

Distribution Agreement

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

Signature:

Yichen Liu

Date

Engineering kinases for dual thymidine and thymidylate kinase activity

By

Yichen Liu

Doctor of Philosophy

Chemistry

Dr. Stefan Lutz

Advisor

Dr. Vince Conticello

Committee Member

Dr. Dale Edmondson

Committee Member

Accepted:

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

Date

Engineering kinases for dual thymidine and thymidylate kinase activity

By

Yichen Liu

B.S. University of Science and Technology of China, 2005

Advisor: Stefan Lutz, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry

2011

Abstract

Engineering kinases for dual thymidine and thymidylate kinase activity

By Yichen Liu

Thymidine kinase (TK) and thymidylate kinase (TMPK) are essential enzymes that function in the salvage pathway for synthesis of thymidine triphosphate, a DNA building block. Their importance also exists in activating nucleoside analogs (NAs), an important category of anti-cancer and anti-virus drugs. Because NAs enter the cell as prodrugs, they require three consecutive phosphorylation steps catalyzed by cellular kinases in the salvage pathway to become biologically active and halt DNA synthesis.

Among the cellular kinases responsible for NA activation, the first two phosphorylation steps usually present the bottleneck because of their poor activity towards NA. Previous engineering efforts to alter and improve specificity of these kinases have largely focused on the first step of phosphorylation catalyzed by deoxyribonucleoside kinase (dNK). However, such approach bears the real risk of simply shifting the bottleneck from the improved first to the second phosphorylation reaction. A more promising strategy is to eliminate both bottleneck steps by creating a dual functioning enzyme that can convert NAs directly to their corresponding diphosphate form. In this dissertation, I developed a conditional auxotroph *E. coli* strain for the selection of dual TK and TMPK activity. As the first selection system reported for selecting two consecutive reactions, this auxotroph strain enables the evaluation of

combinatorial kinase libraries for TK and TMPK activities both independently and concurrently. Employing the dual-function auxotroph strain, I have evaluated chimera libraries of dNK from *Drosophila melanogaster* and TMPK from *Thermotoga maritima* using the non-homologous recombination technique SCRATCHY in combination with the computational method SCHEMA. Separately, more fundamental questions concerning the mechanistic similarities and differences between these two enzymatic reactions were investigated by site-directed mutagenesis and ITC studies of substrate binding. Experiments were conducted on DmdNK and TmTMPK, which exclusively catalyze the first or second phosphorylation reaction, as well as the only known dual-function TK from herpes simplex virus (HSV). The information extracted from these experiments provides new insights regarding the dual function of HSV1-TK and will guide future engineering experiments.

Engineering kinases for dual thymidine and thymidylate kinase activity

By

Yichen Liu

B.S. University of Science and Technology of China, 2005

Advisor: Stefan Lutz, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Chemistry

2011

Acknowledgement

I would like to thank my advisor, Dr. Stefan Lutz, for his support throughout my graduate studies. His strong passion for science, continuous encouragement and helpful advices have made my graduate life a wonderful and unforgettable experience. I have been a great honor to work in his lab. I would also like to thank my committee members, Dr. Dale Edmondson and Dr. Vince Conticello, for their helpful advice regarding my research, and for taking time to meet with me whenever I asked. I also would like to thank Ann Dasher, Steve Krebs and Patti Barnett for all the help I received from them.

The former and current Lutz lab members have been wonderful colleagues and friends to me. Some gave me tremendous help when I first started learning molecular biology: Zhen Qian, Lingfeng Liu, Ying Yu, Monica Gerth. I thank all of my labmates for the times they offered assistance, reagents and patiently sat through all of my practice talks. They made the lab a fun place to work at. I wish them all the best.

I also would like to thank all the friends that I made at Emory: Yuting Li, Li Li, Yi Wang, Huaye Dong, Ana Maria Diaz Burgos. We shared a lot of happy memory together. The friendship that we built is a strong support for me to go through graduate school.

I truly appreciate my parents Xinzhi Liu and Yajuan Wang for their love and support throughout my whole life. I was so lucky to have parents who value education and always support me to pursue my dreams. Without them, I wouldn't have finished.

Finally, a special thanks to James Simmon, my husband. He is my personal consultant, editor, cook, comedian and best friend. I couldn't finish without him supporting me and cheering me up everyday, especially during the difficult time. I want to thank him for his unconditional love.

List of frequently used abbreviations

Abbreviation	Full name
TMP	Thymidine monophosphate
TDP	Thymidine diphosphate
TK	Thymidine kinase
NA	Nucleoside analog
TMPK	Thymidylate kinase
dNK	Deoxyribonucleoside kinase
NMPK	Nucleoside monophosphate kinase
TMPK	Thymidylate kinase
dNK	Deoxyribonucleoside kinase
<i>DmdNK</i>	<i>Drosophila melanogaster</i> dNK
<i>TmTMPK</i>	<i>Thermotoga maritima</i> TMPK
HSV1-TK	Herpes simplex virus type I thymidine kinase
ITC	Isothermal Titration Calorimetry
ITCHY	incremental truncation for the creation of hybrid enzymes
SCRATCHY	Combination of ITCHY and DNA shuffling
TP ₄ A	P ¹ -(5'-Adenosyl) P ⁴ -[5'-(2'-deoxy-thymidyl)] tetraphosphate
TP ₅ A	P ¹ -(5'-Adenosyl) P ⁵ -[5'-(2'-deoxy-thymidyl)] pentaphosphate

Table of Contents

Chapter 1 General introduction.	1
1.1 Nucleoside analog as anti-cancer and anti-virus treatment.	2
1.2 Problems in NA activation by cellular salvage pathway and gene therapy	5
1.3 Deoxyribonucleoside kinase and nucleoside monophosphate kinase	9
1.3.1 dNK family.....	9
1.3.2 NMPK family.....	14
1.3.3 Multifunctional enzymes.....	19
1.4 Engineering multifunctional kinase and the limitation	20
1.5 Aim and scope of the dissertation	23
Chapter 2 Expression and characterization of thymidylate kinase from <i>Thermotoga maritima</i>	24
2.1 Introduction.	25
2.2 Results and discussion.....	30
2.2.1 Structure and active site of <i>Tm</i> TMPK.....	30
2.2.2 <i>Tm</i> TMPK gene isolation and protein purification	34
2.2.3 <i>Tm</i> TMPK is a highly thermostable thymidylate kinase	36
2.2.4 <i>Tm</i> TMPK <i>in vitro</i> catalytic performance	38
2.2.5 Isothermal Titration Calorimetry (ITC) of <i>Tm</i> TMPK.....	40
2.3 Conclusion remarks.....	45
2.4 Materials and methods	46
2.4.1 Materials.....	46
2.4.2 Isolation of <i>Tm</i> TMPK gene.....	46

2.4.3 Overexpression and purification of <i>Tm</i> TMPK	46
2.4.4 Secondary structure characterization and thermal denaturation	47
2.4.5 Spectrophotometric enzyme activity assay	48
2.4.6 Isothermal Titration Calorimetry experiments.....	49
Chapter 3 Design and Construction of a conditional <i>E. coli</i> auxotroph strain for functional selection of dual thymidine and thymidylate kinase activity	50
3.1 Introduction.....	51
3.2 Results and discussion	57
3.2.1 Knock –out of <i>tk</i> gene in <i>E.coli</i>	57
3.2.2 Knock-in <i>Tmtmpk</i> under the P _{BAD} promoter	61
3.2.3 Knock-out <i>Ectmpk</i>	65
3.2.4 Validation of strain YL-1	67
3.2.4.1 YL-1 validation and complementation by a plamid-born kinase.....	67
3.2.4.2 YL-1 growth dependence on arabinose	69
3.2.4.3 YL-1 growth dependence on arabinose with different kinase activity.	69
3.2.4.4 Dual TK and TMPK functional selection using YL-1	70
3.3 Conclusion remarks	74
3.4 Materials and methods	75
3.4.1 Bacterial media, culture conditions and reagents.	75
3.4.2 Construction of <i>tk</i> deletion strain	76
3.4.3 Plasmid construction.....	77
3.4.4 Expression and purification of chromosome-born <i>Tm</i> TMPK	78
3.4.5 Growth on antibiotic plates with various L-arabinose concentrations.....	78

3.4.6 Selection for dual function.....	79
--	----

Chapter 4 Directed evolution of a novel enzyme for dual thymidine and thymidylate

kinase function 80

4.1 Introduction.....	81
-----------------------	----

4.2 Results and discussion.....	85
---------------------------------	----

4.2.1 Construct single crossover library and in-frame selection.....	85
--	----

4.2.2.Construc the first generation multiple crossover library.....	88
---	----

4.2.3 SCHEMA calculation.....	91
-------------------------------	----

4.2.4 Construct the second generation multiple crossover library.....	99
---	----

4.2.5 Selection for dual function using YL-1,a conditional auxotroph strain...	102
--	-----

4.3 Conclusion remarks.....	102
-----------------------------	-----

4.4 Materials and methods.....	103
--------------------------------	-----

4.4.1 Materials.....	103
----------------------	-----

4.4.2 ITCHY library construction.....	104
---------------------------------------	-----

4.4.3 Reading frame selection.....	105
------------------------------------	-----

4.4.4 SCRATCHY library construction.....	106
--	-----

4.4.5 SCHEMA calculation and new SCRATCHY library design.....	106
---	-----

4.4.6 Selection of dual TK and TMPK activity.....	107
---	-----

Chapter 5 Directed evolution of a novel enzyme for dual thymidine and thymidylate

kinase function 109

5.1 Introduction.....	110
-----------------------	-----

5.2 Results and discussion.....	115
---------------------------------	-----

5.2.1 Site-directed mutagenesis study.....	115
--	-----

5.2.1.1 Mutating the P-loop	115
5.2.1.2 Mutating the DRH/ERS motif.....	119
5.2.1.3 Mutating the XEP motif.	122
5.2.1.4 Conclusion of mutagenesis studies	123
5.2.2 Calorimetric study of HSV1-TK, <i>Dmd</i> NK and <i>Tm</i> TMPK.....	127
5.2.2.1 The model of enzyme-substrate ternary complex formation	127
5.2.2.2 Determination of the thermodynamic parameters of thymidine and ATP binding	129
5.2.2.3 Determination of the thermodynamic parameters of TMP binding	135
5.2.2.4 Role of the Magnesium ion.....	138
5.2.2.5 Binding of thymidine triphosphate	139
5.2.2.6 Determination of the thermodynamic parameters of the bi-substrate inhibitors TP4A and TP5A	142
5.2.2.7 Conclusion of ITC experiments	145
5.3 Materials and methods	146
5.3.1 Materials	146
5.3.2 Cloning and mutagenesis	146
5.3.3 Protein expression and purification	148
5.3.4 Thymidine kinase kinetic assay	149
5.3.5 Thymidylate kinase kinetic assay	149
5.3.6 Isothermal Titration Calorimetry	149
Chapter 6 Conclusions and perspectives	151
6.1 Summary	152

6.2 New selection system enables future protein engineering	154
6.3 Future investigation of HSV1-TK's dual function.....	155

List of Figures

Figure 1.1 Deoxyribonucleosides and nucleoside analogs	4
Figure 1.2 The <i>de novo</i> and salvage pathway for dNTP synthesis	7
Figure 1.3 Proposed mechanism of dNK	10
Figure 1.4 Comparison of structures of three type 1 dNKs	12
Figure 1.5 ClustalW sequence alignment of four dNKs	13
Figure 1.6 Structure of four members of the NMPK family	17
Figure 1.7 Proposed mechanism of TMPK	18
Figure 1.8 Structure of HSV1-TK and its comparison with dNK and TMPK	21
Figure 2.1 Crystal structures of TMPKs	28
Figure 2.2 ClustalW alignments of four TMPKs	29
Figure 2.3 Superimposition of crystal structures of three TMPKs	32
Figure 2.4 Active site residues of <i>Tm</i> TMPK	33
Figure 2.5 Gene isolation and expression of <i>Tm</i> TMPK	35
Figure 2.6 Secondary structure and thermostability of <i>Tm</i> TMPK	37
Figure 2.7 ITC of <i>Tm</i> TMPK in the absence of Mg ²⁺	43
Figure 2.8 ITC of <i>Tm</i> TMPK in the presence of Mg ²⁺	44
Figure 3.1 The <i>de novo</i> and salvage pathways for TTP synthesis	54
Figure 3.2 Three step strategy for constructing the conditional auxotroph	56
Figure 3.3 Construction of <i>tk</i> deletion mutants	59
Figure 3.4 Validation of <i>tk</i> knock-out	60
Figure 3.5 Knock-in vector pRDC15: His- <i>Tm</i> TMPK	62
Figure 3.6 Tunable expression of chromosomal <i>Tm</i> TMPK	64

Figure 3.7 <i>Ectmpk</i> knock-out strategy	66
Figure 3.8 Validation of <i>Ectmpk</i> knock-out	68
Figure 3.9 Growth dependence of YL-1 on L-arabinose	71
Figure 3.10 Complementation of YL-1 grown by plasmid-born TMPK	72
Figure 3.11 Dual functional selection in YL-1	73
Figure 4.1 Structure overlay of <i>DmdNK</i> and <i>TmTMPK</i>	84
Figure 4.2 Schematic overview of ITCHY library construction.....	86
Figure 4.3 Crossover distributions in the naïve ITCHY libraries	87
Figure 4.4 Schematic overview of SCRATCHY library construction.....	89
Figure 4.5 Schematic drawing of the sequence of 10 randomly selected SCRATCHY clones	90
Figure 4.6 Scheme of SCHEMA calculations.....	92
Figure 4.7 ClustalW alignments of <i>TmTMPK</i> and <i>DmdNK</i> used in SCHEMA calculations	93
Figure 4.8 Representative of SCHEMA calculation of <i>DmdNK</i> and <i>TmTMPK</i> chimera	96
Figure 4.9 Plot of crossover combinations with low E/m values.....	97
Figure 4.10 Comparison of two generations of primer design	98
Figure 4.11 The sequences of clones from the second generation SCRATCHY libraries.	100
Figure 4.12 The sequences of clones from from the two crossover library	101
Figure 5.1 The structure of three enzymes reveals high homology	114
Figure 5.2 Structure based sequence alignment of dNKs, TMPKs, HSV1-TK and EHV-TK	116

Figure 5.3 Comparison of nucleoside 3'-hydroxyl group hydrogen bonding with active site residues	118
Figure 5.4 Comparison of the DRH/ERS motif interactions with substrates	121
Figure 5.5 Comparison of XEP motif interactions with substrates	125
Figure 5.6 Mechanism of ternary enzyme-substrate complex formation	128
Figure 5.7 ITC of HSV1-TK without Mg^{2+}	131
Figure 5.8 ITC of <i>Dmd</i> NK without Mg^{2+}	132
Figure 5.9 ITC of <i>Tm</i> TMPK without Mg^{2+}	133
Figure 5.10 ITC experiments with TMP	136
Figure 5.11 Bi-substrate inhibitor structures	143

List of Tables

Table 1.1 Natural substrates for the four human dNKs	10
Table 1.2 Natural substrates of the human NMPKs	16
Table 1.3 Comparison of the P-loop and Lid residues of TMPKs.....	18
Table 2.1 <i>Tm</i> TMPK kinetic parameters at 37°C.....	39
Table 2.2 Thermodynamics for the association of nucleotides with <i>Tm</i> TMPK	42
Table 4.1 Primers used for chimeragenesis and cloning.....	108
Table 5.1 Kinetic properties of <i>Dmd</i> NK mutants	120
Table 5.2 Kinetic properties of HSV1-TK mutants	126
Table 5.3 Thermodynamic parameters for thymidine and ATP binding	134
Table 5.4 Thermodynamic parameters for TMP binding	137
Table 5.5 ITC measurements in the presence of Mg ²⁺	140
Table 5.6 ITC measurements of TTP.....	141
Table 5.7 ITC measurements of TP ₄ A and TP ₅ A	144
Table 5.8 Primers used for plasmid construction and mutagenesis	147

Chapter One

General Introduction

1.1 Nucleoside analogs in anti-cancer and anti-virus treatment

Inhibition of gene replication is one strategy that has been effectively exploited in the past by anti-cancer and anti-virus drugs. A common feature of cancer cells and viruses is accelerated gene replication and as a consequence low fidelity of proper nucleotide incorporation. This acceleration and lack of fidelity can be exploited to incorporate replication terminators in dysfunctional cells without inhibition of DNA replication in healthy cells.

An important class of drugs that do this is nucleoside analogs (NAs) (Galmarini et al., 2002; Squires, 2001). From a structural perspective, NAs are pro-drugs that closely resemble natural ribo- and 2'-deoxyribonucleosides with distinct substitutions in the ribose and nucleobase moieties (Figure 1.1). To perform their pharmacological function as DNA/RNA replication terminators, they need to be phosphorylated in three consecutive steps to the corresponding triphosphate anabolites *via* the host's nucleoside salvage pathway (Monnerjahn and Konrad, 2003). Once in the triphosphate form, they become the substrates of low-fidelity polymerases in cancer cells and of reverse transcriptases in viruses. The incorporation of NAs triphosphate into the DNA/RNA causes chain termination and prevents further proliferation of the cancer cell or virus. On the other hand, healthy cells have high fidelity DNA replication machinery that protects the host from the inhibitory effect of NAs.

A major category of these drugs are the 3'-modified NAs, which lack the 3'-hydroxyl group, responsible for forming a phosphodiester bond during the DNA elongation reaction. 3'-Azido-2',3'-dideoxythymidine (AZT or Zidovudine), a thymidine nucleoside analog, is a prime example of this class. It was the first drug approved for the

treatment of HIV (Furman et al., 1986). Human DNA polymerase γ has a strong interaction with AZT triphosphate; also DNA polymerase α is able to incorporate AZT triphosphate into the DNA chain (Samuels, 2006).

A different branch of NAs is represented by 9-(2-hydroxy-ethoxymethyl)-guanine (ACV or acyclovir) and 9-(2-dihydroxypropoxymethyl) guanine (GCV or ganciclovir). They are structurally distinct from other NAs in that the ribose ring is replaced by an open chain structure. Because there is no 3' end, after incorporation into a growing DNA chain, no additional nucleotides can be added to the strand. ACV is an extremely selective and effective anti-herpes drug and shows low cytotoxicity because it is exclusively phosphorylated by the viral thymidine kinase. In addition, it has significantly higher affinity for the viral polymerase over human polymerases. Similar to ACV, GCV is a potent inhibitor of viruses of the herpes family. It selectively and potently inhibits viral DNA replication; but unlike ACV, GCV is not an absolute chain terminator. Elongation of viral DNA is slowed dramatically (Hamzeh and Lietman, 1991).

NAs with an intact 3'-hydroxyl group with structural modifications on the nucleobase comprise the third category of NAs. They impair DNA elongation by causing a strand break or interfering with other enzymes involved in the metabolism of nucleosides. Examples of this type of NAs include (E)-5-(2-bromovinyl)-2'-deoxyuridine (BVVDU or brivudine), a potent inhibitor for HSV-1 and Varicella zoster virus (VZV), and 2-chloro-2'-deoxyadenosine (CdA or cladribine), which is clinically used for the treatment of lymphocytic leukemia (De Clercq, 2005; Rai, 1998).

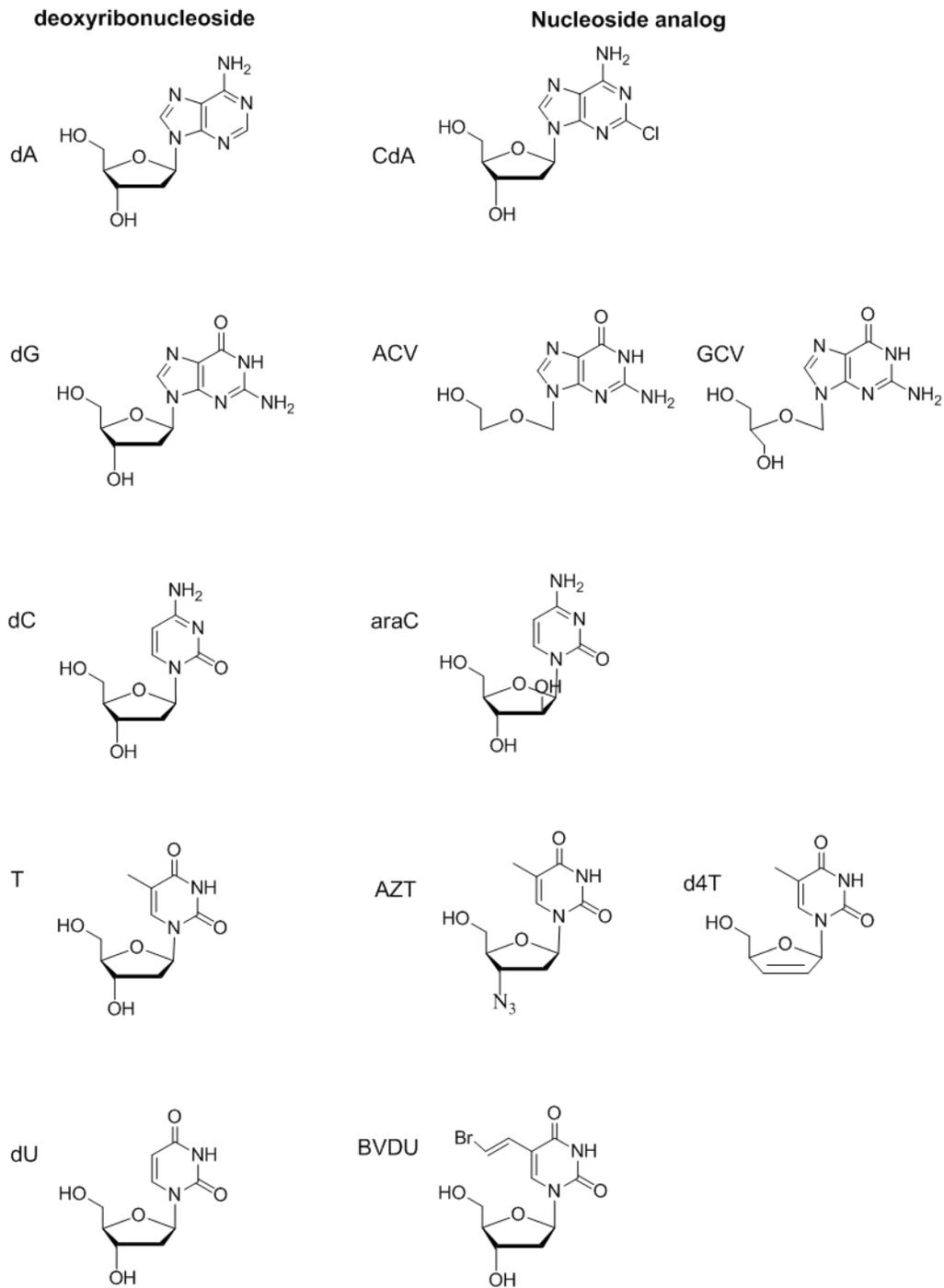


Figure 1.1 Deoxyribonucleosides and nucleoside analogs described in this chapter.

1.2 Problems with nucleoside analog activation

Although these NAs have been proven effective for anti-cancer and anti-viral therapy, their efficacy depends heavily on activation by cellular kinases from the salvage pathway of nucleoside triphosphate syntheses (Figure 1.2). As the name implies, human cells utilize the salvage pathway to recycle nucleobases and ribonucleosides from the degradation of DNA and RNA to generate nucleoside triphosphate for DNA synthesis. First, multiple nucleoside transport proteins either facilitate the diffusion or actively transport the molecules from outside the cell (Baldwin et al., 1999; Kong et al., 2004). In the salvage pathway, deoxyribonucleotides are synthesized from deoxyribonucleosides by three phosphorylation steps: 1) deoxyribonucleoside kinase (dNK) catalyzes the addition of the first phosphate group to the 5' position of the ribose, 2) nucleoside monophosphate kinase (NMPK) converts the deoxyribonucleoside monophosphate to the diphosphate form, 3) nucleoside diphosphate kinase (NDPK) phosphorylates deoxyribonucleoside diphosphate to the triphosphate.

NAs are initially administered in their uncharged nucleoside form, as the negatively charged triphosphate cannot traverse biological membranes. Once transported into the cell by the highly promiscuous membrane transporter, NAs are converted into the monophosphate form by dNK and then phosphorylated again by NMPK. The final step of NA activation is catalyzed by a number of enzymes, including the base nonspecific nucleoside diphosphate kinase (NDPK), pyruvate kinase, phosphoglycerate kinase and creatine kinase (Miller et al., 1992). Since most dNKs and NMPKs have high substrate specificity for their natural nucleoside and nucleotide and low activity towards NAs, either the first or the second phosphorylation step often becomes rate limiting in the

overall activation pathway. Examples of clinically used pro-drugs for which the conversion by dNK is the rate limiting step include d4T and AraC, which are catalyzed by human thymidine kinase 1 and deoxycytidine kinase, respectively (Balzarini et al., 1989; Jordheim et al., 2006). The cases for which the second phosphorylation step is the bottleneck step include AZT monophosphate (AZT-MP) catalyzed by human thymidylate kinase and 6-thioGMP catalyzed by human guanylate kinase (Lavie et al., 1997; Sekulic et al., 2002). The third phosphorylation step is usually not rate limiting, since multiple enzymes can catalyze this reaction.

The inefficiency of NAs activation by dNK and NMPK not only compromises the prodrugs' efficacy in clinical treatment but is also responsible for a series of side effect. The higher cellular concentration that has to be used due to low activity and prolonged half life of unincorporated NAs could potentially lead to their metabolism by other cellular enzymes and pathways that can trigger a change in cellular metabolism, an increase of drug resistance and even alteration of the host genotype (Coulthard et al., 2002; Macchi and Mastino, 2002; Turriziani et al., 2002). A prominent example is the case of anti-HIV drug AZT, for which the conversion of AZT-MP to diphosphate is rate limiting in the overall activation. As a result, the AZT-MP species accumulates up to millimolar concentrations in cells exposed to AZT, while the concentration of the triphosphate form remains very low. In fact, 94% of the AZT metabolites are observed as AZT-MP (Brundiers et al., 1999a). The very low concentration of AZT-TP not only limits the therapeutic efficacy of the drug, but it also causes increased viral proliferation that allows the emergence of AZT resistance HIV strains (Mohri et al., 1993). In addition,

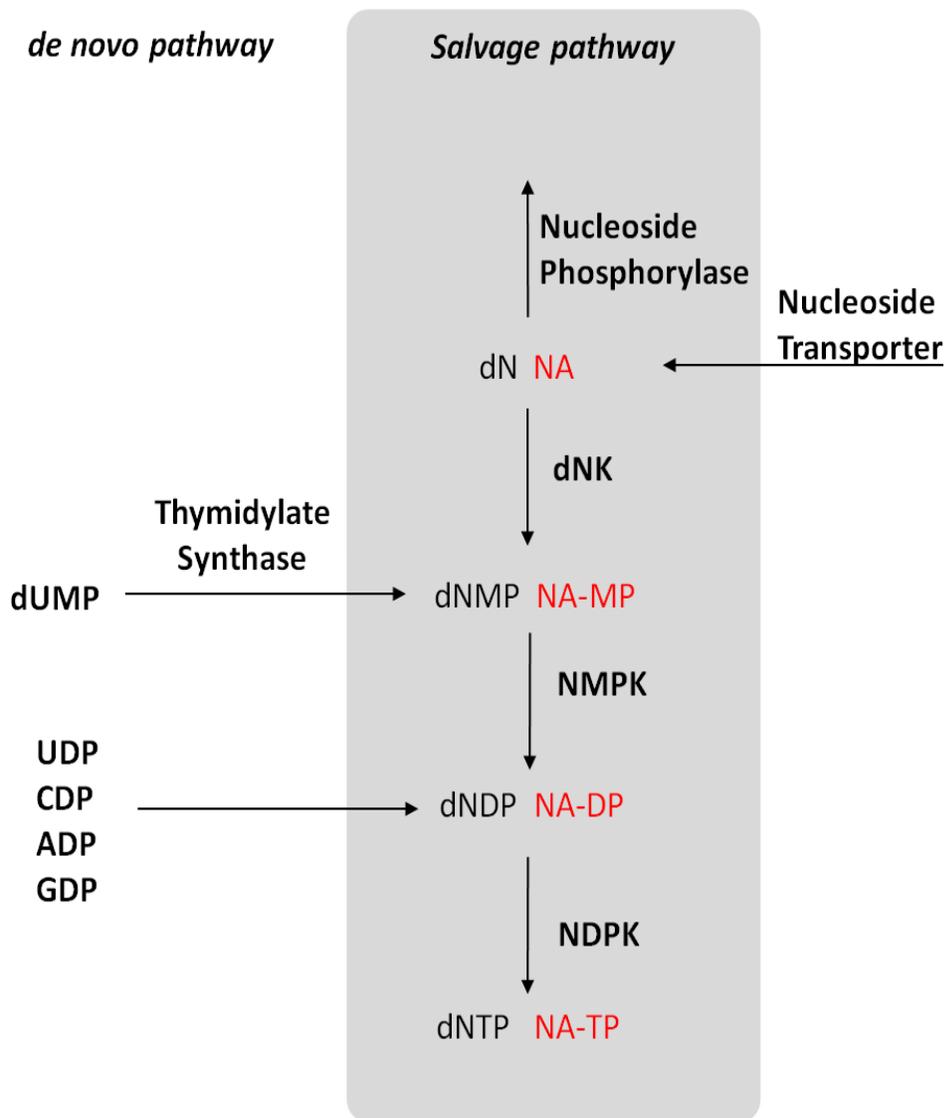


Figure 1.2 The *de novo* and salvage pathway for dNTP synthesis. Nucleoside analogs are activated by the cellular salvage pathway.

the high concentration of AZT-MP has been linked to interference with the host metabolism by suppressing kinase activity and cytotoxic effects through AZT metabolism to the 3-amino derivative (Sales et al., 2001; Tornevik et al., 1995a).

