range of amino acid substitutions, showing Thr (7/16), Ser (4/16), and Asn (3/16) as the most common replacements while Ala and Ile were each found once. In a separate experiment, we tested the same library for growth complementation at 30°C. Interestingly, the lower incubation temperature yielded fewer colonies, forming only 5 cfu. DNA sequence analysis revealed that about half the functional genes carried the previously observed Arg104Met (2/5) substitution while the rest encoded for Arg104Asn (3/5). The reason for the absence of Arg104Gln at lower temperature is unclear. In position 133, these mutants carried the previously identified Asp133Ala, Thr or Ser substitutions. The diversity of amino acid substitutions at position 133 seems unaffected by the change in growth temperature.

Focusing on the functional mutants from the 37°C-selection experiment, we identified and characterized three new amino acid combinations, Arg104Met/Asp133Ser (ssTK1), Arg104Met/Asp133Thr (ssTK2), and Arg104Met/Asp133Asn (ssTK3). As these mutants were among the most frequently selected library members, we hypothesized that such dominance reflects superior function. We therefore overexpressed all three mutants in the pET system and purified them to homogeneity.

The kinetic data for these three new mutants reveal distinct substrate profiles, showing a conspicuous shift towards general pyrimidine kinase activity (Table 2.6). The substitution of Asp133Ser in ssTK1 increases the catalytic turnover for T and dC by 6 and 26-fold over wild-type while showing a substantially lower K_M value for T (19 μM, 183-fold) and only a moderate increase in the binding affinity for dC (4.6 μM, 4.5-fold). The same mutations result in a 5 to 8-fold increase of the apparent binding constants for purines

while leaving their turnover rates largely unaffected (< 2-fold).

Table 2.6   Kinetic parameters of wild-type dCK and variants from site-saturation mutagenesis with natural 2′-deoxyribonucleosides.

substrate	enzyme	K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times$ 10^3
dC	wild-type	1.0 ± 0.2	0.064 ± 0.002	63
	ssTK1	4.6 ± 0.6	1.70 ± 0.06	373
	ssTK1A	5.3 ± 0.5	2.10 ± 0.06	400
	ssTK2	20.9 ± 1.5	2.86 ± 0.07	137
	ssTK2A	71 ± 8	3.54 ± 0.11	50
	ssTK3	11.8 ± 1.5	2.23 ± 0.1	189
T	wild-type	3485 ± 482	0.38 ± 0.03	0.095
	ssTK1	18.9 ± 2.3	2.33 ± 0.08	123
	ssTK1A	5.58 ± 0.58	1.91 ± 0.05	342
	ssTK2	5.8 ± 0.6	1.49 ± 0.04	258
	ssTK2A	3.88 ± 0.22	1.57 ± 0.02	406
	ssTK3	25.6 ± 3.6	2.27 ± 0.10	88
dA	wild-type	81 ± 6	3.10 ± 0.06	38
	ssTK1	398 ± 27	4.85 ± 0.13	12
	ssTK1A	843 ± 80	3.37 ± 0.10	4
	ssTK2	691 ± 128	4.14 ± 0.35	6
	ssTK2A	1437 ± 179	2.67 ± 0.13	2
	ssTK3	452 ± 88	3.47 ± 0.30	8
dG	wild-type	155 ± 10	3.26 ± 0.07	21
	ssTK1	1174 ± 109	4.38 ± 0.23	4
	ssTK1A	739 ± 61	1.72 ± 0.04	2
	ssTK2	1203 ± 179	2.50 ± 0.18	2
	ssTK2A	1164 ± 107	2.29 ± 0.08	2
	ssTK3	324 ± 31	3.59 ± 0.12	11

The experiments were performed in triplets at 37°C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation.

A similar trend can be observed for ssTK2 where the Asp133Thr substitution further enhances the tk activity by reducing the K_M for T to 5.8 µM. Combined with a 5-fold increase in turnover, ssTK2's specificity constant for T surpasses even the rational design triple mutant rTK3 by 2.5-fold, reaching a 2700-fold improvement in k_{obs}/K_M. Interestingly, the methyl side chain in Thr133 seems to interfere with dC binding, increasing its K_M from 1 µM in wild-type to 21 µM in ssTK2. The reduced apparent binding constant is in part compensated by a 45-fold increase in k_{obs}, still doubling the specificity constant for dC in respect to wild-type dCK. Finally, ssTK3 represents a hybrid of rational and random mutagenesis, carrying the consensus substitution Arg104Met of the former in combination with the unconventional Asp133Asn mutation found in the latter. The significantly higher stability of ssTK3 in comparison to epTK16A suggests that Arg104Gln is responsible for the destabilization of the mutant enzymes identified in the random approach. Otherwise, the mutant enzyme shows similar rate changes for pyrimidine and purine substrates as the other two ssTK constructs. In a follow-up experiment, we introduced Ala100Val as a third mutation in ssTK1 and ssTK2 via site-directed mutagenesis, creating ssTK1A and ssTK2A (Table 2.6). At the same time, the third substitution seemed to magnify the trend in substrate specificity change observed for the double mutants. The additional amino acid change raised k_{obs} and K_M for dC in both enzymes to the extent of leaving their specificity constants largely unaffected (< 2-fold). The most dramatic change in the kinetic parameters is a 3-fold raise in the apparent binding constant in ssTK2A. We speculate that the more bulky valine side chain increases the steric rigidity of the adjacent Thr133, orienting its side chain in such a way that steric clash with the exocyclic amino group of dC prevents effective substrate

binding. In contrast, the reduction in side chain flexibility seems to benefit T as an enzyme substrate as Ala100Val results in even lower K_M values. At 3.88 μM, ssTK2A shows a reduction in K_M by almost three orders of magnitude, the lowest apparent binding constant for T of any dCK mutants characterized so far. The effect can be rationalized by a more favorable orientation of the hydroxyl group of Thr for hydrogen bonding with the carbonyl in the 4-position on thymine. Overall, the three mutations in ssTK2A raise the k_{obs}/K_M for T to $4.06 \times 10^5 \, M^{-1}s^{-1}$, more than 4000-fold above wild-type dCK. For purine substrates, the additional substitution in position 100 lowers the specificity constants for dA about 3-fold while leaving dG phosphorylation largely unaffected. The decline in activity for dA can mostly be attributed to increased K_M values. Interestingly, the Ala100Val substitution in rTK3 resulted in exactly the opposite effect, reducing the k_{obs}/K_M for dG but not significantly altering the kinetic parameters for dA.

2.3.5 Nucleoside analog activity of dCK mutants

Beyond our primary study of the effects of active site mutations on the substrate specificity of natural dNs, we evaluated wild-type dCK and six of the dCK mutants for their activities on nucleoside analogs with distinct sugar modifications. Four 2′-deoxycytidine analogs; ddC, 3TC, AraC and gemcitabine were tested as shown in Table 2.7. Three thymidine analogs were also used for specific activity analysis, 2′,3′-dideoxythymidine (ddT), 3′-azido-3′-deoxythymidine (AZT) and 5-(2-bromovinyl)-2′-deoxyuridine (BvdU) (Table 2.7). Although none of the dCK mutants was selected for those particular substrates, our active site changes likely extend beyond nucleobase recognition and affect binding interactions with the sugar moiety of the substrate as well.

Table 2.7 Substrate specificities of wild-type and mutant dCK enzymes.

NA	AraC	gemcitabine	3TC	ddC
wild-type	984 (1)	1648 (1)	88 (1)	461 (1)
r-TK3	3308 (3.4)	4530 (2.7)	149 (1.7)	100 (0.22)
epTK6	4400 (4.5)	6460 (3.9)	815 (9.3)	86 (0.19)
ssTK1	3339 (3.4)	5624 (3.4)	545 (6.2)	66 (0.14)
ssTK1A	3177 (3.2)	6009 (3.6)	270 (3.1)	23 (0.05)
ssTK2	1387 (1.4)	4267 (2.6)	114 (1.3)	15 (0.03)
ssTK2A	599 (0.6)	2262 (1.3)	26 (0.3)	6 (0.01)

NA	ddT	AZT	BvdU
wild-type	6.5 (1)	1.0 (1)	40 (1)
r-TK3	64 (10)	43 (43)	2880 (72)
epTK6	206 (32)	136 (136)	2842 (71)
ssTK1	28 (4.3)	8.5 (8.5)	2639 (66)
ssTK1A	186 (29)	107 (107)	3695 (92)
ssTK2	108 (17)	45 (45)	4552 (114)
ssTK2A	135 (21)	113 (113)	2643 (66)

Activities (μM/min per gram enzyme) were determined by spectrophotometric assay at a constant substrate concentration of 500 μM (standard error: $\pm$ 10%). Relative specificity changes compared to wild-type enzyme are shown in parenthesis.

The kinetic data for AraC and gemcitabine show moderate improvements in activity among the engineered kinases. We attribute the observed 2 to 4-fold activity increases for both NAs, which are already good substrates for the wild-type dCK, mostly to changes in position 104. Substitution of Arg104 might be beneficial to directly or indirectly (by repositioning the neighboring Glu53 and Arg128) reduce unfavorable steric and electrostatic interactions with the additional substituents at the 2′-position of the ribose

[11]. The same rational could also explain the up to 10-fold activity enhancement observed for 3TC [10]. On the contrary, ddC activity decreased between 5 to 80-fold. This significant decline pattern might be due to removing the already limited hydrogen binding interactions to ddC that lacks both hydroxyl groups at the 2′- and 3′-position.

The dramatic improvement in activity for NAs with thymine nucleobases is consistent with the newly acquired T specificity in the dCK mutants. While wild-type dCK shows only residual activity for any analog, the six mutant enzymes show 30-fold higher activity for ddT and a 100-fold enhancement for both AZT and BvdU turnover. Assuming that the enzymes operate under V_{max} conditions at 500 μM substrate, it is interesting to note that the activity changes for the NAs do not correlate with the changes measured for the natural counterparts T and dC. In epTK6, dC activity increases 3.3-fold while 3TC turnover raises 9.3-fold. In contrast, ssTK2A has 55-fold higher activity for dC but a 3-fold decline in 3TC activity. Similarly, epTK6 shows 1.8-fold higher turnover for T and 136-fold activity increase for AZT while ssTK1 shows similar moderate rate increases for T and AZT of 6 and 8.5-fold, respectively. The absence of a clear correlation between data from the native substrates and the NAs supports the notion that the amino acid substitutions in position 104 and 133 extend beyond the nucleobase binding portion of the active site.

Finally, a comparison of the kinetic properties of dCK mutants based on the engineering strategy reveals some interesting differences in the enzyme's response. To introduce tk activity, rational design based on multiple sequence alignments correctly identifies residue Arg104 and Asp133 in dCK as the critical amino acids for the desired changes in substrate specificity. However, faster and more specific mutants could be isolated from

site-saturation mutagenesis libraries, targeting these same two positions, in combination with genetic selection in the tk-deficient *E. coli* KY895. On the other hand, random mutagenesis led to the identification of a generalist, a mutant with substantial activity for all four natural dNs as well as NAs. EpTK6 consistently shows the highest activity for all tested NA substrates.

2.3.6   Biophysical characterization

The effects of the mutations on the overall protein structure and thermal stability for all mutant enzymes compare to wild-type dCK were evaluated by CD spectroscopy (Figure 2.3). The far-UV CD spectra of wild-type and dCK variants were fundamentally identical at ambient temperatures, indicating that no significant secondary structure changes were detected.

CD spectroscopy was used to investigate the thermally induced unfolding and determine the melting point of individual mutant enzymes. T_M is the temperature at which half of the protein exists in the native state and the other half in the denatured state. When the temperature was increased, all constructs showed a sigmoid, symmetrical decrease in ellipticity at 222 nm (Figure 2.4). Our data illustrates a classical two-state denaturation process [35]. The protein denaturation was irreversible for all constructs as no reverse folding was observed after cooling down the samples. The data was fitted to the Boltzmann equation in order to calculate the melting temperature.

Figure 2.3   Far-UV CD spectra of wild-type and mutant dCK enzymes.
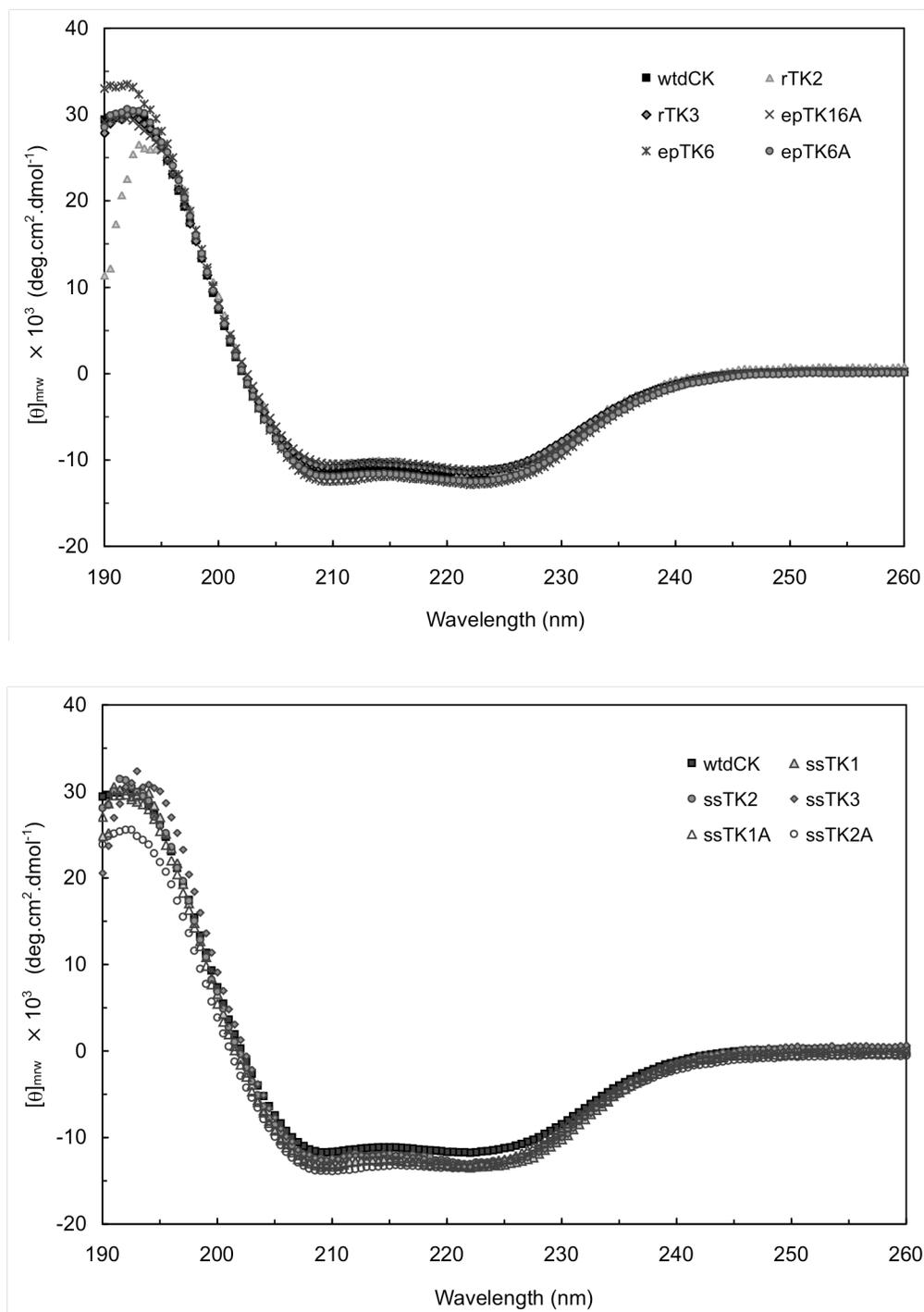
Figure 2.4 Representation of CD thermal denaturation for ssTK2.
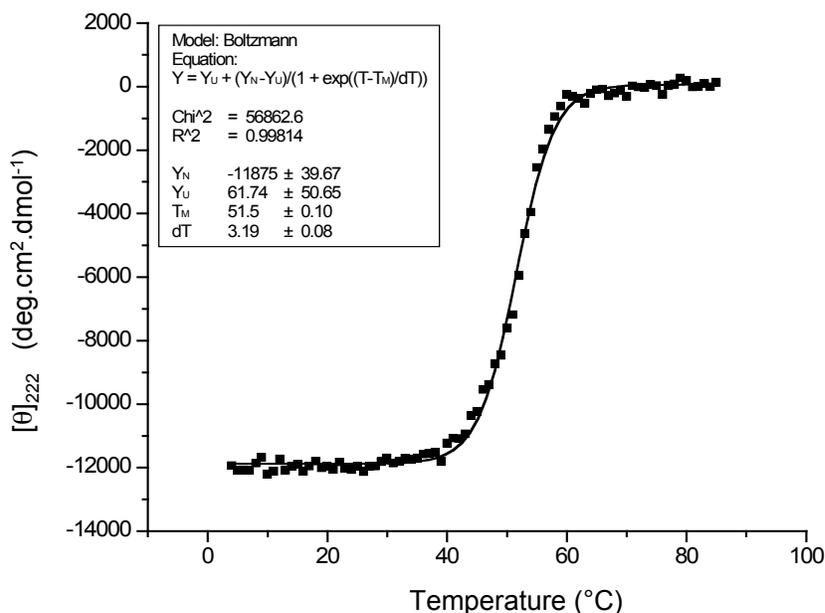
All measurements were performed in duplicate. Recombinant wild-type dCK has an average melting point of 59.5 °C. Denaturation experiments show a decrease in the temperature of unfolding for all mutants. The collected data and all amino acid substitutions are summarized in Table 2.8. Two rational mutants; rTK2 and rTK3 presented a 5 and 9°C drop in T_M, respectively. In contrast to epTK16A, the removal of the three distant substitutions in epTK6 seems to reduce protein stability compared to the parental sequence. While the five mutations in epTK6 cause a 12°C decline in protein stability, epTK6A demonstrates a reduction of its T_M by 15°C. This difference suggests that the three distant mutations actually compensate in part for the protein's destabilization by the mutations in the active site. The conservative nature of the substitutions and their distance to the active site make it difficult to further rationalize these effects.