A promising solution to enhance NAs activation is the co-administration of an exogenous kinase with broad specificity and high catalytic efficiency for NAs through gene therapy. Cancer cells treated with NAs can be transfected with genes encoding target kinases that can efficiently phosphorylate NAs, thus eliminating any bottleneck steps. A more efficient activation process will allow lower administration dose and fewer side effects. This strategy has been proven effective in several cases. For example, HSV1-TK co-administered through a retroviral gene delivery in malignant tumor cells combined with ganciclovir chemotherapy increased the sensitivity of cancer cell to the pro-drug (Hasegawa et al., 1995). However, HSV1-TK, as well as other naturally existing kinases, possesses limited substrate activity and specificity range for NAs because it evolved for natural substrates and not for NAs. Therefore, engineered kinases with improved specificity and higher catalytic efficiency towards NAs would be even more promising candidates for NAs activation. A thorough understanding of the relationship between structure and function in nucleoside and nucleotide kinases is a prerequisite for effective enzyme engineering. In the past several decades, numerous studies have been attributed to the characterization of nucleoside and nucleotide kinases.

1.3 Deoxyribonucleoside kinase and nucleoside monophosphate kinase

1.3.1 *dNK family*

In the salvage pathway, dNK catalyzes the initial phosphoryl transfer from a preferred phosphate donor (in a majority of cases ATP) to a deoxyribonucleoside to form a deoxyribonucleoside monophosphate. A conserved glutamate residue in the active site of dNK acts as a general base, abstracting a proton from the 5'-hydroxyl group of the deoxyribonucleoside thereby activating the 5'-oxygen to be a nucleophile and attack the γ -phosphate of the phosphate donor. Positively charged residues from the enzyme and Mg^{2+} also facilitate the phosphoryl transfer reaction by stabilizing the transition state (Eriksson et al., 2002) (Figure 1.3).

Each organism has different sets of dNKs with unique substrate specificities. In human cells, there are four different deoxyribonucleoside kinases: cytosolic thymidine kinase 1 (TK1), mitochondrial thymidine kinase 2 (TK2), deoxycytidine kinase (dCK) and deoxyguanosine kinase (dGK) (Arner and Eriksson, 1995; Eriksson et al., 2002). These enzymes are specific for deoxyribonucleosides, and their nomenclature is based on the preferred deoxyribonucleoside substrate. However, they also have the ability to phosphorylate other substrates (Table 1.1).

Based on amino acid sequence similarity and global protein structure, dCK, dGK and TK2 belong to the same family (type 1), while TK1 seems to have a different evolutionary origin as it displays little homology to the other three enzymes and therefore belongs to a different family (type 2). The studies described in this dissertation exclusively focus on type 1 dNKs, thus type 2 kinases will not be further discussed.

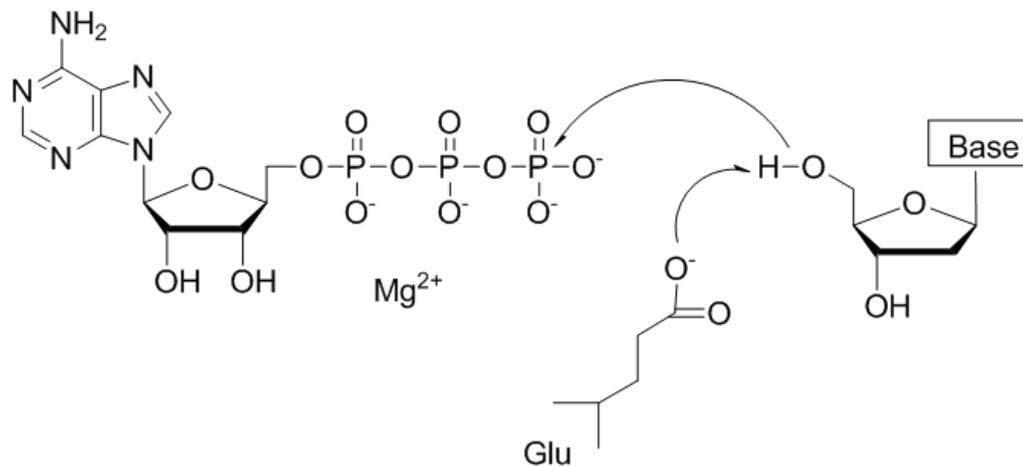


Figure 1.3 Proposed mechanism of dNK (Eriksson et al., 2002).

Table 1.1 Natural substrates for the four human deoxyribonucleoside kinases.

dNK	Natural substrates
TK1	T, dU
TK2	T, dU, dC
dCK	dC, dA, dG
dGK	dG, dA

Each of the type 1 kinases possesses different substrate specificities. For example, human dCK has broad substrate specificity and phosphorylates dC, dG and dA (Table 1.1), while human dGK only catalyzes the purine deoxyribonucleosides phosphorylations. Although dGK has been shown to share 48% sequence homology with dCK, it cannot phosphorylate pyrimidines (Wang et al., 1993). Similarly, human TK2 exclusively phosphorylates pyrimidine nucleosides. A much broader substrate specificity was found for the deoxynucleoside kinase from *Drosophila melanogaster* (*DmdNK*), which phosphorylates all four natural deoxynucleosides, with a preference for pyrimidines (Munch-Petersen et al., 2000).

The three-dimensional structures of dNKs are available for human dGK (PDB: 1JAG; (Johansson et al., 2001)), human dCK (PDB: 2A30; (Godsey et al., 2006)) and *DmdNK* (PDB: 1OT3; (Mikkelsen et al., 2003)). All of these dNKs form dimers in the crystal structure, and the monomer of each enzyme shares a common protein fold of five central β -sheets surrounded by nine α -helices (Figure 1.4). Two conserved motifs of dNKs, the P-loop and the Lid region, are particularly important for their function (Figure 1.5). The conserved P-loop (GXXXXGKS/TT) is responsible for binding the phosphate donor and orienting the α - and β -phosphate groups through interactions involving the conserved glycine residues and the side chain hydroxyl of conserved serine or threonine residues. The Lid (RXXXXRXXE) has been shown to undergo a conformational change upon substrate binding, bringing the positively charged residues close to the reaction center to facilitate phosphate transfer.

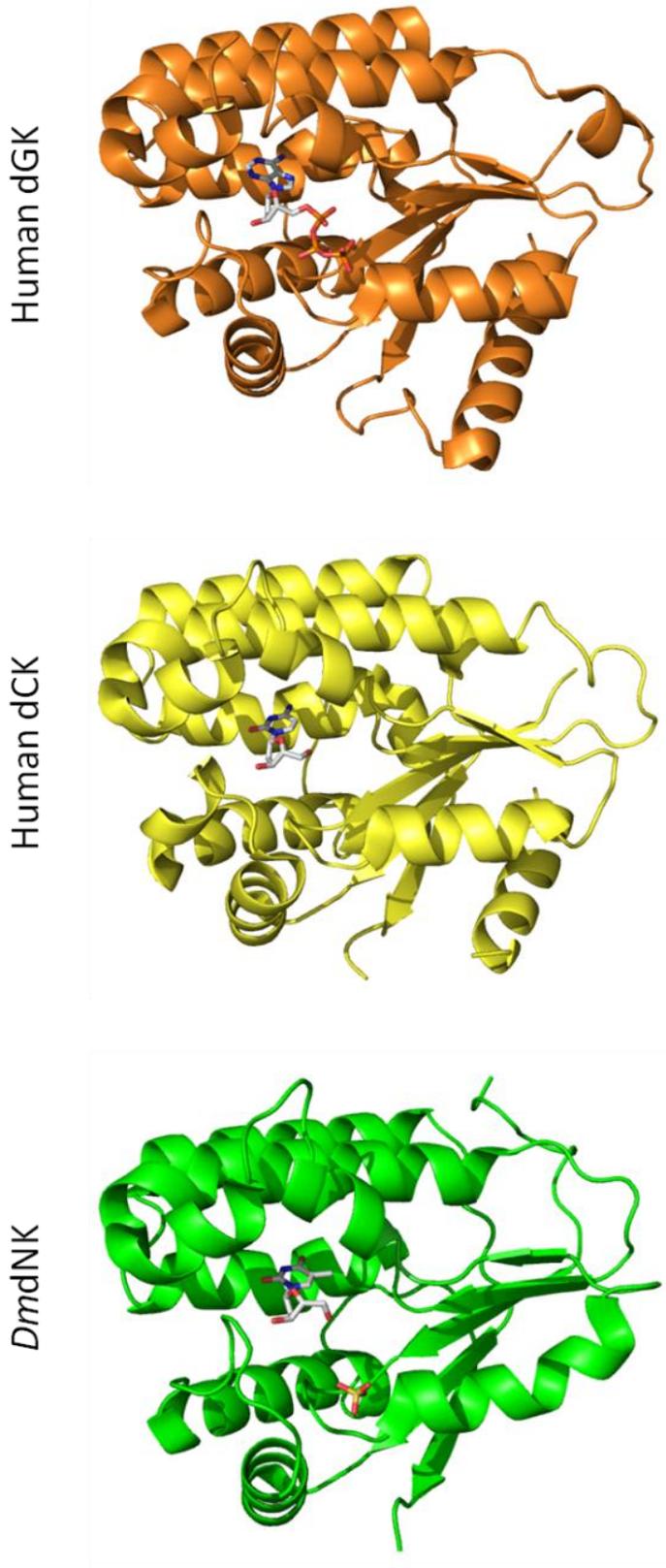


Figure 1.4 Comparison of structures of three type 1 dNKs: *DmdNK* with Thymidine and sulfate (PDB: 1OT3; (Mikkelsen et al., 2003)). Human dCK with dC (PDB: 2A30; (Godsey et al., 2006)). Human dGK with dATP (PDB: 1JAG; (Johansson et al., 2001)).

P-loop

```

hTK2  --MGAFQQRPSDDKEQEKEKKSVICVEGNIASGKTTCLEFFSN-ATDVEVLTEPVS KWRN
Dmd   MAEAAASCARKGT-KYAEGTQPFTVLIENIGSGKTTYLNHF EKYKNDICLLTEPVEKWRN
hdGK  MAKSPLEGVSSSRGLHAGRGRRLSIEGNIAVGKSTFVKLLTKTYPEWHVATEPVATWQN
hdCK  MATPPKRSCP SFSASSEGTRIKKISIEGNI AAGKSTFVNI LKQLCEDWEVVPEPVARWCN
      . . : : ***** **:* :: : : : : .*** * *

hTK2  VR-----GHNPLGLMYHDASRWGLTLQTYVQLTMLDRHTRPQ-----V
DmdNK VN-----GVNLELMYKDPKKWAMPFQSYVTLTMLQSHTAPT-----N
hdGK  IQAAG---TQKACTAQSLGNLLDMMYREPARWSYTFQTFSL SRLKVQLEPFPEKLLQAR
hdCK  VQSTQDEFEELTMSQKNGGNVLQMMYEKPERWSFTFQTYACL SRIRAQLASLNGKLDKAE
      : .          * * : **... : * . . : * : * : : : .

hTK2  SSVRLMERSIHSARYIFVENLYRSGKMPEVDYVVLSEWFDWILRNMDVSVDL--IVYLRT
DmdNK KKLKIMERSIFSARYCFVENMRRNGSLEQGMYNLTLEEWYKFI EESIHVQADL--IIYLRT
hdGK  KPQVIFERSVYSDRYIFAKNLFENDSLSDIEWHIYQDWH SFLLWEFASRITLHGFIYLQA
hdCK  KPVLFERSVYSDRYIFASNLYESECMNETEWTIYQDWH DWMNNQFGQSLELDGIIYLQA
      . : : : ** : . * * * . : * : . : : : . : * . : : : * : : ** : :

      Lid

hTK2  NPETCYQLKKRCREEEKVIPLEYLEAIHHLHEEWLIKGS LFP----MAAPVLVIEADH
DmdNK SPEVAYLRIRQRARSEESCVP LKYLQELHELHEDWLIH-QRRP----QCKVLVLDADL
hdGK  SPQVCLRLYQRAREEEEGIELAYLEQLHGQHEAWLIHKT TLHF EALMNPVLVLDVND
hdCK  TPETCLRLIYLRGRNEEQGIPL EYLEKLHYKHESWLLHRTLKTNFDY LQEVPI LTLDVNE
      . * . . . * : * * . * . : * * * : * * * * : : * . : : :

hTK2  HMERMLELFEQNRDRI LTPENRKHCP-----
DmdNK NLENIGTEYQRSESSIFDAISSNQPS PVLVSPSKRQRVAR
hdGK  DFSEEVTKQEDLMREVNTFVKNL-----
hdCK  DFKD---KYESLVEKVK EFLSTL-----
      . : . : : .

```

Figure 1.5 ClustalW sequence alignment of four dNKs (human TK2, *Dmd*NK, human dGK and human dCK). Symbols below the sequence indicate the degree of conservation in an alignment column: strictly conserved (stars), highly conserved (two dots), or moderately conserved (one dot). Conserved P-loop and Lid regions are marked by a red box.

1.3.2 *NMPK family*

Nucleoside monophosphate kinases catalyze the phosphorylation of nucleoside monophosphates to nucleoside diphosphates. The kinetic mechanism of NMPK has been shown to be random bi-bi, in which the nucleotides bind independently of each other (Cheng and Prusoff, 1973; Li et al., 1996; Rhoads and Lowenstein, 1968). The reaction proceeds by direct phosphate transfer based on the observed stereo-inversion of the phosphoryl group (Richard et al., 1980). However, whether the phosphoryl transfer in NMPK follows an associative or dissociative mechanism remains controversial (Yan and Tsai, 1999).

There are four groups of NMPKs in human cells: TMP kinase (TMPK), UMP-CMP kinase (UMP-CMPK), five isoforms of adenylate kinase (AK), and several guanylate kinases (GMPK) (Van Rompay et al., 2000). The NMPKs have high substrate specificity for phosphate acceptors but are highly promiscuous for phosphate donors (Table 1.2). For example, TMPK phosphorylates TMP and dUMP to their diphosphate form. Its preferred phosphate donors are ATP, dATP, GTP and dGTP, but it can use several other phosphate donors as well (Huang et al., 1994; Lee and Cheng, 1977). UMP-CMPK phosphorylates CMP, dCMP and UMP with highest efficiency and dUMP, AMP, and dAMP with lower efficiency. The preferred phosphate donors are ATP and dATP, but the enzyme can use several other donors (Van Rompay et al., 1999).

The structure of all NMPKs can be divided into three parts, the CORE, NMP binding and the Lid domain (Yan and Tsai, 1999) (Figure 1.6). The CORE domain includes the central parallel β -sheet and the surrounding α -helices. The conserved P-loop, which plays an important role in binding the phosphate donor, resides in this domain. The

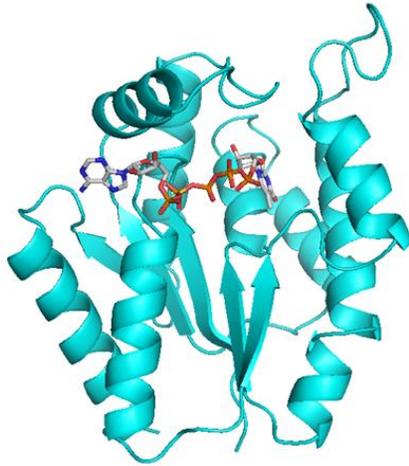
NMP binding domain is mostly conserved among NMPKs except GMPK. In GMPK, this domain contains 29 residues that form four β -strands and a short α -helix; but in other NMPKs this domain is constituted by 47 residues in the format of helices. The Lid domain, responsible for closing the active site during the catalytic cycle, displays the most variety in terms of sequence and length. For example, the Lid region in AKs is much longer and in several cases forms secondary structures (Bellinzoni et al., 2006), while the Lid domain in other NMPKs exists in the form of loop structure.

Among all NMPKs, the structure of TMPK shows the highest homology to the dNK structures and is the focus of this dissertation. The reaction mechanism of TMPK is different from dNK in that there is no residue in the active site that functions as a general base. The enzyme positions both of phosphate donor and receptor to close proximity and at the correct orientation to facilitate the phosphoryl transfer (Figure 1.7). Aligning the P-loop and the Lid region sequences of TMPKs reveals a disparity that can be used to further categorize them into two types (Table 1.3). Type I TMPKs (for example, yeast and human) have a basic residue in their P-loop sequence in addition to an invariant lysine residue, which can interact with the γ -phosphate group of ATP, and lack such residues in the Lid region. On the contrary, type II TMPKs (for example, *E. coli* and *T. maritima*) have a glycine in the P-loop and basic residues in the Lid region that interact with ATP. This sequence arrangement is more similar to other NMPKs, such as adenylate kinase or uridylate kinase (Lavie et al., 1998b).

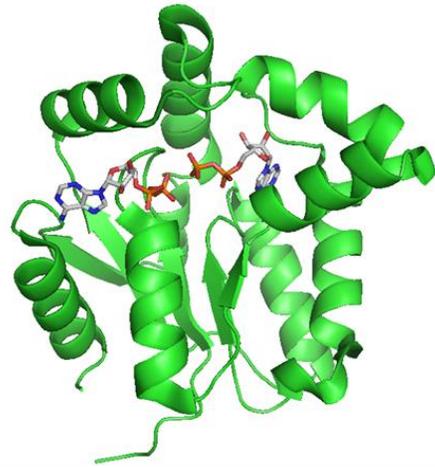
Table 1.2. Natural substrates of the human NMPKs.

Monophosphate kinase	Natural substrate
TMPK	TMP, dUMP
UMP-CMPK	CMP, dCMP,UMP, dUMP, AMP, dAMP
AKs	AMP, dAMP, CMP, dCMP
GMPK	GMP, dGMP

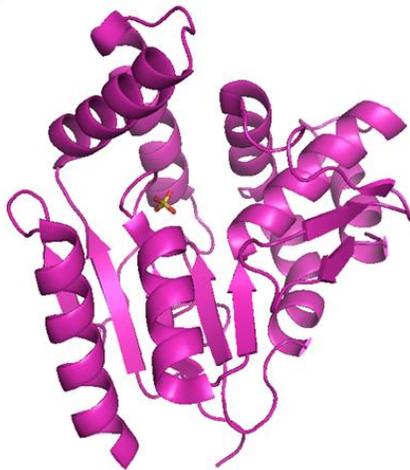
E.coli Thymidylate kinase



Mycobacterium tuberculosis
Adenylate kinase



E.coli CMP kinase



E.coli Guanylate Kinase

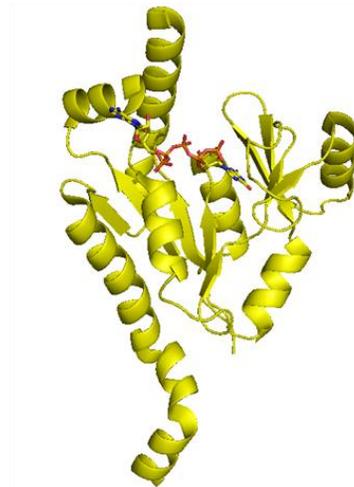


Figure1.6 Structure of four members of the NMPK family: *E.coli* TMPK (PDB 4TMK; (Lavie et al., 1998b)). *Mycobacterium tuberculosis* AK (PDB 2CDN;(Bellinzoni et al., 2006)). *E.coli* CMPK (PDB 1KDO;(Bertrand et al., 2002)). *E.coli* GMPK (PDB 2F3R; (Hible et al., 2006)).

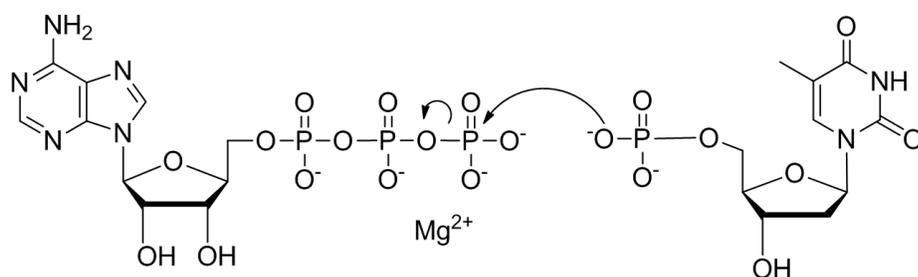


Figure 1.7 Proposed mechanism of TMPK ((Ostermann et al., 2000).

Table 1.3 Comparison of the P-loop and Lid residues of representative members from Type I and II TMPK families, as well as AK and UMPK.

	<u>P-loop</u>	<u>Lid</u>
Type I		
Human TMPK	EGVD R AGKST	KRGAF A HERY
Yeast TMPK	EGLD R TGKTT	AEKSGFGDERY
Plasmodium falciparum TMPK	EGLD R SGKST	AQ N RSYGE E IY
Type II		
<i>E. coli</i> TMPK	EGLEGAGKTT	LKRAR A R G E L DRI
<i>Thermotoga maritima</i> TMPK	EGIDGSGKST	LKRKGELNN R FEKR
Mycobacterium tuberculosis TMPK	EGVDGAGKRT	GERSR G R A Q R DPG
Pig Adenylate kinase	VGGPGSGKGT	TKRLL K R G ETSGRV
<i>Dictyostelium</i> Uridylate kinase	LGGPGSGKGT	TQRL L K R GESSGRS

1.3.3 Multifunctional enzymes

A particularly interesting enzyme in the context of NAs activation is HSV1-TK, a special member in the type I dNK family. It was the first deoxyribonucleoside kinase to be found capable of catalyzing two consecutive phosphoryl transfers on a single nucleoside substrate, therefore functioning both as TK and TMPK (Wild et al., 1997).

HSV1-TK can utilize a broad range of substrates, including pyrimidines and pyrimidine analogs as well as acyclic purine analogs.

HSV1-TK shares a similar core structure with both dNKs and TMPKs (Wild et al., 1997). It also contains five central β -strands and surrounded by α -helices (Figure 1.8). However, HSV1-TK is a much larger protein than dNK and TMPK, due to two additional elements: the 45 residues at the N-terminal, 29 residues at the C-terminal and 72 residues between α 8 and β 5, comprising 4 helices and a 15-residue loop (Wild et al., 1997). These differences are mainly located at the dimer interface but not at the catalytic center. More importantly, HSV1-TK shares key catalytic residues with dNKs (Johansson et al., 2001).

Because of the uniqueness of HSV1-TK, research has been done to investigate the determinant of the multifunctionality possessed by this enzyme. The most prominent example is the crystallography study of HSV1-TK and EHV-TK (Equine herpes virus thymidine kinase, a homolog of HSV1-TK) with the bi-substrate inhibitors TP₄A and TP₅A. These two inhibitors differ from each other in the length of their phosphate linkers, therefore, mimicking the state when both substrates (ATP and thymidine or TMP) are bound to the active site. In theory, the enzymes would use different sets of residues for each reaction, since the first step reaction requires a general base and the second reaction

does not. However, an overlay of the two crystal structures shows that all the active site residues are superimposed on each other. Even the general base residue, Glutamate 83, shows the same orientation in the two structures (Gardberg et al., 2003a). Overall, the mechanism of how HSV1-TK catalyzes both reactions with the same set of active site residues is still unclear.

1.4 Engineering multifunctional kinases and its limitations

Based on the known structure and function of dNK and NMPK, protein engineering experiments have been carried out towards the goal of improving NAs activation. Specifically, engineering dNK has achieved fairly successful results. For example, by combining a FACS-based assay and directed evolution approaches, such as random mutagenesis and DNA shuffling, a mutant of *DmdNK* was evolved to specifically phosphorylate ddT, while the catalytic efficiency for the natural substrate, 2'-deoxyribonucleosides, was lowered up to 48,000 fold (Liu et al., 2009). TMPK has also been engineered by both rational design and directed evolution approaches. In the study by Levie *et. al.*, yeast and human TMPKs were engineered by inserting a Lid region from bacterial TMPK, which results in a significant change of AZT-MP acceptance such that it is phosphorylated faster by the TMPK mutant (Brundiers et al., 1999b) .

In principle, increased efficiency of NAs phosphorylation in either step (TK or TMPK catalyzed reaction) could shift the bottleneck of the activation to the other. The long-term challenge is to eliminate the bottleneck reactions all together by creating a dual-functioning biocatalyst. While there have been numerous enzyme engineering studies involving HSV1-TK, the natural dual functioning enzyme, none of them has

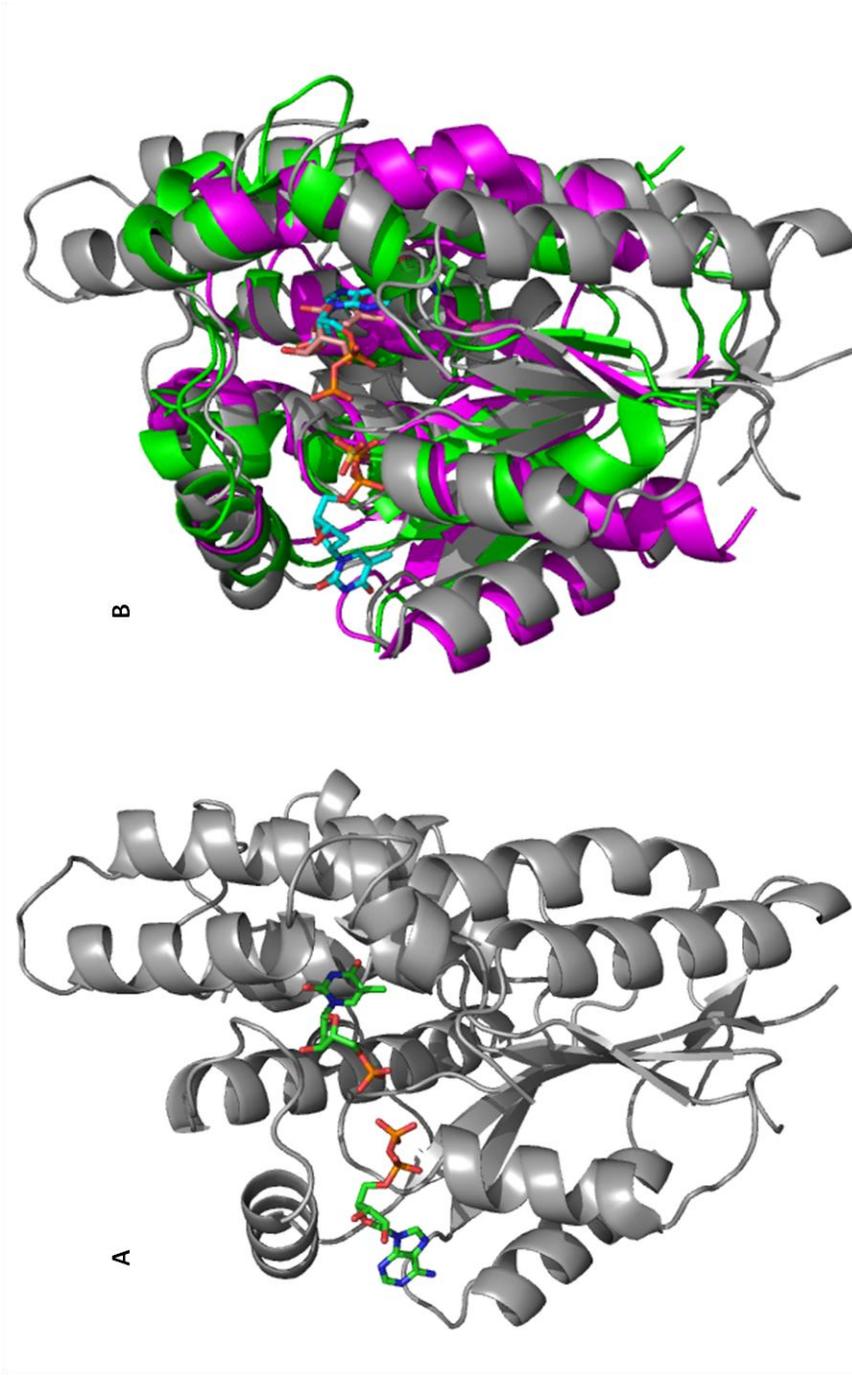


Figure 1.8 Structure of HSV1-TK and its comparison with dNK and TMPK. (A) Structure of HSV1-TK with ADP and TMP (PDB 1VTK; (Wild et al., 1997)). (B) Structure overlay of HSV1-TK (grey (Wild et al., 1997)), *DmdNk* (green (Mikkelsen et al., 2003)) and *TmTMPK* (magenta (Yoshikawa, 2009)).

focused on improving the second phosphorylation step. The fact that very little engineering work has been reported on the multifunctional system is probably due to the lack of appropriate techniques to efficiently screen and select functional members from directed evolution libraries.

Previously, selection and screening strategies for TK function have been developed. For example, *E. coli* auxotroph strain KY895 carries a lethal mutation in chromosomal *tk*; it has been successfully used to select for mutations with wild type TK function (Hiraga et al., 1967). Under the selection condition, *E. coli* KY895 grows in the presence of 5-fluoro-2'-deoxyuridine (FdU) and uridine, the former will become phosphorylated and inhibit thymidylate synthase, shutting down the *de novo* pathway. At the same time, uridine inhibits thymidine phosphatase, which prevents dephosphorylation of FdU monophosphate. Cell survival solely depends on the salvage pathway, relying on intake of thymidine by the nucleoside transporter and its subsequent phosphorylation by exogenous TK expressed from a plasmid.

Unlike TK's exclusive function in the salvage pathway, TMPK exists at the junction of the *de novo* and salvage pathways of TTP synthesis. As an essential gene, *tmpk* presents a special challenge in creating an auxotroph *E. coli* strain. While *tk* deficient strains can survive utilizing the *de novo* pathway, mutations or deletions that destroy TMPK function are lethal to *E. coli*. Therefore, creating an efficient selection system for dual TK and TMPK function is a major roadblock to achieving success in engineering of multifunctional enzymes with both TK and TMPK activity for better NAs activation.

1.5 Aim and scope of the dissertation

This dissertation focuses on the engineering of kinases with dual TK and TMPK function as well as the investigation of the structural basis for the dual function of HSV1-TK. In chapter 2, I report on the characterization of the TMPK from *Thermotoga maritima*, a thermophilic enzyme with robust TMPK function and high thermostability, which in the following studies serves as a representative for the TMPK family and a target for enzyme engineering.

In chapter 3, the development and validation of a conditional auxotrophic *E. coli* strain for selection of dual TK and TMPK function is described. The establishment of a successful selection method is required for the directed evolution experiments described in chapter 4 and will be widely applicable to future engineering works focusing on dual TK and TMPK function.

In chapter 4, the non-homologous recombination technique SCRATCHY in combination with the computational program SCHEMA were applied to generate hybrid enzyme libraries between *DmdNK* and *TmTMPK*. The results showed that although *DmdNK* and *TmTMPK* are structurally similar to the dual function enzyme HSV1-TK, chimeras made of the two enzymes do not possess dual function. To better understand the structure basis for the dual function of HSV1-TK, I carried out site-directed mutagenesis experiments that targeted non-conserved residues in conserved catalytic regions. In addition, Isothermal Titration Calorimetry (ITC) was also used to investigate the thermodynamics of substrate recognition and inhibitor binding in *DmdNK*, *TmTMPK* and HSV1-TK.

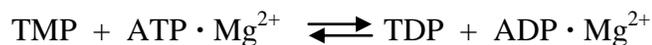
Chapter Two

Expression and characterization of thymidylate kinase from *Thermotoga maritima*

2.1 Introduction

Thymidylate kinase (TMPK) catalyzes the phosphorylation of thymidine monophosphate (TMP) to thymidine diphosphate (TDP) using ATP as its preferred phosphate donor (Jong and Campbell, 1984). As the first enzyme after the merging of the *de novo* and the salvage pathways in thymidine triphosphate (TTP) synthesis, it is essential for cell proliferation and survival (Liang et al., 1995; Su and Sclafani, 1991). Its fundamental role in nucleotide metabolism is emphasized by the fact that TMPK genes have been identified in all sequenced genomes, including the smallest microbial genomes, which are thought to consist of a minimal set of genes required for cellular life (Fraser et al., 1995; Tamas et al., 2002). In addition, a knock-out of the single TMPK encoding gene in yeast is lethal (Sclafani and Fangman, 1984).

Physiologically, TMPK catalyzes the reversible phosphorylation reaction according to the following scheme:



In the enzyme active site, TMPK binds the phosphate donor and acceptor in an orientation that the terminal phosphate groups face each other. The kinetic mechanism of TMPK, along with other NMPKs, has been shown to be random bi-bi, in which the nucleotides bind independently of each other (Cheng and Prusoff, 1973; Li et al., 1996; Rhoads and Lowenstein, 1968). It is unlikely that there is a phosphoenzyme intermediate because the stereocourse is inversion (Richard et al., 1980). However, whether the phosphoryl transfer in TMPK follows an associative or dissociative mechanism remains controversial (Yan and Tsai, 1999).

In addition to the biological function of TTP synthesis, this enzyme is also required for the phosphorylation of nucleoside analogs (NAs), which are anti-viral and anti-cancer prodrugs. These compounds enter cells as uncharged nucleosides and require cellular kinases to phosphorylate them to their triphosphate form in order to inhibit DNA and RNA synthesis. For most NAs this conversion is the rate determining step in the overall activation pathway. Medical examples include the clinically used d4T (2',3'-didehydro-2',3'-dideoxythymidine) and AZT (3'-azido-3'-deoxythymidine). For the latter one, the phosphorylation of AZT-monophosphate to AZT-diphosphate is the rate-limiting step in the overall activation, which causes toxic accumulation of AZT-MP in the cell (Lavie et al., 1998b). Therefore, understanding the mechanism of phosphoryl transfer in TMPK and how TMPK functions towards natural substrates and NAs are important for drug design and potentially for its therapeutic application.

The crystal structures of several members of the TMPK family have been solved. They reveal a conserved protein fold similar to other NMPKs. Each monomer of the dimeric protein adopts a common fold of a five parallel β -stranded core surrounded by nine α -helices. The highly hydrophobic dimer interface is composed of three parallel α -helices from each monomer that stack against each other (Figure 2.1). There are three loops critical to their function: the P-loop, which binds and positions the α - and β -phosphate groups of ATP; the DRH/Y motif, which binds and positions the magnesium ion; and the Lid region, which is a flexible stretch that closes the active site when the enzyme binds substrates.

In this chapter, a novel TMPK from the hyperthermophilic bacterium *Thermotoga maritima* (*Tm*TMPK) was isolated and characterized. Following the publication of the

Thermotoga maritima genomic sequence (accession number AE000512), the hypothetical *Tm*TMPK gene sequence was proposed according to sequence alignment with other TMPKs (Zacherl et al., 1997) (Figure 2.2). *Tm*TMPK shares highest sequence identity to *Ec*TMPK (38%) over the human and yeast TMPK. The reason behind our interest in *Tm*TMPK is that, although the crystal structure of *Tm*TMPK was published in the Protein Data Bank as part of the structural genomic/proteomic initiative project (RIKEN, Japan PDB 3HJN), a detailed biochemical study of *Tm*TMPK has not been reported. The enzyme has the potential to provide novel functional and structural insight. Furthermore, enzymes from hyperthermophilic organisms often have high thermal stability, which promises potential for enzyme engineering and future use in gene therapy.

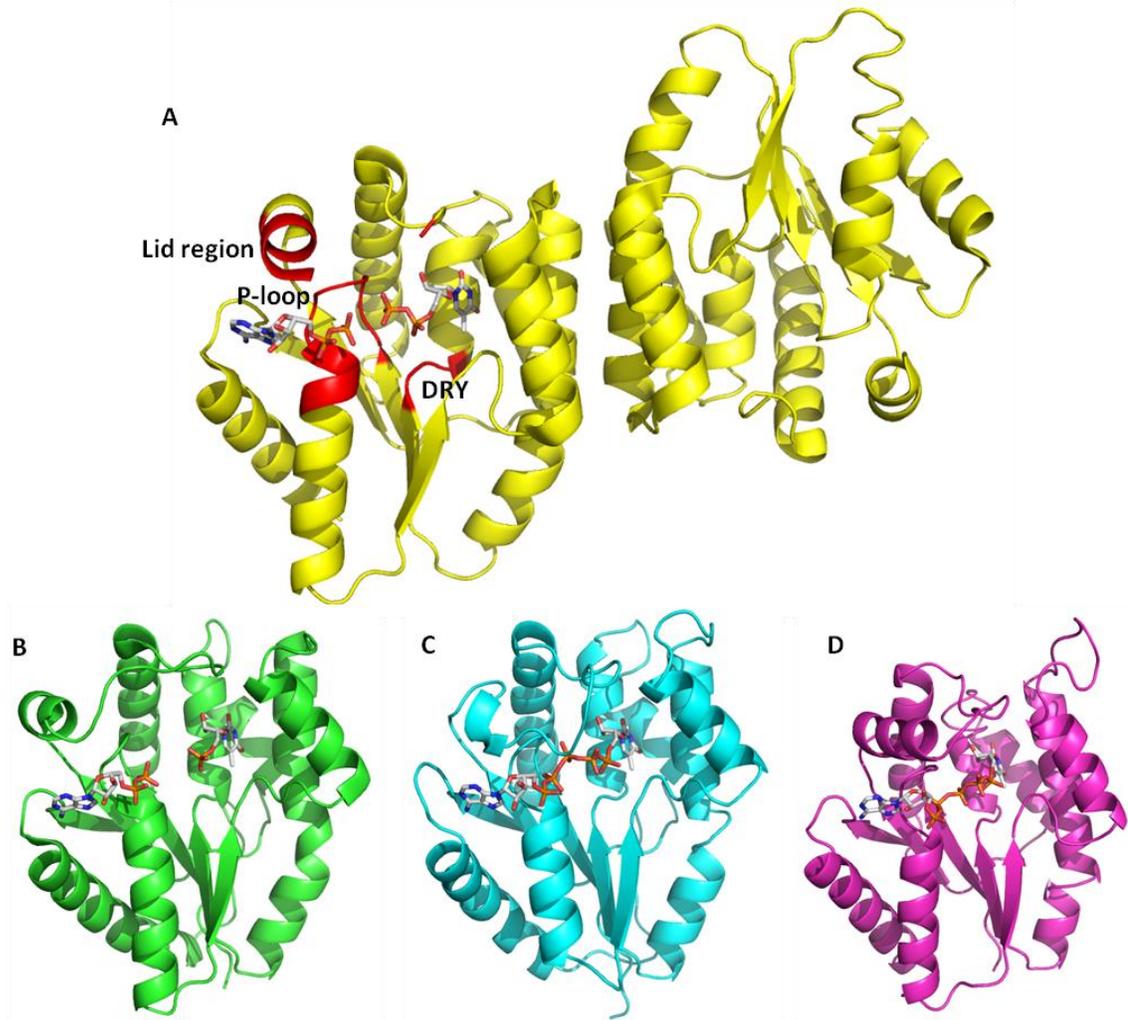


Figure 2.1 Crystal structures of TMPKs. (A) Dimeric *Tm*TMPK in complex with ADP and TDP (PDB: 3HJN; (Yoshikawa, 2009)); (B) Single subunit of human TMPK with TMP and ADP in the active site (PDB: 1E2D); (Ostermann et al., 2000)); (C) subunit of *Saccharomyces cerevisiae* TMPK with bi-substrate inhibitor TP₅A (PDB: 3TMK; (Lavie et al., 1998a)); (D) *E. coli* TMPK subunit with bi-substrate inhibitor TP₅A (PDB: 4TMK); (Lavie et al., 1998b)).

CLUSTAL 2.1 multiple sequence alignment



Figure 2.2 ClustalW alignments of four TMPKs (human TMPK (hTMPK), *Saccharomyces cerevisiae* TMPK (yTMPK), *E. coli* TMPK (EcTMPK) and *Thermotoga maritima* TMPK (TmTMPK)). Symbols below the sequence indicate the degree of conservation in an alignment column: strictly conserved (stars), highly conserved (two dots), or moderately conserved (one dot). Conserved P-loop, DRH/Y and Lid regions are marked in red box.

2.2 Results and discussion

2.2.1 Structure and active site of *Tm*TMPK.

Although the crystal structure of *Tm*TMPK has been solved as part of the structural genomic/proteomic project (RIKEN, Japan), a detailed analysis of the *Tm*TMPK sequence and structure has not been reported. Superimposing the crystal structure of *Tm*TMPK (PDB code:3HJN; (Yoshikawa, 2009)) with *E. coli* TMPK (PDB code: 4TMK; (Lavie et al., 1998b)) and human TMPK (PDB code: 1E2D; (Ostermann et al., 2000)) reveals a global structure shared by all three enzymes (Figure 2.3). The β -sheet core overlaps well, but the surrounding α -helices are slightly offset relative to each other. Loop regions important for substrate binding and catalysis also overlap well.

In the active site, the reaction product thymidine diphosphate (TDP) is bound in a deep cleft within which the nucleobase forms hydrogen bonds with multiple residues and is similar to the binding interaction of TMP in *Ec*TMPK (Figure 2.4). Located at the α 3 helix in *Tm*TMPK, Phe64 forms a base stacking interaction with the thymine base, and the side chain of Arg68 forms a H-bond with oxygen O4 of thymine. The deoxyribose of TDP is anchored by a H-bond between its 3'-hydroxyl group and the carboxylate side chain of Asp9 from the P-loop. This acidic residue in the P-loop of *Tm*TMPK is equivalent to Asp15 in hTMPK, Asp14 in *y*TMPK and Glu12 in *Ec*TMPK, a signature residue in several NMPKs (Lavie et al., 1998b). The α -phosphate of TDP is bound to Arg89 from the DRH/Y motif and Arg47 while the β -phosphate group of TDP interacts with Lys13, Lys139 and Mg^{2+} .

In the phosphate donor site, *Tm*TMPK contains the highly conserved P-loop residues among TMPK family members (GX₄GKS/T) (Lavie et al., 1998b), which bind

the phosphate groups of ADP. The terminal phosphate in ADP occupies the position of the β -phosphate of ATP and binds to backbone nitrogens from Lys13 to Thr16, and to the Thr16 side chain. Mg^{2+} interacts with both the phosphate group of ADP and TDP, orienting the phosphate groups and presumably facilitating the phosphoryl transfer reaction by stabilizing the transition state.

In TMPKs, the phosphate groups of the donor and the acceptor also interact with the highly basic Lid region. The closure of the Lid region upon substrate binding in TMPK brings arginine residues to the reaction center to stabilize the transition state of the phosphoryl-transfer reaction. In *Ec*TMPK the Lid region harbors three arginines (¹⁴⁸KRARAR¹⁵³), but only Arg153 interacts with the phosphates in the crystal structure. Arg149 stacks against the adenine base and Arg151 points toward the solvent (Lavie et al., 1998b). From the sequence comparison, the *Tm*TMPK Lid region also contains three basic residues (¹³⁷KRKGEL¹⁴²), although they are adjacent to each other, instead of being each separated by one-residue. In the *Tm*TMPK crystal structure, part of the Lid region is invisible most likely due to its high flexibility. However, a similar interaction pattern is observed in *Ec*TMPK. Arg138 stacking against the adenine base, similar to Arg149 in *Ec*TMPK; and Lys139 interacts with both phosphate groups from ADP and TDP, similar to Arg153 in *Ec*TMPK. Whether Lys140 adopts a similar orientation as Arg151 in *Ec*TMPK (pointing to the solvent) is unknown.

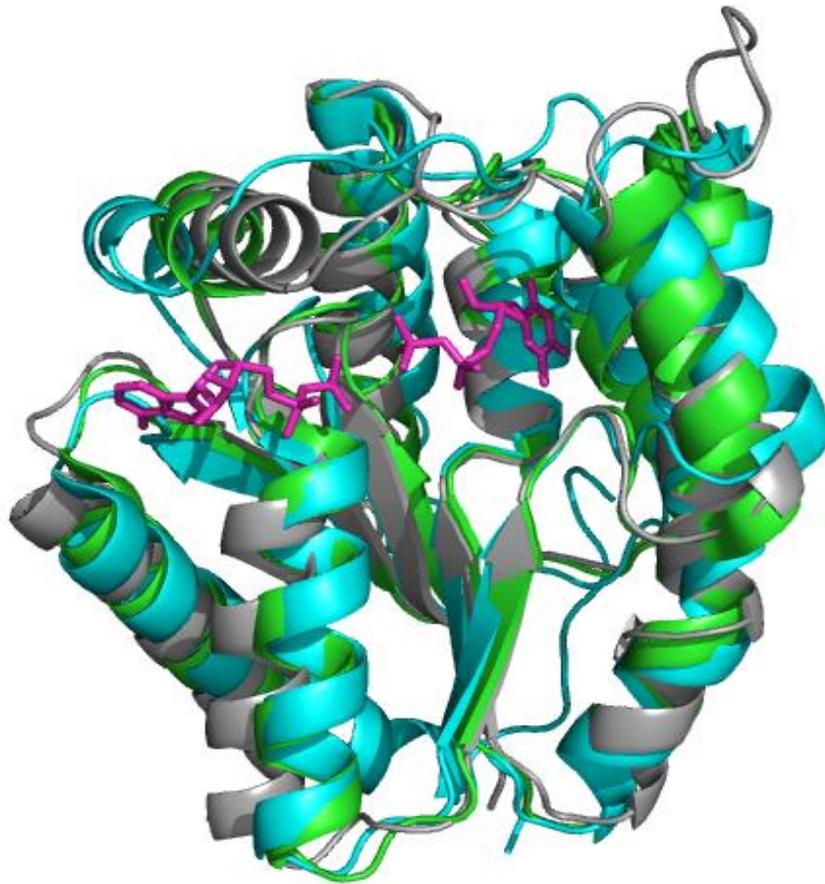


Figure 2.3 Superimposition of crystal structures of three TMPKs: *Tm*TMPK (green, PDB code: 3HJN; (Yoshikawa, 2009)); *Ec*TMPK (grey, PDB code: 4TMP; (Lavie et al., 1998b)) and hTMPK (blue, PDB code: 1E2D; (Ostermann et al., 2000)). Substrate ADP and TDP (magentas) are from the *Tm*TMPK crystal structure.

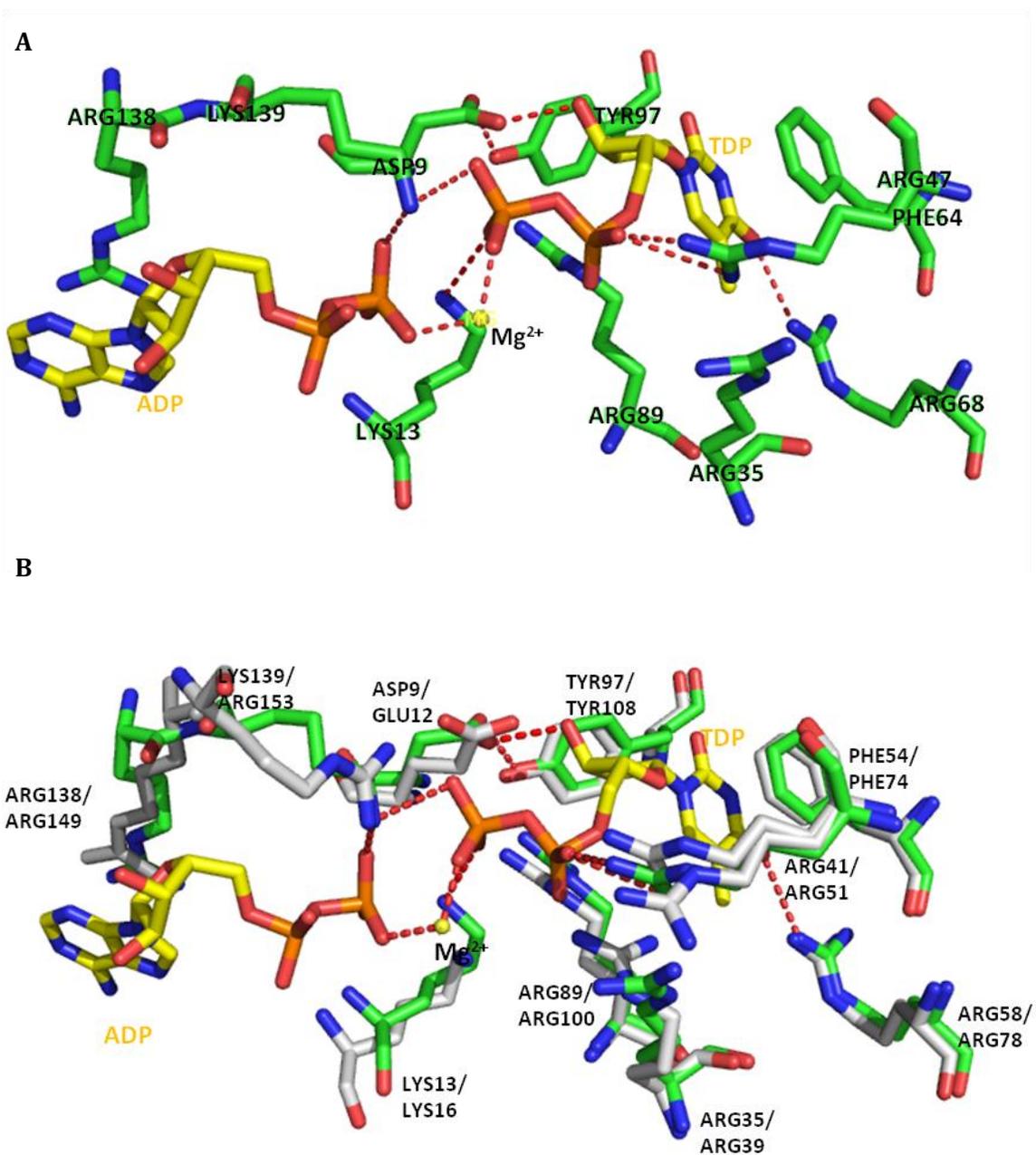


Figure 2.4 Active site residues of TmTMPK. (A) Interactions of ADP and TDP with residues in *Tm*TMPK (marked by dashed lines). (B) Overlay of the *Ec*TMPK (grey) with *Tm*TMPK (green) active site residues.

2.2.2 TmTMPK gene isolation and protein purification

The hypothetical gene of *TmTMPK* was amplified from the genomic DNA (accession number AE000512) using primers designed to flank the coding region. Analysis on 1% agarose shows a single band of 600 bp, which is consistent with the size of the coding region (594 bp) (Figure 2.5A). Confirmed by DNA sequencing, the gene is identical to the TM1099 open reading frame from the *Thermotoga maritima* genomic sequence. A protein of 197 amino acids with molecular mass of 22,850 Daltons is predicted from translation of the open reading frame.

To facilitate protein purification, the gene was subcloned into the *NdeI* and *SpeI* sites of a pET14b vector. *TmTMPK* was expressed with an N-terminal affinity tag of six histidine residues, and then purified using a Ni-NTA resin. This protein was expressed at a relatively high level under the induction condition. SDS-PAGE analysis of cell lysate and insoluble fractions shows that the majority of the protein was in the soluble fraction, but a small amount existed in the insoluble portion. Purification with Ni-NTA resin yielded protein of > 95% purity. The calculated molecular mass of *TmTMPK* with the His-tag is 25,014 Daltons, which is consistent with the size of the purified protein band in SDS-PAGE analysis (Figure 2.5B). The yield of protein production is 1.5 g /Liter culture.

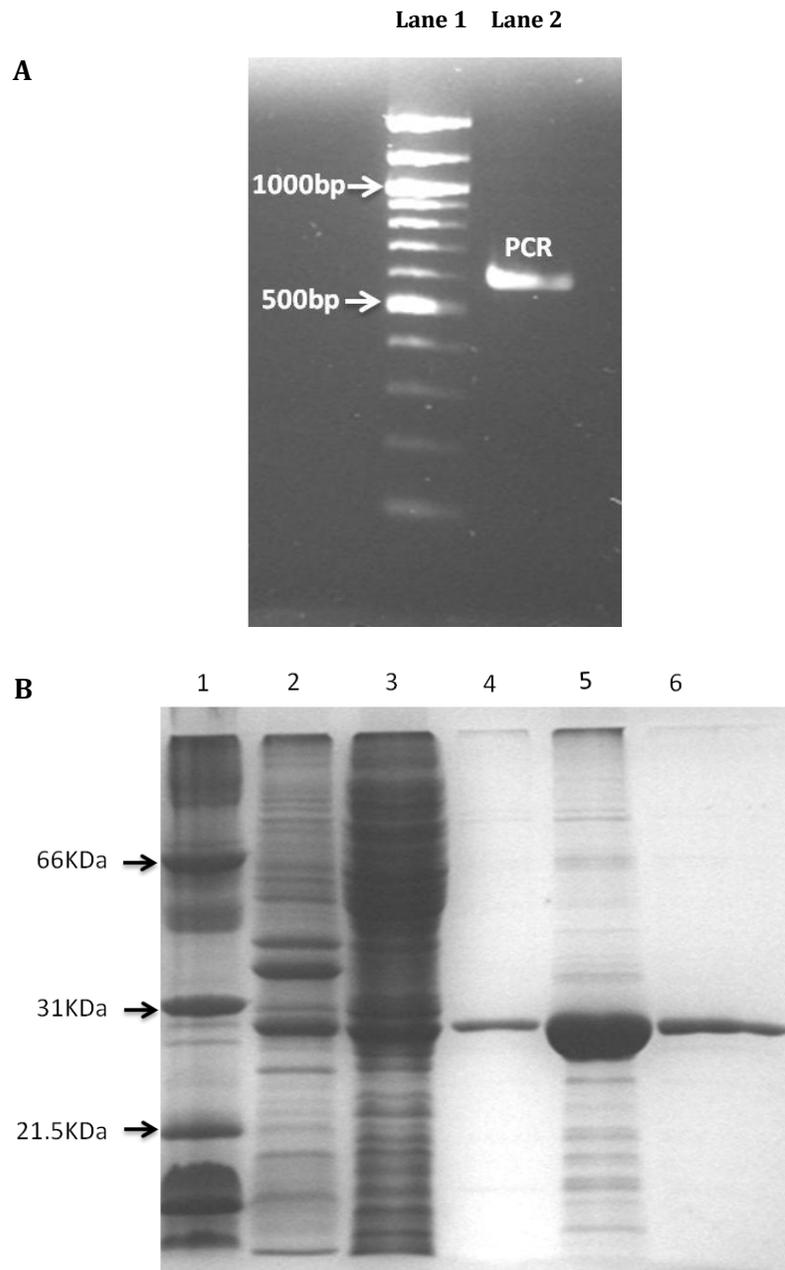


Figure 2.5 Gene isolation and expression of *TmTMPK*. (A) Electrophoresis of amplified *TmTMPK* on 1% agarose gel. Lane 1: size marker; lane 2: PCR product. (B) SDS-PAGE analysis of *TmTMPK* purification from *BL21(DE3)pLysS* cells. Lane 1: molecular weight markers; lane 2: cell lysate; lane 3: insoluble fraction; lane 4, 5 and 6: elution fractions.

2.2.3 *Tm*TMPK is a highly thermostable thymidylate kinase

Protein secondary structure information was investigated by far-UV circular dichroism (CD) (Figure 2.6 A). The CD spectrum of the protein exhibits strong negative ellipticity at 209 and 223 nm suggesting high helical content. To further characterize the thermal stability of *Tm*TMPK, the change in molar ellipticity upon temperature increases was recorded following the global minimum at 209 nm. Shown in Figure 2.6 B, minimal change in ellipticity was observed up to 100°C, at which temperature the buffer reaches its boiling point. The results indicate that at neutral pH, no thermal denaturation was induced over the temperature range from 20 to 100°C. Therefore, *Tm*TMPK has high intrinsic stability against thermal unfolding.

In order to further understand the thermal stability of *Tm*TMPK, Differential Scanning Calorimetry (DSC) studies were performed. The thermal transition midpoint (T_m) of *Tm*TMPK is 99°C, at which 50% of the protein sample is unfolded (Figure 2.6 C). The transition was accompanied by heavy protein precipitation, producing a distorted exothermic transition. Because it is an irreversible transition, no other thermodynamic parameters were calculated. This T_m also agrees well with the far-UV CD melting experiment results, as the ellipticity at 209 nm started to increase as the temperature approaches 99°C. Interestingly, *Tm*TMPK exhibits a low-amplitude endothermic transition with a midpoint of 79°C, perhaps indicating a structural rearrangement at this temperature. A similar behavior was also observed in another enzymes from *Thermotoga maritima* (Brown et al., 1993).