Table 2.8   The effects of amino acid substitutions on protein stability.

| | amino acid residue | | | | | | | |
mutant	47	100	104	133	163	202	242	T_M (°C)
wild-type	D	A	R	D	N	L	F	59.5 ± 0.3
rTK2			M	A				54.5 ± 0.1
rTK3		V	M	A				51.0 ± 0.1
epTK16			Q	N		H		nd
epTK16A			Q	N				45.5 ± 0.1
epTK6	E		Q	G	I		L	47.5 ± 0.1
epTK6A			Q	G				44.2 ± 0.2
ssTK1			M	S				54.5 ± 0.1
ssTK1A		V	M	S				46.1 ± 0.2
ssTK2			M	T				51.5 ± 0.1
ssTK2A		V	M	T				48.4 ± 0.3
ssTK3			M	N				55.6 ± 0.1

The three site-saturated proteins (ssTK1, ssTK2 and ssTK3) show equal or higher temperatures of unfolding than the proteins from rational and random mutagenesis with transition midpoints at 54.5°C, 51.5°C and 55.6°C, respectively. The introduction of Ala100Val substitution into ssTK1 and ssTK2 lowered the overall thermostability similar to the drop observed for rTK3 (9°C) by 13 and 11°C, respectively.

2.4   Discussion

Initial work by Lavie and coworkers suggested that the introduction of tk activity in dCK was relatively straightforward. The authors hypothesized that the wild-type enzyme's inability to phosphorylate T was related to steric clashes and adverse electrostatics of the thymine nucleobase in the active site. Supporting their hypothesis, the authors identified

amino acid position 100, 104, and 133 as responsible for the unfavorable interactions. Inspired by the broad substrate specificity of *Dm*-dNK, they mutated the three residues in dCK to the corresponding amino acids in the fruit fly kinase improved the dCK activity for T by three orders of magnitude. As our subsequent kinetic study of these mutants shows, the enhanced catalytic activity is largely due to a lower K_M value. These findings are consistent with the model of an enlargement of the active site binding pocket. Nevertheless, our experiments also reveal the drop in catalytic activity for purine nucleosides associated with the expansion of the active site which seems at first counterintuitive. The moderate decline in activity for dA can be rationalized by the removal of a hydrogen-bonding interaction between the nucleoside's exocyclic amino group and the side chain carboxylate on Asp133. The 10-fold drop in the specificity constant for dA can be accounted for by the higher Michaelis-Menten constants, a trend also seen with dC which, according to the crystal structure, benefits from the same non-covalent interaction. In contrast, the reduced specificity constant for dG by an additional order of magnitude is more difficult to explain. Given the similar trends in the kinetics of dA and dC, we expected the size difference between pyrimidines and purines to be of lesser importance to catalysis. This led us to predict that an expanded active site and favorable enzyme-substrate interactions that greatly improves tk activity could be beneficial for dG phosphorylation as well. However, our steady-state kinetics does not support such hypothesis. Potential steric conflicts arising from dG's exocyclic amino group are ruled out based on the structural data for clofarabine, showing a sufficiently large binding pocket for substituents in the 2-position of the purine base [13]. Furthermore, multiple sequence alignments indicate a high degree of conservation in that

region of the active site among all members of the type 1 dNK family, providing no indication of special accommodation for dG in dedicated purine kinases such as dGK. A review of our current dG binding model, based on crystallographic information for dA and its derivatives, which places the heterocycle in anti-conformation relative to the sugar moiety, might be necessary. Alternative binding modes might include a syn-orientation of the guanine base in the bound state, as well as relocation of protein side chains and backbone elements. The data analysis would also greatly benefit from crystal structure information of dG bound in the phosphoryl acceptor site of dCK or other type 1 dNK family members.

Another interesting test of the functional relevance of position 100, 104, and 133 in dCK is the comparison of our kinetic data with previously published results for the reverse experiment in *Dm*-dNK. Introducing Val84Ala, Met88Arg, and Ala110Asp in *Dm*-dNK which corresponds to position 100, 104, and 133 in dCK, Knecht and coworkers observed a substantial decline in T and dC kinase activity while purine substrates were largely unaffected and, in the case of dG, even slightly improved [15]. Overall, these findings suggest that the current model captures the critical interaction of the enzyme with pyrimidine deoxyribonucleosides, providing a relatively accurate predictive framework, but is less reliable to forecast effects on purine substrates. The data confirm that residues 104 and 133 are critical for substrate specificity while alterations in position 100 generally enhance the trend in specificity seen in the double mutants.

Naturally, the results from the rational design study did raise the question whether a) Arg104Met and Asp133Ala are the only two mutations leading to tk activity in dCK and b) there are alternative amino acid substitutions in these two positions or elsewhere in the

protein that could accomplish the same change in substrate specificity. We addressed these questions via two complementary techniques; whole-gene random mutagenesis and site-saturation mutagenesis at positions 104 and 133. In searching for mutations throughout the entire dCK sequence that allowed for T phosphorylation in the *E. coli* auxotroph, our random mutagenesis library yielded two unique variants with three and five amino acid substitutions respectively. Besides amino acid changes in positions distant to the active site, both variants showed mutations in positions 104 and 133. Amino acid Arg104 was changed to Gln in both mutants while Asp133 was replaced by either Asn or Gly. In subsequent experiments, the reversion of the distant mutations to WT showed little change in catalytic performance, clearly linking the gains in tk activity to the two active site residues.

Interestingly, the kinetic analysis of these two variants obtained by random mutagenesis shows very distinct substrate specificity profiles compared to the rational design constructs. Even though both groups of enzymes were selected via in vivo genetic complementation, the mutants from the error-prone PCR experiment behave less like pyrimidine kinases but instead show a more balanced substrate profile than rTK2 and rTK3. The Arg104Gln/Asp133Gly mutations in epTK6A improve tk activity 400-fold, mostly by lowering the K_M value for T, but cause less than 6-fold change in the catalytic parameters for dC, dA and dG. With only two amino acid changes, we have been able to create a human dNK with respectable turnover rates for all four natural dNs. The ease of such transformation is intriguing in light of the recently put forward idea of enzyme evolution via a generalist as intermediate state of functional divergence [36]. Along the same lines, evidence that changes in position 104 and 133 are not limited to enzymes

with broader substrate specificity was found in ssTK2A whose substrate profile indicates a complete inversion, turning the dCK into a thymidine kinase with no significant catalytic activity for the other three natural 2′-deoxynucleosides (see below). Our experiment also highlights an often overlooked problem with random mutagenesis. Although methods such as error-prone PCR can introduce random nucleotide substitutions in any given DNA sequence, the simultaneous mutation of more than one nucleotide per codon is a low probability event. Consequently, random mutagenesis libraries are biased in favor of amino acid substitutions that require fewer nucleotide changes per codon. The situation is further exacerbated by experimental limitations in library preparation, together explaining the absence of Arg104Met mutants in our random mutagenesis library.

Given the perceived importance of positions 104 and 133 in dCK with regards to nucleobase specificity, we chose site-saturation mutagenesis of those two residues to overcome the shortcomings of random mutagenesis and obtain a complete set of functional amino acid combinations. The results from this study help elucidate the functional role of these two positions. Position 104 tolerates very limited diversity, accommodating only Met and Gln at 37 °C, as well as Asn at 30 °C selection conditions. In the case of Met104, the loss of hydrogen bonding interactions with Glu53 can account for the moderate 5 °C decline of T_M in Arg104Met. Nevertheless, the crystal structure of *Dm*dNK suggests that the new residue should fit well into a hydrophobic region near Phe126 below the substrate binding pocket. The consequences of the Arg104Gln substitution are less certain. A comparison of epTK16A and ssTK3 which share Asp133Asn but carry either Arg104Met or Gln links the observed decline in protein

stability to the Gln residue. We speculate that Gln104 assumes a similar orientation, possibly interacting with Gln108 on the adjacent helical turn. Although accommodating Gln104's polar side chain is less favorable, the strikingly similar trend in the kinetic parameters for epTK16A and ssTK3 argues against a direct involvement of the amino acid in that position in substrate binding or catalysis. As such, our kinetic data are consistent with the proposed orientation of the side chain, points away from the active site.

In position 133, the substitution of Asp with small hydrophobic residues such as Ala or Gly seems a suitable compromise for substrate binding and catalysis with all four natural dNs. The amino acid change eliminates unfavorable electrostatic interactions for T binding while only slightly compromising the Michaelis-Menten constants for dC and dA, which benefit from an additional hydrogen bonding interaction in the wild-type enzyme. In the case of dG, all mutations except Asp133Gly appear detrimental to enzyme activity. An argument for steric constraints seems unsupported by the kinetic data and the structural data for dCK with bound clofarabine as discussed above. In the absence of direct structural evidence, we hypothesize that the introduction of Gly133 can result in higher backbone flexibility, which could aid conformational changes in the enzyme along the reaction coordinate.

Site-saturation mutagenesis in combination with selection for tk activity also identified a second category of functional substitutions in position 133; small polar amino acids such as Ser, Thr, and Asn. In contrast to the rational design mutants with pyrimidine kinase specificity and the random mutagenesis variants with broad dNK specificity, the introduction of Ser and Thr resulted in a complete inversion of dCK's substrate

specificity. Best demonstrated by ssTK2, Thr133 raises the apparent binding affinity for dC and the purine substrates, which we attribute to unfavorable interactions between the amino acid side chain and the nucleobase moiety of the substrate. At the same time, the hydroxyl side chain can establish hydrogen bonding interactions with the thymine base, explaining its dramatically lower K_M value for T. The introduction of Ala100Val in ssTK2A further amplifies the trend, resulting in a dCK variant that, at physiological substrate concentrations in the low micromolar range, functions primarily as a thymidine kinase.

In summary, our study demonstrates that positions 104 and 133 are responsible for the nucleobase specificity of dCK. Substitutions in position 133 can directly influence substrate specificity via steric and electrostatic effects, allowing for tailoring of dCK's substrate profile from "generalist" kinases to thymidine kinases. In contrast, position 104 seems to play only an indirect role in substrate specificity by shaping the substrate-binding pocket of the mutant enzymes. Even though the two residues in this study are part of the active site region near the phosphoryl acceptor's nucleobase, our experiments with NAs confirm that the effects of mutagenesis are not locally restricted. The seven antiviral and cancer prodrugs in our tests (Table 2.7) carry native thymine or cytosine moieties, yet have undergone various degrees of modifications in the ribose portion except BvdU. The results of our activity measurements clearly show in some cases substantial rate enhancements for the phosphorylation of these NAs. Furthermore, the previously discussed "generalist" behavior of error-prone PCR mutant epTK6 extends to these substrates. The Arg104Gln/Asp133Gly mutations yield the most active kinase for NAs among all mutants.

Our experiments demonstrate the simplicity by which the substrate specificity of dCK, and presumably other members of the type 1 dNK family, can be manipulated. These results are encouraging in light of various hypotheses about the evolutionary trajectory of dNKs. Nevertheless, a major limitation in the current study is the inability to select or screen for nucleoside kinase activity other than T. Our experiments rely on the genetic complementation of the tk-deficient *E. coli* strain KY895 for library analysis, inevitably biasing the results towards tk activity. The absence of screening and selection protocols that allow direct testing for dC, dA and dG activity is a major problem and severely restricts progress in the field. Future work would greatly benefit from the development of novel screening techniques, not only allowing for more comprehensive tests on the enzyme's residues responsible for nucleobase specificity, but also offering the opportunity to expand the active site probing to residues responsible for recognizing and discriminating the sugar moiety of the substrate. As many nucleoside analog prodrugs carry key modifications in the ribose portion, such a protocol would be invaluable to directly identifying dNKs with higher NA activity and even generate orthogonal NA kinases.

References

1. Arner ESJ, Eriksson S: **Mammalian deoxyribonucleoside kinases.** *Pharmacology & Therapeutics* 1995, **67:**155-186.

2. Galmarini CM, Mackey JR, Dumontet C: **Nucleoside analogues and nucleobases in cancer treatment.** *Lancet Oncol* 2002, **3:**415-424.

3. Johansson M, Karlsson A: **Differences in kinetic properties of pure recombinant human and mouse deoxycytidine kinase.** *Biochem Pharmacol* 1995, **50:**163-168.

4. Kewn S, Veal GJ, Hoggard PG, Barry MG, Back DJ: **Lamivudine (3TC) phosphorylation and drug interactions in vitro.** *Biochem Pharmacol* 1997, **54:**589-595.

5. Kierdaszuk B, Krawiec K, Kazimierczuk Z, Jacobsson U, Johansson NG, Munch-Petersen B, Eriksson S, Shugar D: **Substrate/inhibitor specificities of human deoxycytidine kinase (dCK) and thymidine kinases (TK1 and TK2).** *Adv Exp Med Biol* 1998, **431:**623-627.

6. Lotfi K, Mansson E, Spasokoukotskaja T, Pettersson B, Liliemark J, Peterson C, Eriksson S, Albertioni F: **Biochemical pharmacology and resistance to 2-chloro-2'-arabino-fluoro-2'-deoxyadenosine, a novel analogue of cladribine in human leukemic cells.** *Clin Cancer Res* 1999, **5:**2438-2444.

7. Shewach DS, Liotta DC, Schinazi RF: **Affinity of the antiviral enantiomers of oxathiolane cytosine nucleosides for human 2'-deoxycytidine kinase.** *Biochem Pharmacol* 1993, **45:**1540-1543.

8. Eriksson S, Munch-Petersen B, Johansson K, Eklund H: **Structure and function of cellular deoxyribonucleoside kinases.** *Cellular and Molecular Life Sciences* 2002, **59:**1327-1346.

9. Johansson K, Ramaswamy S, Ljungcrantz C, Knecht W, Piskur J, Munch-Petersen B, Eriksson S, Eklund H: **Structural basis for substrate specificities of cellular deoxyribonucleoside kinases.** *Nat Struct Biol* 2001, **8:**616-620.

10. Sabini E, Hazra S, Konrad M, Burley SK, Lavie A: **Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase.** *Nucleic Acids Res* 2007, **35:**186-192.

11. Sabini E, Ort S, Monnerjahn C, Konrad M, Lavie A: **Structure of human dCK suggests strategies to improve anticancer and antiviral therapy.** *Nat Struct Biol* 2003, **10:**513-519.

12. Soriano EV, Clark VC, Ealick SE: **Structures of human deoxycytidine kinase product complexes.** *Acta Crystallogr D Biol Crystallogr* 2007, **63:**1201-1207.

13. Zhang Y, Secrist JA, 3rd, Ealick SE: **The structure of human deoxycytidine kinase in complex with clofarabine reveals key interactions for prodrug activation.** *Acta Crystallogr D Biol Crystallogr* 2006, **62:**133-139.

14. Knecht W, Munch-Petersen B, Piskur J: **Identification of residues involved in the specificity and regulation of the highly efficient multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster.** *Journal of Molecular Biology* 2000, **301:**827-837.

15. Knecht W, Sandrini MP, Johansson K, Eklund H, Munch-Petersen B, Piskur J: **A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines.** *Embo J* 2002, **21:**1873-1880.

16. Usova EV, Eriksson S: **Identification of residues involved in the substrate specificity of human and murine dCK.** *Biochem Pharmacol* 2002, **64:**1559-1567.

17. Godsey MH, Ort S, Sabini E, Konrad M, Lavie A: **Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: illuminating the role of main-chain reorganization.** *Biochemistry* 2006, **45:**452-461.

18. Sabini E, Hazra S, Ort S, Konrad M, Lavie A: **Structural basis for substrate promiscuity of dCK.** *J Mol Biol* 2008, **378:**607-621.

19. Mikkelsen NE, Johansson K, Karlsson A, Knecht W, Andersen G, Piskur J, Munch-Petersen B, Eklund H: **Structural basis for feedback inhibition of the deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase.** *Biochemistry* 2003, **42:**5706-5712.

20. Gentry GA: **Viral thymidine kinases and their relatives.** *Pharmacol Ther* 1992, **54:**319-355.

21. Piskur J, Sandrini MP, Knecht W, Munch-Petersen B: **Animal deoxyribonucleoside kinases: 'forward' and 'retrograde' evolution of their substrate specificity.** *FEBS Lett* 2004, **560:**3-6.

22. Sambrook J, Russell DW: *Molecular cloning: a laboratory manual.* 3rd ed. edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.

23. Lutz S, Ostermeier M, Benkovic SJ: **Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides.** *Nucleic Acids Research* 2001, **29**.

24. Igarashi K, Hiraga S, Yura T: **A deoxythymidine kinase deficient mutant of Escherichia coli. II. Mapping and transduction studies with phage phi 80.** *Genetics* 1967, **57:**643-654.

25. Munir KM, French DC, Loeb LA: **Thymidine kinase mutants obtained by random sequence selection.** *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90:**4012-4016.

26. Munir KM, French DC, Dube DK, Loeb LA: **Herpes thymidine kinase mutants with altered catalytic efficiencies obtained by random sequence selection.** *Protein Eng* 1994, **7:**83-89.

27. Mani RS, Usova EV, Eriksson S, Cass CE: **Hydrodynamic and spectroscopic studies of substrate binding to human recombinant deoxycytidine kinase.** *Nucleosides Nucleotides Nucleic Acids* 2003, **22:**175-192.

28. Gill SC, von Hippel PH: **Calculation of protein extinction coefficients from amino acid sequence data.** *Anal Biochem* 1989, **182:**319-326.

29. Munch-Petersen B, Knecht W, Lenz C, Sondergaard L, Piskur J: **Functional expression of a multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster and its C-terminal deletion mutants.** *J Biol Chem* 2000, **275:**6673-6679.

30. Schelling P, Folkers G, Scapozza L: **A spectrophotometric assay for quantitative determination of kcat of herpes simplex virus type 1 thymidine kinase substrates.** *Analytical Biochemistry* 2001, **295:**82-87.

31. Stratagene LJ, CA (2004): **GeneMorph® II random mutagenesis kit.** *in Instruction Manual***:**2-4.

32. Bosley AD, Ostermeier M: **Mathematical expressions useful in the construction, description and evaluation of protein libraries.** *Biomol Eng* 2005, **22:**57-61.

33. Hayes RJ, Bentzien J, Ary ML, Hwang MY, Jacinto JM, Vielmetter J, Kundu A, Dahiyat BI: **Combining computational and experimental screening for rapid optimization of protein properties.** *Proc Natl Acad Sci U S A* 2002, **99:**15926-15931.

34. LaBean TH, Kauffman SA: **Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics.** *Protein Sci* 1993, **2:**1249-1254.

35. Waldner JC, Lahr SJ, Edgell MH, Pielak GJ: **Nonideality and protein thermal denaturation.** *Biopolymers* 1999, **49:**471-479.

36. Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS: **The 'evolvability' of promiscuous protein functions.** *Nat Genet* 2005, **37:**73-76.