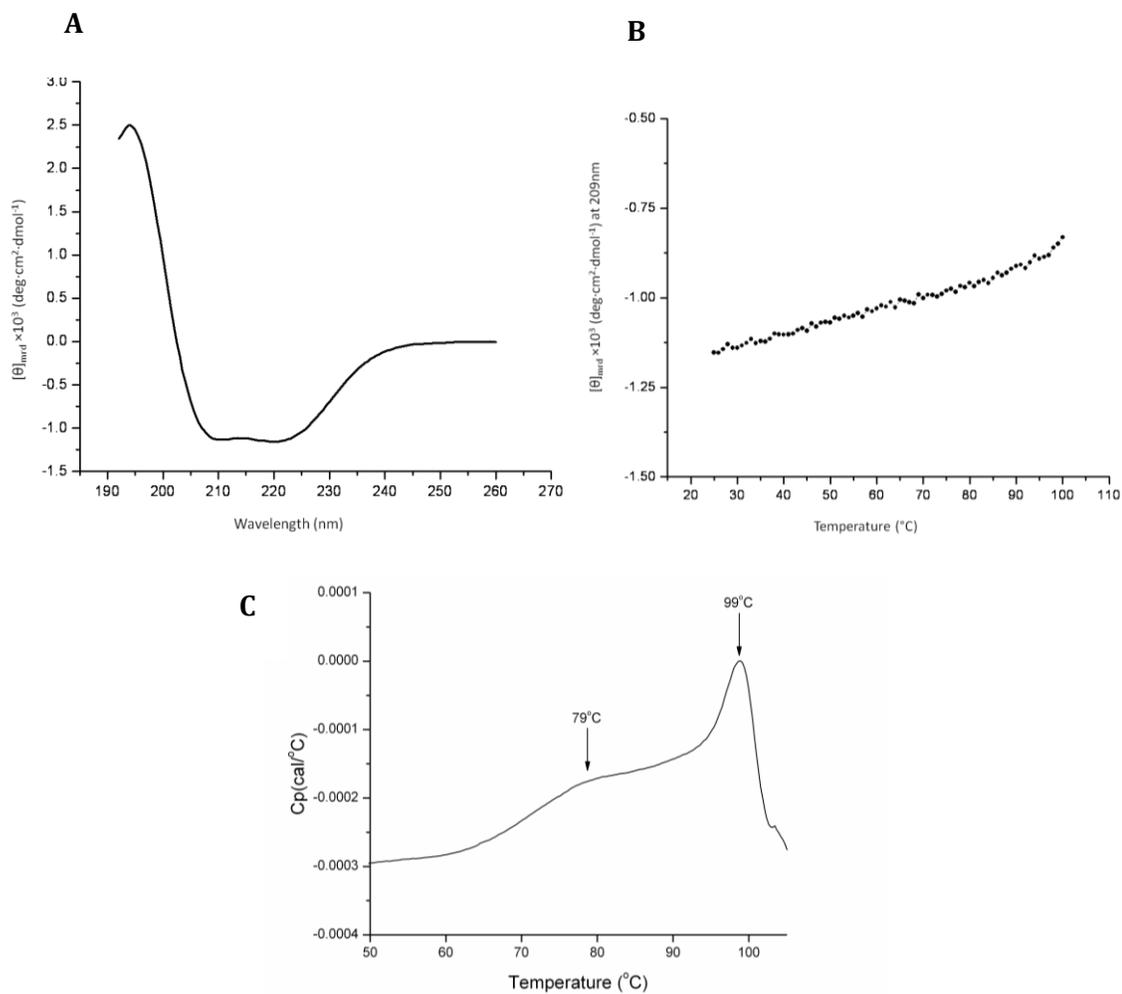


Figure 2.6 Secondary structure and thermostability of *TmTMPK*. (A) Far-UV CD spectrum of *TmTMPK* at room temperature. Negative ellipticities occur at 209 and 223 nm, indicating high helical content. (B) CD spectrum of *TmTMPK* as a function of temperature. Minimum change of ellipticity at 209 nm is observed with increasing temperature. (C) Differential scanning calorimetry thermogram of *TmTMPK*.

2.2.4 *Tm*TMPK *in vitro* catalytic performance

To examine the activity and substrate specificity of *Tm*TMPK, a coupled enzyme assay measuring steady-state kinetics was employed. This method has been applied to numerous TKs and several TMPKs for characterization of catalytic efficiency (Brundiens et al., 1999a; Lavie et al., 1998a; Lavie et al., 1998b). A drawback of this method for TMPK is that TDP, the product of the reaction catalyzed by TMPK, can also be utilized by the coupled pyruvate kinase which leads to an intrinsic background reading. To measure the absolute TMPK activity, a HPLC assay was also developed (chapter 5). However, the K_M of *Tm*TMPK is below the detection limit of the HPLC method ($K_M = 50 \mu\text{M}$). Therefore, the coupled enzyme assay was used for these studies.

First, *Tm*TMPK activity towards natural phosphate acceptors (TMP, dAMP, dCMP, dGMP) was tested. The catalytic efficiency (k_{cat}/K_M) for TMP is similar to *Ec*TMPK, both the K_M and k_{cat} value are 5 fold lower than *Ec*TMPK (*Ec*TMPK $K_M = 15 \mu\text{M}$; $k_{\text{cat}} = 19.8 \text{ s}^{-1}$). Nevertheless, *Tm*TMPK is a better catalyst due to its thermal stability compared to *Ec*TMPK, whose activity drops 40% in 2 hours upon incubation on ice. The same incubation has no effect on the activity of *Tm*TMPK.

Results for all four natural phosphate acceptors showed that *Tm*TMPK is highly specific for its phosphate acceptor, as it only phosphorylates TMP among the four nucleoside monophosphates. The results are not surprising, since separate nucleoside monophosphate kinases for all four nucleosides were found in the *Thermotoga maritima* genome (Zacherl et al., 1997). It is possible that there is no overlapping substrate specificity between nucleoside monophosphate kinases.

Table 2.1 *Tm*TMPK kinetic parameters at 37°C.

	Substrate	K_M (μM)	k_{cat} (S^{-1})	k_{cat}/K_M ($\text{M}^{-1} \text{s}^{-1}$)
P-acceptors	TMP	2.2±0.3	2.4±0.08	1.1×10 ⁶
	dCMP	n.d.	n.d.	n.d.
	dAMP	>10 ³	<0.10	<10 ²
	dGMP	n.d.	n.d.	n.d.
P-donors	ATP	6.4±0.7	0.85±0.02	1.3×10 ⁵
	GTP	270±20	0.15±0.003	5.6×10 ²
	UTP	15±4	0.039±0.003	2.6×10 ³
	CTP	484±70	0.13±0.007	2.7×10 ²

Values represent the average of three experiments.

n.d. = not detectable.

Next, the specificity towards natural phosphate donors was investigated. *Tm*TMPK has broader substrate specificity for nucleoside triphosphates. It exhibits the highest catalytic efficiency (k_{cat}/K_M) with ATP, the preferred phosphate donor. CTP, GTP and UTP can also serve as phosphate donors although with much lower activity. The order of efficacy is: ATP>UTP>GTP>CTP.

2.2.5 Isothermal Titration Calorimetry (ITC) of *Tm*TMPK

To characterize substrate recognition and probe the reaction mechanism, ITC studies of *Tm*TMPK were performed. In order to prevent enzymatic turnover during the ITC measurement, EDTA was added to chelate Mg^{2+} and a non-hydrolyzable substrate analog, Adenosine 5'- γ -thiotriphosphate (ATP- γ S), was used instead of ATP. Representative titrations were shown in Figure 2.7.

First, *Tm*TMPK was titrated with TMP or ATP- γ S individually. Both titrations were characterized by a significant exothermic heat change, with ΔH of -16.8 kcal/mol for TMP and -11.0 kcal/mol for ATP- γ S. Calculated from ITC data fitting software, the dissociation constant K_d of TMP and ATP- γ S are also similar to each other (Table 2.2). Next, to investigate the order of binding for *Tm*TMPK, the enzyme was pre-incubated with one substrate and titrated with the other substrate. However, these experiments did not show typical heat changes of binding. The result is contrary to previous ITC studies of TMPK from *Streptococcus pneumoniae*, which showed that TMPK adopts an ordered binding pathway in which ATP binds to the enzyme first followed by TMP binding (Petit and Koretke, 2002).

Since Mg^{2+} is required for the phosphorylation reaction catalyzed by TMPK and it could influence the substrate binding by either direct participation in the bonding network of the binding pocket or by an indirect structural rearrangement, all measurements were repeated in the presence of 10 mM $MgCl_2$. In the presence of Mg^{2+} , titration of TMP shows similar thermodynamic parameters. In contrast, titration of ATP- γ S does not show any binding events. Titration of ATP- γ S to *Tm*TMPK pre-incubated with TMP shows a weak binding curve. However, the curve deviates from the sigmoid function; therefore, it cannot be fitted in the software to generate the thermodynamic parameters. To achieve a good binding curve in this experiment requires high enzyme concentration, at which protein aggregation is unavoidable. Other techniques measuring binding affinity, such as surface plasmon resonance (SPR), could be helpful to measure the weaker binding in this scenario. Overall, the ITC experiment suggests that Mg^{2+} plays an important role in substrate binding in *Tm*TMPK. It may help the enzyme adopt the appropriate structure to recognize the substrate. In the presence of Mg^{2+} , *Tm*TMPK displays an ordered binding in which TMP binds to the active site first and ATP second.

Table 2.2 Thermodynamics for the Association of nucleotides with *Tm*TMPK

		K_d (μ M)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)
No Mg^{2+}	TMP	4.5 ± 0.3	-16.8 ± 1.7	-8.4	-8.5
	ATP	7.3 ± 0.1	-11.0 ± 2.2	-2.5	-8.1
	E·TMP + ATP			No binding	
with Mg^{2+}	TMP	3.3 ± 0.2	-10.2 ± 0.9	-1.5	-8.7
	ATP			No binding	
	E·TMP + ATP			Weak binding	

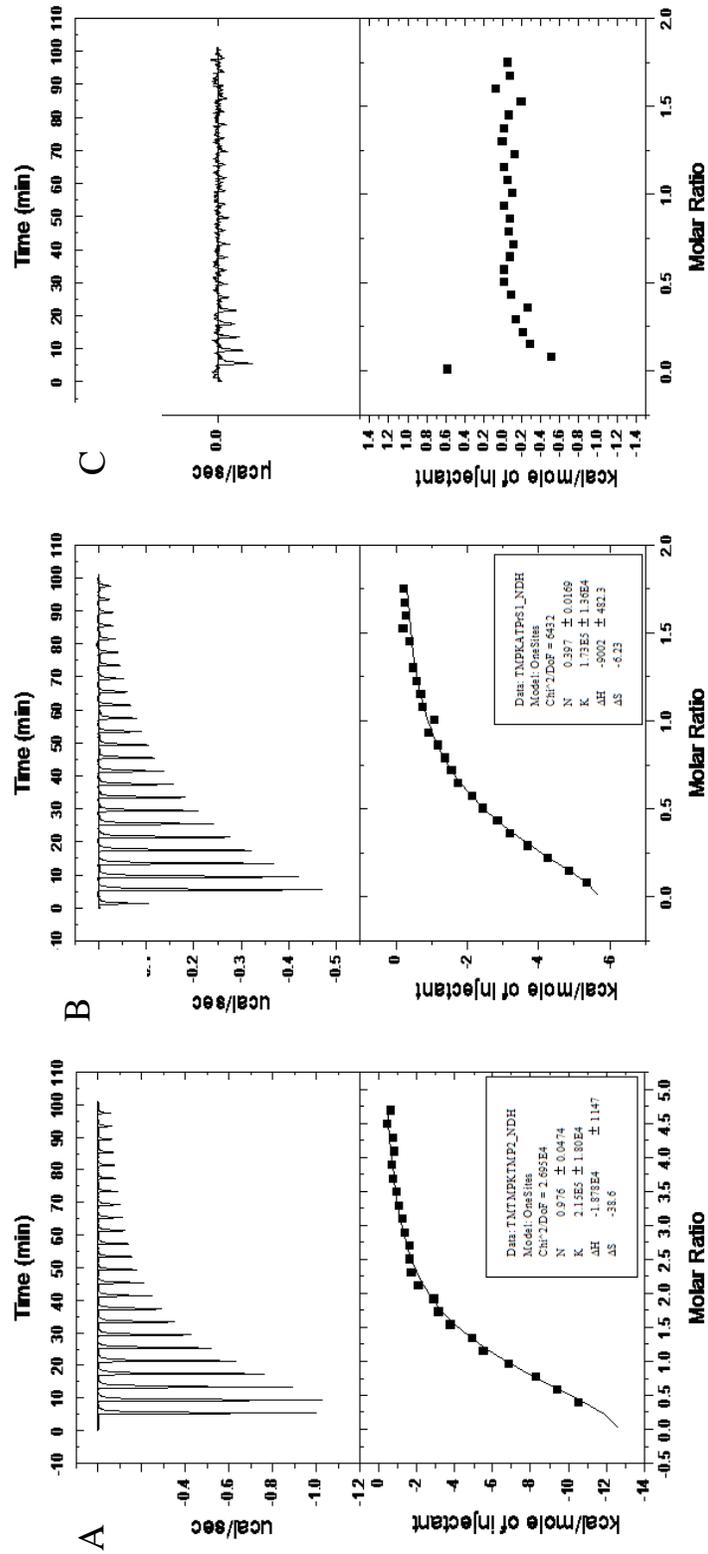


Figure 2.7 ITC of *TmTMPK* in the absence of Mg^{2+} . Titrations were performed at 25 °C in 50 mM Tris-HCl, 4 mM EDTA, pH 8.0. 25 μ M *TmTMPK* was titrated with 1 mM TMP (A) or 1.0 mM ATP- γ S (B). *TmTMPK* (25 μ M) in the presence of 1 mM TMP was titrated with 1 mM ATP- γ S (C). The upper panels are untreated data shown as differential power signals (μ cal /s). The lower panels are binding isotherms.

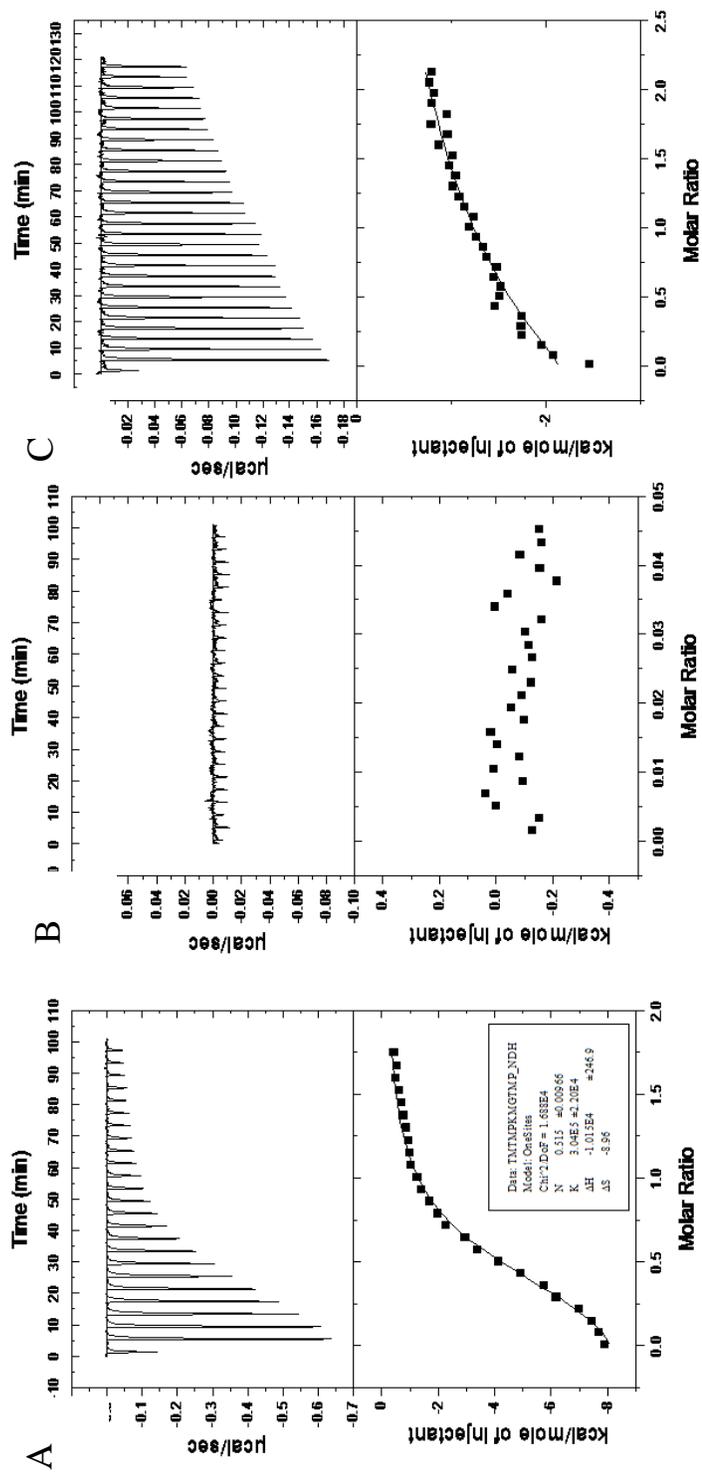


Figure 2.8 ITC of *Tm*TMPK in the presence of Mg^{2+} . Titrations were performed at 25 °C in 50 mM Hepes, 10 mM $MgCl_2$, pH 8.0. 25 μM *Tm*TMPK was titrated with 1 mM TMP (A) or 1.0 mM ATP- γ S (B). *Tm*TMPK (25 μM) in the presence of 1 mM TMP was titrated with 1 mM ATP- γ S (C). The upper panels are untreated data shown as differential power signals ($\mu\text{cal/s}$). The lower panels are binding isotherms.

2.3 Conclusion remarks

In summary, we have isolated the *Tm*TMPK gene and demonstrated that it encodes an active enzyme by expression of the enzyme in *E. coli* and characterization of its catalytic performance *in vitro*. Its high thermostability was proved by both far-UV CD and DSC studies. The melting temperature of 99°C makes it the most thermostable TMPK thus far. With its high catalytic activity, *Tm*TMPK presents one of the best candidates for protein engineering.

Several interesting behaviors were observed in this enzyme. First, in the DSC study, the broad peak at 79°C indicates a likely conformational change. A previous study on the thymidine kinase from *Thermotoga maritima* (*Tm*TK) revealed a conformational change at 70°C, which alters its substrate specificity from specific to promiscuous. Isolated from the same organism, *Tm*TMPK could also display a temperature switch for substrate specificity. Therefore, temperature dependent kinetic studies will be valuable to determine whether this structural reorganization correlates with the catalytic performance. In addition, earlier studies of *Ec*TMPK have shown a superior catalytic efficiency for the nucleoside analog AZT monophosphate compared to human and yeast TMPKs. With the highest sequence homology to *Ec*TMPK, *Tm*TMPK has the potential to be capable of efficiently phosphorylating nucleoside analogs. Kinetic analysis of *Tm*TMPK with nucleoside analogs will not only enable us to evaluate the potential of *Tm*TMPK's application in gene therapy but will also enhance our understanding of the structure-function relationship in this enzyme.

2.4 Materials and methods

2.4.1 Materials

Enzymes were from New England Biolabs (Beverly, MA) unless otherwise indicated. *Pfu* Turbo DNA polymerase (Stratagene, La Jolla, CA) was used for cloning. Pyruvate kinase and lactate dehydrogenase were from Roche Applied Science (Indianapolis, IN). All other reagents were from Sigma-Aldrich (St. Louis, MO) unless noted.

2.4.2 Isolation of *Tm*TMPK gene

The putative *Tm*TMPK gene was isolated by amplifying from *T.maritima* genomic DNA (ACTT#43589D) with primers including flanking restriction sites (5' *Nde*I and 3' *Spe*I). The primers used were *Tm*TMPK_for (5'-CGCCCATATGTTCAT AACGTTCTGA GGG-3') and *Tm*TMPK_rev (5'-CGCACTAGTCTAAACATCGAG TTTCC-3') (restriction site underlined). The PCR product was digested with *Nde*I and *Spe*I and subcloned into pET14b (Novagen, Madison, WI). The resulting construct was confirmed by DNA sequencing, and designated pET14b-*Tm*TMPK.

2.4.3 Overexpression and purification of *Tm*TMPK

E. coli BL21 (DE3) pLysS cells (Novagen, Madison, WI) were transformed with pET14b-*Tm*TMPK and grown at 37°C in LB broth containing carbenicillin (100 µg/ml) and chloramphenicol (34 µg/ml). At OD₆₀₀ ~ 0.5, isopropyl-b-D-thiogalactopyranoside (IPTG) was added to 0.4 mM (final concentration) to induce protein expression. The cells

were continually grown for 4 hours and then harvested by centrifugation at 3000g at 4°C. The cell pellets were stored at -20°C.

For protein purification, cell pellets were resuspended in the lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0) supplied with protease inhibitor cocktail, rLysozyme and Benzonase (Novagen, Madison, WI) according to the manufacturers' protocols. Cells were then lysed by sonication on ice (6 x 10 sec bursts at 15 W), and the soluble cell lysate was separated by centrifugation (10,000g, 4°C, 10 min) before loaded onto Ni-NTA resin (QIAGEN, Valencia, CA). Lysate with 5–10 mg of protein was incubated with 1 ml of resin at 4°C for 1 h, and the resin was separated by centrifugation and loaded to a polypropylene column (Bio-Rad, Hercules, CA). After three wash steps with 5 column volumes of lysis buffer containing 2 mM MgCl₂ and 10 mM ATP, 10 mM and 15 mM imidazole, the target proteins were eluted with lysis buffer containing 150 mM imidazole. Fractions were analyzed using 12% SDS-PAGE and samples containing target protein were pooled and buffer exchanged via ultra-filtration (Amicon Ultra-15, MWCO: 10 kDa, 4000g at 4 °C) into storage buffer (50 mM potassium phosphate, 150 mM NaCl, 5 mM MgCl₂, pH 8.0). Protein concentrations were quantified by measuring A₂₈₀ (*Tm*TMPK, $\epsilon = 14,650 \text{ M}^{-1} \cdot \text{cm}^{-1}$, calculated according to Gill and von Hippel (S.C. Gill, 1989)). Aliquots were flash frozen and stored at -80°C.

2.4.4 Secondary structure characterization and thermal denaturation

Far-UV circular dichroism (CD) spectra were obtained using a J-810 spectropolarimeter (Jasco, Inc., Easton, MD). Spectra were recorded at 22°C from 260-

190 nm (0.5 nm increments) using a 0.1 mm pathlength cell, 60 nm/min scan rate, 4 s response time, and 2 nm bandwidth. Proteins were analyzed at 1 mg/ml concentration. Each scan was repeated three times and taken the average of the three. Spectra were corrected for buffer absorbance and converted to mean residue ellipticity ($[\theta]_{\text{mrw}}$). Thermal denaturation on far-UV CD was monitored from 20 to 100°C following the change in ellipticity at 209 nm with a temperature gradient of 1 °C/min. The data is the average of two independent experiments.

Differential Scanning Calorimetry (DSC) was measured with the VP-DSC from MicroCal Inc. (Northampton, MA). Experiments were performed in buffer containing Tris-HCl 50 mM, NaCl 150 mM at pH 8.0. Protein samples at concentration of 0.5 mg/ml were equilibrated for 30 min at 25 °C, heated to 110°C at a constant scan rate of 1 °C/min, and the data collected with a 16 s filter. Data was analyzed by Origin (Origin Lab, Northampton, MA).

2.4.5 Spectrophotometric enzyme activity assay

The kinase activity of recombinant enzymes was determined using a spectrophotometric coupled-enzyme assay (Munch-Petersen et al., 2000; Schelling et al., 2001). Briefly, nucleoside monophosphate and nucleoside triphosphate at 1 to 5000 μM were prepared in reaction buffer containing 50 mM Tris-HCl, 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, pH 8.0. Assays were performed at 37 °C, measuring the absorbance change at 340 nm in the presence of 0.5-1.0 μg of 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 triplicates, and kinetic data was determined by nonlinear regression analysis using the Michaelis–Menten equation in Origin7 (OriginLab, Northampton, MA).

2.4.6 Isothermal Titration Calorimetry experiments

ITC experiments were performed using a VP-ITC titration microcalorimeter from Microcal Inc.(Northampton, MA, U.S.A.). Experiments were performed at 25 °C in buffer A (50 mM Tris-HCl, 4 mM EDTA, pH 8.0) or buffer B (50 mM HEPES, 10 mM MgCl₂, pH 8.0). Solutions of the *Tm*TMPK were filled in the sample cell and titrated with substrate solution. A typical titration experiment consisted of a first control injection of 1 µl followed by 24 injections, each 5.5 µl of 15 s duration, with a 4 min interval in between. Raw data were collected and integrated using Origin (OriginLab, Northampton, MA). The data were fitted to a single-site binding model by a non-linear regression analysis to yield binding constants (K_b), enthalpy of binding (ΔH) and the stoichiometry of binding.

Chapter Three

Design and construction of a conditional *E. coli* auxotroph strain for functional selection of dual thymidine and thymidylate kinase activity

3.1 Introduction

Nucleoside analogs (NAs) have emerged as a widely used class of drugs in anti-cancer and anti-viral therapy. They enter target cells in a non-phosphorylated state (pro-drug) and require conversion to the triphosphate form by several cellular nucleoside and nucleotide kinases in the salvage pathway before they can target DNA polymerase or reverse transcriptase to halt DNA/RNA synthesis. For most NAs, this activation is the bottleneck step in their therapeutic application, which limits their overall therapeutic efficacy. The slow conversion not only leads to low effective concentrations of the active triphosphate form in the cell but also induces side effects, such as accumulation of cytotoxic intermediates (Lavie et al., 1997). Therefore, improving NAs activation presents a challenge for engineering nucleoside and nucleotide kinases.

Thymidine kinase (TK) and thymidylate kinase (TMPK) are key enzymes in the thymidine triphosphate (TTP) salvage pathway. Their biological role is to phosphorylate thymidine to thymidine monophosphate (TMP) and to phosphorylate TMP to thymidine diphosphate (TDP), consecutively. They are both required to phosphorylate several NAs but often prove inefficient. One example of prodrugs used in the clinic for which the conversion to the monophosphate is rate limiting is d4T (2'-3'-dideohydro-2'-3'-dideoxythymidine), which is phosphorylated by human TK 1 (Balzarini et al., 1989). On the other hand, the most prominent example that monophosphate kinase is critical in NAs activation is the case of 3'-azido-3'-deoxythymidine (AZT) (Arnon Lavie, 2004). The final NA activation step, phosphorylating to the triphosphate form, is usually more efficient, because a number of enzymes can catalyze this step, including nucleotide diphosphate kinase (NDK), pyruvate kinase and creatine kinase (Miller et al., 1992).

Since phosphorylating NAs to the triphosphate form is crucial to its successful application, previous research has focused on applying a variety of approaches to engineer nucleoside and nucleotide kinases including directed evolution (Gerth and Lutz, 2007; Lutz et al., 2009). Engineering nucleoside kinases has been relatively successful; however, few successful examples of engineering nucleoside monophosphate kinases were found. The reason behind the unsuccessful application lies in the lack of an efficient selection and/or screening method, which is critical for accurate assessment of targeted function in libraries of kinase variants.

A number of selection and screening strategies for TK function have been developed. For example, the *E. coli* auxotrophic strain KY895 carries a lethal mutation in chromosomal *tk*; it has been successfully used to select for mutations with wild type TK function (Hiraga et al., 1967). Under the selection condition, *E. coli* KY895 grows in the presence of 5-fluoro-2'-deoxyuridine (FdU) and uridine, the former will become phosphorylated and inhibit thymidylate synthase, shutting down the *de novo* pathway. At the same time, uridine inhibits thymidine phosphatase, which prevents dephosphorylation of FdU monophosphate (Figure 3.1). Cell survival solely depends on the salvage pathway, relying on intake of thymidine by the nucleoside transporter and its subsequent phosphorylation by exogenous TK expressed from a plasmid. In terms of NAs activity selection, cells harboring dNK mutants were replica-plated onto NA containing media and selected for negative growth (Christians et al., 1999). The drawback of this method is that it depends on cytotoxicity of NAs in *E. coli*, which can be caused by metabolism of NAs through other pathways not by TK. In addition, to selecting enzymes with orthogonal NA activity, a FACS based assay has been developed by our group. In this

assay, cells were fed fluorescent NAs, which are trapped inside the cell once phosphorylated (Liu et al., 2009).

Unlike TK's exclusive function in the salvage pathway, TMPK exists at the junction of the *de novo* and salvage pathways of TTP synthesis (Figure 3.1). As an essential gene, *tmpk* presents a special challenge in creating an *E. coli* strain auxotroph. While *tk* deficient strains can survive by utilizing the *de novo* pathway, mutations or deletions that destroy TMPK function are lethal to *E. coli*. In the literature, there are only two reported selection systems for nucleoside monophosphate kinases. One is a conditional guanylate kinase (GMPK) deficient *E. coli* strain that requires a functional, plasmid-born GMPK for growth under selection conditions (Stolworthy et al., 2003). The other selection system is the in-frame deletion of *E. coli tmpk*, in which a kanamycin resistant gene replaced *tmpk* in its natural operon (Chaperon, 2006). However, this system requires transformation with plasmid that expresses target TMPK before gene recombination. For each TMPK to be tested, several steps of recombination have to be repeated, which compromises its use for selection of directed evolution libraries.