Chapter 3

Site-saturation mutagenesis studies within the putative sugar-binding site residues of human dCK to enhance the prodrug activation

3.1   Introduction

In its biological environment, human deoxycytidine kinase (dCK) is a key enzyme of the salvage pathway, phosphorylating 2′-deoxycytidine (dC), 2′-deoxyadenosine (dA) and 2′-deoxyguanosine (dG) to their corresponding monophosphate forms in the presence of ribonucleoside triphosphates as phosphoryl donors and magnesium. In the cell, other cellular kinases subsequently convert these 2′-deoxyribonucleoside monophosphates to their diphosphate and triphosphate forms and the resulting products eventually incorporate into DNA during repair and synthesis by polymerases [1].

Human dCK also activates various nontoxic chemotherapeutic and antiviral nucleoside analog (NA) prodrugs by catalyzing the initial, often rate-determining phosphorylation step [2]. Clinically relevant NAs mostly differ from their endogenous counterparts in respect to various modifications at the sugar portion of 2′-deoxyribonucleoside (dN). These modifications could be present at the 2′- and/or 3′- position of the ribose moiety of natural substrates such as different 2′-substitutions in chemotherapeutic agents like dFdC (2′,2′-difluorodeoxycytidine) and AraC (1-β-D-arabinosylcytosine) or 3′-substitutions in NAs used against HIV including ddC (2′,3′-dideoxycytidine) and 3TC (2′-deoxy-3′-thiacytidine) (Figure 1.5). Therefore, understanding the fundamentals of substrate recognition in the active site and the structural characteristics that allow efficient phosphorylation of these prodrugs by dCK has been a prime interest for protein engineers to promote the current and future NA prodrug therapies.

The previous study in Chapter 2 has shown that positions 104 and 133 are responsible for nucleobase specificity of dCK. While progress made in understanding the detailed

aspects of the nucleobase binding at the acceptor site of dCK, the most important question awaiting to be unraveled is how to improve the specificity towards modified ribose in NA prodrugs.

To address this issue, we propose that substituting the specific amino acid positions that mold the sugar-binding segment of dCK's active site could facilitate the identification of enzyme variants with higher NA activity and shed some light on this concept. In this chapter, we try to find answers through implementing site-saturation mutagenesis at multiple amino acid sites of dCK to improve AZT (3′-azido-3′-deoxythymidine) phosphorylation. We were inspired by successful engineering attempts on the structurally related deoxyribonucleoside kinases (dNKs) from *Drosophila melanogaster* (*Dm*-dNK) and *Herpes simplex* virus 1 thymidine kinase (HSV1-TK) both members of the type 1 kinase family along with dCK [3-6]. Despite the limited sequence homology between dCK and HSV1-TK, as well as *Dm*-dNK (dCK-HSV1-TK, 16% identity; dCK-*Dm*-dNK, 30% identity), the crystal structures of all three kinases reveal a common fold (see also Chapter 1) [7-9]. Moreover, the two non-human kinases can phosphorylate AZT with low K_M values of <10 μM and share a similar hydrogen-bonding network with the 3′-hydroxyl group of the dN compared to dCK [10, 11].

Guided by the crystal structures and earlier research findings, we have selected six amino acid positions Asn80, Leu141, Met85/Tyr86 and Glu196/Glu197 (corresponding residues are highlighted in Figure 3.1A and 3.1B for dCK and *Dm*-dNK, respectively) nearby the ribose moiety in the active site of dCK for our site-saturation mutagenesis experiments to find a variant with improved turnover rate for AZT phosphorylation.

Figure 3.1   Comparison of the residues that surround the sugar portion of the 2′-deoxyribonucleoside in dCK and *Dm*-dNK with bound substrates.
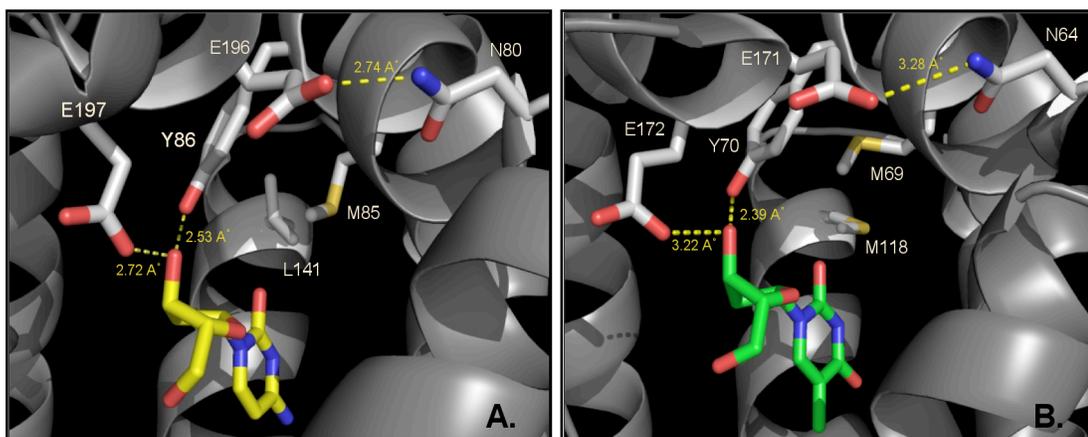
Detailed view of the active site residues where the sugar portion binds: dCK in the presence of dC (**A**; PDB code: 1P60 [7]) and *Dm*-dNK with T (**B**; PDB code: 1OT3 [12]). Hydrogen-bonding interactions between the substrate's 3′-hydroxyl group and dCK's Y86-E197 pair (Y70 and E172 in *Dm*-dNK) are highly conserved. The hydrogen-bonding interaction between N80 and E196 (N64 and E171 in *Dm*dNK) is conserved and important for the closed-active state of dCK. The residues in close proximity M85 and L141 (M69 and M118 in *Dm*-dNK) are also shown in sticks.

Initial HSV1-TK studies including site-directed mutagenesis and DNA family shuffling of the closely related HSV type1 and type 2 TK genes have led to the identification of several key residues; Tyr101, Arg176 and Glu225 (Tyr86, Leu141 and Glu197 in dCK, respectively) that transitions the enzyme's substrate specificity from thymidine (T) to AZT [3, 4]. Independently, two different positions in *Dm*-dNK, Asn64 and Glu171 (Asn80 and Glu196 in dCK), were shown to be important for AZT preference [5, 11]. Both studies used a negative screening strategy to identify kinases with AZT activity. Using the tk-deficient *E. coli* strain KY895, phosphorylation of AZT by a mutant kinase results in the accumulation of cytotoxic product inside the cell, which triggers an SOS response and eventually kills the host [13]. Library members with AZT activity can be

identified by replica-plating colonies on growth media containing NA and monitoring for its failure to grow.

We have constructed site-saturation libraries of ssTK2A (see in chapter 2), a dCK variant with thymidine kinase activity, at the selected residue positions that are separated into two distinctive groups on the basis of the proximal distance in the protein structure. Recently, a directed evolution approach has been successfully exploited to probe the protein sequence space by randomizing two or three amino acids simultaneously, whose side chains reside spatially close and point in the direction of the active site, for expanding the substrate acceptance of an enzyme [14]. As a result the adjacent residues in dCK, M85/Y86NNK and E196/E197NNK, were saturated simultaneously at both sites in order to gain more tolerance for the conformational changes occurring from the different side chain orientations. Positions Leu141 and Met85/Tyr86 as well as Asn80 and Glu196/Glu197 were combined correspondingly to prepare triple libraries at the indicated three amino acid sites because of their close spatial proximity in the substrate binding pocket. Therefore, the positions were divided into two independent groups: Group-I Asn80/Glu196/Glu197 and Group-II Leu141/Met85/Tyr86 with single, double and triple combinatorial libraries in each group.

Herein, we have utilized the six site-saturated libraries for the *in vivo* screening system to evaluate the possibility of further evolving ssTK2A kinase with increased AZT affinity. In addition, the same libraries were exploited for thymidine kinase activity via the established *in vivo* complementation assay to investigate the conservation degree of Tyr86-Glu197 pair, which stabilizes the 3′-hydroxyl group of ribose moiety with hydrogen-boding interactions. Furthermore, we kinetically characterized the mutants

derived from these experiments to investigate the significance of selected residues that distinguish the natural dNs from their modified analogs. Our results indicate that this particular screening system might not be compatible with dCK for the desired enhanced NA activation outcome.

3.2   Materials and Methods

3.2.1   Materials

Pfu Turbo DNA polymerase (Stratagene, La Jolla, CA) was used for all cloning. Enzymes were purchased from New England Biolabs (Beverly, MA), unless otherwise indicated. All reagents were purchased from Fisher (Pittsburg, PA) and Sigma & Aldrich (St. Louis, MO), unless noted. Plasmid DNA was isolated using the QIAprep Spin Miniprep kit and PCR products were purified with the QIAquick PCR Purification kit (Qiagen, Valencia, CA) according to the manufacturer's protocols. Pyruvate kinase and lactate dehydrogenase were purchased from Roche Biochemicals (Indianapolis, IN). Primers were purchased from Integrated DNA Technologies (Coralville, IA). All constructs were confirmed by DNA sequencing.

3.2.2   Construction of site-saturation mutagenesis libraries

Previously constructed ssTK2A served as a template for the new library design in this chapter. Site-saturation libraries of ssTK2A were generated by overlap PCR, using hand-mixed skewed primers N80NNS_f/r, E196/E197NNK_f/r, L141NNK_f/r, and M85/Y86NNK_f/r (Table 3.1). The degenerate codons (N = 1:1:1:1 mixture of dA, dG,

dC, & T; S = 1:1 mixture of dG & dC; K = 1:1 mixture of dG & T) were introduced during automated primer DNA synthesis, using hand-mixing 2′-deoxyribonucleoside phosphoramidites to minimize library biases. PCR reactions were run under following conditions: 95°C for 5 min; 30 cycles of 94°C for 30 sec, 52°C for 30 sec, 72°C for 80 sec; 1 cycle of 72°C for 10 min. The resulting PCR products of six individual libraries were purified, digested with *Nde*I and *Spe*I restriction enzymes, and ligated into pDIM-PGX [15].

Table 3.1   Primers used for library construction and sequencing.

Primer	Sequence
G-1	5'-GCGCATATGGCCAGCTCTGAGGGGACCCGC-3'
G-2	5'-GCGACTAGTTCACAAAGTACTCAAAAACTCTTTG-3'
P-1	5'-CGCGCAATTAACCCTCACTAAAG-3'
P-2	5'-GAATAAGGGCGACACGGAAATG-3'
P-3	5'-GCGAAATTAATACGACTCACTATAGGG-3'
P-4	5'-GCTAGTTATTGCTCAGCGG-3'
N80NNS_f	5'-TCAGATGATGTATGAGAAACCTGAACGATGG-3'
N80NNS_r	5'-GTTTCTCATACATCATCTGAAGAACSNNCCCACCATTT-3'
E196/E197NNK_f	5'-TTACGGGGAAGAAATNNKNNKCAAGGCATTCCTCTTG-3'
E196/E197NNK_r	5'-TTCTTCCCCGTAAATATATTCTATG-3'
L141NNK_f	5'-GAATCTGAATGCATGAATGAGACAG-3'
L141NNK_r	5'-CATGCATTCAGATTCATAKNNATTAGATGCAAAAAT-3'
M85/Y86NNK_f	5'-GTTCTTCAGATGNNKNNKGAGAAACCTGAACG-3'
M85/Y86NNK_r	5'-CATCTGAAGAACATTCCCACCATTTTTC-3'

Competent *E. coli* strain KY895 [16] cells were transformed with the individual ligation mixtures and grown overnight at 37°C on small LB-agar plates containing 100 μg/ml of ampicillin in order to use in the *in vivo* screening assay. The plates were stored at 4°C

until use. The same transformation procedure was followed with the library construction intended for the functional thymidine kinase selection. After overnight growth at 37°C, the cells were recovered with 2YT medium [17], supplemented with glucose (2% w/v) and glycerol (15% v/v). The recovered libraries were divided into aliquots; one portion of the preparation was isolated in plasmid form and the rest was flash frozen in liquid nitrogen to store at -80°C for further analysis. The number of independent clones in each library was estimated by counting the number of colonies from an aliquot of transformants. All constructs were confirmed by DNA sequencing. Plasmid-specific primers P-1/P-2 for pDIM were used for the DNA sequence analysis.

3.2.3 *In vivo* screening for AZT sensitivity

Determination of the effective AZT concentration range for the screen:

The pDIM plasmids carrying the genes for wild-type dCK as negative control, *Thermotoga maritima* thymidine kinase [18] (*Tm*TK) as positive control and ssTK2A as a criterion for the screened mutants were transformed into tk-deficient *E. coli* KY895 cells. All three constructs were grown overnight at 37°C in LB-ampicillin (100 μg/ml) medium followed by 100-fold dilution in 10% (w/v) glycerol. 2 μl of these dilutions were added into the freshly prepared 1 ml LB-medium containing 100 μg/ml ampicillin and various concentrations of AZT. The cell growth of each clone was examined in the presence of 0, 50, 250 and 500 μM AZT by OD_{600} readings after incubation at 37°C for 16 h. The required AZT concentration to kill all the *E. coli* cells also known as a dose producing 100% Lethality (LD_{100}) was identified for the selected kinases.

Screening for the active library members:

The individual colonies from each library plate were picked randomly and inoculated into the 96-well culture plates (USA Scientific, Ocala, FL), which were filled with 250 μl of LB medium containing 100 μg/ml of ampicillin. The plates were incubated overnight with shaking at 37°C. Individual colonies were diluted 200-fold in 10% (v/v) glycerol and 2 μl drops of the dilutions were spotted on M9 minimal plates supplemented with 40 μg/ml isoleucine, 40 μg/ml valine, 0.4% (w/v) glucose, 100 μg/ml ampicillin for screening [5]. The plates were prepared by mixing various concentrations of AZT (0, 1, 2.5, 5, 10 and 25 μM) with the medium at the lowest temperature possible right before pouring them. The degree of colony growth for processed clones including all the controls was visually inspected from the different AZT plates after 24 hour incubation at 37 °C. The active clones with no growth on AZT plates below 25 μM, but normal growth on plates lacking AZT, were picked and cultured in LB containing 100 μg/ml of ampicillin. The plasmids were isolated and introduced into KY895 cells again. These clones were retested two times and the LD_{100} values for isolated colonies were compared with the control plasmids to verify the observed AZT sensitivity. The resulting candidates were analyzed for the mutations by DNA sequencing.

3.2.4 In vivo complementation for thymidine kinase activity

The aliquots of the frozen site-saturated libraries were thawn for each library and 20 μl cell suspension was used to inoculate 50 ml LB medium containing ampicillin (100 μg/ml) and incubated at 37°C. When library cultures reached an OD_{600} of ~0.2, proper dilutions were prepared with LB media as listed in Table 3.5. 1 ml of the resulted

solutions were plated on thymidine kinase selection plates [10, 19] as outlined in chapter 2. All library plates were incubated at 37°C for 16 h and randomly picked colonies were sequenced for mutation analysis.

3.2.5   Expression and purification of the mutant enzymes

Genes of interest were subcloned into expression plasmid, pET-14b (Novagen, Madison, WI) using *Nde*I and *Spe*I sites and subsequently transformed into *E. coli* strain BL21(DE3)pLysS (Novagen). Cell cultures were grown at 37°C in 400 ml 2YT media containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) to an $OD_{600} \sim 0.5$. Protein expression was induced with 0.1 mM IPTG at 28°C for 3 h. Enzymes were purified by a gravity flow Prepcolumn (BioRad, Carlsbad, CA) using Ni-NTA agarose resin (Qiagen), as described previously [20]. Amicon Ultra-4 ultrafiltration units (Millipore, Bedford MA; MWCO: 10 kDa) were used to concentrate the collected protein fractions and exchange the purified proteins into storage buffer (50 mM Tris-HCl, pH 8; 0.5 M NaCl, 5 mM $MgCl_2$, 2 mM DTT). The protein yields were determined as >90% pure by SDS-PAGE. Protein concentrations were quantified by measuring A_{280}, using molar extinction coefficients calculated by the prediction models [21, 22]. Purified protein aliquots were flash frozen in liquid nitrogen and stored at -80°C until further use. The his-tag fusion mutant, C10, was insoluble and to improve its solubility the gene was subcloned into pMal-c2X vector (NEB, Beverly, MA) and subsequently expressed as N-terminal maltose-binding protein (MBP) fusion [23]. The same cloning procedure was followed, as indicated above. The MBP fusion dCK mutant was purified by gravity flow

column using Amylose resin (NEB, Beverly, MA), following the manufacturer's protocol.

3.2.6 Enzyme activity assay

Spectrophotometric assays to determine the thymidine and NA phosphorylation were performed as described previously [20]. Assays were performed at 30°C or 37°C (as indicated), measuring the absorbance change at 340 nm in the presence of $0.28 - 3.60$ μg enzyme per reaction. All experiments were performed in triplicate. The kinetic parameters were calculated from a non-linear least-squares fit of the initial velocity data to the Michaelis-Menten equation using Origin 7 (OriginLab, Northampton, MA).

3.3   Results

3.3.1   Site-saturation libraries

Site-saturation mutagenesis approach was used to introduce combinations of amino acid substitutions within the dCK active site residues that mold the sugar-binding pocket. We generated six small site-saturated libraries using the ssTK2A mutant [20] as a template for AZT screening. Two single-site libraries at positions Asn80 and Leu141, two double amino acid libraries at positions Met85/Tyr86 and Glu196/Glu197, and finally two combined triple-site libraries at positions Asn80/Glu196/Glu197 and Met85/Tyr86/Leu141 were prepared to cover the full randomization of all the positions with the adequate sizes as listed in Table 3.2. The assembled naïve library sizes for each site maintain the coverage between 10-fold to 5000-fold of the theoretical diversity. If the

library size is smaller than 10^4, sampling a 10-fold excess of members should represent almost 100% of the theoretical variants [24]. All the possible codon optimization for each site is maintained with the advantage of including only one stop codon by degenerate NNK primers (only NNS for Asn80) [25].