In addition, selection or screening assays for engineering enzymes with dual TK and TMPK function have not yet been developed. Because of the success in engineering nucleoside kinases, enhancing the initial step of NAs phosphorylation could shift the rate-limiting step of NAs activation to the second step, the conversion of NA monophosphate to its diphosphate form. A single enzyme performing both TK and TMPK function will not only eliminate the bottleneck step but also minimize the accumulation of any cytotoxic intermediates. A novel *E. coli* strain that can be used to select TK and TMPK function will provide a powerful means to identify new and hopefully better dual

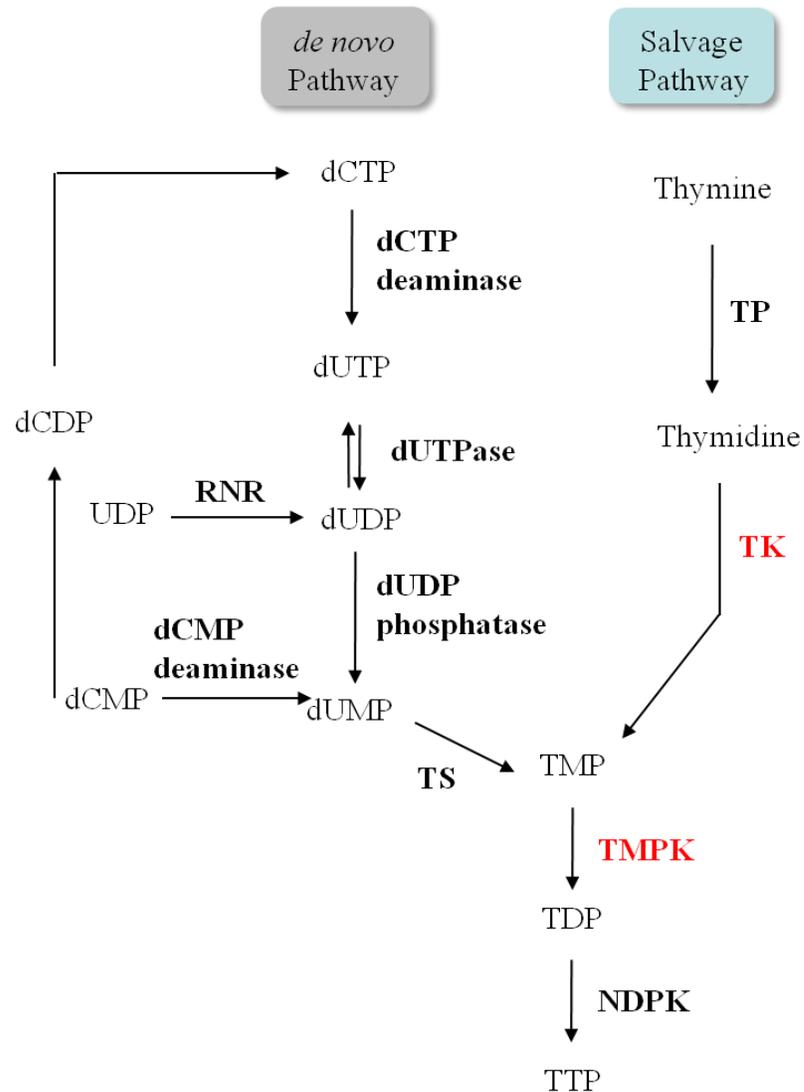


Figure 3.1 The *de novo* and salvage pathways for TTP synthesis. TP: thymidine phosphotase; TK: thymidine kinase; TMPK: thymidylate kinase; NDPK: nucleoside diphosphate kinase; RNR: ribonucleotide reductase; TS: thymidylate synthase.

function enzymes that can improve overall activation efficiency of NAs. In this chapter, a conditional auxotroph *E. coli* strain for dual TK and TMPK function selection was constructed. The strategy consists of three steps: first, knock-out of *E. coli* chromosomal *tk*; second, knock-in of *tmpk* from *Thermotoga maritima* (*TmTMPK*) under the control of a tunable arabinose promoter; third, knock-out of *E. coli* chromosomal *tmpk* with the expression of *TmTMPK* (Figure 3.2). The most significant advantage of this strain is that the expression of TMPK is tunable, which allows selection pressure for the second step reaction to be systematically adjusted in directed evolution experiments. We have evaluated the strain by transforming plasmids expressing *DmdNK* or *TmTMPK*, as well as a duet vector expressing both kinases simultaneously. The results demonstrated that: 1) the growth of the auxotroph strain depends on L-arabinose concentration; 2) its growth can be complemented by an exogenous kinase; 3) it can be used for selection of enzymes with dual TK and TMPK activities.

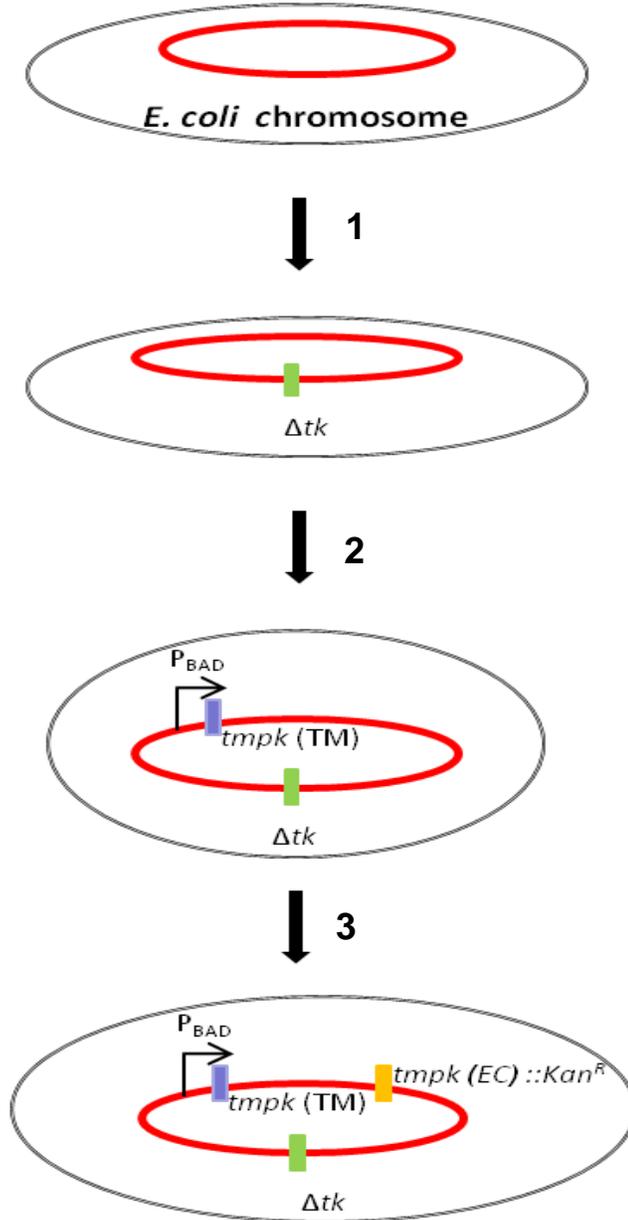


Figure 3.2 Three step strategy for constructing the conditional auxotroph. (1) knock-out of *tk* on *E. coli* chromosome; (2) knock-in of *Tmtmpk* under the control of a tunable arabinose promoter; (3) knock-out of *E. coli tmpk*.

3.2 Results and discussion

3.2.1 Knock-out of *tk* gene in *E. coli*

E. coli strain BW27783 was chosen as the parent strain in order to construct a strain with homogeneous TMPK expression under the arabinose inducible promoter P_{BAD} . In normal *E. coli*, genes under the control of P_{BAD} are expressed in an all-or-non fashion, caused by the high-capacity low-affinity L-arabinose transporter *araE* (Siegele and Hu, 1997). Changing the L-arabinose concentration causes a variation of the fraction of the cells that are fully induced, instead of the level of gene expression in individual cells. In strain BW27783, *araE* was placed under the control of an arabinose-independent, constitutive promoter, which allows homogeneous gene expression under P_{BAD} (Khlebnikov et al., 2001).

To knock out *tk* in BW27783, λ red recombination technology was applied (Figure 3.3) (Datsenko and Wanner, 2000). First, using the pKD4 plasmid as a template, the kanamycin resistant gene (*Kan^R*) flanked by FRT sequence was amplified with primers introducing homologous regions upstream and downstream of *tk*. The linear DNA was transformed into the host strain carrying the λ red expression plasmid pKD46 to facilitate homologous recombination. Strains with correct recombination were isolated by growing colonies on media containing kanamycin. To verify the correct recombination sequence at the targeted chromosomal location, PCR with primers designed at the upstream and downstream genes of *tk* was carried out (Figure 3.4A). Because the linear DNA integrated into the *E. coli* chromosome is 900bp larger than *tk*, a PCR product at higher molecular weight (3.1kb) indicates correct gene recombination. Since kanamycin resistance will be used as the selection marker in the third step, the selected strain was

transformed with the FLP recombinase plasmid pCP20 to eliminate *Kan^R* at the FRT site (Datsenko and Wanner, 2000). The upstream and downstream primers were used again to verify the correct chromosome structure. A PCR product at a lower molecular weight (1.6kb) corresponds to the short scar sequence left at the location after the FRT site deletion (Figure 3.4A). To confirm that the strain has lost kanamycin resistance, three colonies were streaked on LB plates and then on kanamycin containing media (Figure 3.4B). As expected, all three colonies grew on LB plates but not on kanamycin containing media. Next, to confirm that the *tk* deletion strain has lost TK activity, both the strain and strain harboring plasmid expressing TK (pDIM-*DmdNK*) were streaked onto TK functional selection media (Black and Loeb, 1993). The *tk* deletion strain alone does not survive TK functional selection, but the strain harboring pDIM-*DmdNK* is able to grow (Figure 3.4C). The BW27783 strain with *tk* deletion strain is denoted as Δtk .

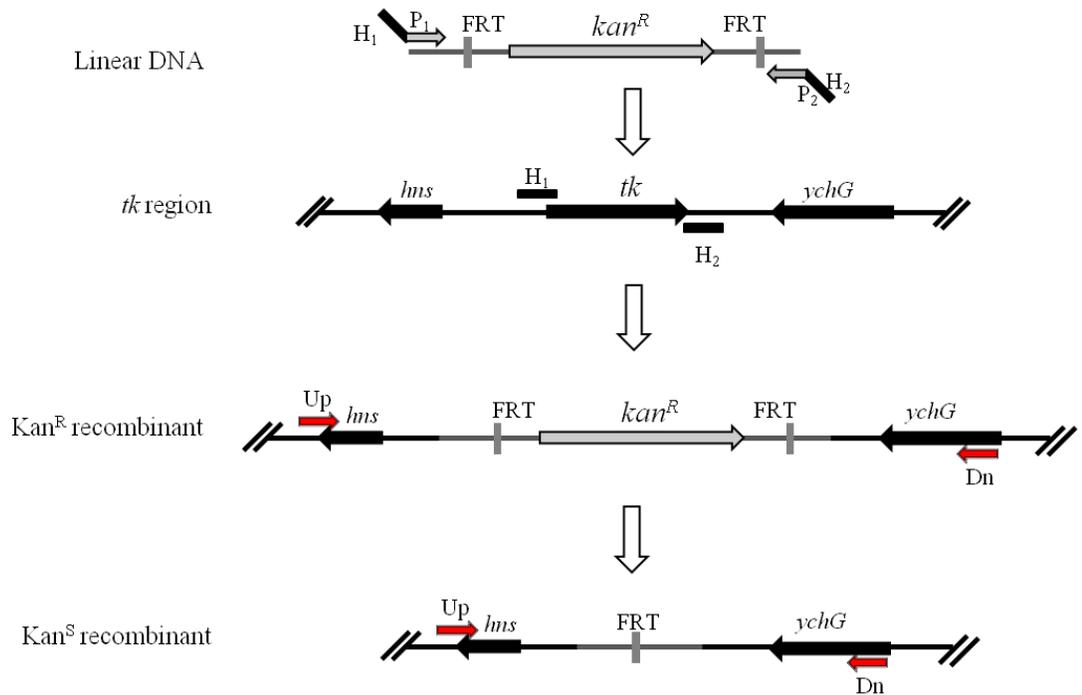


Figure 3.3 Construction of *tk* deletion mutants. Gene knockout primers have 20-nt 3' ends for priming upstream (P₁) and downstream (P₂) of the FRT sites flanking *Kan^R* in pKD4 (Datsenko and Wanner, 2000) and 50-nt 5' ends homologous to upstream (H₁) and downstream (H₂) chromosomal sequences for *tk* deletion. Recombination created between the resistance cassette and neighboring upstream (*hms*) and downstream (*ychG*) sequences were verified by PCR with primers upstream and downstream of *tk* (Up and Dn). Structures created after excision of the resistance gene are also verified by PCR with Up and Dn primers and by DNA sequencing.

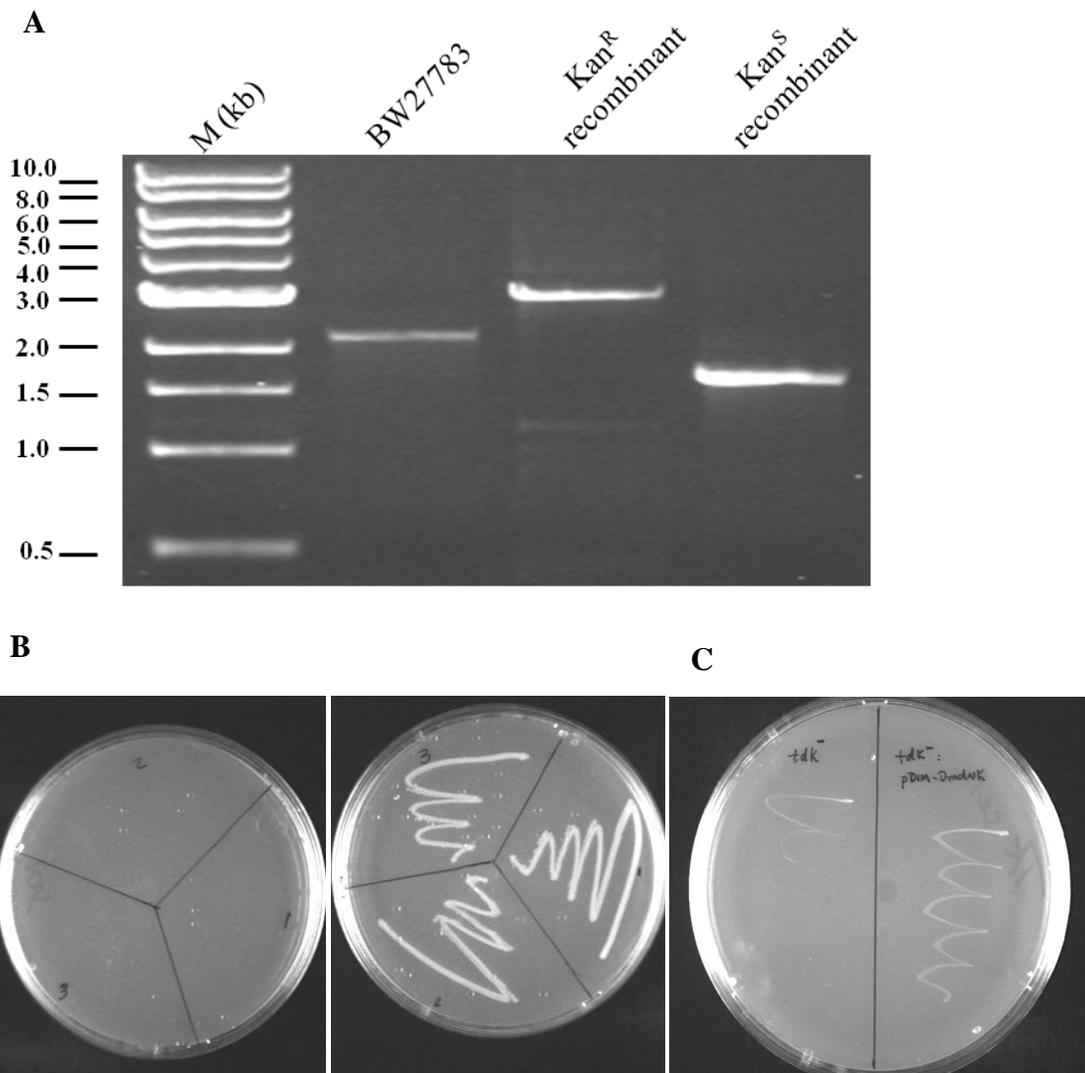


Figure 3.4 Validation of *tk* knock-out. (A) PCR product with Up and Dn primers. Lane 1: BW27783 (expected size 2.2kb). Lane2: Kan^R recombinant (expected size 3.1kb). Lane 3: kanamycin sensitive (Kan^S) recombinant (expected size 1.6kb) (B) Strains streaked on kanamycin and LB plates to test for loss of resistance. Left: three colonies streaked on kanamycin plates. Right: the same three colonies on LB plates. (C) TK functional selection plate streaked with left: Δtk and right $\Delta tk:pDIM-DmdNK$.

3.2.2 Knock-in *Tmtmpk* under the P_{BAD} promoter

To integrate *Tmtmpk* under a P_{BAD} promoter, a knock-in plasmid pRDC15 was used (Arigoni et al., 1998). It allows chromosomal replacement of the *araBAD* gene cluster by a gene of interest thereby placing the target gene under the control of the arabinose promoter. *Tmtmpk* with an N-terminal his-tag was cloned into a pRDC15 plasmid and transformed into the BW27783 Δtk strain (Figure 3.5). The cells were plated onto chloramphenicol (Cm) containing media and incubated at 43°C, the nonpermissive temperature of the plasmid replicon, to select for integration by allowing replication of plasmid sequences by the chromosomal origin. A small amount of the same transformants was incubated at 30°C as a dilution control. The numbers of colonies on both plates were used to calculate the integration efficiency, which is 3.6×10^{-2} integrants/CFU. Four integrants were picked from the 43°C plate, suspended in 1 ml LB broth, plated onto 5% sucrose media and incubated at 30°C. When shifted to the permissive temperature, the plasmid will be excised from the chromosome. The counterselectable *sacB* marker was used to select for the loss of plasmid sequences. Because the media is supplemented with sucrose, expression of *sacB* is lethal to *E. coli*. The excision frequency was 0.1 excisants/CFU. Finally, sucrose resistant colonies were replica-plated onto sucrose/Cm media to identify any CFU that grew as a result of a mutation in *sacB* rather than loss of the pRDC15:His-*TmTMPK* plasmid. 76 out of 80 colonies lost the excised plasmid. PCR of 24 qualified colonies identified 4 candidates with a correctly sized PCR product. DNA sequencing also confirmed the correct sequence.

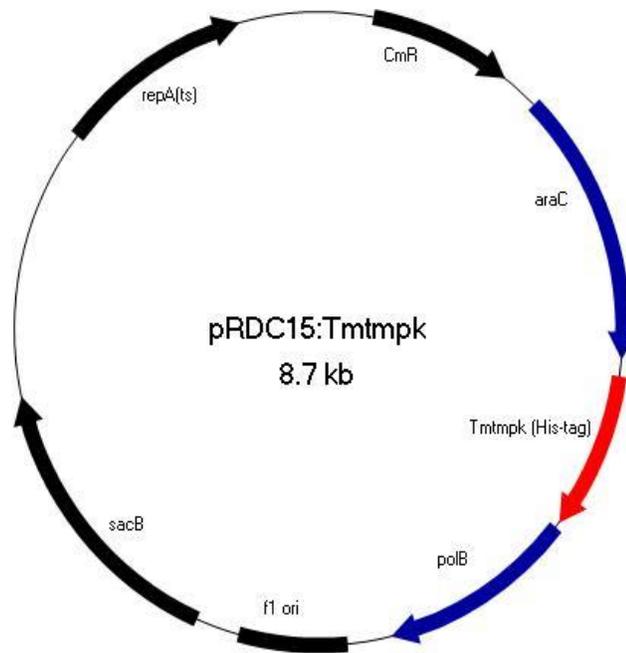


Figure 3.5 Knock-in vector pRDC15: His-*Tm*TMPK. *repA*(Ts): temperature sensitive replication origin; it has a permissive temperature of 30°C but is inactive 43°C. *CmR*: chloramphenicol resistance used as selection marker. *araC*: arabinose gene activator, located upstream of *araBAD* on the *E.coli* chromosome. *polB*: DNA polymerase II, located downstream of *araBAD*. *sacB*: levansucrase from *Bacillus subtilis*, an enzyme that catalyzes the hydrolysis of sucrose and levan elongation.

To verify the phenotype of the knock-in strain and confirm the tunable expression under the arabinose promoter in this modified strain, different arabinose concentrations were used to induce *Tm*TMPK expression from the chromosome. The N-terminal his-tag integrated with *Tm*TMPK allows easy purification with a Nickel-NTA resin. Both total cell lysate and purified protein were analyzed by SDS-PAGE (Figure 3.6). With increasing arabinose concentration, enhanced intensity of the band corresponding to *Tm*TMPK was observed in both the total protein and the purified protein SDS-PAGE. Although the last three lanes of the total protein seem to show similar amount of *Tm*TMPK in the sample, the increasing trend is more obvious on the SDS-PAGE of purified protein. In the un-induced sample, there is no band corresponding to *Tm*TMPK, which indicates no detectable leaky expression. For the five tested arabinose concentrations, the amount of purified *Tm*TMPK increases upon increasing arabinose concentration, which reflects the tunable expression level under the P_{BAD} promoter.

The major drawback of this assay is the lack of a loading control. The amount of protein loaded on each gel could be different than *in vivo* concentrations. Also, during the many steps of purification, the overall volume of solution could change even with careful execution. However, for the purpose of assessing the tunable expression by the arabinose promoter *in vitro*, the results of this assay indicates the homologous expression of *Tm*TMPK from the chromosome. No leaky expression was observed in the absence of arabinose, which is important for TMPK functional selection. The activity of the purified protein was also tested by a coupled enzyme assay at a substrate concentration of 100 μ M TMP (Schelling et al., 2001). The activity is similar to the enzyme purified from pET14b in BL21(DE3) cells.

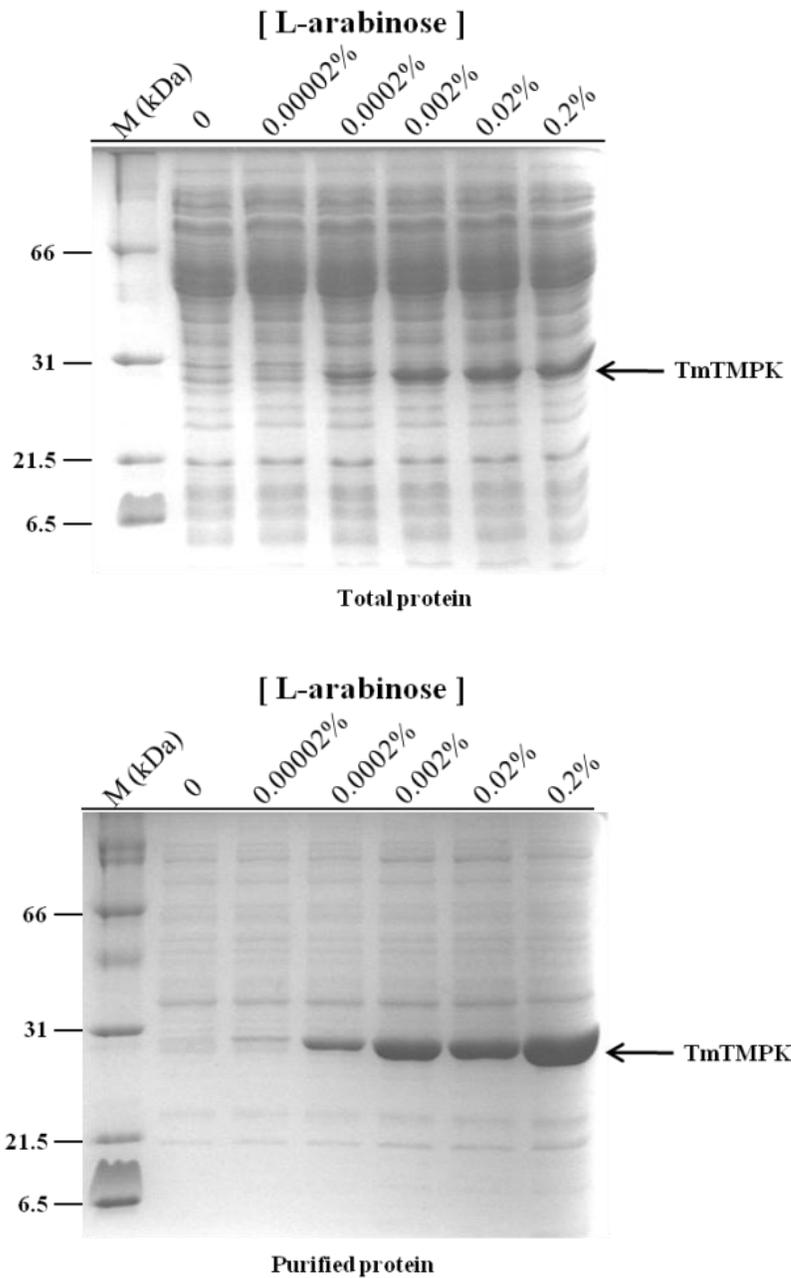


Figure 3.6 Tunable expression of chromosomal *TmTMPK*. The protein was expressed under the control of an arabinose promoter at different arabinose concentrations. Upper: SDS-PAGE analysis of total cell lysate. Lower: SDS-PAGE analysis of purified *TmTMPK*.

3.2.3 Knock-out *Ectmpk*

The *E. coli tmpk* gene is located at 24.9 min in a putative five-gene operon on the chromosome. It is the center one among the five genes, and its N- and C- termini overlap with the adjacent genes (in different open reading frames) (Figure 3.7 A and B). Because the five genes are under the same promoter and the adjacent genes are also essential, it is important to keep the overall reading frames intact when deleting *tmpk*. On the knock-out plasmid, a kanamycin resistant gene was sandwiched between *yceG* and *holB*, the adjacent genes of *tmpk*, to facilitate homologous recombination (Figure 3.7C). All the recombination experiments were performed in the presence of arabinose, which induces constant expression of *TmTMPK* inside the cell to ensure an adequate supply of TDP. The plasmid pKOV-KanR was transformed into the knock-in strain and incubated at 43°C on Cm/Kan/Ara (arabinose) media to select for integration. The observed integration efficiency was 5.0×10^{-5} integrants/CFU. Three integrant colonies were diluted and plated onto Kan/Sucrose/Ara to induce excision with the frequency of 0.25 excisants/CFU. 40 excisants were streaked onto both Kan/Sucrose/Ara and Cm/Sucrose/Ara plates to select for Cm sensitive strains. Seven strains survived the selection and their correct assembly was confirmed by DNA sequencing. One of them was selected for the following experiments and named strain YL-1.