Table 3.2   The site-saturation library sizes and number of clones screened in in vivo screening

Library	Theoretical size	Library size	Number of clones screened
N80	20	3.0×10^3	134
E196/E197	400	1.1×10^5	320
N80/E196/E197	8000	7.5×10^4	332
L141	20	5.0×10^3	90
M85/Y86	400	2.1×10^5	160
M85/Y86/L141	8000	2.0×10^5	257

Randomly picked colonies were sequenced to evaluate the codon distributions for each amino acid site as following; 47 and 46 DNA sequences for N80NNS and E196/E197NNK libraries, respectively and 18 DNA sequences for each L141NNK and M85/Y86NNK libraries. In our sample sets, we found 20 and 19 /17 out of 20 natural amino acids for the N80 and E196/E197 libraries, respectively. In contrast, nucleotide distributions for positions M85/Y86 and L141 were not well balanced, which could be due to low sample numbers or unbalanced nucleotide mix during the primer synthesis. The sequenced 18 constructs of the M85/Y86 and L141 library members included 8/13 and 15 amino acids, respectively out of 20 natural amino acids. The codon distributions for all six positions are summarized in Table 3.3.

Table 3.3   Codon distributions in position 80, 85, 86, 141, 196 & 197 for site-saturation mutagenesis.

		dA	dG	dC	T			dA	dG	dC	T
N (ideal)		0.25	0.25	0.25	0.25	*N (ideal)*		0.25	0.25	0.25	0.25
S (ideal)		0	0.5	0.5	0	*K (ideal)*		0	0.5	0	0.5
position 80	N	0.3	0.27	0.28	0.15	**position 85**	N	0.28	0.39	0.17	0.16
	N	0.4	0.18	0.23	0.19		N	0	0.12	0.33	0.55
	S	0	0.45	0.55	0		K	0	0.5	0	0.5
position 196	N	0.26	0.17	0.35	0.22	**position 86**	N	0.32	0.12	0.28	0.28
	N	0.26	0.27	0.22	0.26		N	0.33	0.28	0.22	0.17
	K	0	0.46	0.04	0.5		K	0	0.39	0	0.61
position 197	N	0.25	0.35	0.2	0.2	**position 141**	N	0.22	0.17	0.17	0.44
	N	0.22	0.37	0.24	0.17		N	0.39	0.11	0.17	0.33
	K	0	0.46	0.02	0.52		K	0.22	0.06	0.44	0.28

3.3.2   *In vivo* screening of mutants for AZT sensitivity

Before starting screen experiments, we decided to determine the sub-lethal concentration of AZT for the growth plates for the *E. coli* KY895 cells [16]. AZT might be lethal to the cells by itself via interactions between the highly active 3′-azido groups and the intracellular environment, thus interfering with the cellular metabolism. For this purpose, we used KY895 cells carrying the wild-type dCK plasmid with no ability to convert AZT to its monophosphate form in the presence of 0, 50, 250 and 500 μM concentrations of AZT in the LB-medium. The AZT concentration in which zero reading was observed at OD_{600} was 250 μM for wild-type dCK. Simultaneously, ssTK2A was used as a standard measure to adjust the proper AZT dilutions for the screening plates in order to identify the better candidates than the starting dCK variant. Meanwhile a thymidine kinase from our lab which efficiently activates AZT with a K_M value of ~1 μM, *Tm*TK [18], was used as a positive control to compare and verify the sensitization effect of ssTK2A variants. The measured zero readings for AZT concentration was 50 μM for ssTK2A.

After setting up the 50 μM limit associated with the prior ssTK2A results, the following dilutions; 1, 2.5, 5, 10 and 25 μM AZT, were used to prepare the screening plates. The KY895 cells transformed with wild-type dCK (negative control) displayed at AZT concentrations up to 25 μM per plate. In contrast, the cells transformed with *Tm*TK (positive control) did not grow on plates with 1 μM or higher concentrations of AZT. Finally, KY895 cells transformed with ssTK2A did not grow on plates with 25 μM AZT but showed normal growth at 10 μM AZT or below.

The numbers of screened clones were 134, 320 and 332 colonies for the N80, E196/E197 and N80/E196/E197 libraries, respectively and 90, 160 and 257 library members for the individual L141, M85/Y86, and M85/Y86/L141 libraries, respectively (Table 3.2). According to the survival degree of the transformed individual bacteria on AZT containing plates, we found four positive hits with 100% sensitization at either 5 μM or 10 μM after screening several hundred clones from each site-saturated libraries. To verify the functionality, the selected clones were tested three times on the same plate by spotting them in different sections of the plate and comparing the sensitization to the negative and positive controls as illustrated in Figure 3.2.

Two of these mutants were from the E196/E197 double library: C9 (Glu196Cys/Glu197Asp) and C10 (Glu196Val/Glu197Gly) with a LD_{100} value of 10 μM and 5 μM, respectively. The remaining two clones were from the M85/Y86/L141 triple library: D4 (Met85Val/Tyr86Met/Leu141Leu) and D5 (Met85Val/Tyr86His/Leu141Leu) with a LD_{100} value of 5 μM and 10 μM, respectively. Both variants showed silent codon mutations in position Leu 141.

Figure 3.2   A representation of colony sensitization on AZT plates for screening.
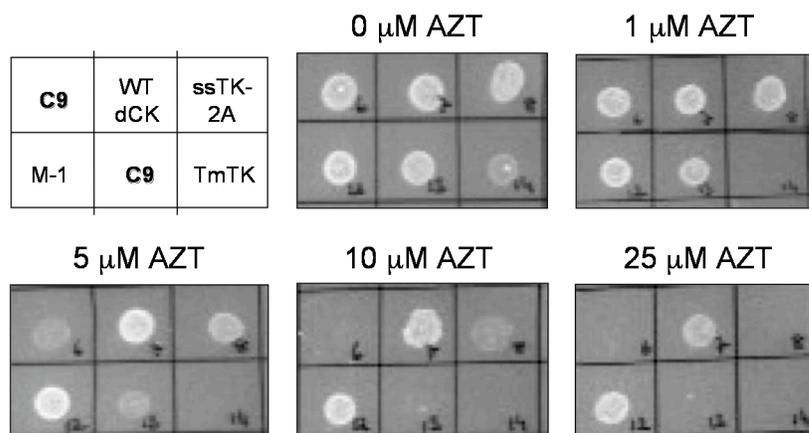
Various plasmids, shown in the first diagram, were transformed into KY895 and the corresponding cells were examined for growth, white colony, on the plates containing AZT. The concentrations of AZT are indicated per plate, accordingly. The selected mutant, C9 (C196/D197) showed poor growth at 5 μM and no growth on both 10 μM and 25 μM plates where spotted at each sections (LD_{100} : ~10 μM). Unlike the C9 mutant, some of the screened mutants like M-1 showed growth at all AZT concentrations, as shown in the picture above.

Table 3.4   Summary of the amino acid substitutions for the ssTK2A variants found from the AZT screening

Library	Mutants	Residue substitution	AZT (μM) / Cell viability				
			0 μM	1 μM	5 μM	10 μM	25 μM
N80			+	+	+	+	+
E196/E197	**C9**	C196/D197	+	+	+	−	−
	C10	V196/G197	+	+	−	−	−
N80/E196/E197			+	+	+	+	+
L141			+	+	+	+	+
M85/Y86			+	+	+	+	+
M85/Y86/L141	**D4**	V85/M86/L141	+	+	−	−	−
	D5	V85/H86/L141	+	+	+	−	−

The cell viability of the selected mutants for each plate at indicated AZT concentrations are represented as (+) growth and (-) no growth on the AZT plates.

All the selected ssTK2A variants were retested for two more times. Table 3.4 summarizes the screening results of the promising candidates for AZT sensitivity. The selected mutants were isolated and further characterized in vitro for AZT activity.

3.3.3 *In vivo* functional selection for thymidine kinase activity

The highly conserved Tyr-Glu pair is important for positioning and binding of 3′-OH group in 2′-deoxribonucleosides and also discriminating against ribonucleosides through the proximity of the Tyr86 hydroxyl group to the unoccupied 2′-sugar position. To obtain more insight into these interactions and to test the necessity of the conserved amino acids for the kinase activity, we have selected the individual library members that retained thymidine kinase activity by complementation of the auxotroph *E. coli* strain KY895. Numerous colonies survived the functional selection as listed in Table 3.5. The main reason for this result is possibly due to using ssTK2A as a template which displays a low K_M value and high catalytic efficiency of $4.0 \times 10^5\,\mathrm{M^{-1}s^{-1}}$ for thymidine. The accumulated substitutions in the selected variants determined by DNA sequence analysis (Table 3.6).

Table 3.5   Thymidine kinase selected site-saturation libraries.

Library	Naïve library size	Plated library size	Number of TK-selected	Sequenced clones
N80	3.0×10^3	3.0×10^3	~ 200	7
E196/E197	1.1×10^5	9.2×10^4	> 1000	5
N80/E196/E197	7.5×10^4	1.3×10^5	96	6
L141	5.0×10^3	4.0×10^3	105	10
M85/Y86	2.1×10^5	2.0×10^5	100	10
M85/Y86/L141	2.0×10^5	2.0×10^5	> 1000	9

According to the sequencing results, Glu197 is conserved in all 11 selected sequences of the E196/E197 and N80/E196/E197 libraries. This result explains the requirement of selected Glu197 for the natural dNK activity. Glu196 can be substituted with Gly (3 out of 11), Pro, Ser, Ala (each 1 out of 11) and Glu in 5 sequences. Based on our analysis of the N80 and N80/E196/E197 libraries, position 80 is changed into Asp (6 out of 13), Ser (5 out of 13) and Gly (2/13). In the second group, the single-site L141 and M85/Y86 libraries tolerate a broader range of amino acid substitutions with some dependency on the shape and size complementary in the combination libraries. Variants in the L141 library includes Val (4/9), Thr (2/9), Ser, Asn, Ile (each 1 out of 9 sequences). Within the M85/Y86 library, position 85 is mainly substituted with Phe (6/10) while Y86 tolerates more variety; Ala (2/10), Gly, Ser, Met, Ile (each 1 out of 10 sequences). Fewer times, we found M85 to Leu (2/10) along with Y86 to Trp and Met (each 1 out of 10 sequences).

Table 3.6   Amino acid changes in the site-saturated mutants with thymidine kinase activity.

N80	E196/E197		N80/E196/E197			L141	M85/Y86		M85/Y86/L141		
80	**196**	**197**	**80**	**196**	**197**	**141**	**85**	**86**	**85**	**86**	**141**
N	**E**	**E**	**N**	**E**	**E**	**L**	**M**	**Y**	**M**	**Y**	**L**
3 S	2 G	E	2 D	E	E	4 V	2 F	A	2 M	Y	I
2 D	S	E	2 S	E	E	2 T	F	G	M	Y	V
2 G	A	E	D	P	E	S	F	S	M	Y	S
	E	E	D	G	E	N	F	M	M	Y	C
						I	F	I	M	Y	Q
							L	W	M	L	F
							L	M	V	H	L
							V	M	A	T	Y
							C	Y			

Finally, the triple library (M85/Y86/L141) was largely unaffected in positions 85/86; M85 and Y86 remain the same in 7 and 6 sequences out of 9, respectively. At the same

time, Leu141 was substituted with Ile, Val, Ser, Cys, and Gln (1 each). Interestingly, we also obtained unique sequences as follows; M/L/F, A/T/Y and V/H/L in the 85/86/141 library. All the amino acid combinations will be discussed in the next section related to double or triple mutation for the site-saturated libraries. M85F/Y86A (D8) mutant was isolated and characterized to analyze the kinetic performance.

3.3.4 *In vitro* characterization of the selected mutants

To further evaluate the *in vivo* activity for selected variants C9, C10, D4 and D5 from the AZT screen, as well as D8, selected from the in vivo complementation assay, were subcloned into pET-14b and expressed in the *E. coli* BL21 as described in the methods section. We obtained soluble proteins for all of them except C10. Alternately, we used an N-terminal maltose binding protein (MBP) tag to obtain soluble MBP-C10 fusion protein. Nevertheless no activity was detected for this mutant with any substrates.

Figure 3.3   The specific activities of ssTK2A and mutants for AZT and T.
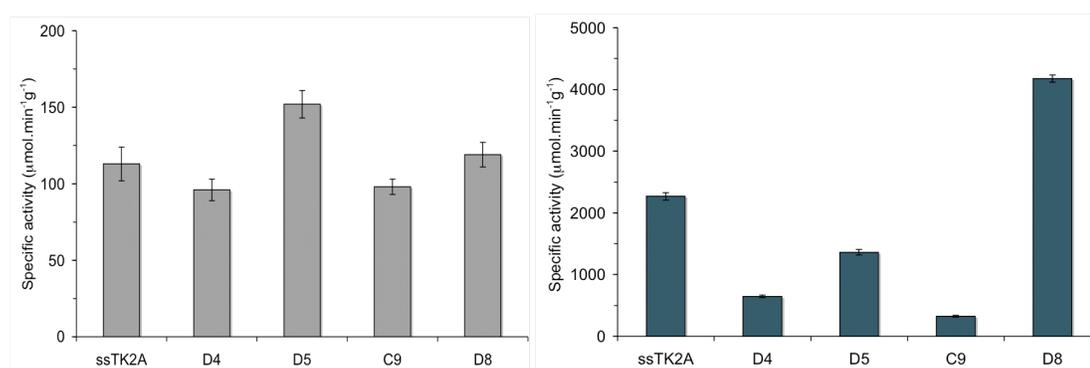
Specific activities (μM/min per g enzyme) were determined by spectrophotometric assay at a substrate concentration of 500 μM AZT (grey scale) and thymidine (blue scale). The experiments were performed in triplets at 30 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation.

The activity of C9, D4, D5, D8 and ssTK2A was detected at 500 μM AZT and T to conclude the kinetic effects. All mutants could phosphorylate AZT more or less in the same range between 0.9 to 1.4-fold compare to ssTK2A (Figure 3.3). D4, D5 and C9 retained 30, 60 and 14% of ssTK2A's activity for thymidine. More interestingly, D8 possessed 1.8-fold greater activity towards T (Figure 3.3). Purified mutants from the M85/Y86/L141 libraries were also tested towards uridine (1-β-D-ribofuranosyluracil) to investigate the 2′-OH group acceptance by the small amino acid substitutions at position 86, yet no activity was detected (data not shown).

Table 3.7   Kinetic parameters of ssTK2A and D8 mutant with T and AZT.

enzyme	T			AZT		
	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times$ 10^3	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times$ 10^3
ssTK2A	3.88 ± 0.22	1.57 ± 0.02	406	1361 ± 124	0.28 ± 0.01	0.205
D8	167 ± 13	3.05 ± 0.08	18	995 ± 119	0.18 ± 0.03	0.185

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation.

A detailed kinetic analysis was performed for ssTK2A and the D8 mutant (Table 3.7). The catalytic efficiency of D8 for T phosphorylation was 23-fold lower than ssTK2A, due to a dramatic increase in K_M value by 43-fold. AZT kinetics more or less was unaffected compare to ssTK2A. Unlike for thymidine, AZT turnover was slow (k_{obs}<1s^{-1}) for both kinases with high K_M values of 1361 μM and 995 μM, respectively.

3.4   Discussion

Despite the progress made in understanding detailed aspects of dCK ligand binding, several key questions regarding the improvement of NA activation remain uncertain [7, 26, 27]. In our attempt to rationalize this property for NAs, we have prepared a series of dCK site-saturated mutants at the highly conserved positions, proposed to be critical for AZT phosphorylation and characterized the biochemical properties of the selected mutants from the screening.

The structure/sequence-based design of human kinase mutant with enhanced NA activity rather than exogenous kinases (e.g. HSV1-TK) is becoming more attractive for gene therapy applications as it minimizes the risk for immunogenic reactions caused by exogenous enzymes [28, 29]. Three independent mutational studies regarding HSV1-TK and *Dm*-dNK on the well-studied paradigm of AZT activation suggested some critical residue substitutions for the increase in specificity for AZT compare to the wild type sequences via *in vivo* screening method with tk-deficient *E. coli* KY895 [11, 30]. Alignment of the dCK amino acid sequence with the *Dm*-dNK and HSV1-TK sequences revealed a close evolutionary relationship [31], justifying our strategy for using prior data as guidelines for our studies.

Such structural and sequence similarities hint that the mutation positions obtained from AZT sensitive HSV1-TK and *Dm*-dNK mutants should be applicable at the equivalent sites for dCK as well. To prove such concept, site-saturation mutagenesis was utilized at the highly conserved residues Glu197 plus Tyr86. These residues stabilize the deoxyribose by forming two hydrogen bonds the 3′-hydroxyl group on substrate sugar,

which is replaced by a more bulky azido group in AZT. The Glu225Leu mutant (Glu197; dCK) and the double mutant Tyr101His/Arg176Trp (Tyr86 and Leu141 in dCK) of HSV1-TK were shown to each increase AZT specificity by 200-fold [3, 4]. We also included the adjacent Glu196 and Met85, as well as Leu141 and Asn80, located in close proximity to the sugar-binding portion of the active site, with the intention of improving the activity of dCK towards medically relevant NAs.

Figure 3.4   The selected residues for dCK mutagenesis and the corresponding residues in related type 1 kinases.

```
dCK       80 : N      85/86   : MY    141 : L    196/197 : EE
Dm-dNK    64 : N      69/70   : MY    118 : M    171/172 : EE
hdGK      94 : N      99/100  : MY    155 : L    210/211 : EE
hTK2      62 : N      67/68   : MY    116 : L    169/170 : EE
HSV1-TK   95 : E     100/101  : IY    176 : R    224/225 : GE
```

All six amino acids are structurally aligned for human dCK, dGK, TK2, Dm-dNK and HSV1-TK. Identical residues for all kinases are in red and semi-conserved residues are in blue. The amino acid numbers are given according to the related kinase.

We generated six small, focused libraries rather than randomization at six amino acid positions simultaneously, since the theoretical diversity of this library is expected to be on the order of 64×10^6 for a 100% coverage at six codons. Moreover, grouping the spatially close amino acids for the construction of combinatorial libraries should compensate the disfavorable side chain orientations resulted from the simultaneous substitutions per residue. Therefore, the selected positions were divided into two main independent groups: Leu141, Met85/Tyr86 as Group-I and Asn80, Glu196/Glu197 as Group-II amino acids.

In an effort to find variants of dCK with improved activity for AZT, a triple dCK mutant (ssTK2A) from our previous study was used as a template for the second-generation mutagenesis studies for the following reasons: *(i)* low K_M value of 3.88 μM for T, meaning that active site can accommodate the thymine base, and *(ii)* relatively stable (T_M of 48.4°C), providing sufficient robustness for additional mutations.