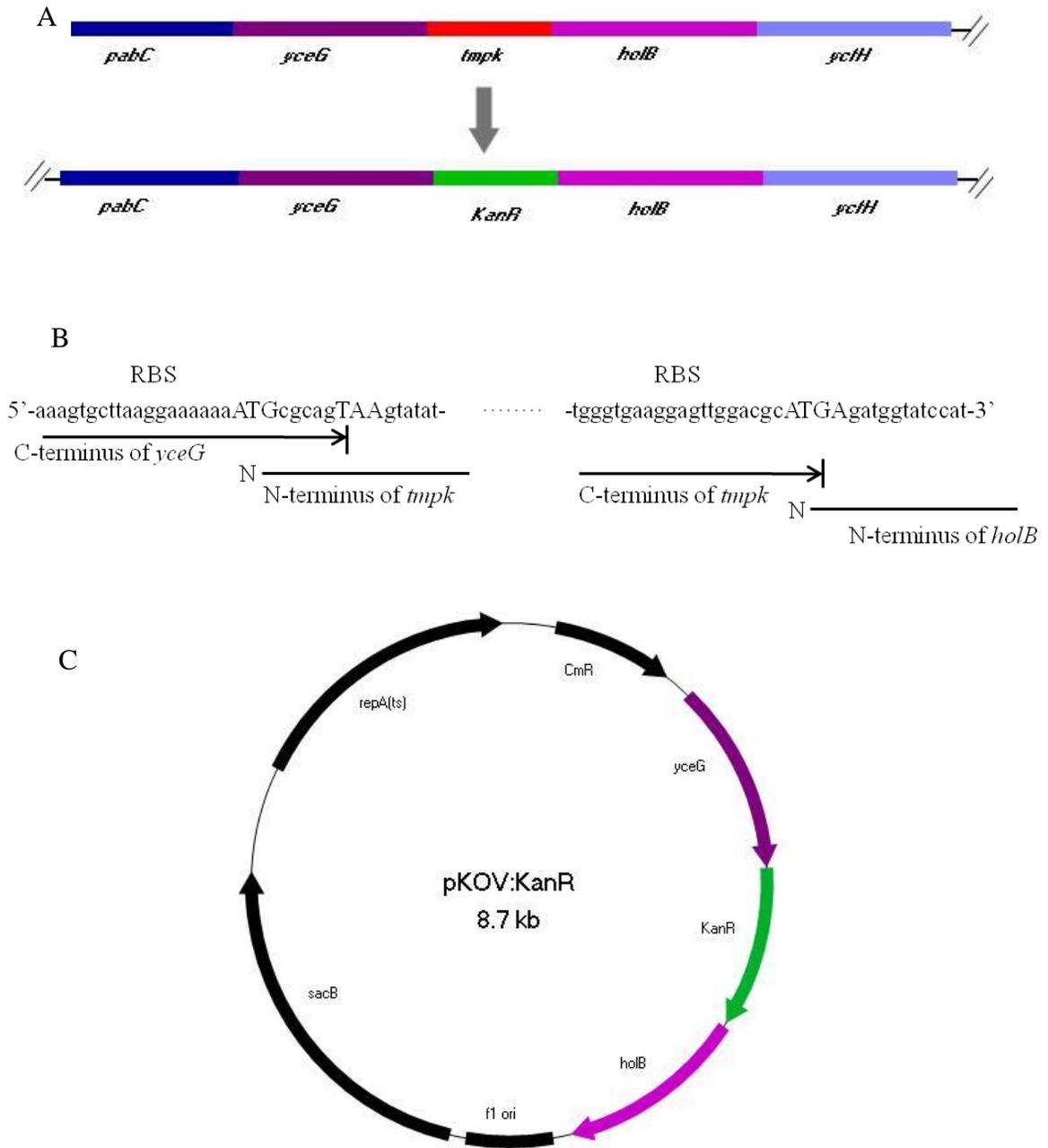


Figure 3.7 *Ectmpk* knock-out strategy. (A) Gene structure of the *tmpk* operon and the resulting gene structure after replacement with *Kan^R*. (B) Overlapping sequence at the N- and C- termini of the *tmpk* sequence. (C) Knock-out vector pKOV-*KanR*. *Kan^R* was sandwiched between *yceG* and *holB*, which are the adjacent genes of *tmpk*. The remaining features of this plasmid are similar to pRDC15.

3.2.4 Validation of strain YL-1.

3.2.4.1 YL-1 validation and complementation by a plasmid-born kinase

First, to verify the strain's growth dependence on arabinose, it was streaked onto LB, LB/Kan and LB/Kan/Ara plates (Figure 3.8A). Strain YL-1 survives kanamycin selection but only in the presence of arabinose. The results confirmed that: 1) the recombination occurs as expected since the strain possesses kanamycin resistance, 2) *tmpk* is an essential gene, 3) growth of the strain requires arabinose-induced *TmTMPK* expression.

Next, YL-1 was tested to determine whether a plasmid-born TMPK can complement growth. The plasmid pDIM-*TmTMPK* was transformed into the YL-1 strain. Its leaky lac promoter allows constitutive expression of the target gene in *E. coli* (Ostermeier et al., 1999). As a negative control, pDIM-GPX containing a non-kinase gene was used as a negative control to test whether transformation of the plasmid causes any phenotypical change. YL-1 strains harboring both plasmids were streaked onto Kan/Ara/Carb and Kan/Carb (Figure 3.8 B). Without arabinose, only strains harboring pDIM-*TmTMPK* survived the selection condition. The negative control pDIM-GPX did not show any growth. Both transformants grew when arabinose was supplied. The results confirm the predicted phenotype of the strain: growth of the YL-1 depends on arabinose and the dependence can be substituted with a plasmid-born TMPK.

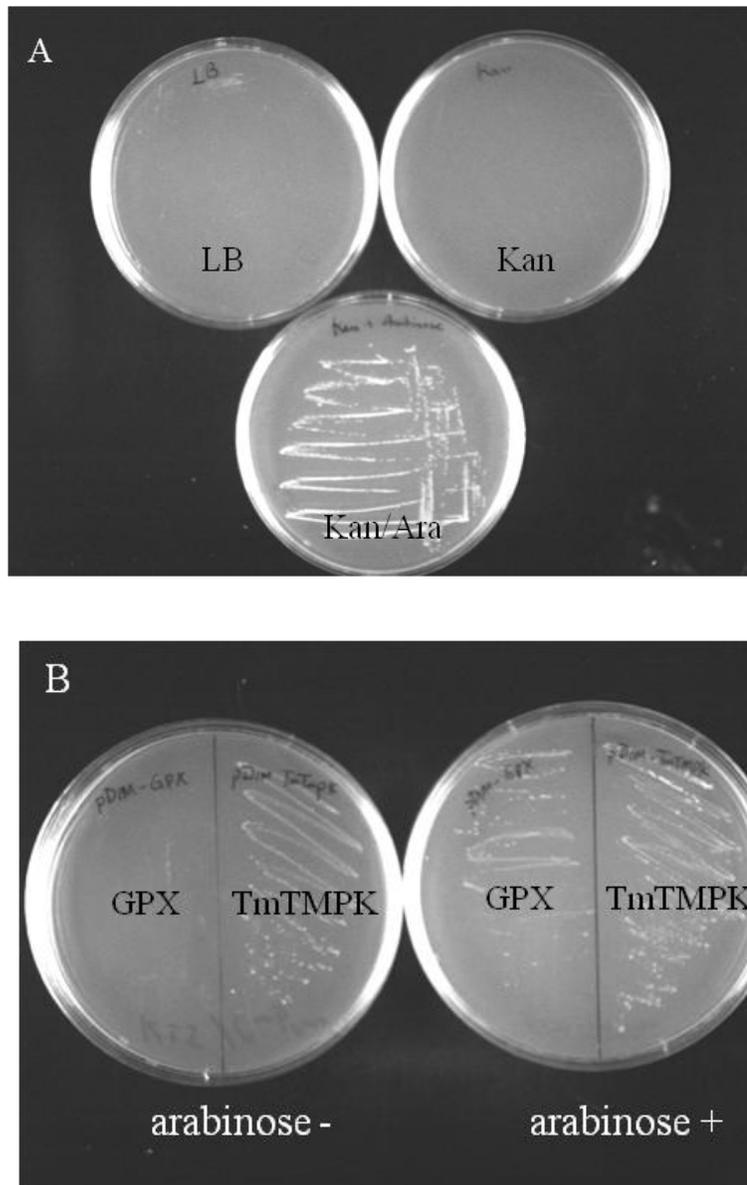


Figure 3.8 Validation of *Ectmpk* knock-out. (A) Strain YL-1 streaked on LB, Kan and Kan/Ara media. The cells show kanamycin resistance and only grow when arabinose is present (B) Straining strain YL-1 harboring pDIM-*TmTMPK* and pDIM-GPX on Kan/Carb (left) and Kan/Carb/Ara (right) media. In the absence of arabinose, only YL-1 harboring pDIM-*TmTMPK* grows; with arabinose both strains grow.

3.2.4.2 YL-1 growth dependence on arabinose

To test whether strain growth responds to arabinose in a dose-dependent pattern, a spot assay was applied to quantify cell growth. A range of different densities of cells were spotted onto separate Kan/Ara plates, each with different arabinose concentrations (Figure 3.9A). For a majority of cell densities (from 10^3 to 10^8 cfu/ml), the plotting spot integrated density against arabinose concentration consistently displays an increasing trend (Figure 3.9B). At higher cell concentrations (10^8 and 10^7 cfu/ml), cell growth was observed even when no arabinose was present and glucose, which inhibits arabinose induction, was added to LB media (Schleif, 2000). This false positive growth is probably due to cells scavenging of each other. When a small number of cells were spotted (10^2 and 10^3 cfu/ml), the curve slopes only slightly, making it difficult to assess the response to lower arabinose concentrations. Comparing this series of cell densities, optimal cell densities (10^4 and 10^5 cfu/ml) were selected for further study.

3.2.4.3 YL-1 growth dependence on arabinose with different kinase activity

Because plasmid-born TMPK can complement the growth dependence of arabinose, the growth dependence of YL-1 harboring kinases with different catalytic efficiency was tested (*Tm*TMPK has higher TMPK catalytic efficiency than HSV1-TK) and show different growth dependences on arabinose. YL-1 cells harboring pDIM-*Tm*TMPK, pDIM-HSV1TK and pDIM-GPX were spotted on Kan/Carb/Ara plates supplied with various arabinose concentrations. (Figure 3.10A and B).

For *Tm*TMPK, the strain shows almost constant growth independent of the arabinose concentration, which indicates the plasmid-born *Tm*TMPK complemented the

TMPK function almost completely. The strain no longer depends on the arabinose-induced, chromosomal copy of TMPK. YL-1 expressing HSV1-TK shows an increase in growth when the arabinose concentration increases. The difference in catalytic efficiency in these two enzymes clearly causes the different growth response to arabinose. Interestingly, the strain harboring GPX grew less than YL-1 itself, which is probably due to the extra burden the cell has to carry to express a non-functional protein. The tunable expression of TMPK under arabinose regulation provides a useful tool for finely controlling the selection pressure for directed evolution experiments.

3.2.4.4 Dual TK and TMPK functional selection using YL-1

The last validation experiment is to confirm that YL-1 can be used to select for dual TK and TMPK activity. The YL-1 strain was transformed with pET14b-*DmdNK*, pET14b-*TmTMPK* and pETduet-*DmdNK-TmTMPK* (a vector that expresses two target genes simultaneously). The dual TK and TMPK functional selection uses the TK selection condition in the absence of arabinose. Cells will be viable only if both TK and TMPK enzymes are expressed and functional inside the cell. When transformants containing any of the three plasmids were streaked onto Kan/Carb/Ara, cells were viable as expected (Figure 3.11 A). However, when plated on the dual functional selection plate, only pETduet-*DmdNK-TmTMPK* was able to complement the TK and TMPK deficiency. No growth was observed in the other cells where only one kinase was expressed.

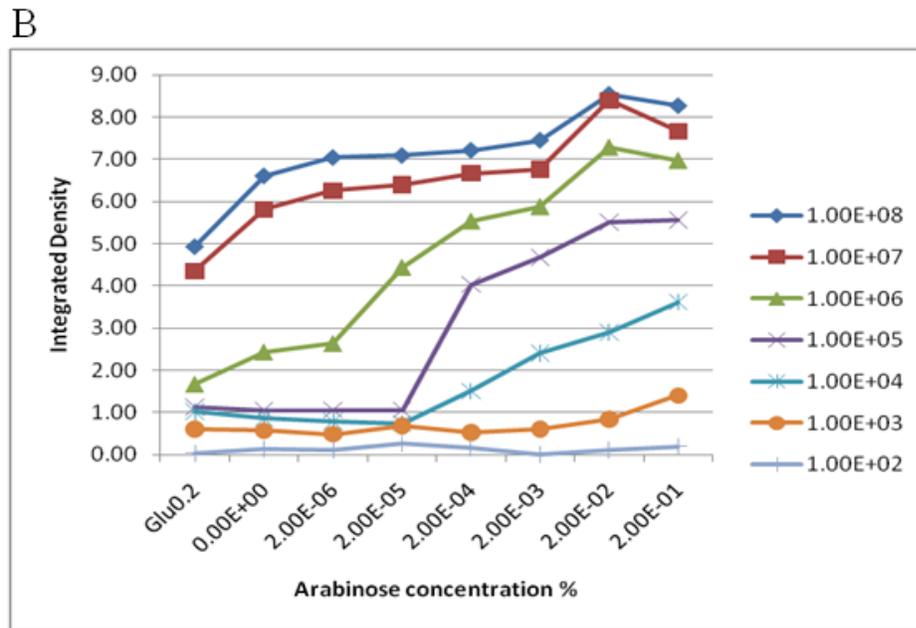
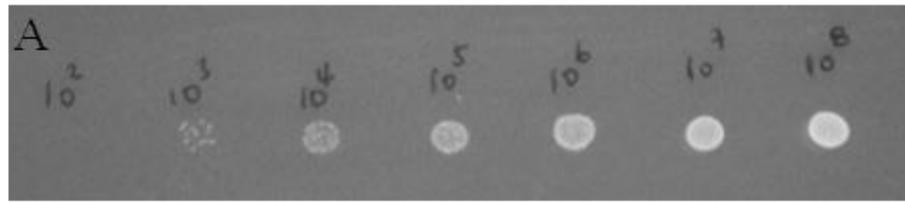
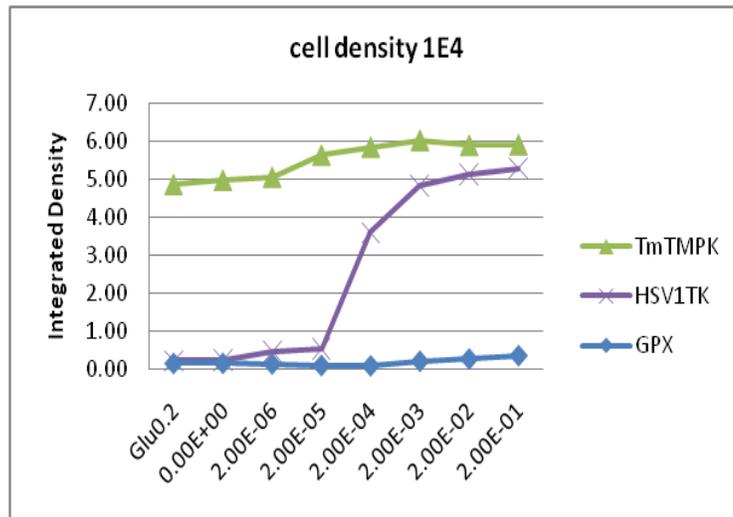


Figure 3.9 Growth dependence of YL-1 on L-arabinose. (A) Picture of representative spot assay plate. Cell densities (cfu/ml) were labeled above each cell spot. (B) Correlation of cell growth with arabinose concentration at different cell densities.

A



B

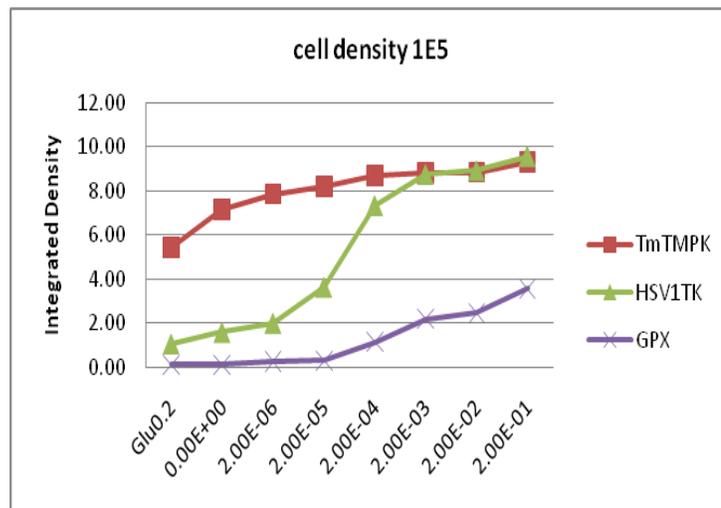


Figure 3.10 Complementation of YL-1 grown by plasmid-born TMPK. Comparison of strain growth when expressing *Tm*TMPK, HSV1-TK and GPX at cell density of 10^4 CFU/ml (A) and 10^5 CFU/ml (B).

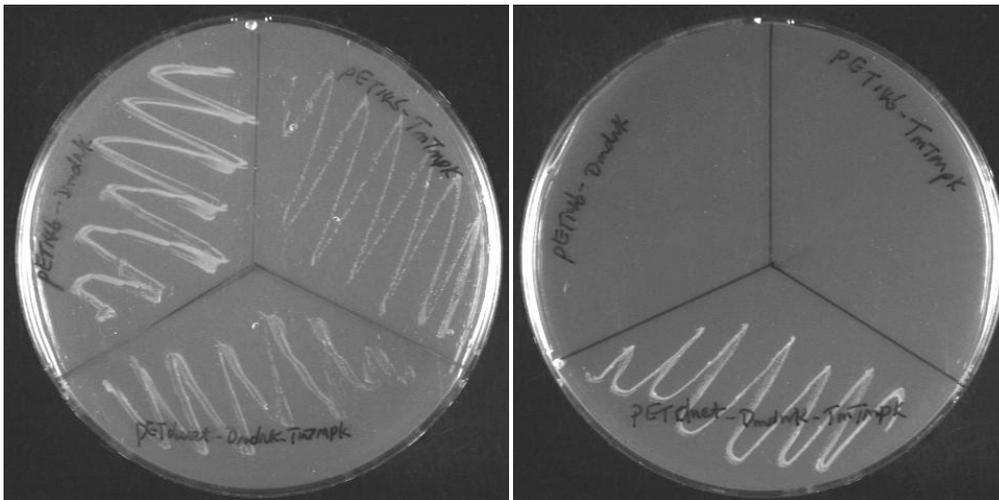
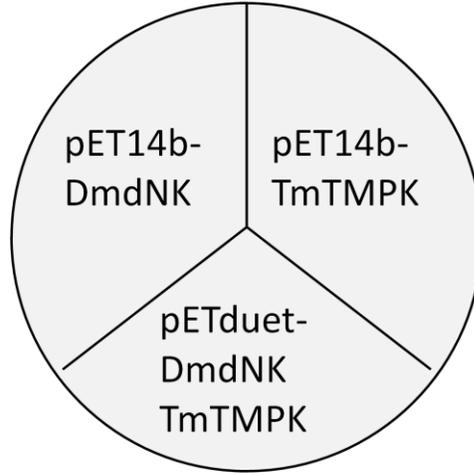


Figure 3.11 Dual functional selection in YL-1. Functional complementation of YL-1, a new conditional auxotroph under dual TK and TMPK selection conditions. YL-1 harboring pET14b-*DmdNK*, pET14b-*TmTMPK* and pETduet-*DmdNK-TmTMPK* were streaked on Kan/Carb/Ara (left) and dual selection media (right) and incubated at 37°C overnight. Only pETduet-*DmdNK-TmTMPK* was able to complement the auxotroph strain under the dual selection condition.

3.3 Conclusion Remarks

In this chapter, a conditional *E. coli* auxotroph strain (YL-1) for selection of dual TK and TMPK function was constructed. This is the first system reported to be able to select for consecutive functions in the same cellular pathway. By placing TMPK under the arabinose promoter, the selection pressure for the second step phosphorylation can be easily tuned by changing the arabinose concentration. It provides a rapid method to identify not only individual TK or TMPK function, but also those with dual function. This novel strain will be beneficial for directed evolution experiments that focus on engineering novel enzymes with dual TK and TMPK function, which could directly phosphorylate NAs to their corresponding diphosphate anabolites, as well.

To control the selection pressure even more delicately, YL-1 with a pKTS selection vector was utilized (Neuenschwander et al., 2007). In this system, the expression of the target protein is controlled by a tunable tetracycline promoter and a degradation tag, which reduces the cellular concentration of the target protein and allows a systematic change of selection pressures in directed evolution experiments. To allow tetracycline-induced protein expression, the tetracycline resistant gene Tn10 was introduced to YL-1 by bacterial conjugation with the XL1-blue strain (Stratagene, La Jolla CA). The resulting strain is denoted YL-T. After transforming the strain with pKTS-*Tm*TMPK, pKTS-HSV1-TK and pKTS-*Dmd*NK (as a negative control) and applied to the TMPK selection condition, only *Tm*TMPK was able to complement the growth. HSV1-TK was not able to complement growth even at very high tetracycline concentrations (1 ng/ml). The reason for HSV1-TK's failure to complement cell growth might be that the active degradation of HSV1-TK initiated by the degradation tag causes

the cellular enzyme concentration to be too low to produce an adequate amount of thymidine diphosphate for cell survival. Therefore, the pKTS system seems to have a protein dependency and cannot be applied in our system.

3.4 Materials and methods

3.4.1 Bacterial media, culture conditions and reagents.

Luria-Bertani (LB) medium was used for bacterial growth. SOC medium (2% Bacto Tryptone (Difco), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) was used for recovering cells after transformation by electroporation. For selection of antibiotic resistance, 100 µg/ml carbenicillin, 50 µg/ml and 25 µg/ml kanamycin, 20 µg/ml chloramphenicol were used. PCR reactions were performed with *Pfu* Turbo DNA polymerase (Stratagene, La Jolla CA). For thymidine kinase function selection, media containing 2% casamino acids, 1.5% noble agar (both Difco, BD Biosciences, Sparks, MD), 0.5% NaCl, 0.2% glucose, 50 µg/ml carbenicillin, 10 µg/ml 5'-fluorodeoxyuridine (Acros Organics, Morris Plains, NJ), 12.5 µg/ml uridine and 2 µg/ml thymidine was used. Bacteria always grow at 37 °C unless noted. Culture media, chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) and New England Biolabs (Ipswich, MA), unless noted.

3.4.2 Construction of *tk* deletion strain.

E. coli BW27783 strain (Khlebnikov et al., 2001) was chosen as parent because of its property of homologous expression regulated by the arabinose promoter. The gene deletion method is based on Wanner Red technology (Datsenko and Wanner, 2000). The kanamycin resistant gene flanked by FRT sequence (FLP recognition target) was amplified from template plasmid pKD4 with primer pair tdkHP1_FOR (5'-ATAAACTCCAGCCAAC TTTATTTTCATATCATTGAGGGCCTGTGGCTGATGTGTGTAGGCTGGAGCTGCTT-3') and tdkPH2_REV (5'-ATATGAATATCCTCCTTAGTGCTCACCCCTGTCAGTAAAGAAATTCTTATTAATCGTGGCGATGCCTTTC-3'), which contains homologous regions adjacent to *tk* on *E. coli* chromosome. The PCR product was digested with *DphI* to eliminate template plasmid and transformed into parent strain carrying Red plasmid pKD46, which assists in gene recombination. Strains with correct chromosomal structure were selected by plating on agar that contains kanamycin, and colonies were verified by colony PCR with primers- Up (5'- AGCAGCAGCCGCGCTTTCTT-3') and Down (5'- CGCCAGACCGCACAGACAGG-3'), which are locus specific primers flanking the recombination site. DNA sequencing analysis also revealed desired chromosomal structure. To prepare the strain for the future *tmpk* knock-out, the kanamycin resistance marker needs to be eliminated. FLP helper plasmid pCP20 was transformed into selected strains to facilitate Kanamycin-resistance gene elimination. PCR and DNA sequencing were applied again to verify the correct chromosomal structure.

3.4.3 Plasmid construction.

To knock in *Tmtmpk*, gene replacement method using vector pRDC15 was applied (Arigoni et al., 1998). pRDC15 is a derivative of pKO3 containing the *araC* and *repA*^{ts} for selection of plasmid integrants at the arabinose operon in *E. coli* chromosome. *TmTMPK* was amplified from pET14b-*TmTMPK* using primer pair- T7 forward primer (5'- TAATACGACTCACTATAGGG-3') and SphITmTMPK reverse primer (5'- GGCCGCATGCTTCTAAACATCGAGTT-3'). The PCR product was digested with *SphI* and *XbaI*, and cloned into digested pRDC15 vector. The insert also keeps the N terminal His-tag for easy detection of *TmTMPK* expression. The resulting construct was designated pRDC15: His-TmTMPK.

To knock out *E.coli tmpk*, gene replacement method using vector pKOV was applied (Link et al., 1997). pKOV vector was digested with *NotI* and *XbaI*, and cloned with a DNA cassette consisting Kanamycin resistance gene flanked by *yceG* and *holB* — genes adjacent to *tmpk* on the chromosome. The DNA cassette was constructed by overlap extension PCR. First, two separate gene fragments were PCR amplified from *E. coli* strain BW27783 using primer pairs NotIyceN (5'-GCTTGCGGCCGCATGAAAA AAGTGTTATTGATAATCT-3'); KanRN_rev (5'-GTCATTGAGGGGCTGGAAGC ATGCTAAGTTATGAGCCATATTCAACGGGA-3') , and KanRC_for (5'- TGCTCG ATGAGTTTTTCTAAATGGCATGCTGGGTGAAGGAGTTGGACGC-3'); *holBCXbaI*(5'-CCGGTCTAGATTAAAGATGAGGAACCGGTAGCACA-3'), respectively. A third gene fragment was amplified from pET23d using primers KanRN_for (5'- GTCATTGAGGGGCTGGAAGCATGCTAAGTTATGAGCCATA TTCAACGGGA-3') and KanRC_rev (5'-GCGTCCAACCTCCTTCACCCAGCATG

CCATTTAGAAAAACTCATCGAGCA -3'). The internal primers (KanRN_rev, KanRC_for, KanRN_for and KanRC_rev) were designed to incorporate overlapping regions, so that the three fragments can be assembled by overlap extension PCR. The DNA cassette was digested by *NotI* and *XbaI* and cloned into digested vector pKOV. The resulting plasmid was confirmed by DNA sequencing and designated pKOV: Kan^R.

3.4.4 Expression and purification of chromosome-born TmTMPK

Six different concentrations of L-arabinose were added at the time of inoculation (2×10^6 , 2×10^5 , 2×10^4 , 2×10^3 , 2×10^2 and 0.2, (v/v %)). The bacterial cultures grew at 37°C overnight. OD₆₀₀ was measured and cells were pelleted by centrifugation. For purification, pellets were resuspended in five volumes of 20 mM Tris-HCl, pH 8.0, containing lysozyme, protease inhibitor cocktail and nuclease benzonase. Sonicate 3 minutes on ice. The resulting homogenate was centrifuged at 12,000g for 10 minutes. The soluble portion was separated from the pellet and loaded onto Ni-NTA resin. The washing and elution steps were according to the manufactory's protocol. The purified protein was analyzed on SDS-PAGE, and its kinetic performance was measured by coupled enzyme assay on UV spectrum.

3.4.5 Growth on antibiotic plates with various L-arabinose concentrations.

Spot assay was performed to determine the level of growth dependence of cells expressing thymidylate kinases with different activities. To verify that changes in thymidylate kinase expression level are responsible for the observed differences in cell

growth, the strain was also transformed with two different thymidylate kinases, as well as a plasmid with no kinase as negative control. Mid-log phase *E.coli* cells were adjusted to $OD_{600}=1.0$ and serially diluted in ten-fold increments into $1\times$ PBS buffer. $2\mu\text{l}$ of each dilution was spotted onto LB/Kan plates supplemented with increasing concentrations of L-arabinose, and the plates were incubated at 37°C for 18 hr. When the cells harbor pDIM plasmids, additional Carbenicillin was added to the media. To quantify the growth spot, ImageJ (Abramoff, 2004) was used to measure the integrated optical density of each spot.

3.4.6 Selection for dual function.

The YL-1 cells were prepared to be electrocompetent (Russell, 2001), and transformed with pET14b-DmdNK, pET14b-TmTMPK and pETduet-DmdNK-TmTMPK and plated onto Kan/Carb/Ara plates. A single colony of each strain was picked and streaked onto dual functional selection plate, which contains 2% casamino acids, 1.5% noble agar (both Difco, BD Biosciences, Sparks MD), 0.5% NaCl 0.2% glucose, 50 $\mu\text{g}/\text{ml}$ carbenicillin, 25 $\mu\text{g}/\text{ml}$ kanamycin, 20 $\mu\text{g}/\text{ml}$ 5'-fluorodeoxyuridine (Acros Organics, Morris Plains, NJ), 12.5 $\mu\text{g}/\text{ml}$ uridine and 2 $\mu\text{g}/\text{ml}$ thymidine. The cells were incubated at 37°C for three days.

Chapter Four

Directed evolution of a novel enzyme for dual thymidine kinase and thymidylate kinase function

4.1 Introduction

This chapter describes using non-homologous recombination to explore the possibility of building a multifunctional enzyme based on nucleoside kinase and nucleotide kinase protein scaffolds. Our interest originates from the observation of Herpes Simplex Virus type 1 thymidine kinase's (HSV1-TK) natural ability to catalyze two consecutive phosphorylations on a single nucleoside substrate (Chen and Prusoff, 1978). Although not equally efficient, the enzyme was shown to bind both thymidine and thymidine monophosphate (TMP) in the same active site and to perform the phosphorylation of the two substrates using ATP as the preferred phosphate donor (Wild et al., 1997). The dual function of this enzyme makes it a promising candidate for nucleoside analog (NA) activation, allowing the direct processing of NAs to the NA diphosphate without accumulation of potentially cytotoxic intermediates.

This hypothesis was demonstrated by expression of wild-type HSV1-TK in HIV-infected cells that are also treated with AZT. The results shows that co-administration of HSV1-TK lowers the virus count 2 to 3-fold compared to cells only treated with AZT (Guettari et al., 1997). However, wild type HSV1-TK has limited substrate specificity for NAs and low catalytic efficiency for the second phosphorylation reaction. Previously, various protein engineering techniques have been applied on HSV1-TK to improve its NA activation efficiency. These approaches, including site-directed, cassette, and random mutagenesis, have led to the identification of several HSV1-TK mutants that display ten to a hundred-fold improvements in substrate specificity (Black et al., 1996; Dipak K. Dube, 1991; Kokoris and Black, 2002; Munir et al., 1993). However, in all cases, the

improvement in specificity is accompanied by a significant reduction of enzyme activity, which compromises the *in vivo* applicability.