Our toxicity experiments on *E. coli* KY895 cell culture revealed that AZT is cytotoxic at concentrations above 250 μM in the absence of kinase activity. Expression of our kinase template, ssTK2A, further reduced the toxicity to ~50 μM. We performed the screening of randomly picked library members using 1 μM, 2.5 μM, 5 μM, 10 μM and 25 μM dilutions of AZT on minimal plates. Four dCK mutants came through the negative genetic selection, showing sensitization at 5 μM and 10 μM AZT. Unfortunately, none of the screened variants showed sensitivity at either 2.5 or 1 μM AZT like the positive control, *Tm*TK. Two mutants from the E196/E197 library, C9 and C10 (LD_{100} : 10 μM and 5 μM), and the remaining mutants from the M85/Y86/L141 library, D4 and D5 (LD_{100} : 5 μM and 10 μM) all showed higher LD_{100} values than the previously reported HSV1-TK and *Dm*-dNK. These results are in complete agreement with the slight changes in velocity for AZT phosphorylation, indicating that the elevated amounts of AZT may possibly produce internal toxic effects by other pathways and as a result would interfere with the cell growth during the screening.

Furthermore, the solubility of C10 was severely diminished probably due to Val196/Gly197 mutations in the LID region, which led to additional flexibility and damaged the integrity of the whole protein. This result is contradictory to the kinetically characterized HSV1-TK Glu225Leu mutant (Glu197; dCK) carrying a Gly residue in

position 224 naturally (Glu196; dCK). More so, the result could be an interesting case for the unpredictable outcome of similar mutations in different kinases.

Since the genetic complementation for tk activity strictly conserves the Glu residue at position 197, the amino acid is confirmed to be required for the natural 2′-deoxyribonucleoside phosphorylation. In contrast, the observed exchangeability of Tyr86 from the selected TK mutants suggests that, rather than forming a hydrogen bond to the 3′-OH group of ribose moiety, the main task for this residue is to maintain the preference for deoxyribonucleosides over ribonucleosides. Its orientation in close proximity to the 2′-position of the sugar moiety causes steric hindrance for any substituents larger than fluorine, for example gemcitabine (low K_M of 22 µM for 2′-fluorine) versus cytidine (high K_M of 383 µM for 2′-OH) [7]. We propose that certain mutant enzymes with small residue substitutions in Tyr 86 (e.g. Ala, Ser, His and Met) should present better binding for any large groups at 2′-position thus an improved activation of the substrate. When we tested this hypothesis using uridine, the kinetics of the mutant enzymes was inconsistent with the proposed scenario. Only residual phosphorylation was detected.

Our kinetic findings for ssTK2A and D8 mutant using AZT were not promising in comparison to Dm-dNK, HSV1-TK and their respective N64D and E225L mutants, as shown in Table 3.8. One potential reason that this system failed to provide the expected results is weak AZT binding constant for ssTK2A (K_M of 1361 µM) versus the tighter AZT binding properties of Dm-dNK and HSV1-TK (K_M of 8.3 µM and 5.2 µM, respectively). The mutations in D8 cause a noticeable raise in the binding constant for T due to loss of a hydrogen bond to the 3′-OH of the deoxyribose ring by Tyr86Ala while leaving the AZT binding mostly unaffected. The other reason for this outcome could be

that the less strict screening methods can give rise to false positives or unpredictable evolvability.

Table 3.8   Kinetics of T and AZT phosphorylation by purified mutant proteins compare to *Dm*-dNK and HSV1-TK.

enzyme	T			AZT		
	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times\ 10^3$	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) $\times\ 10^3$
ssTK2A	3.88 ± 0.22	1.57 ± 0.02	406	1361 ± 124	0.28 ± 0.01	0.205
D8	167 ± 13	3.05 ± 0.08	18	995 ± 119	0.18 ± 0.03	0.185
Dm-dNK[a]	1.2	14.2	12000	8.3	0.036	4.3
N64D[a]	23.2	0.82	35	11.1	0.037	3.3
HSV-1 TK[b]	0.20 ± 0.05	0.350 ± 0.010	1750	5.2 ± 1.7	0.056 ± 0.013	10.8
E225L[b]	12.3 ± 1.2	0.016 ± 0.003	1.3	17.0 ± 1.2	0.032 ± 0.004	1.9

The experimental data for *Dm*-dNK and HSV1-TK were adopted from (a) Welin et. al [32] (b) Pilger et. al [3].

In summary, our study demonstrates that Glu197 is required for the natural dN phosphorylation. Substitutions in position 197 can directly abolish activity due to a reduction of hydrogen bond to the 3′-OH and polarization at the 5′-O on the sugar. Using small residue variants at Tyr86 could not support our speculation for the removal of steric effect on 2′-modifications with any experimental evidence. The outcome of the AZT screen protocol rely on the negative genetic complementation of the tk-deficient *E. coli* strain KY895 for library analysis, making this approach rely on the naturally toxic material thus creating false positives. Our experiments suffered from this side effect of AZT due to the necessity for large quantities.

We conclude that the hydrogen-bonding network manipulations for the sugar portion are more complicated and capricious than the successful alterations for nucleobase binding in the active site. In addition, we believe that a single functional mutation for one dNK could be detrimental in other family members, even though they are closely related. In the future, novel screening methods with positive selection for improved NA activation would be the best choice for such mutational studies on human kinases to be custom-made in suicide gene therapy or several other enzyme-prodrug therapies.

References

1.   Van Rompay AR, Johansson M, Karlsson A: **Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases.** *Pharmacology & Therapeutics* 2000, **87:**189-198.

2.   Arner ESJ, Eriksson S: **Mammalian deoxyribonucleoside kinases.** *Pharmacology & Therapeutics* 1995, **67:**155-186.

3.   Pilger BD, Perozzo R, Alber F, Wurth C, Folkers G, Scapozza L: **Substrate diversity of herpes simplex virus thymidine kinase. Impact Of the kinematics of the enzyme.** *J Biol Chem* 1999, **274:**31967-31973.

4.   Christians FC, Scapozza L, Crameri A, Folkers G, Stemmer WP: **Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling.** *Nat Biotechnol* 1999, **17:**259-264.

5.   Welin M, Skovgaard T, Knecht W, Zhu C, Berenstein D, Munch-Petersen B, Piskur J, Eklund H: **Structural basis for the changed substrate specificity of Drosophila melanogaster deoxyribonucleoside kinase mutant N64D.** *Febs J* 2005, **272:**3733-3742.

6.   Eriksson S, Munch-Petersen B, Johansson K, Eklund H: **Structure and function of cellular deoxyribonucleoside kinases.** *Cellular and Molecular Life Sciences* 2002, **59:**1327-1346.

7.   Sabini E, Ort S, Monnerjahn C, Konrad M, Lavie A: **Structure of human dCK suggests strategies to improve anticancer and antiviral therapy.** *Nat Struct Biol* 2003, **10:**513-519.

8.   Wild K, Bohner T, Folkers G, Schulz GE: **The structures of thymidine kinase from Herpes simplex virus type 1 in complex with substrates and a substrate analogue.** *Protein Science* 1997, **6:**2097-2106.

9.   Johansson K, Ramaswamy S, Ljungcrantz C, Knecht W, Piskur J, Munch-Petersen B, Eriksson S, Eklund H: **Structural basis for substrate specificities of cellular deoxyribonucleoside kinases.** *Nat Struct Biol* 2001, **8:**616-620.

10.   Munir KM, French DC, Loeb LA: **Thymidine kinase mutants obtained by random sequence selection.** *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90:**4012-4016.

11.   Knecht W, Munch-Petersen B, Piskur J: **Identification of residues involved in the specificity and regulation of the highly efficient multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster.** *Journal of Molecular Biology* 2000, **301:**827-837.

12.   Mikkelsen NE, Johansson K, Karlsson A, Knecht W, Andersen G, Piskur J, Munch-Petersen B, Eklund H: **Structural basis for feedback inhibition of the deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase.** *Biochemistry* 2003, **42:**5706-5712.

13.   Mamber SW, Brookshire KW, Forenza S: **Induction of the SOS response in Escherichia coli by azidothymidine and dideoxynucleosides.** *Antimicrob Agents Chemother* 1990, **34:**1237-1243.

14.   Reetz MT, Bocola M, Carballeira JD, Zha D, Vogel A: **Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test.** *Angew Chem Int Ed Engl* 2005, **44:**4192-4196.

15. Lutz S, Ostermeier M, Benkovic SJ: **Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides.** *Nucleic Acids Research* 2001, **29**.

16. Igarashi K, Hiraga S, Yura T: **A deoxythymidine kinase deficient mutant of Escherichia coli. II. Mapping and transduction studies with phage phi 80.** *Genetics* 1967, **57:**643-654.

17. Sambrook J, Russell DW: *Molecular cloning: a laboratory manual.* 3rd ed. edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.

18. Lutz S, Lichter J, Liu L: **Exploiting temperature-dependent substrate promiscuity for nucleoside analogue activation by thymidine kinase from Thermotoga maritima.** *J Am Chem Soc* 2007, **129:**8714-8715.

19. Munir KM, French DC, Dube DK, Loeb LA: **Herpes thymidine kinase mutants with altered catalytic efficiencies obtained by random sequence selection.** *Protein Eng* 1994, **7:**83-89.

20. Iyidogan P, Lutz S: **Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity.** *Biochemistry* 2008, **47:**4711-4720.

21. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T: **How to measure and predict the molar absorption coefficient of a protein.** *Protein Sci* 1995, **4:**2411-2423.

22. Gill SC, von Hippel PH: **Calculation of protein extinction coefficients from amino acid sequence data.** *Anal Biochem* 1989, **182:**319-326.

23. Fox JD, Waugh DS: **Maltose-binding protein as a solubility enhancer.** *Methods Mol Biol* 2003, **205:**99-117.

24. Denault M, Pelletier JN: **Protein library design and screening: working out the probabilities.** *Methods Mol Biol* 2007, **352:**127-154.

25. LaBean TH, Kauffman SA: **Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics.** *Protein Sci* 1993, **2:**1249-1254.

26. Sabini E, Hazra S, Konrad M, Burley SK, Lavie A: **Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase.** *Nucleic Acids Res* 2007, **35:**186-192.

27. Sabini E, Hazra S, Ort S, Konrad M, Lavie A: **Structural basis for substrate promiscuity of dCK.** *J Mol Biol* 2008, **378:**607-621.

28. Lal S, Lauer UM, Niethammer D, Beck JF, Schlegel PG: **Suicide genes: past, present and future perspectives.** *Immunol Today* 2000, **21:**48-54.

29. Vassaux G, Lemoine NR: **Gene therapy for carcinoma of the breast: Genetic toxins.** *Breast Cancer Res* 2000, **2:**22-27.

30. Black ME, Loeb LA: **Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants in vitro.** *Biochemistry* 1993, **32:**11618-11626.

31. Harrison PT, Thompson R, Davison AJ: **Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinase.** *J Gen Virol* 1991, **72 (Pt 10):**2583-2586.

32. Welin M, Skovgaard T, Knecht W, Zhu CY, Berenstein D, Munch-Petersen B, Piskur J, Eklund H: **Structural basis for the changed substrate specificity of**

Drosophila melanogaster deoxyribonucleoside kinase mutant N64D. *Febs Journal* 2005, **272:**3733-3742.

Chapter 4

Engineering the phosphate donor-binding region of human dCK to investigate the preference for UTP over ATP

4.1   Introduction

Deoxycytidine kinase (dCK) partakes in the salvage pathway via catalyzing the initial phosphorylation of 2′-deoxycytidine (dC), 2′-deoxyadenosine (dA) and 2′-deoxyguanosine (dG) to their monophosphate forms using ribonucleoside triphosphates as phosphoryl donors. Subsequently, additional enzymes provide 2′-deoxyribonucleoside triphosphates (dNTPs), which are precursors for DNA synthesis and repair [1]. In addition to its biological importance, dCK is also considered to be the rate-determining enzyme in the overall pathway that converts the pharmacologically inactive nucleoside analog (NA) prodrugs to their active triphosphate anabolites. Given its crucial role in phosphorylating several antiviral (e.g., ddC and 3TC against HIV) and anticancer (e.g., Gemcitabine and AraC) NAs, dCK has gathered attention for several structural and mutational studies.

The recently solved crystal structures of several deoxyribonucleoside kinases (dNKs) and multiple sequence alignments have suggested that dCK belongs to a large type 1 dNK family [2]. While sharing a similar architecture, all of the family members vary widely in their substrate profiles and catalytic efficiencies. The type 1 family includes two additional human dNKs, the purine-specific deoxyguanosine kinase (dGK) and pyrimidine-specific thymidine kinase 2 (TK2), as well as the fruit fly dNK from *Drosophila melanogaster* (*Dm*-dNK) that phosphorylates both purine and pyrimidine substrates with the highest catalytic turnovers among all known type 1 dNKs. Understanding the structural parameters and functional properties of dCK and other related dNKs that allows for efficient NA phosphorylation is the prime interest of

research in the kinase field.

Human dCK is a unique enzyme, as it not only phosphorylates the naturally occurring D-nucleosides, but also efficiently activates the non-physiological chiral substrates such as L-nucleosides and L-NAs. An early study showed that dCK is capable of activating both D and L enantiomers of dC, dA and dG with similar efficiencies [3]. Furthermore, the relaxed enantioselectivity of the enzyme has been supported by the recently solved crystal structures of dCK in complex with the antiviral L-NAs like 3TC ((−)-L-2′,3′-dideoxy-3′-thiacytidine) and FTC ((−)-L-2′,3′-dideoxy-5-fluoro-3′-thiacytidine) [4, 5].

In addition to being highly promiscuous towards the phosphoryl acceptor, studies with the purified enzyme have shown that human dCK can utilize the endogenous ribonucleoside and 2′-deoxyribonucleoside triphosphates as phosphoryl donors except dCTP, which acts as a feedback inhibitor [6-9]. Further studies have demonstrated that UTP is the preferred phosphoryl donor for dCK under physiological conditions rather than ATP, which is the universal donor for all other type 1 dNKs [10-13]. Recent crystallography studies have illuminated the distinct conformational changes induced by the UDP-bound state of dCK instead of ADP in the following structural segments; phosphoryl donor base-sensing region (residues 240-254), LID region (residues 179-207) and an insert region (residues 54-90) as indicated in Figure 4.1 [14, 15]. The initial crystal structure of wild-type dCK has revealed the significance of several arginine residues in the LID region for ATP binding and also suggested to be important for stabilizing the transferred γ-phosphoryl group during catalysis [16]. While LID region forms hydrogen-bonding interactions between the arginine residues and the phosphate groups, the phosphoryl donor base-sensing loop participates in hydrogen-bonding

interactions with the phosphoryl donor base. Furthermore, to evaluate the functional significance of the flexible insert region, a dCK variant lacking the residues from 65 to 79 was constructed and kinetically characterized with natural substrates using ATP and UTP as phosphoryl donors [17]. The truncated dCK variant showed reduced purine affinity and unchanged dC binding compare to wild-type enzyme. The mechanism by which the loop controls the specificity for purines versus pyrimidine is unclear. However, the authors have speculated that the insert may affect the conformational flexibility of its flanking α-helices (α2 and α3), which supplies residues also contribute to the substrate binding in the active site.

Figure 4.1   The monomeric subunit of dCK in the presence of dC and ADP.
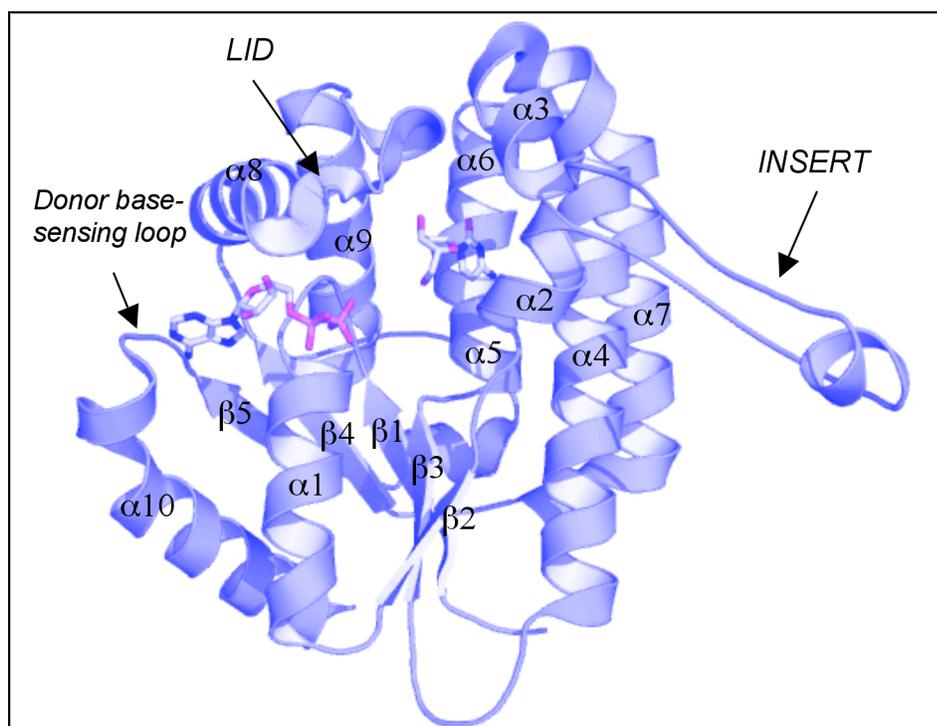
Ribbon diagram of dCK (PDB code 1P60) [16]: dC (phosphoryl acceptor) and ADP (phosphoryl donor) are represented in sticks. The structural features in dCK; LID region as well as the flexible insert region and the donor base-sensing loop are indicated with arrows.

The crystal structure of the truncated insert variant and several crystal structures of the full-length dCK have been solved as ternary complexes with the natural substrates and both ADP and UDP in the donor binding site [14, 15, 17, 18]. These structures reveal a major main-chain rearrangement in the phosphoryl donor base-binding loop (residues 240-247) between the UDP-bound and ADP-bound forms. In the donor base-binding loop, the amino acid positions Glu240, Asp241 and Phe242 are speculated to be responsible for the observed conformational changes in the C-terminal region and consequently engage in the unusual phosphoryl donor preference for dCK (Figure 4.2).