Since improving the existing HSV1-TK has achieved only moderate success, an alternative is to build a novel multifunctional enzyme “from scratch.” Dual function could arise from the exchange of protein fragments between parents that share similarities in sequence and structure but catalyze individual phosphorylation steps. Structure overlay suggests a striking similarity of the catalytic core of type 1 deoxyribonucleoside kinase (dNK) and TMPK, demonstrated by the overlay of the crystal structure of *DmdNK* (PDB: 1OE0) and *Thermotoga maritima* TMPK (*TmTMPK*) (PDB: 3HJN) (Fig 4.1). Although the amino acid sequence identity is only 17% by ClustalW alignment, analysis of the active site based on crystal structures showed similar substrate binding interactions. Both enzymes bind the phosphate donor with the P-loop and phosphate acceptor in a deeper active site cleft; both also use a magnesium ion to stabilize the negatively charged phosphate group in the transition state.

Given the highly similar overall structure of dNK and TMPK, and the similar active site interactions, we explored the chimeragenesis approach in this study. In the past, homologous recombination methods have proven beneficial in identifying regions of functional importance in protein frameworks and generating enzymes with new catalytic activities. Christians *et al.* used DNA shuffling to recombine the closely related HSV1-TK and HSV2-TK and identified two mutants with enhanced activity towards AZT (Christians *et al.*, 1999). However, the sequence homology between dNK and TMPK is below the threshold at which DNA shuffling techniques are effective (>70% sequence identity between parent genes) (Lutz and Benkovic, 2000). Therefore, non-

homologous recombination techniques such as ITCHY and SCRTCHY are the best tools to achieve the combinatorial diversification between dNK and TMPK (Lutz et al., 2001a; Lutz et al., 2001b). Previously, these techniques have been applied to generate hybrid TKs with novel activities towards d4T using hTK2 and *DmdNK* as parent genes (Gerth and Lutz, 2007). Nevertheless, engineering multifunctional enzyme presents a bigger challenge because none of the parents possesses both functions. Another reason behind the fact that very little engineering work has been reported on multifunctional kinase systems has to do with the lack of appropriate and efficient selection or screening systems. With the conditional auxotroph strain YL-1 (chapter 3), an efficient screen for directed evolution libraries with dual TK and TMPK function is available.

In this chapter, several generations of ITCHY and SCRATCHY libraries were constructed to create hybrid enzymes from *DmdNK* and *TmTMPK* as the parent genes. *DmdNK*'s robust TK activity and *TmTMPK*'s high thermal stability make them promising candidates for enzyme engineering. To improve the existing SCRATCHY technique, the SCHEMA algorithm (Voigt et al., 2002) was also applied to optimize the crossover location to enable proper protein folding. Libraries with one, two and three crossovers were constructed and screened for dual function with the conditional auxotroph YL-1.

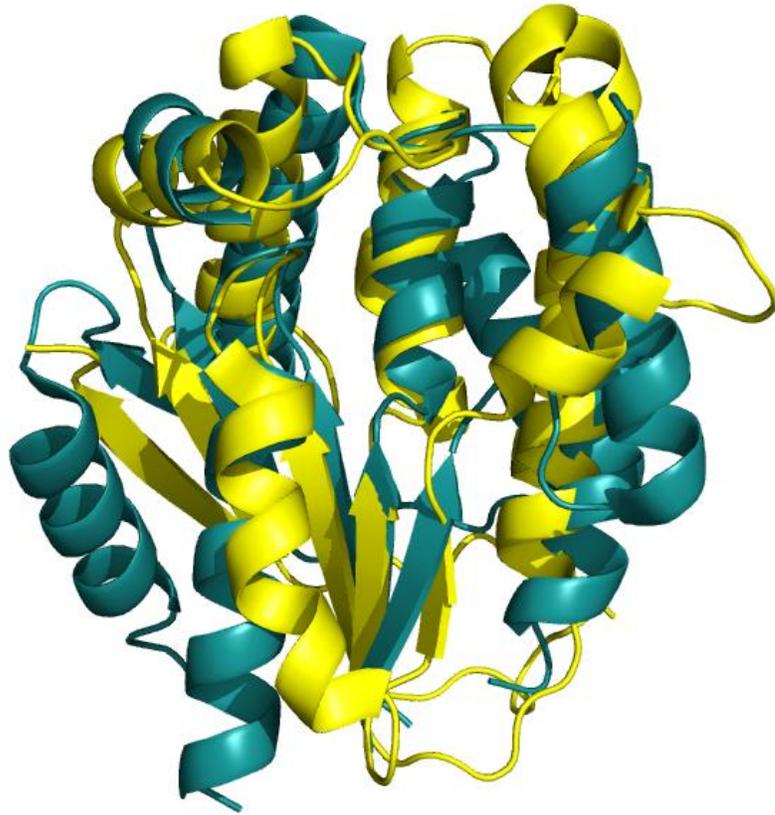


Figure 4.1 Structure overlay of *DmdNK* (yellow; PDB: 1OE0) and *TmTMPK* (blue; PDB: 3HJN) (Mikkelsen et al., 2003; Yoshikawa, 2009). The core structure, the central β -sheet, as well as the P-loop and Lid regions are superimposable.

4.2 Results and discussion

4.2.1 Construct single crossover library and in-frame selection

Two single-crossover ITCHY libraries of *DmdNK* and *TmTMPK* were generated; one with the *DmdNK* gene at the N-terminus and the *TmTMPK* gene at the C-terminus, and the other one with the reverse orientation (Figure 4.2). The theoretical diversity of each library is 4.0×10^5 (675 bp \times 594 bp). The DPX (*DmdNK-TmTMPK*) library contains 3.18×10^5 independent members, and the PDX (*TmTMPK-DmdNK*) library contains 2.70×10^5 members. Therefore, the size of the libraries was comparable to the theoretical size although not able to cover all diversity. The distribution of crossovers between the parental gene fragment, as well as, the fragment size of the unselected library were analyzed by DNA sequencing of randomly selected library members (Figure 4.3). The random distribution over the sampled sequence space indicates no apparent bias towards a particular region within each parent gene.

Theoretically, crossover can occur at any base, which leads to the prediction that approximately two thirds of the library members will contain frame-shifts. The out-of-frame genes will translate into truncated or non-functional proteins. To maximize the meaningful multi-crossover hybrid library members before the next recombination step, both ITCHY libraries were cloned into the pInSAlect vector and plated on carbenicillin media to select for clones with correct reading frames (Gerth et al., 2004). 29% of the DPX library members and 31% of the PDX library members survived this selection, consistent with the prediction that one third of library members are in-frame. Twenty randomly picked members after selection were sequenced by DNA sequencing and all of them were confirmed to have correct reading frames.

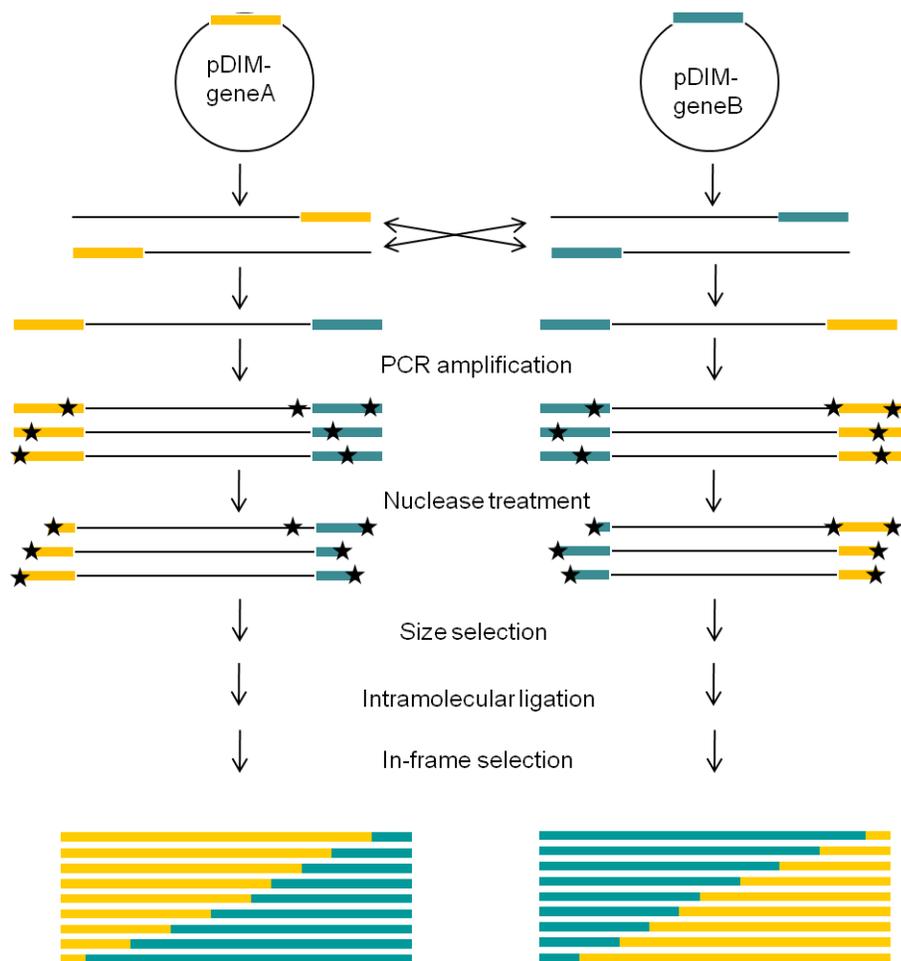


Figure 4.2 Schematic overview of ITCHY library construction (Lutz et al., 2001a), the in-frame selection is carried out by cloning naïve ITCHY libraries to pInSAlect vectors (Gerth et al., 2004).

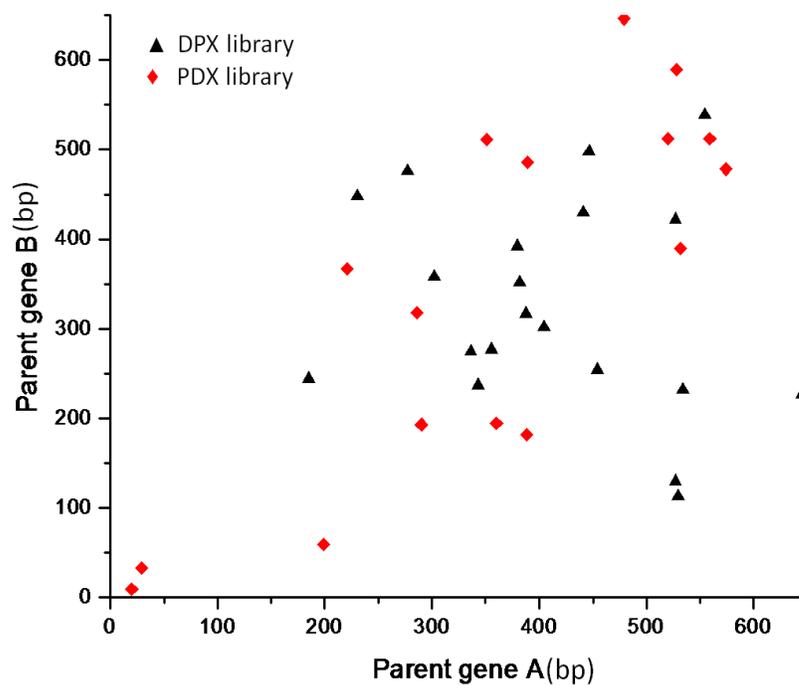


Figure 4.3 Crossover distributions in the naïve ITCHY libraries. The DPX library is plotted as black triangles, each triangle represent a single library member, of which the X axis value represents the last base of the first parent gene and the Y axis value represents the first base of the second parent gene. Similarly, PDX library is plotted as red diamonds, and each diamond represents an individual member of the library.

4.2.2 Construct the first generation multiple crossover library

The ITCHY technique is limited to constructing single crossover hybrids between two parent genes. To further increase the diversity in each library member, SCRATCHY was used to increase the number of crossovers per sequence. Following the enhanced SCRATCHY protocol (Kawarasaki et al., 2003), which allows regulation of the average crossover number per gene, we decided to start by constructing three crossovers libraries (Figure 4.4). The gene sequence was divided into three portions of equal length with short overlap regions. Each section was amplified with skewered primers, resulting in fragments of 200 bp on average, each containing one crossover. The fragments were then reassembled by overlap extension PCR to generate two triple crossover libraries, denoted DPDP (*DmdNK-TmTMPK-DmdNK-TmTMPK*) and PDPD (*TmTMPK-DmdNK-TmTMPK-DmdNK*) according to their gene structure. The size of each library was $> 10^6$. 20 of each library members were randomly picked and sequenced to validate whether the library possesses any bias (Figure 4.5). No particular bias towards a specific sequence of parent genes was found for both libraries, and all clones have three crossovers at different locations throughout the gene sequence.

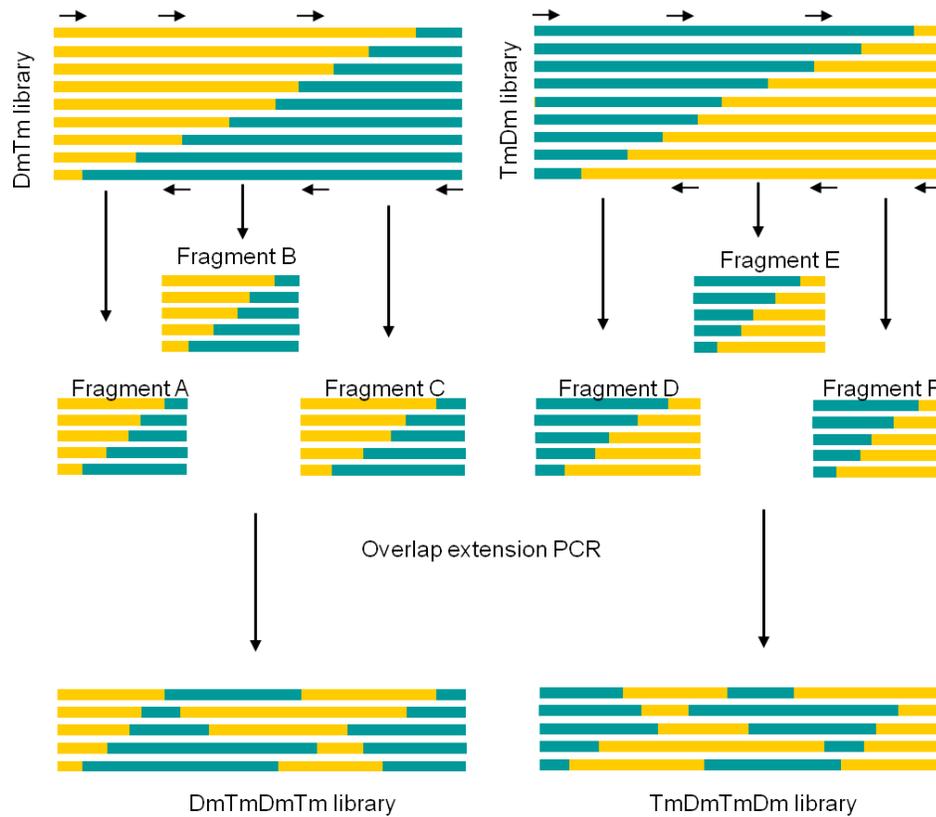


Figure 4.4 Schematic overview of SCRATCHY library construction (Kawarasaki et al., 2003). The ITCHY libraries are divided into arbitrarily defined sections, and amplified by a skewed set of primers. The pooled chimeric fragments are allowed to assemble by overlap extension to construct a library of chimeric genes with 3 crossovers.

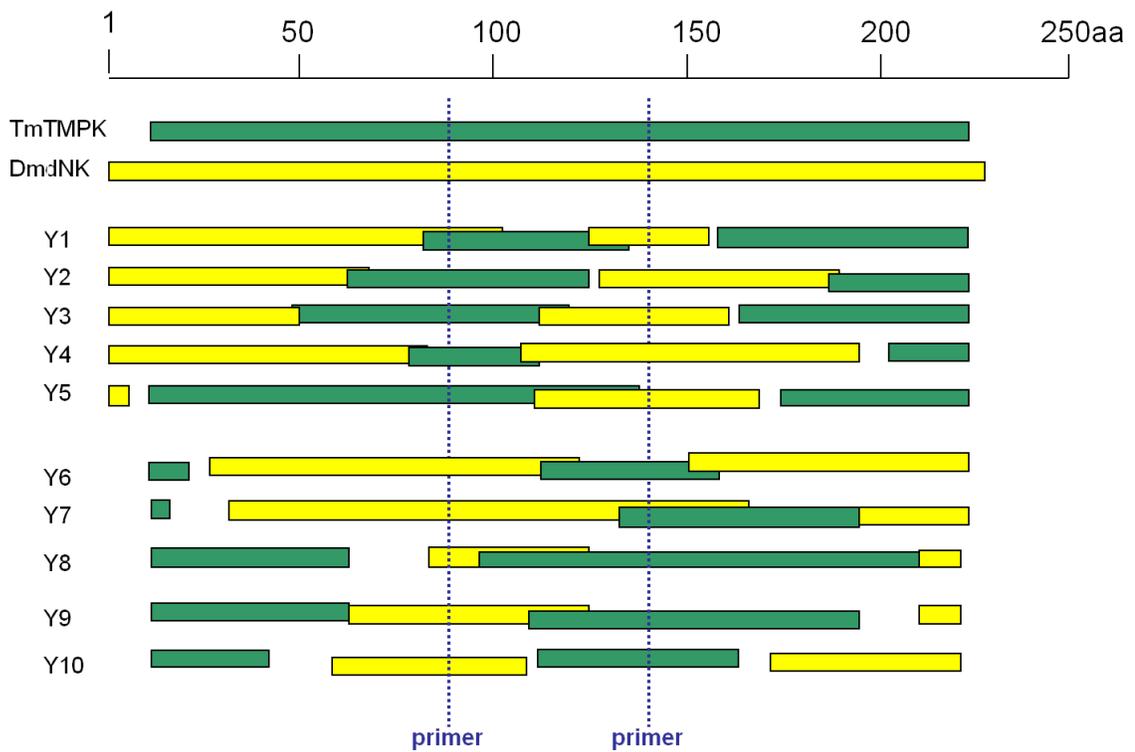


Figure 4.5 Schematic drawing of sequences of 10 randomly selected SCRATCHY clones. The DNA sequences of individual chimeric progeny are depicted by sequential bars. Yellow and green bars represent the *DmdNK* and *TmTMPK* derived sequences, respectively. The positioning of the bars corresponds to their respective parental sequences (depicted on the top). Deletion and redundancy of the parental sequences at non-homologous crossover points are depicted as bars with a gap and overlapping bars, respectively. The dashed blue lines indicate the annealing sites of skewed primers.

4.2.3 SCHEMA calculation

A potential pitfall of the multiple crossover library construction exists in the primer design. Because the crossover location is vital to its proper folding and enzymatic activity, the skewed primers render the crossovers in between these regions but eliminate any diversity within the primer region itself, which is 30bp each on average. Because the design of these primers is arbitrary, it could potentially block the critical regions that are ideal for crossovers to occur. In theory, the crossover should happen where a majority of interactions within the protein structures can be maintained and therefore interruptions are minimized.

To optimize primer location, SCHEMA calculations were introduced to guide primer design (Figure 4.6). SCHEMA is a computational algorithm that identifies the fragments of proteins that can be recombined without disturbing the integrity of the 3D structure (Voigt et al., 2002). When recombination leaves these fragments undisturbed, the hybrid proteins are more likely to fold properly and be functional. In the past, this algorithm has been applied to several systems and has been shown effective in predicting crossover location in creating chimera proteins (Heinzelman et al., 2010; Meyer et al., 2006; Otey et al., 2006).

Starting with the multiple sequence alignment of the parent genes (*DmdNK* and *TmTMPK*) by ClustalW (Figure 4.7) and each protein structure (PDB: 1OE0 and PDB: 2Z0H), SCHEMA generates a protein contact map, which calculates the distance between all possible residue pairs of a three-dimensional protein structure using a binary two-dimensional matrix. Combining the contact map with the desired number of crossovers

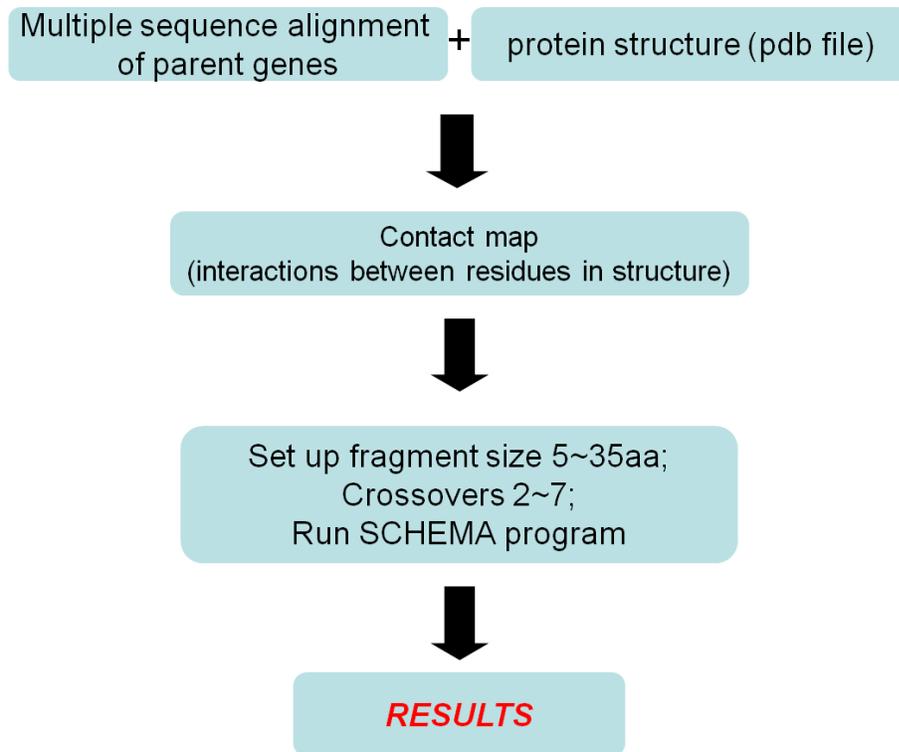


Figure 4.6 Scheme of SCHEMA calculations (Voigt et al., 2002).

```

TmTMPK  -----MFITFEGIDGSGKSTQIQLLAQYLEKRGKKVILKREPGGTETGEKIRKILL
DmdNK   MKYAEGTQPFTVLIIEGNIGSGKTTYLNHFKEYKN---DICLLTEPVEKWRNVNGVNLLLE
          : : :** *****: * : : : * : . : * ** . . : : *

TmTMPK  EEEVTPKAE LFLF L A S R N L L V T E I K Q Y L S E G Y A V L L D R -- Y T D S S V A Y Q G F G R N -- L G K -
DmdNK   LMYKDPKKWAMPFQS Y V T L T M L Q S H T A P T N K K L K I M E R S I F S A R Y C F V E N M R R N G S L E Q G
          ** : * : . * : : : : : : : * : : : : * * * :

TmTMPK  --EIVEELNDFATD--GLIPDLTFYIDVDVETALKRKGE L N R F E K -----REF L E R V R
DmdNK   MYNTLEEWYKFIEESIHVQADLI I Y L R T S P E V A Y E R I R Q R A R S E E S C V P L K Y L Q E L H E L H
          : : ** . * : : . ** : * : . . * . * : * : * * : : * . . . :

TmTMPK  EGYLV L A R - E H P E R I V V L D G K R S I E E I H R D V V R E V K R R W K L D V -----
DmdNK   EDWLIHQRRPQ S C K V L V L D A D L N L E N I G T E Y Q R S E S I F D A I S S N Q Q P
          * . : * : * : : : * * . . . : * * : * . . : .

```

Figure 4.7 ClustalW alignments of *Tm*TMPK and *Dmd*NK used in SCHEMA calculations. Symbols below the sequence indicate the degree of conservation in an alignment column: strictly conserved (stars), highly conserved (two dots), or moderately conserved (one dot).

and a minimum fragment length, which specifies the minimum number of residues a fragment can contain, SCHEMA generates an energy plot of all the possible crossover location combinations. Representative results are shown in Figure 4.8. Each data point represents a combination of crossover locations and corresponds to a distinct energy level (E) and mutation level (m). Although the absolute value of energy is arbitrary, comparing the relative energy value of each data point to others provides more useful information. Overall, the diagonal and upward distribution from low to high mutation level reflects the trade off of energy and sequence diversity. The data points with higher mutation level also have higher energy, which makes sense because a further deviation from the parent sequence will more likely have significant disruption in interactions within the protein structure, therefore more likely to result in improperly folded proteins.

To obtain more information about the crossover locations, we manually selected data points with the lowest E/m ratio across the whole range of available m values. These data points represent combinations that have relatively low energy but also fairly high diversity, thus likely generating a new function. The crossover locations in each combination were plotted against the energy level (Figure 4.9). Interestingly, the crossover location clustered in a few specific regions (around residue 100, 140 and 190) instead of randomly distributed in the sequence. The first crossover location at residue 10 was insignificant because it is very close to the N-terminus and far from the active site. Recombination at this position is unlikely to create a new function. Locating these regions in the crystal structure reveals that the optimal crossovers occur either in a loop region or at the terminus of a secondary structure (α -helix or β -sheet). The result is not surprising since to minimize the structural interruption, changes should be made in more

flexible regions, and the more rigid structure should be left untouched to maintain the interactions between residues.

Next, to investigate whether the primer design in the first generation library was desirable, the arbitrary primers positions were located on the plot. The results show that the first generation primers did create undesired constructs (Figure 4.10). Especially, the third set of primers resides in the region (marked in cyan) where the desirable crossovers are located. Such design could generate library members that are unlikely to fold correctly and thus will likely lack function. Based on the SCHEMA calculation, the second-generation multiple crossover libraries with primers in regions with no predicted crossover were assembled to maximize the probability for properly folded chimeras.

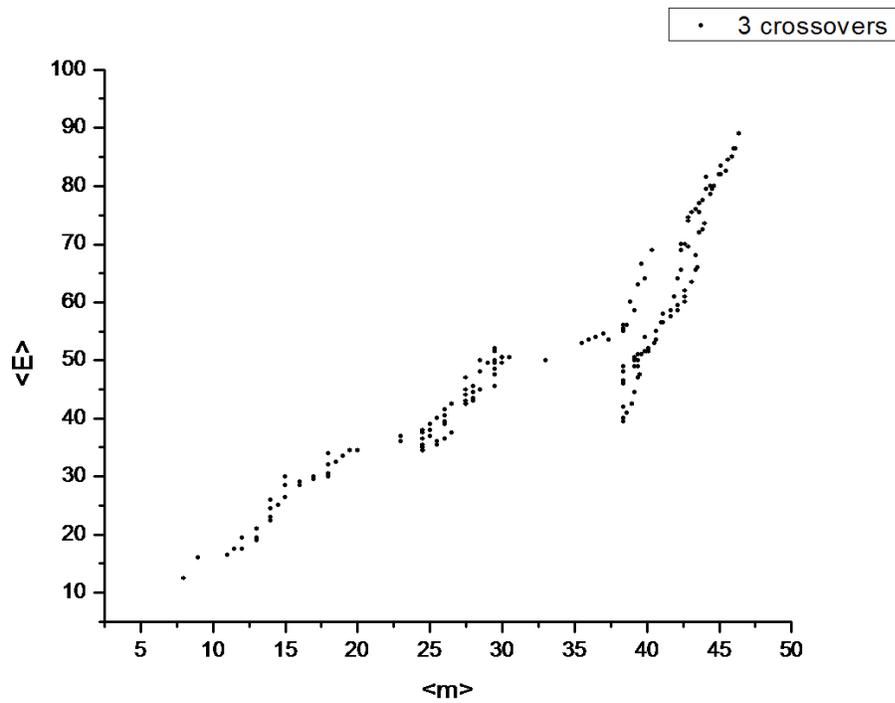


Figure 4.8 Representative of SCHEMA calculation of *Dmd*NK and *Tm*TMPK chimera. E , SCHEMA disruption energy; m , average numbers of amino acid changes needed to convert the protein into one of the parents.