Figure 4.2   Comparison of the hydrogen-bonding interactions between the ADP or UDP donor base and the loop residues in dCK.
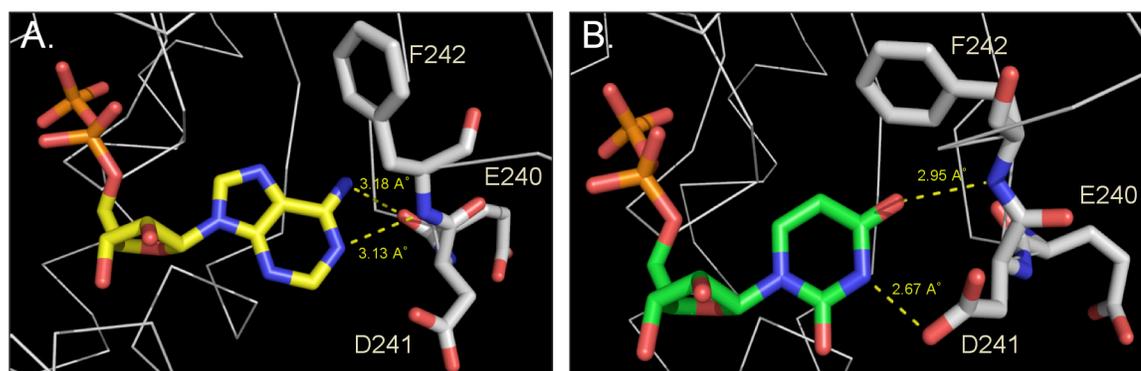
The conformational differences in E240, D241 and F242 with ADP (yellow) versus UDP (green) bound states are shown in sticks. (**A**) Close-up of the adenine base-enzyme interaction (PDB code 1ZI3) [15]. (**B**) Close-up of the uracil base-enzyme interaction (PDB code 1ZI6) [15]. The yellow dashed lines and labels represent the specific hydrogen bonds.

In the case of ADP, the backbone carbonyl of Glu240 forms a hydrogen bond to the amino group (N6) and the main chain amide of Phe242 donates a hydrogen bond to N1 of adenine base. Upon binding of UDP, the carboxylic group of Asp241 forms a hydrogen

bond to N3 position of uracil and the backbone amide of Phe242 forms another hydrogen bond to the carbonyl group (O4) in position 4 of the uracil base (Figure 4.2A and 2B). Thus, substituting the Asp241 and Phe242 amino acid positions at the donor recognition loop could indirectly allow modifying the wild-type dCK activity.

Interestingly, the equivalent C-terminal region was found to impact the substrate specificity of two structurally and functionally related kinases, human TK2 and *Dm*-dNK [19]. Recombination of these two kinases highlights a unique crossover site at the C terminus of *Dm*-dNK, which is located on the loop region before β5 strand and right after α8 helix (a 52-residue long segment) (Figure 4.3). Specifically, swapping this C-terminal region including the phosphoryl donor base-sensing loop to a mostly hTK2 protein frame results in intermediate activity properties originated from both parental enzymes, which emphasizes the importance of this subdomain for substrate specificity. In this study, two different chimeras including predominantly TK2 sequence with the last 52-residue segment from *Dm*-dNK have shown to possess purine activity contrary to the human parental kinase that completely lacks in purine phosphorylation. Additionally, the selected chimeras have demonstrated novel activation properties towards d4T, in which both parental enzymes are deficient. Even though the authors did not test the effects of the substitution of phosphoryl donor specificity, their findings also support the hypothesis of a functional relevance of the *Dm*-dNKs' C-terminal region.

In this chapter, we have probed the role of the donor base-sensing loop in regards to phosphoryl donor and acceptor specificity in dCK via preparing a number of rationally designed chimeras. For chimeragenesis, sequence elements from *Dm*-dNK were introduced successively at the secondary structure elements before and after the

phosphoryl donor base-sensing loop. In addition to the rationally designed chimera study, alanine scanning mutagenesis at positions Asp241 and Phe242 was implemented to investigate the specific binding interactions of both adenine and uridine phosphoryl donors to these residues in dCK enzyme. Both experimental methods yielded chimeras and alanine variants with shifted donor preference from UTP to ATP compare to the wild-type dCK enzyme by mainly changing the binding constants for both donors. Further studies of NA phosphorylation of one selected chimera with enhanced ATP affinity yet showed no significant improvement in activity. We conclude that the C-terminal sequences at the phosphoryl donor binding site influence the donor specificity but has no major impact on either phosphoryl acceptor binding or catalytic efficiencies.

4.2   Materials and Methods

4.2.1   Chemical reagents and bacterial strains

All reagents were purchased from Fisher (Pittsburg, PA) and Sigma & Aldrich (St. Louis, MO). Enzymes were purchased from New England Biolabs (Beverly, MA) unless indicated otherwise. *Pfu* Turbo DNA polymerase (Strategene, La Jolla, CA) was used for all cloning. PCR reactions for sequencing were performed using *Taq* DNA polymerase (New England Biolabs). Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). PCR samples were purified using the QIAquick PCR Purification kit and plasmid DNA was isolated using the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) according to the manufacturer's protocols. T4 DNA ligase was obtained from Promega (Madison, WI). Pyruvate kinase and lactate dehydrogenase were

purchased from Roche Biochemicals (Indianapolis, IN). All DNA manipulations were performed in *E. coli* DH5α-E (Invitrogen, Carlsbad, CA) using standard methodologies and *E. coli* strain BL21(DE3)pLysS (Novagen, Madison, WI) was used for protein expression. All constructs were confirmed by DNA sequencing.

4.2.2   Construction of site-directed chimeras at the C-terminal.

The single and double-crossover chimeras were generated using human dCK and a truncated variant of *Dm*-dNK (Dm-dNKΔ15) that lacks the last 15 amino acids of its C terminus. Throughout the chapter, the *Dm*-dNK represents this variant. The single-crossover dCK-*Dm*dNK chimeras at the three different positions were constructed as indicated: The corresponding dCK fragments were created using *dCK* as template and the following primer pairs; hC_f/hC216Dm_r, hC_f/hC232Dm_r and hC_f/hC237Dm_r. The complementary *Dm*-dNK gene fragments were PCR amplified by using hC216Dm_f/Dm_r, hC232Dm_f/Dm_r and hC237Dm_f/Dm_r primer pairs. The resulting gene fragments from both parents were fused by overlap-extension PCR, using gene specific primers hC_f/Dm_r (Table 4.1). Double-crossover chimeras were constructed by the same strategy, using the gene fusions mentioned above as template. All final gene fusions were digested with *Nde*I and *Spe*I restriction enzymes and ligated into pET-14b vector (Novagen, Madison, WI) for protein overexpression. All constructs were confirmed by DNA sequencing using the plasmid-specific P-3/P-4 primer pairs.

4.2.3   Site-directed mutagenesis of dCK

The Asp241Ala and Phe242Ala mutations, as well as the double alanine

Asp241Ala/Phe242Ala substitution in dCK were introduced by primer overlap extension PCR, using the corresponding AP_f/D241A_r, AP_f/F242A_r, and AP_f/(D241A/F242A_r) primer pairs (Table 4.1), respectively. The resulting PCR products were cloned into pET-14b using *Nde*I and *Spe*I restriction sites as described above, and the gene sequences were confirmed by DNA sequencing.

Table 4.1 Primers used for plasmid construction and sequencing.

Primer	Sequence
hC_f	5'-GCGCATATGGCCAGCTCTGAGGGGACCCGC-3'
hC_r	5'-GCGACTAGTTCACAAAGTACTCAAAAACTCTTTG-3'
Dm_r	5'-CGCACTAGTTCAGGGCTGTTGGTTACTTGA-3'
P-3	5'-GCGAAATTAATACGACTCACTATAGGG-3'
P-4	5'-GCTAGTTATTGCTCAGCGG-3'
hC216Dm_f	5'-AGCTGGCTGATACACCAGAGGCGC-3
hC216Dm_r	5'-CTGGTGTATCAGCCAGCTTTCATGTTT-3'
hC232Dm_f	5'-TATCTTCAAGAGGTGCCTGTCCTAGTCCTCGATGCC-3'
hC232Dm_r	5'-GGCATCGAGGACTAGGACAGGCACCTCTTGAAGATA-3'
hC237Dm_f	5'-CCTATCTTAACACTGGATGCCGATCTAAACCTGGAA-3'
hC237Dm_r	5'-TTCCAGGTTTAGATCGGCATCCAGTGTTAAGATAGG-3'
Dm244hC_f	5'-CTGGAAAACATTGGCACCAAATATGAAAGTCTGGTTG-3'
Dm244hC_r	5'-CAACCAGACTTTCATATTTGGTGCCAATGTTTTCCAG-3'
AP_f	5'-AAAGACAAATATGAAAGTCTGGTTG-3'
D241A_r	5'-ACTTTCATATTTGTCTTTAAAGGCTTCATTAAC-3'
F242A_r	5'-ACTTTCATATTTGTCTTTTGCGTCTTCATTAACATC-3'
D241A/F242A_r	5'-ACTTTCATATTTGTCTTTTGCGGCTTCATTAACATC-3'

4.2.4   Protein overexpression and purification

For in vitro characterization, wild-type dCK, *Dm*-dNK and all mutant kinases were overexpressed as fusion proteins with a N-terminal hexa-histidine tag. The pET14b

plasmids, carrying the different genes, were transformed into *E. coli* BL21 (DE3) pLysS (Novagen, Madison, WI). Cell cultures were grown at 37°C in 400 ml 2YT media containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) to an $OD_{600} \sim 0.5$. Protein expression was induced with 0.1 mM IPTG for 4 h at 25°C for all chimeras, 3 h at 30°C for dCK alanine mutants and 3 h at 37°C for wild-type dCK and *Dm*-dNK. All enzymes were purified by gravity flow column purification using Ni-NTA agarose resin (Qiagen), as described in chapter 2. The purified proteins were concentrated in an Amicon Ultra-4 ultrafiltration unit (Millipore, Bedford, MA; MWCO:10 kDa; 5000 *g* at 4°C) and buffer-exchanged into storage buffer (50 mM Tris-HCl pH 8, 0.5 M NaCl, 5 mM $MgCl_2$, 2 mM DTT). Yields varied from 5-15 mg per liter of culture, depending on the enzyme, and >95% purity was determined for all enzymes by SDS-PAGE. Protein concentrations were determined by A_{280} measurements, using molar extinction coefficients calculated as described by Gill & von Hippel [20, 21]. Aliquots of purified proteins were flash frozen in liquid nitrogen and stored at -80°C until further use.

4.2.5 Steady-state kinetic assays.

Spectrophotometric assays were performed in 50 mM Tris-HCl, pH 7.5, 0.1 M KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.21 mM phosphoenolpyruvate, 0.18 mM NADH, and 2 units/ml pyruvate kinase and 2 units/ml lactate dehydrogenase [22, 23]. Assays were performed at 37°C in triplicate, measuring the absorbance change at 340 nm in the presence of 0.19 – 3.8 μg enzyme per reaction. For phosphoryl acceptor measurements, ATP or UTP was kept constant at 1 mM for wild-type dCK and all variants. While ATP was used at 1 mM, UTP was kept constant at 5 mM for *Dm*-dNK. The acceptor

concentrations were varied between 1 and 100 μM for dC and 1 and 1000 μM for the NAs. For phosphoryl donor measurements, dC was kept at 100 μM, while donor concentrations varied from 1 μM to 5000 μM. Kinetic data were fit to the Michealis-Menten equation using Origin7 program (OriginLab, Northampton, MA).

4.3   Results

4.3.1   Rational design of chimera and purification

The three single-crossover chimeras at amino acid positions 216 (at the end of α9), 232 (before β5) and 237 (after β5), near the C terminus of dCK were chosen based on the promising results from a prior study on recombination of TK2 and *Dm*-dNK [19]. The functional chimeras from the recombination study reflect the predominant occurrence of crossovers beyond α8 helix in *Dm*-dNK (α9 in dCK), which indicates a "hot spot" for domain swapping (Figure 4.3).

Figure 4.3   Ribbon diagram of parental enzymes.
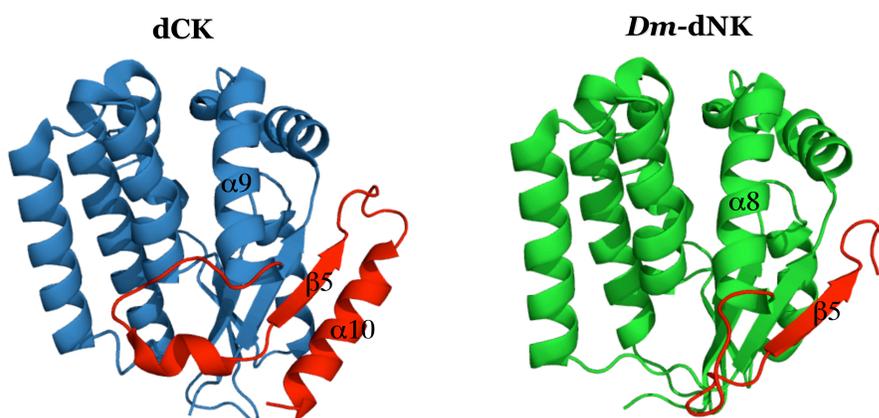
The monomeric subunits of dCK [16], and *Dm*-dNK [24] are colored in blue and green, respectively. The secondary structure elements used in the chimera design are highlighted in red. The specific crossover positions in dCK (α9/β5/α10) and the corresponding positions in *Dm*-dNK (α8/β5), are indicated at the related secondary structure.

Unfortunately, the last helical segment of *Dm*-dNK, α9, after the β5 strand could not be evaluated due to the loss of electron density of the entire helix in all solved *Dm*-dNK crystal structures. Most importantly, a single-crossover chimera of hTK2 with the C-terminal subdomain from *Dm*-dNK showed higher catalytic activity and broadened substrate specificity compare to the wild-type TK2 enzyme. Predicting a similar effect for dCK, we rationally created a chimera, hC216Dm, with the same crossover at the end of α9 using dCK as the predominant scaffold and substituting the C-terminal subdomain derived from *Dm*-dNK. We also constructed hC232Dm and hC237Dm chimeras (Figure 4.4) whereat crossovers included the adjacent β5 sheet for exploring the entire structural segments and their significance on enzyme activity.

Figure 4.4   Schematic representation of parental enzyme sequences compare to the constructed rational chimeras.
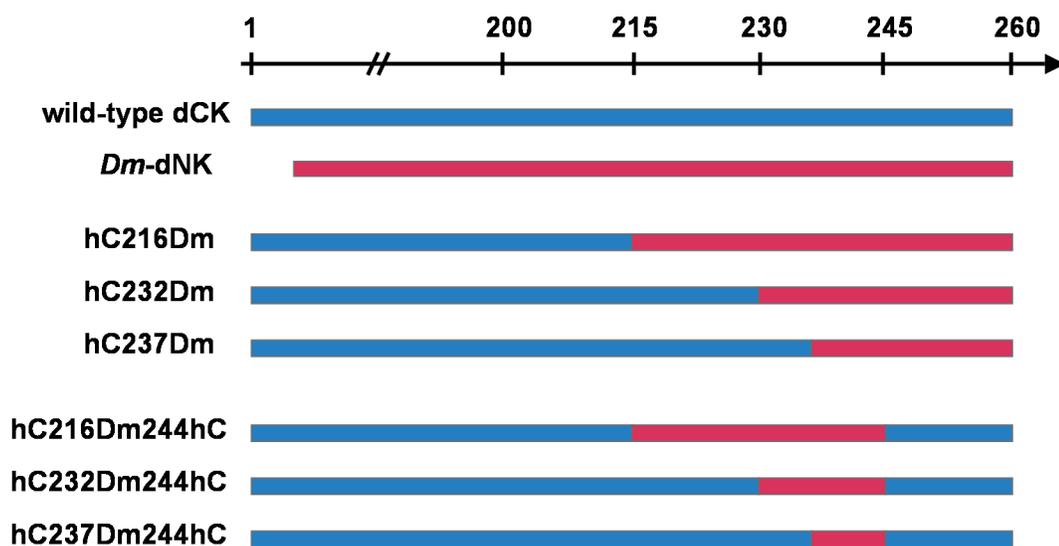

Sequence segments from dCK and *Dm*-dNK are represented as blue and red bars, respectively. The indicated amino acid numbering is according to the wild-type dCK sequence. All the single and double-crossover chimeras are colored as the parental enzymes, with swapped segments from dCK and *Dm*-dNK.

However, the overexpression of all three single-crossover chimeras yielded insoluble protein even at low induction temperatures and they could not be characterized further. Inspection of the individual swapped subdomains in each chimera by multiple sequence alignment (Figure 4.5) has pointed the helix α10 that could impede with the proper folding due to very low sequence homology in this region. Besides the sequence dissimilarity, the different charge distribution in helix α10 for both parental kinases suggests that this could also cause the complete destabilization of all single-crossover chimeras. Therefore, we created the double-crossover chimeras: hC216Dm244hC, hC232Dm244hC and hC237Dm244hC (Figure 4.4) by substituting back the α10 helix from dCK using the previously prepared chimeras as templates. The newly constructed double-crossover chimeras yielded soluble protein after overexpression in the pET-system at 25°C and were purified via immobilized metal affinity chromatography.

In summary, we have rationally swapped the 10-residue segment, which includes the donor base-sensing loop region from *Dm*-dNK for all human hybrids. This small region substitutes the critical dCK amino acid positions, Glu240/Asp241/Phe242, in enzyme donor hydrogen bonding network with *Dm*-dNK residues, Leu209/Asn210/Leu211, respectively. Our results revealed that the dCK's phosphoryl donor preference for UTP could be altered towards ATP by substituting the donor-binding loop sequence from a kinase like *Dm*-dNK with different donor preference properties compare to dCK. The removal of the established interactions from the charged Glu240 and Asp241 residues and the bulky phenyl side chain of Phe242 alter the wild-type sequence's preference for uracil base to the more favorable adenine binding via the new sequence motif. The observed donor shift from these experiments further analyzed by specifically focusing on

only Asp241 and Phe242 shown to be the conformationally most effected residues upon UTP binding. We applied alanine-scanning mutagenesis to completely eliminate the side chain interactions for both positions.

Figure 4.5 Multiple sequence alignment of human dCK, dGK and *Dm*-dNK.
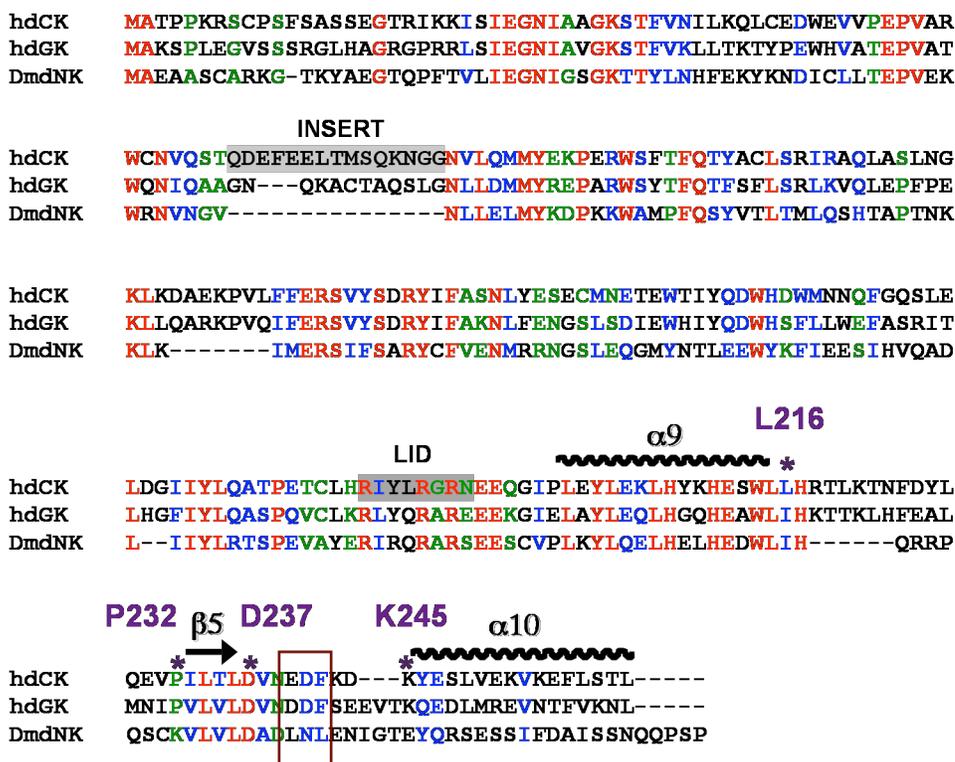
Sequence alignment of human dCK, dGK and *Dm*-dNK. The flexible insert region and the highly conserved LID region motif for type 1 dNKs are shaded in gray boxes. Absolutely conserved residues are in red; identical residues are in blue; similar residues are in green. The α9/β5/α10 subdomains are indicated above the corresponding sequences. The individual crossover sites for single chimeras are marked with asterisks; L216, P232 and D237. The amino acid numbering indicates the last dCK residue included in the corresponding chimeras. K245 residue indicates the second crossover site for double-crossover chimeras including the dCK α10 helix. The amino acid positions Glu240, Asp241 and Phe242 that recognize the donor base by specific interactions are framed in brown box.

4.3.2 Activity measurements of chimeras

The kinetic properties of the three designed chimeras and parental enzymes were determined for ATP and UTP (phosphoryl donors) using dC as the phosphoryl acceptor at a constant concentration of 100 μM (Table 4.2).

Table 4.2 Kinetic analysis of parental and chimeric enzymes with ATP and UTP as phosphoryl donors.

Phosphate donor	Phosphate acceptor	Enzyme	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) x 10^3
ATP	100 μM dC	wild-type dCK	20.3 ± 1.6	0.059 ± 0.001	2.9
		hC216Dm244hC	7.8 ± 1.0	0.055 ± 0.002	7
		hC232Dm244hC	7.0 ± 1.0	0.035 ± 0.001	5
		hC237Dm244hC	7.3 ± 0.9	0.039 ± 0.001	5.4
		Dm-dNK	114 ± 9	7.5 ± 0.2	66
UTP	100 μM dC	wild-type dCK	10.0 ± 1.1	0.068 ± 0.002	6.8
		hC216Dm244hC	12.9 ± 1.0	0.058 ± 0.001	4.5
		hC232Dm244hC	9.5 ± 0.9	0.037 ± 0.001	3.9
		hC237Dm244hC	9.1 ± 1.5	0.032 ± 0.001	3.5
		Dm-dNK	1040 ± 122	14.4 ± 0.7	14

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation. The phosphoryl acceptor, dC, concentration was kept constant at 100 μM.

For the parental enzymes, *Dm*-dNK possesses high catalytic turnover rates for ATP and UTP by 120- and 210-fold over wild-type dCK while showing a substantial decrease in binding affinity for ATP (114 μM, 5.7-fold) and UTP (1040 μM, 100-fold) compare to dCK that has moderate K_M values of 20 μM and 10 μM, respectively. Overall, *Dm*-dNK shows 22-fold and 2-fold increase in k_{obs} / K_M for ATP and UTP, respectively.

Since all the double-crossover chimeras possess primarily dCK sequence, they reflect a dCK-like catalytic performance with catalytic efficiencies in the same range as the dominant parental enzyme. The kinetic data for the three chimeras reveal a shift towards ATP, largely due to a ~3-fold decrease in apparent binding constants and a moderate increase in k_{obs} / K_M values (1.7 to 2.4-fold) compare to dCK. Interestingly, the most active chimera, hC216Dm244hC that includes a 25-residue sequence from *Dm*-dNK (the major segment that is swapped), shows unchanged turnover rate for ATP. In contrast, the other two chimeras, hC232Dm244hC with a 15-residue segment and hC237Dm244hC with a 10-residue segment, show a decline in k_{obs} values by 1.5-fold. The same pattern is observed for the turnover rates for UTP while the K_M values mostly stay unchanged in respect to wild-type dCK. Consequently, a moderate decrease in catalytic activity by ~1.7-fold is observed for each chimera in the case of UTP.

We also determined the kinetic constants for dC as an acceptor, using the double-crossover chimeras and parental enzymes in the presence of constant ATP and UTP concentrations (1 mM) as phosphoryl donors (Table 4.3). For *Dm*-dNK experiment, UTP at 5 mM was used due to the high K_M value of 1040 μM that was measured in the previous experiment. For ATP, all chimeras show dCK-like catalytic performance while phosphorylating dC with a less than 2-fold change in all kinetic parameters. The hC216Dm244hC chimera presents similar catalytic efficiency to wild-type dCK with a K_M value of 1.32 μM and a ~1.3-fold increase in k_{obs} value. The other two chimeras maintain the low apparent binding affinities (~1 μM) for dC as a general trend but show a slight decline for the k_{obs} values by 1.3- and 2.6-fold for ATP and UTP, respectively. Among all chimeras hC216Dm244hC is the most active enzyme with the most efficient

activity parameters for dC, ATP and UTP.

Table 4.3 Kinetic analysis of parental and chimeric enzymes with dC as phosphoryl acceptor.

Phosphate donor	Phosphate acceptor	Enzyme	K_M (µM)	k_{obs} (s⁻¹)	k_{obs}/K_M (M⁻¹s⁻¹) x 10³
1 mM ATP	dC	wild-type dCK	1.00 ± 0.17	0.064 ± 0.002	63
		hC216Dm244hC	1.32 ± 0.05	0.080 ± 0.001	79
		hC232Dm244hC	< 1	0.052 ± 0.001	> 52
		hC237Dm244hC	< 1	0.045 ± 0.001	> 45
		Dm-dNK	2.3 ± 0.4	16.9 ± 0.6	7500
1 mM UTP	dC	wild-type dCK	1.26 ± 0.09	0.100 ± 0.001	81
		hC216Dm244hC	< 1	0.060 ± 0.001	> 60
		hC232Dm244hC	< 1	0.038 ± 0.001	> 38
		hC237Dm244hC	1.07 ± 0.07	0.037 ± 0.001	35
*(5 mM UTP)		*Dm*-dNK	6.0 ± 1.2	7.8 ± 0.4	1300

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation. The phosphoryl donors, ATP and UTP, concentrations were kept constant at 1 mM. (*) For *Dm*-dNK, UTP was kept at 5 mM.

For *Dm*-dNK, the striking performance difference compare to dCK-like activity is mostly resulting from the higher catalytic turnover rates (264-fold and 78-fold for ATP and UTP, respectively) while K_M values in the single digits micromolar range for dC 2.3 µM and 6 µM using ATP and UTP, respectively.

4.3.3 Nucleoside analog activity of wild-type dCK and hC216Dm244hC chimera

We selected the most active chimera hC216Dm244hC to test the effects of swapped domains for the NA phosphorylation by evaluating this particular chimera's activity, as

well as wild-type dCK, towards four NAs with distinct sugar modifications; dFdC, ddC, AraC and 3TC (Table 4.4). The kinetic data for 3TC shows slight changes in K_M and k_{obs} values, compensating each other and thus resulting in no change in the overall specific activity. While wild-type dCK shows high efficiency towards dFdC and AraC, the chimera shows a 2-fold increase in the turnover rates for both NAs but also a raise in their K_M values by 2.8-fold. Therefore the specificity constants decline by 1.5-fold for dFdC and AraC compare to the wild-type dCK. For the case of ddC, we measured a 2.3-fold activity loss due to worse binding affinity and lower turnover rates. Overall, better ATP binding properties observed for this chimera does not correlate with the activation of the tested NAs, resulting in no significant improvements for any prodrug.

Table 4.4   Kinetic analysis of wild-type dCK and hC216Dm244hC chimera with various nucleoside analogs as phosphoryl acceptors.

Phosphate donor	Phosphate acceptor	wild-type dCK			hC216Dm244hC		
		K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) x 10^3	K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) x 10^3
1 mM ATP	dFdC	1.6 ± 0.1	0.62 ± 0.01	389	4.5 ± 0.4	1.10 ± 0.03	241
	ddC	204 ± 10	0.36 ± 0.01	1.8	348 ± 25	0.28 ± 0.01	0.8
	AraC	4.0 ± 0.3	0.48 ± 0.01	117	11.5 ± 0.8	0.96 ± 0.02	84
	3TC	6.5 ± 0.3	0.04 ± 0.001	6.2	5.6 ± 0.6	0.036 ± 0.001	6.4

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation. 1 mM ATP was used as phosphoryl donor.

4.3.4   Alanine-scanning mutagenesis of dCK, purification of mutants and activity measurements

We have used alanine-scanning mutagenesis at positions Asp241 and Phe242 which are located in the donor base-sensing loop to further explore the hydrogen bond network of

the phosphoryl donor base to these residues by side chain and main chain interactions. The rationale behind this experimental method has been addressed in the introduction section. The two single mutants, Asp241Ala and Phe242Ala, as well as the double-alanine mutant Asp241Ala/Phe242Ala were overexpressed in the pET-system at 30°C and purified via metal affinity chromatography. The catalytic properties of the three alanine mutants were determined for ATP and UTP (phosphoryl donors) using dC as phosphoryl acceptor at a constant concentration of 100 µM (Table 4.5).

Table 4.5   Kinetic analysis of wild-type dCK and alanine variants from rational design with ATP and UTP as phosphoryl donors.

Phosphate donor	Phosphate acceptor	Enzyme	K_M (µM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) x 10^3
ATP	100 µM dC	wild-type dCK	20.3 ± 1.6	0.059 ± 0.001	2.9
		D241A	16.4 ± 2.4	0.076 ± 0.002	4.7
		F242A	16.9 ± 2.8	0.076 ± 0.002	4.5
		D241A/F242A	9.2 ± 1.0	0.096 ± 0.002	10.4
UTP	100 µM dC	wild-type dCK	10.0 ± 1.1	0.068 ± 0.002	6.8
		D241A	19.5 ± 1.9	0.105 ± 0.002	5.4
		F242A	22.1 ± 3.0	0.093 ± 0.003	4.2
		D241A/F242A	13.1 ± 1.3	0.089 ± 0.002	6.8

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation. The phosphoryl acceptor, dC, concentration was kept constant at 100 µM.

The substitutions in single alanine mutants result in slight changes in ATP binding and turnover rates with a 1.6-fold improvement for overall activity compare to wild-type dCK. Most noticeable, the double-alanine mutant shows a 3.6-fold increase in k_{obs} / K_M

resulted from a tighter binding value of 9 μM (2-fold improvement) and 1.6-fold higher

turnover rate. In the case of UTP, both single mutants show a moderate reduction of the

binding affinity from a K_M value of 10 μM for the wild-type enzyme to 19.5 μM

(D241A) and 22 μM (F242A) and a ~1.4-fold increase in k_{obs} for each variant, resulting

in a slight decrease in catalytic activity. For the double mutant D241A/ F242A, the

substitutions raise the K_M values slightly, an effect that is compensated by a 1.3-fold

increase in turnover, resulting in an unaffected specificity constant for UTP compare to

wild-type dCK. Subsequently, we tested dC (phosphoryl acceptor) activity for all alanine

mutants using constant concentrations of ATP and UTP at 1 mM as donors (Table 4.6).

Table 4.6 Kinetic analysis of wild-type dCK and alanine variants from rational design with dC as phosphoryl acceptor.

Phosphate donor	Phosphate acceptor	Enzyme	K_M (μM)	k_{obs} (s^{-1})	k_{obs}/K_M (M^{-1}s^{-1}) x 10^3
1 mM ATP	dC	wild-type dCK	1.00 ± 0.17	0.064 ± 0.002	63
		D241A	1.46 ± 0.11	0.079 ± 0.001	54
		F242A	1.19 ± 0.13	0.081 ± 0.001	68
		D241A/F242A	< 1	0.104 ± 0.001	> 104
1 mM UTP	dC	wild-type dCK	1.26 ± 0.09	0.100 ± 0.001	81
		D241A	1.00 ± 0.10	0.115 ± 0.003	115
		F242A	1.52 ± 0.17	0.110 ± 0.003	72
		D241A/F242A	< 1	0.101 ± 0.001	> 101

The experiments were performed in triplets at 37 °C and kinetic parameters were determined by non-linear fit to the Michaelis-Menten equation. The phosphoryl donors, ATP and UTP, concentrations were kept constant at 1 mM.

When ATP is used as phosphoryl donor, all the alanine variants show unaffected binding affinities for dC in the ~1 μM range. The maximum change is observed for the double-alanine D241A/F242A mutant with a 1.6-fold increase in k_{obs} as where the remaining two single mutants show slight improvements around 1.2-fold compare to wild-type enzyme. Generally for UTP, all the alanine mutants show comparable kinetic properties to dCK with minor improvements (<2-fold) for dC activity due to small changes in K_M values.

4.4   Discussion

Human dNKs are essential for supplying active forms of NAs through the phosphorylation of NA prodrugs into the corresponding monophosphate forms for the antiviral and anticancer therapy. Especially, dCK offers an excellent framework to engineer because of its importance in phosphorylating several clinically administered NAs derived from dC and dA. In addition to the promiscuous substrate specificity at the phosphoryl acceptor site, dCK utilizes several nucleotides as phosphoryl donors. While this property is a common feature for all type 1 dNKs, the preferred donor under physiological conditions was assumed to be ATP until UTP was shown to be the actual physiological donor for wild-type dCK [11]. The authors from the previous structural studies hypothesized that the preference for UTP was specifically related to the different hydrogen-bonding network (due to the distinct nature of bound donor base) that formed with Asp241 and Phe242 at the enzyme's C terminus, resulting in conformational changes at the phosphoryl donor base-sensing loop [17].

We utilized rationally designed chimeras of dCK to probe the role of this donor base-

sensing loop (between β5 and α10) by swapping the interested regions from *Dm*-dNK. The ATP preference of this kinase with dissimilar sequence for the same region (Figure 4.4) along with its high catalytic performance for all natural dNs and numerous NA prodrugs are the reason to use *Dm*-dNK in the chimera creation. A truncated variant of *Dm*-dNK (Dm-dNKΔ15) that lacks the last 15 amino acids of its C terminus was used as the other parent. This variant offers a similar length in size for the α10 helix of dCK thus provides an advantage for structural alignment at the C terminus. Separate studies showed that Dm-dNKΔ15 has the same catalytic performance compare to the wild-type *Dm*-dNK enzyme [25].

In the meantime, an independent chimeragenesis study between *Dm*-dNK and hTK2 in our laboratory yielded chimeras with novel and broadened substrate profile by substituting the similar regions from the C terminus of *Dm*-dNK (crossover at the loop between α9 and β5) to TK2 enzyme [19]. Due to these promising findings, we expanded our crossover sites and constructed three single-crossover chimeras as following; hC216Dm, hC232Dm and hC237Dm (Figure 4.4). None of the single-crossover chimeras were soluble and this outcome is believed to be a result of the different charge distribution and the high sequence dissimilarity in helix α10. We suggest that the last helix from *Dm*-dNK probably causes incompatible interactions with the neighboring α1 helix in the dCK structure. Unfortunately, our assumptions cannot be related to any structural data for the region after β5. Because all the *Dm*-dNK structures have been determined using the variant that lacks the last 20 amino acids at the C-terminal and in the solved structures, the last 16 residues including the donor-binding loop and the entire helix have no ordered density (Figure 4.3) [24, 26, 27]. Consequently, we introduced a

second crossover, swapping back the α10 from dCK to create the double-crossover chimeras hC216Dm244hC, hC232Dm244hC and hC237Dm244hC (Figure 4.4). All of the double-crossover hybrids were soluble and could be characterized kinetically for both phosphoryl donors, ATP and UTP, as well as dC, the most preferred phosphoryl acceptor. In the case of the phosphoryl donor, all chimeras showed 3-fold lower binding affinity constants for ATP and no significant change in UTP binding compare to the wild-type dCK. Also their kinetic efficiencies closely reflected a dCK-like pattern since it is the predominant parent. The observed switch in the donor preference could be due to the substituted subdomains that contain the donor base-sensing loop from *Dm*-dNK. We predict that the universally swapped 10 amino acid segment for all hybrids that corresponds to the donor base-sensing loop region is the sole reason for the reversed ATP preference effect. Given the low sequence homology between dCK and *Dm*-dNK in regards to this small loop, the aforementioned Glu240/Asp241/Phe242 amino acid positions in dCK are substituted with corresponding *Dm*-dNK residues that are Leu209/Asn210/Leu211, respectively. Removing the negatively charged residues at the 240 and 241 positions in addition to the bulky phenyl side chain could alter the distinctive hydrogen-bonding network, hence favoring adenine binding over the wild-type sequence's preference for the uracil.

As for the parental enzyme kinetics, our apparent binding constants for ATP and UTP in dCK are in agreement with previous reports from several groups [10, 13, 28]. However, our kinetic data for *Dm*-dNK shows consistently higher K_M values for ATP than a previously reported study [25]. In that report, K_M values for ATP were 5 μM and 3 μM for the recombinant *Dm*-dNK and a truncated *Dm*-dNK mutant lacking the last 20

residues, respectively. For UTP, only a relative activity analysis using *Dm*-dNK has been reported a 50% activity with UTP relative to ATP [24]. In the case of dC as phosphoryl acceptor, the kinetic parameters are comparable to the earlier reported values [25]. In order to verify our results we used different batches of materials, repeated the measurements several times to omit all the possible sources for experimental errors. As a result, we believe that indeed our findings are correct for *Dm*-dNK.

Interestingly, further kinetic analysis of these chimeras with dC as acceptor, using ATP and UTP, shows similar specificity constants but reduced turnover rates in the presence of both donors. hC216Dm244hC was the most active chimera for dC phosphorylation. This particular chimera possesses the 25 aa C-terminal subdomain including the loop region right after α9 helix, β5 sheet and the loop after β5 as the biggest sequence portion from *Dm*-dNK and preference for ATP donor resembling *Dm*-dNK. The observed high activity of hC216Dm244hC could be rationalized by the added conservation of the secondary structural elements in the substituted segments, which then promotes beneficial interactions. Therefore, the overall protein integrity and catalytic activity could be maintained comparable to the parental enzyme or better. Upon testing for NA activity, hC216Dm244hC displayed equal or lower activity (relative to dCK) for the four tested compounds. The reason for this outcome could be due to the removal of the incompatible α10 helix from *Dm*-dNK, consequently resulting in curtailed enzyme activity opposed to the hTK2/*Dm*-dNK example. As swapping the entire 52 aa C terminus region of *Dm*-dNK to TK2 enzyme initiated novel activities with high efficiency properties, possibly caused by the small conformational changes in the overall protein structure.

Independently, the functional significance of Asp241 and Phe242 residues for donor

binding was further analyzed by alanine scanning mutagenesis at both sites. Our data on single alanine mutants for 241 and 242 suggests that the removal of the hydrogen bond from Asp241 to uracil in D241A mutant yields a less favorable binding of UTP. Additionally, the F242A mutants' analogous outcome hints the codependence of Asp241 to the adjacent Phe242 residue to form favorable bonding interactions with the uracil base. The substitution of the bulky phenylalanine side chain with a small alanine group could constrain the required main chain arrangement for positioning the carboxylic oxygen of Asp241 in close proximity to the N3 of uracil base. On the other hand these single mutants showed no distinguishable effect for ATP binding. The most striking and unexpected outcome was found for the double-alanine mutant D241A/F242A with a tighter binding value for ATP compare to the wild-type enzyme and at the same time in the similar range for UTP. This result could be explained by eliminating the charged group from Asp241 and bulky phenyl group from Phe242 cause a more flexible/less constrained loop region that provides improved main chain interactions to form favorable hydrogen bonds for both donor base entities. The double alanine mutations at these sites also improve the dC phosphorylation results enhancing both affinity and catalytic performance in comparison to the two single-alanine mutants and dCK as well. In summary, our study indicates that positions 241 and 242 are indeed important for phosphoryl donor preference of dCK. Substitutions in these residues can directly influence the donor specificity via electrostatic interactions and/or main chain conformational adjustments, allowing for us to understand the properties for both phosphoryl acceptor and phosphoryl donor binding sites in the enzyme structure.

Finally, our experiments demonstrate an elegant way to investigate the structure-function

relationship in this small loop region and to explore the molecular aspects of dCK's unusual donor preference by either introducing novel frameworks from other type 1 dNKs or deciphering the inherited sequence responsible for the desired function over the course of natural evolution of these kinases. Even though we could not attain the anticipated novel or improved activity towards nucleoside analog prodrugs, a better understanding for the complete picture of catalysis would be beneficial to the field of kinase engineering.

References

1. Arner ESJ, Eriksson S: **Mammalian deoxyribonucleoside kinases.** *Pharmacology & Therapeutics* 1995, **67:**155-186.

2. Eriksson S, Munch-Petersen B, Johansson K, Eklund H: **Structure and function of cellular deoxyribonucleoside kinases.** *Cellular and Molecular Life Sciences* 2002, **59:**1327-1346.

3. Wang J, Choudhury D, Chattopadhyaya J, Eriksson S: **Stereoisomeric selectivity of human deoxyribonucleoside kinases.** *Biochemistry* 1999, **38:**16993-16999.

4. Sabini E, Hazra S, Konrad M, Burley SK, Lavie A: **Structural basis for activation of the therapeutic L-nucleoside analogs 3TC and troxacitabine by human deoxycytidine kinase.** *Nucleic Acids Res* 2007, **35:**186-192.

5. Sabini E, Hazra S, Konrad M, Lavie A: **Nonenantioselectivity Property of Human Deoxycytidine Kinase Explained by Structures of the Enzyme in Complex with l- and d-Nucleosides.** *J Med Chem* 2007, **50:**3004-3014.

6. Cheng YC, Domin B, Lee LS: **Human deoxycytidine kinase. Purification and characterization of the cytoplasmic and mitochondrial isozymes derived from blast cells of acute myelocytic leukemia patients.** *Biochim Biophys Acta* 1977, **481:**481-492.

7. Datta NS, Shewach DS, Hurley MC, Mitchell BS, Fox IH: **Human T-lymphoblast deoxycytidine kinase: purification and properties.** *Biochemistry* 1989, **28:**114-123.

8.   White JC, Hines LH: **Role of uridine triphosphate in the phosphorylation of 1-beta-D-arabinofuranosylcytosine by Ehrlich ascites tumor cells.** *Cancer Res* 1987, **47:**1820-1824.

9.   Sarup JC, Johnson MA, Verhoef V, Fridland A: **Regulation of purine deoxynucleoside phosphorylation by deoxycytidine kinase from human leukemic blast cells.** *Biochem Pharmacol* 1989, **38:**2601-2607.

10.   White JC, Capizzi RL: **A critical role for uridine nucleotides in the regulation of deoxycytidine kinase and the concentration dependence of 1-beta-D-arabinofuranosylcytosine phosphorylation in human leukemia cells.** *Cancer Res* 1991, **51:**2559-2565.

11.   Shewach DS, Reynolds KK, Hertel L: **Nucleotide specificity of human deoxycytidine kinase.** *Mol Pharmacol* 1992, **42:**518-524.

12.   Krawiec K, Kierdaszuk B, Eriksson S, Munch-Petersen B, Shugar D: **Nucleoside triphosphate donors for nucleoside kinases: donor properties of UTP with human deoxycytidine kinase.** *Biochem Biophys Res Commun* 1995, **216:**42-48.

13.   Hughes TL, Hahn TM, Reynolds KK, Shewach DS: **Kinetic analysis of human deoxycytidine kinase with the true phosphate donor uridine triphosphate.** *Biochemistry* 1997, **36:**7540-7547.

14.   Sabini E, Hazra S, Konrad M, Lavie A: **Elucidation of Different Binding Modes of Purine Nucleosides to Human Deoxycytidine Kinase.** *J Med Chem* 2008.

15.   Sabini E, Hazra S, Ort S, Konrad M, Lavie A: **Structural basis for substrate promiscuity of dCK.** *J Mol Biol* 2008, **378:**607-621.

16. Sabini E, Ort S, Monnerjahn C, Konrad M, Lavie A: **Structure of human dCK suggests strategies to improve anticancer and antiviral therapy.** *Nat Struct Biol* 2003, **10:**513-519.

17. Godsey MH, Ort S, Sabini E, Konrad M, Lavie A: **Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: illuminating the role of main-chain reorganization.** *Biochemistry* 2006, **45:**452-461.

18. Soriano EV, Clark VC, Ealick SE: **Structures of human deoxycytidine kinase product complexes.** *Acta Crystallogr D Biol Crystallogr* 2007, **63:**1201-1207.

19. Gerth ML, Lutz S: **Non-homologous recombination of deoxyribonucleoside kinases from human and Drosophila melanogaster yields human-like enzymes with novel activities.** *J Mol Biol* 2007, **370:**742-751.

20. Gill SC, von Hippel PH: **Calculation of protein extinction coefficients from amino acid sequence data.** *Anal Biochem* 1989, **182:**319-326.

21. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T: **How to measure and predict the molar absorption coefficient of a protein.** *Protein Sci* 1995, **4:**2411-2423.

22. Schelling P, Folkers G, Scapozza L: **A spectrophotometric assay for quantitative determination of kcat of herpes simplex virus type 1 thymidine kinase substrates.** *Analytical Biochemistry* 2001, **295:**82-87.

23. Iyidogan P, Lutz S: **Systematic exploration of active site mutations on human deoxycytidine kinase substrate specificity.** *Biochemistry* 2008, **47:**4711-4720.

24. Mikkelsen NE, Johansson K, Karlsson A, Knecht W, Andersen G, Piskur J, Munch-Petersen B, Eklund H: **Structural basis for feedback inhibition of the

deoxyribonucleoside salvage pathway: Studies of the Drosophila deoxyribonucleoside kinase.** *Biochemistry* 2003, **42:**5706-5712.

25. Munch-Petersen B, Knecht W, Lenz C, Sondergaard L, Piskur J: **Functional expression of a multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster and its C-terminal deletion mutants.** *J Biol Chem* 2000, **275:**6673-6679.

26. Johansson K, Ramaswamy S, Ljungcrantz C, Knecht W, Piskur J, Munch-Petersen B, Eriksson S, Eklund H: **Structural basis for substrate specificities of cellular deoxyribonucleoside kinases.** *Nat Struct Biol* 2001, **8:**616-620.

27. Mikkelsen NE, Munch-Petersen B, Eklund H: **Structural studies of nucleoside analog and feedback inhibitor binding to Drosophila melanogaster multisubstrate deoxyribonucleoside kinase.** *Febs J* 2008, **275:**2151-2160.

28. Krawiec K, Kierdaszuk B, Kalinichenko EN, Rubinova EB, Mikhailopulo IA, Eriksson S, Munch-Petersen B, Shugar D: **Striking ability of adenosine-2'(3')-deoxy-3'(2')-triphosphates and related analogues to replace ATP as phosphate donor for all four human, and the Drosophila melanogaster, deoxyribonucleoside kinases.** *Nucleosides Nucleotides Nucleic Acids* 2003, **22:**153-173.

Chapter 5

Conclusions and future perspectives

5.1 Summary

In this dissertation, I have used a variety of protein engineering techniques to explore the substrate specificity of human dCK. Chapter 2 focuses on the features of active site mutations on dCK's catalytic performance via using rational design and combinatorial protein engineering. Characterization of the selected mutants suggested that especially two amino acid positions are key players in modulating the substrate specificity and catalytic activity of dCK towards natural substrates as well as nucleoside analog prodrugs. The concluded ease of dCK evolvability from Chapter 2 inspired the following site-saturation mutagenesis approach. Chapter 3 focuses on site-saturation libraries of active site residues, targeting the sugar specificity in dCK. I conclude that the choice of NA and related screening method for the desired modified sugar specificity were not in accordance with the limited traits of dCK. Finally, in Chapter 4, I have explored the function of the C-terminal donor base-sensing loop, responsible for the phosphoryl donor specificity in dCK, via chimeragenesis of human dCK and *Dm*-dNK in combination with alanine scanning mutagenesis to redesign this loop region at the particular sites.

5.2 Modulating the nucleobase specificity of dCK

Among the characterized mammalian salvage pathway enzymes, dCK is unique in its broad specificity for dC, dA and dG except T (Table 1.1) and its ability to facilitate the phosphorylation of structurally different, natural and unnatural substrates with in a single active site. In this dissertation, I have used combination of several protein engineering techniques to explore the determinants of promiscuous substrate specificity of dCK and established the critical role of two amino acid positions 104 and 133 in the active site

region located near the phosphoryl acceptor's nucleobase for the enzymes' specificity and activity. I was able to select mutants that would serve as either a catalyst with 'generalist' performance through broad specificity for all natural 2′-deoxyribonucleosides or modulated the specificity towards a 'specialist' catalyst relative to the amino acid entity in these two positions. While Arg104Gln/Asp133Gly mutations yield a generalist, Arg104Met/Asp133Thr mutations create a specialist via improving the T specificity by 2700-fold and raise the apparent binding affinity for dC and the purine substrates.

An independent similar study on *Dm*-dNK showed that mutations at the same corresponding residues in the active site could convert the multisubstrate *Dm*-dNK specificity from pyrimidines to purines [1]. However, the observed switch in preference is mainly due to abolishing the pyrimidine activity rather than improving the purine activity. In contrast, my results demonstrated that the specificity of dCK towards pyrimidines can be modulated by improving the binding affinity for T and increasing the catalytic efficiency for the case of dC as a substrate. Moreover, the observed mutations generate novel catalytic activities towards thymidine analogs like AZT. In conclusion, dCK sets an excellent model for kinase engineering to gain insight into the relationship between structure and function in the type 1 dNKs.

5.3   Improving the specificity for NA prodrugs

In addition to probing structural elements that modulate the activity of dCK towards 2′-deoxyribonucleosides, the ultimate goal of this dissertation is to engineer variants that possess improved catalytic activity towards medically relevant NA prodrugs. There have been many studies on type 1 dNKs explaining how mutations at the specific amino acid

positions affect the sugar-binding segment of the active site pocket and could improve the specificity towards NA prodrugs modified in ribose ring [2-5]. Therefore, several residues were selected for the second-generation engineering of dCK, which are predicted to be responsible for discriminating against particular modifications at the 2′- and 3′-position of the ribose moiety of NAs. An effective site-saturation mutagenesis approach inspired by the combinatorial active site saturation test (CASTing) was utilized to introduce mutations with small library sizes in order to cover all the possibilities at each site as described in Chapter 3 [6]. However, the choice of screening method for focused libraries did not yield any dCK variants with improved specificities or activities towards NAs. Future studies might generate the desired enhanced NA activation properties if novel screening and selection systems will be developed. The absence of the suitable techniques severely hinders the current research for beneficial developments on dCK. As a matter of fact, engineering efforts for all dNK family members are affected by this problem.

5.4 "It is not all about the active site"

There are many reports in the literature that describe the importance of certain domains and structural elements for determining the specificity and activity in numerous dNKs rather than active site residues. For instance, it has been suggested that the C-terminal region, the last 20 amino acid residues of *Dm*-dNK, has an impact on the substrate specificity by reducing the affinity towards dideoxy- and purine-ribonucleosides compare to the full-length enzyme [7]. Alternately, an independent study on recombination of hTK2 and *Dm*-dNK verified the importance of this C terminus region from *Dm*-dNK by

reporting chimeras with novel NA activity and broadened dN specificity and activity [8]. Lavie and coworkers have shown that a flexible insert region near the N terminus in dCK causes a drop in binding affinity of purines when truncated [9]. Structural studies from the same group suggested that during catalysis, several structural adjustments take place upon binding of phosphoryl acceptors, as well as phosphoryl donor, causing some main chain adjustments at the donor binding loop and C terminus region of dCK [10]. The authors hypothesized that communication between the phosphoryl acceptor and donor sites affects the destiny of various binding modes in regards to the nature of substrates. The aforementioned observations were tested in Chapter 4 by combining the rationally designed chimeras at different locations and site-directed mutagenesis at the donor-base binding loop for the case of dCK. I conclude that the specific amino acid mutations could alter the donor preference of dCK from UTP to ATP. However, the additive functionality of the C-terminal region of *Dm*-dNK does not seem to be integrated into dCK's activity. In many kinase engineering experiments, there are limitations in reproducing interesting functional properties in other dNK family members. As the desired effects for the analogous enzymes fail to arise, lessons could be learned for future experimental design.

5.5   Future perspectives

The work presented in this dissertation combines the protein engineering methodologies to engineer the specificity and catalytic activity of human dCK towards natural substrates as well as clinically important NAs. Even though promising variants have been obtained, further dCK engineering has been hampered by the lack of screening or selecting techniques for nucleoside kinase activity other than T. The major caveat could be only tackled by developing positive selection approaches for dC, dA and dG activity.

Currently, progress in the entire protein engineering field strictly relies on this common problem. In the past 10 years, the published studies have focused on novel protocols/techniques for how to create more diverse, efficient and cost-effective mutant libraries. While tremendous improvement has been made in library design, the development of novel screening and selection approaches has been neglected.

In future experiments, one might try to engineer dCK towards dA, dG and related purine NAs that have been less studied up to now. At present, there are only a few purine derivatives in use as cancer and antiviral drugs such as clofarabine and fludarabine. The high binding constants of purines compare to dC could be used as starting point for further improvement of dCK by directed evolution. When variants with the desired low binding constants are observed, the candidates may be used in designing an improved version of purine NAs.

This work emphasizes the different yet complementary engineering of dCK. Nevertheless, the proposed novel screen and/or selection systems will be all important to the future direction of the kinase field. There are multitudes of biochemical and mutagenesis studies that could be performed in vitro and *in vivo* to better understand the NA binding and activation and their therapeutic potential.

References

1.   Knecht W, Sandrini MP, Johansson K, Eklund H, Munch-Petersen B, Piskur J: **A few amino acid substitutions can convert deoxyribonucleoside kinase specificity from pyrimidines to purines.** *Embo J* 2002, **21:**1873-1880.

2.   Pilger BD, Perozzo R, Alber F, Wurth C, Folkers G, Scapozza L: **Substrate diversity of herpes simplex virus thymidine kinase. Impact Of the kinematics of the enzyme.** *J Biol Chem* 1999, **274:**31967-31973.

3.   Christians FC, Scapozza L, Crameri A, Folkers G, Stemmer WP: **Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling.** *Nat Biotechnol* 1999, **17:**259-264.

4.   Welin M, Skovgaard T, Knecht W, Zhu C, Berenstein D, Munch-Petersen B, Piskur J, Eklund H: **Structural basis for the changed substrate specificity of Drosophila melanogaster deoxyribonucleoside kinase mutant N64D.** *Febs J* 2005, **272:**3733-3742.

5.   Egeblad-Welin L, Sonntag Y, Eklund H, Munch-Petersen B: **Functional studies of active-site mutants from Drosophila melanogaster deoxyribonucleoside kinase. Investigations of the putative catalytic glutamate-arginine pair and of residues responsible for substrate specificity.** *Febs J* 2007, **274:**1542-1551.

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

7. Munch-Petersen B, Knecht W, Lenz C, Sondergaard L, Piskur J: **Functional expression of a multisubstrate deoxyribonucleoside kinase from Drosophila melanogaster and its C-terminal deletion mutants.** *J Biol Chem* 2000, **275:**6673-6679.

8. Gerth ML, Lutz S: **Non-homologous recombination of deoxyribonucleoside kinases from human and Drosophila melanogaster yields human-like enzymes with novel activities.** *J Mol Biol* 2007, **370:**742-751.

9. Godsey MH, Ort S, Sabini E, Konrad M, Lavie A: **Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: illuminating the role of main-chain reorganization.** *Biochemistry* 2006, **45:**452-461.

10. Sabini E, Hazra S, Ort S, Konrad M, Lavie A: **Structural basis for substrate promiscuity of dCK.** *J Mol Biol* 2008, **378:**607-621.