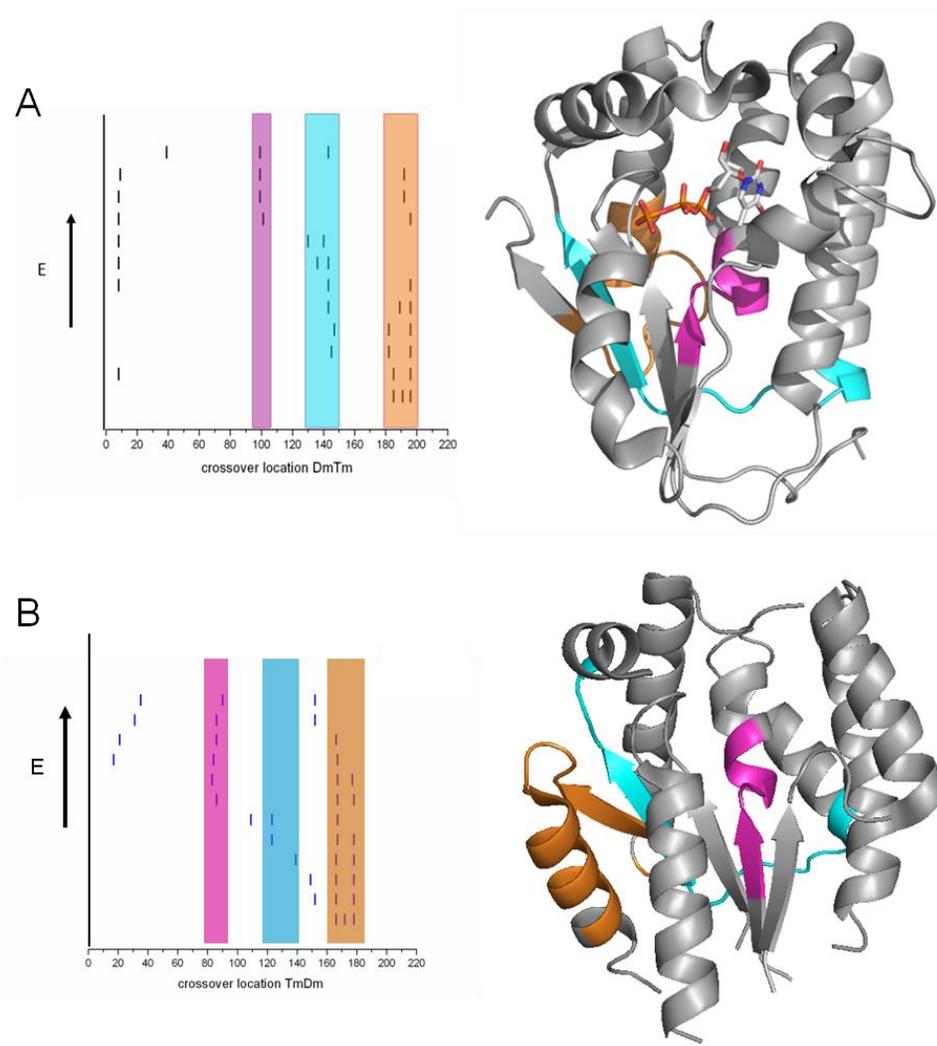


Figure 4.9 Plot of crossover combinations with low E/m values. (A) Location of the crossovers in the crystal structure of *DmdNK* (PDB:10E0). (B) Location of the crossovers in the crystal structure of *TmTMPK* (PDB: 2Z0H).

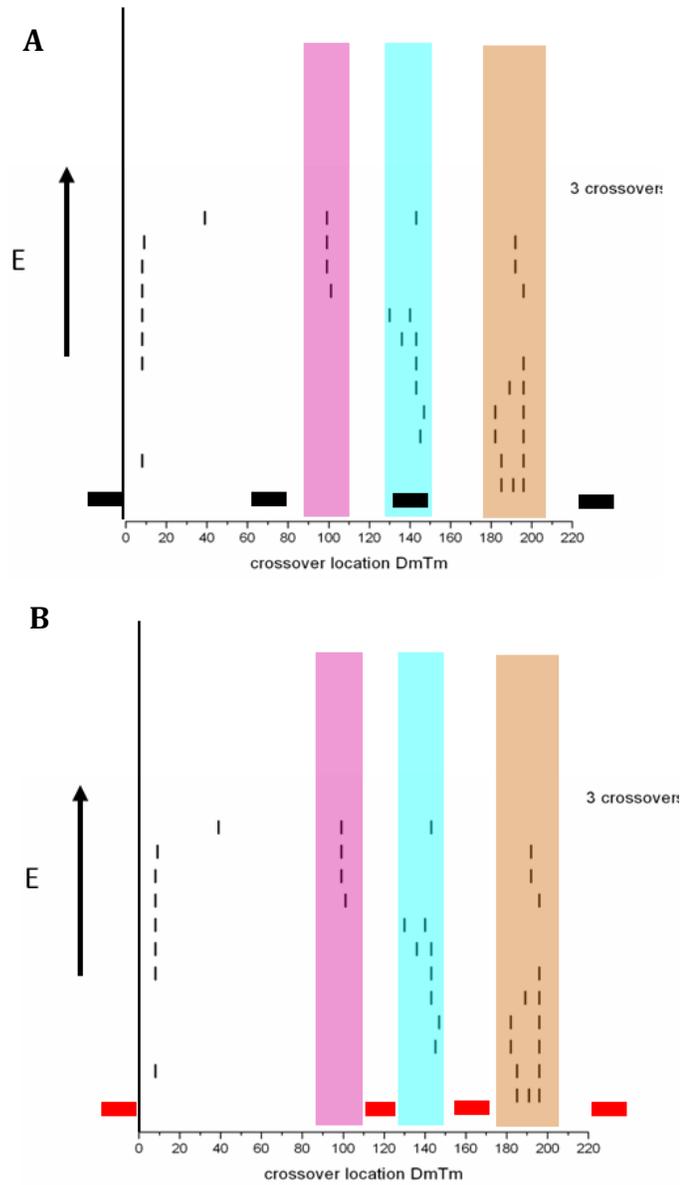


Figure 4.10 Comparison of two generations of primer design. (A) First generation primer sequence location (B) Second generation of primer designed flanking calculated crossover regions.

4.2.4 Construct the second generation multiple crossover library

Amplified with the second-generation primers, the three fragments are 340bp, 170bp and 180bp in length, respectively. Using the same protocol, the fragments were reassembled by overlap extension PCR to generate two triple crossover libraries, denoted 2nd-DPDP (*DmdNK-TmTMPK-DmdNK-TmTMPK*) and 2nd-PDPD (*TmTMPK-DmdNK-TmTMPK-DmdNK*). The size of each library was $> 10^6$. Twenty randomly selected members of each naive library members were sequenced to validate if there is bias in the library. No bias was found for each library. All clones have three crossovers at the predicted regions (Figure 4.11).

The SCHEMA calculation also suggested that the combination of crossover locations around residue 140 and 190 has a lower SCHEMA disruption energy than any combinations with a crossover location at residue 100. Interestingly, these two crossover positions in the protein crystal structures show that they sandwich the Lid region. The resulting chimera will have a swapped lid domain between the two parent genes. Hence, additional libraries with two crossovers were constructed to only include crossovers near residue 140 and 190. The protocol was similar except that the wild type parent genes were used as a template to amplify the first fragment, which has no crossover. All the fragments were reassembled by overlap extension PCR. The two libraries were named DPD (*DmdNK-TmTMPK-DmdNK*) and PDP (*TmTMPK-DmdNK-TmTMPK*). Twenty randomly selected members of each library were sequenced to validate if there is bias in the library. No bias was found for each library, and all clones have two crossovers at the predicted regions (Figure 4.12).

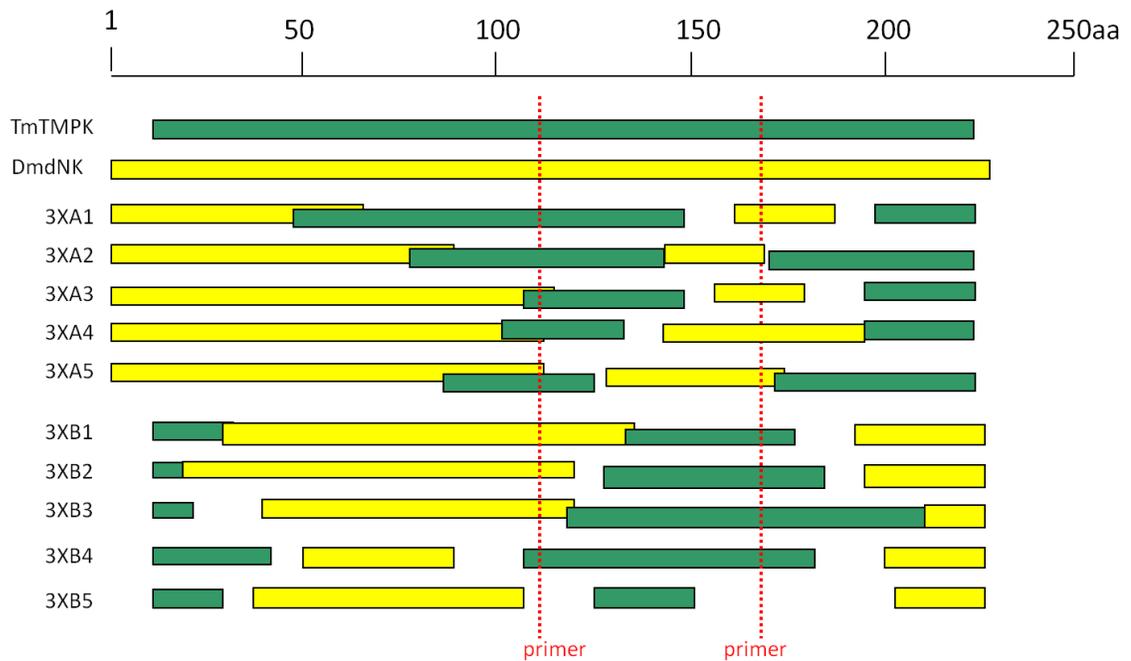


Figure 4.11 The sequences of clones from the second generation SCRATCHY libraries. The DNA sequences of individual chimeric progeny are depicted by sequential bars. Yellow and green bars represent the *DmdNK* and *TmTMPK* derived sequences, respectively. The positioning of the bars corresponds to their respective parental sequences (depicted on the top). Deletion and redundancy of the parental sequences at non-homologous crossover points are depicted as bars with a gap and overlapping bars, respectively. The dashed red lines indicate the new primer based on SCHEMA calculations.

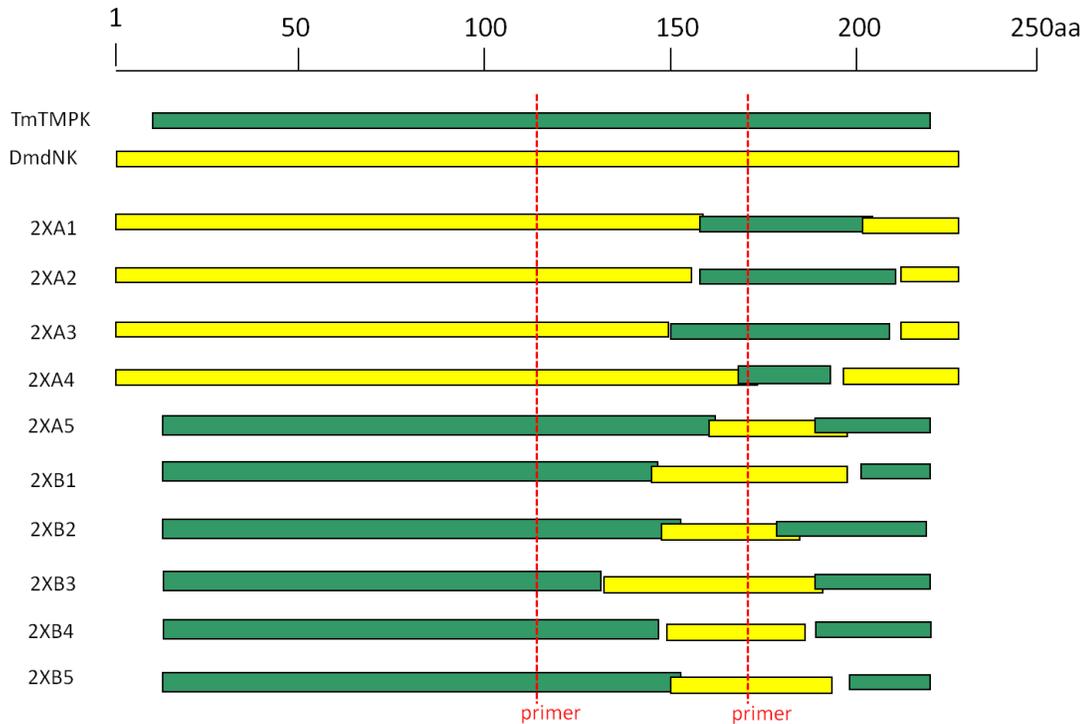


Figure 4.12 The sequences of clones from from the two crossover library. The DNA sequences of individual chimeric progeny are depicted by sequential bars. Yellow and green bars represent the *DmdNK* and *TmTMPK* derived sequences, respectively. The positioning of the bars corresponds to their respective parental sequences (depicted on the top). Deletion and redundancy of the parental sequences at non-homologous crossover points are depicted as bars with a gap and overlapping bars, respectively. The dashed red lines indicate the new primer based on SCHEMA calculations.

4.2.5 Selection for dual function using YL-1, a conditional auxotroph strain

Functional complementation of the conditional auxotroph strain YL-1 (chapter 3) was used to select chimeras with dual TK and TMPK activity. Two million of each library members from the one crossover libraries (ITCHY) and the two and three crossover libraries (SCRATCHY) were transformed into YL-1 and plated onto the selection media to cover all the diversities in the library. The cells were incubated at 30°C for seven days. Unfortunately, no functional mutants were found. One possibility is that

4.3 Conclusion remarks

Although no functional mutants from the libraries are found, the importance of this study lies in the methodology improvement. Non-homologous recombination is a remarkable method for studying sequence-function relationships on a significantly larger scale than accessible by point mutagenesis. Without the requirement of high homology of protein sequence, it permits the search of much larger protein sequence space compared to the traditional homologous recombination methods, such as DNA shuffling. However, swapping protein segments also introduces a huge change in protein structure, thus increasing the likelihood of creating improperly folded variants. In theory, combining SCHEMA calculations and SCRATCHY allows for a more rational design of chimeric protein libraries using the available protein structural information. SCHEMA calculations in this study suggest that the desirable crossover locations in the protein structure have a preference for specific regions instead of randomly distributed throughout the whole sequence space. Extracting such information is valuable in efficiently designing and

constructing recombination libraries since it is impossible to survey every residue as a crossover location practically.

The reason for the failure of this experiment lies in several folds. Firstly, the selection pressure may be too high to identify chimeras with weak dual TK and TMPK function. Further study can be done to investigate other vectors capable of producing higher protein expression levels in order to ease the selection pressure. In addition, different concentrations of arabinose can be added to the media to select for mutants with lower TMPK activity. Secondly, the mechanisms of the two functions selected are different. Although both add a phosphate group to the substrate, one results in an ester, and the other produces an anhydride. The only example of a dual functioning enzyme (HSV1-TK) is not fully understood in terms of the determinants of its dual function. Further investigation on the mechanism of TK and TMPK dual activity will be elucidated in chapter 5.

4.4 Materials and methods

4.4.1 Materials

Pfu Turbo polymerase from Stratagene (La Jolla, CA) was used for all cloning, unless otherwise indicated. The α -phosphorothioate dNTP (α S dNTP) and exonuclease III were purchased from Promega (Madison, WI). Plasmid DNA was isolated using the Qiaprep Spin Miniprep kit and PCR products were purified with the Qiaquick PCR Purification kit (Qiagen, Valencia, CA). Enzymes were from New England Biolabs (Beverly, MA), and all other reagents were from Sigma-Aldrich (St. Louis, MO), unless noted.

4.4.2 *ITCHY* library construction

Truncated *DmdNK* and full length *TmTMPK* were used to create the single-crossover libraries (DPX and PDX). This truncated *DmdNK* was created by removing the first 12 amino acids and last 15 amino acids, introduced a new methionine start codon in place of lysine13 and stop codon TGA replacing proline236. The kinetic properties of this truncated enzyme are identical to full length enzyme (Gerth et al., 2004). A modified thio-*ITCHY* protocol was used to construct the libraries. pDIM plasmids encodes parent genes were linearized by *NdeI* or *SpeI* to serve as template for two different libraries (Figure 4.2) . The assembly PCR is catalyzed by *Taq* DNA polymerase (Promega) and the protocol is described below. The template and primer combinations used in constructing PDX library were: pDIM-*TmTMPK* linearized with *NdeI*; pDIM-*DmdNK* linearized with *SpeI*; *TmTMPK_for* primer and D201 reverse primer (Table 4.1). The template and primer combinations used in DPX library were pDIM-*DmdNK* linearized with *NdeI*; pDIM-*TmTMPK* linearized with *SpeI*; Dm31 forward primer and *TmTMPK* reverse primer (Table 4.1). Thermocycling conditions were: 95°C for 2 min; 3 cycles of 94°C 10 sec and 68°C 1 min; add primers and α S dNTPs mix (α S dNTPs: dNTP= 1:9 molar ratio); 30 cycles of 94°C 10 sec, 60 °C 1 min, 68°C 3.5 min; 1 cycle 68°C 5 min. The PCR products were purified by Qiaquick PCR Purification kit, and treated with exonuclease III (120 U/ μ g DNA) at 37°C 30 min and quenched by PB buffer and PCR purification. The single strand overhang was removed by incubating with Mung bean nuclease (2.5 U/ μ g DNA) at 30 °C 30 min. The purified DNA was blunt-ended by Klenow fragment treatment (0.5 U/ μ g DNA) at 37°C 15 min and dNTPs at 37°C 10 min.

After purification again, the DNA was cyclized by intramolecular ligation using T4 ligase (Promega) overnight at 4°C.

The ligation mixture was used to transform *E. coli* DH5 α . Libraries were plated on LB agar containing carbenicillin (100 μ g/ml). After overnight growth at 37 °C, the colonies were recovered by scraping the plate with recovery medium (2YT medium supplied with glucose (2% w/v) and glycerol (15% v/v)). The recovered libraries were divided into aliquots; one portion was saved at -80°C as frozen stock, and another portion was isolated in plasmid form.

4.4.3 Reading frame selection

The single-crossover libraries in pDIM vector were amplified by PCR to replace the stop codon with codon for glycine, using the T3 forward primer and the reverse primers DmdNKNS_rev and TmTMPKNS_rev (Table 4.1). The PCR products were digested with *Nde*I and *Spe*I, and cloned into pInSAlect vector (Gerth et al., 2004). The ligation products were used to transform *E. coli*. DH5 α . The cells were plated on LB agar containing chloramphenicol (34 μ g/ml) and incubated at 37°C. Colonies were recovered in recovery media. Reading frame selection was carried out by plating the libraries on LB agar containing carbenicillin (100 μ g/ml) and incubating at 22°C. Libraries were recovered again and the plasmid pools were isolated.

4.4.4 *SCRATCHY library construction*

The sequence of chimeric ITCHY library was arbitrarily divided into three overlapping sections; each of them was amplified by skewed primers (Table 4.1). To generate product that contains one crossover, the forward primer is designed to be specific to one parent gene and the reverse primer to be specific to the other parent gene. The pInSALect-PDX library plasmids were used to generate three segments with the following primers: fragment A, T3 and PR1; fragment B, DF2 and PR2; fragment C, DF3 and PGX. The pInSALect-DPX library plasmids were used to generate three segments with the following primers: fragment D, T3 and DR1; fragment E, PF2 and DR2; fragment F, PF3 and PGX. The amplifications were carried out using the following thermal cycling conditions: 95°C for 2 min; 30 cycles of 95°C for 30 sec, 55 °C for 30 sec, 72 °C for 45 sec; 1 cycle of 72°C for 10 min.

4.4.5 *SCHEMA calculation and new SCRATCHY library design*

The SCHEMA disruption energy (E) for a chimera was calculated according to

$$E = \sum_i \sum_{j>i} C_{ij} \Delta_{ij}$$

Where $C_{ij} = 1$ if any side-chain heavy atoms or main-chain carbons in amino acid residues i and j are within 4.5 Å (Meyer et al., 2006). The crystal structure of *DmdNK* (PDB code: 1OE3) and *TmTMPK* (2ZOH) were used to calculate C_{ij} . The Δ_{ij} was calculated from the sequence alignment of the parent proteins. $\Delta_{ij} = 0$ if amino acids i and j in the chimera are found at the same positions in any parental protein sequence, otherwise the interaction is broken and $\Delta_{ij} = 1$. The sequence of *DmdNK* and *TmTMPK*

were aligned using ClustalW. Python scripts for calculating E are downloaded from the Arnold lab website <http://www.che.caltech.edu/group/fha/>.

New primers were designed to optimize crossover positions in the libraries. The pInSALect-PDX library plasmids were used to generate three segments with the following primers: fragment A, T3 and N-PR1; fragment B, N-DF2 and N-PR2; fragment C, N-DF3 and PGX (Table 4.1). The pInSALect-DPX library plasmids were used to generate three segments with the following primers: fragment D, T3 and N-DR1; fragment E, N-PF2 and N-DR2; fragment F, N-PF3 and PGX (Table 4.1). The amplifications were carried out using the following thermal cycling conditions: 95°C for 2 min; 30 cycles of 95°C for 30 sec, 56 °C for 30 sec, 68 °C for 45 sec; 1 cycle of 68°C for 10 min.

4.4.6 Selection for dual TK and TMPK activity

The conditional auxotrophic strain YL-1 (chapter 3) was transformed with library plasmids and plated on LB agar containing kanamycin (25 µg/ml), carbenicillin (50 µg/ml) and L-arabinose 0.2% (w/v). For library selection, the cells were recovered and plated on dual function selection plate containing 2% casamino acids, 1.5% noble agar (both Difco, BD Biosciences, Sparks MD), 0.5% NaCl 0.2% glucose, 50 µg/ml carbenicillin, 25 µg/ml kanamycin, 20 µg/ml 5'-fluorodeoxyuridine (Acros Organics, Morris Plains, NJ), 12.5 µg/ml uridine and 2 µg/ml thymidine. The cells were incubated at 30°C for seven days.

Table. 4.1 Primers used for chimeragenesis and cloning

Name	Sequence
TmTMPK_for	5'-CGCCATATGTTTCATAACGTTTCGAGGG-3'
D201	5'-CGAGGACTAGGACCTTGCACGACTGCG-3'
Dm31	5'-CCACATCGGCAGC GGGAAGACCA-3'
TmTMPK reverse	5'-CGCACTAGTCTAAACATC GAGTTTCC-3'
T3	5'-ATTAACCCTCACTAAAGGGA-3'
DmdNKNS_rev	5'-CGCACTAGTTCCGGGCTGTTGGTACTTGAGAT-3'
TmTMPKNS_rev	5'-CGCACTAGTTCCAACATCGAGTTTCCACCTTCT-3'
PR1	5'-CGAGGCGAGA AACAGGAAGAG-3'
DF2	5'-GCTGACCATGCTGCAGTCGCACAC-3'
PR2	5'-GTAGAAAGTCCAGGTCAGGAATCAGTC-3'
DF3	5'-GCAGGCGGACCTCATCATATATC-3'
PGX	5'-AGCGCGGCGGGTGTGGTG G-3'
DR1	5'-GTGTGCGAGTGCAGCATGGTCAGC-3'
PF2	5'-CTCTTCC TGTTTCTCGCCTCG-3'
DR2	5'-GATATATGATGAGGTCCGCCTGC-3'
PF3	5'-GACTGATTCTGACCTGACTTTCTAC-3'
N-PR1	5'-CGATCTCTTTTCCCAGGTTTCTTCC-3'
N-DF2	5'-GGCTCGCTG GAGCAGGGCATGTA-3'
N-PR2	5'-CACGTTTCGAGAACTCTCTTTTTTC-3'
N-DF3	5'-GCTGCGTGCCGCTTAAGTA CCTTC-3'
N-DR1	5'-TACATGC CCTGCTCCAGCGAGCC-3'
N-PF2	5'-GGAAGAAACCTGGGAAAA GAGATCG-3'
N-DR2	5'-GAAGGTACTTAAGCGGCACGCAGC-3'
N-PF3	5'-GAAAAAAGAGAGTTTCTCGAACGTG-3'

Chapter Five

Investigating the structural basis for substrate specificity of HSV1-TK, *DmdNK* and *TmTMPK* through site-directed mutagenesis and Isothermal Titration Calorimetry

5.1 Introduction

In the salvage pathway of thymidine triphosphate (TTP) synthesis, thymidine kinase (TK) and thymidylate kinase (TMPK) are the key enzymes that catalyze the first and second phosphorylation reactions, producing thymidine monophosphate (TMP) and thymidine diphosphate (TDP), respectively. This pathway is also used for activation of several anti-viral and anti-cancer pro-drugs (nucleoside analogs), such as acyclovir, ganciclovir and AZT. Either of these two phosphorylation steps can be the “bottle neck” of nucleoside analog activation, and both scenarios have been identified for different nucleoside analogs (Monnerjahn and Konrad, 2003). The inefficient nucleoside analog activation not only compromises the drug efficacy but also causes the accumulation of cytotoxic intermediates that could lead to cell death (Tornevik et al., 1995b). The central role of TK and TMPK has led to interest in investigating their substrate specificities and reaction mechanisms.

Humans have two thymidine kinases: cytosolic thymidine kinase 1 (hTK1) and mitochondrial thymidine kinase 2 (hTK2). Due to amino acid sequence, structure and substrate specificity, hTK2 belongs to the larger type 1 dNK family that also includes human dCK and dGK. The crystal structures of several type 1 dNKs have now been solved and share high structural similarity with the TMPK family despite only about 13% amino acid identity in the sequence alignment. In these structures, the monomer of each enzyme comprises 8-10 α -helices surrounding a five-stranded parallel β -sheet core. More importantly, the key domains for catalysis and substrate binding in the active sites are highly similar (Figure 5.1). It has been proposed that TK and TMPK share a common but

distant ancestor in evolution (Wild et al., 1997). The proposed phylogenetic relationship of the two enzyme families from ClustalW alignment is illustrated in Figure 5.1 D.

A special member of the type 1 dNK family is the thymidine kinase from the Herpes Simplex virus (HSV1-TK). It was the first deoxyribonucleoside kinase found capable of catalyzing two consecutive phosphoryl transfers on a single nucleoside substrate (Chen and Prusoff, 1978). Although sharing a similar core structure with other dNKs and TMPKs, HSV1-TK is a much larger protein due to additional elements - 45 residues at the N-terminus, 29 residues at the C-terminus and 72 residues between $\alpha 8$ and $\beta 5$, comprising 4 α -helices and a 15-residue loop (Wild et al., 1997). These differences are primarily located at the dimer interface and leave the catalytic center unaffected. The existence of the dual functioning HSV1-TK is another indication of the close evolutionary relationships between type 1 dNK and TMPK families.

It is interesting that TK and TMPK display no overlapping substrate specificity despite their highly similar structures. TK is known to only phosphorylate nucleosides and not the monophosphate, and none of the TMPKs has been reported to have catalytic activity towards thymidine. HSV1-TK and equine herpes virus TK are the only known enzymes that have both TK and TMPK functions; however, the reason for such dual function remains unclear. Wild *et.al.* showed that superposition of crystal structures of HSV1-TK with thymidine and TMP in the active site only reveal minimal differences. The phosphate of TMP pushes the backbone of Arg222 outward by 1.5 Å (Wild et al., 1997). In a study by Lavie and co-workers, HSV1-TK was crystallized with two bisubstrate inhibitors, TP₄A and TP₅A, in order to understand the structural basis for the dual function (Gardberg et al., 2003b). In each protein complex, one of the phosphate

groups does not form any interactions to the protein but functions as a linker so that the overall structure mimics the transition state of the phosphoryl transfer reaction. HSV1-TK accommodates both inhibitors in a highly similar fashion; the only difference observed is the movement of the phosphate group but not the active site residues.

This chapter focuses on investigating the determinant of substrate specificity and the reaction mechanism by comparing three enzymes – dNK from *Drosophila melanogaster*, TMPK from *Thermotoga maritima* and HSV1-TK. As a well studied type 1 dNK, *DmdNK* possesses a high catalytic rate and broad substrate specificity; it can phosphorylate all four native deoxyribonucleosides (Munch-Petersen et al., 1998). *TmTMPK* was selected to represent the TMPK family, as it has high catalytic efficiency for its natural substrate TMP as well as high thermal stability (chapter 2). HSV1-TK as a dual functional enzyme was shown to bind nucleoside and nucleoside monophosphate in the same active site and phosphorylate both substrates with the second step being less efficient than the first one (Maga et al., 1994; Wild et al., 1997).

First, by comparing the structure and sequence of the three enzymes, information was extracted that may help understand the contributing factors that result in these distinct substrate specificities. Next, by site-directed mutagenesis, the roles of non-conserved active site residues on substrate specificity were explored. In addition, because the two phosphorylation reactions of thymidine and TMP have different mechanisms (one results in an ester and the other results in an anhydride), isothermal titration calorimetry (ITC) was also used to explore substrate, product and inhibitor binding to the three enzymes. The results help to understand the difference in substrate recognition in TK and TMPK.

Separately, the contributing factors that control how HSV1-TK catalyzes both reactions with the same set of active site residues were also explored. Understanding how HSV1-TK catalyzes both reactions will not only facilitate the improvement of HSV1-TK's second step reaction but will also guide further engineering of a novel enzyme with improved dual function.

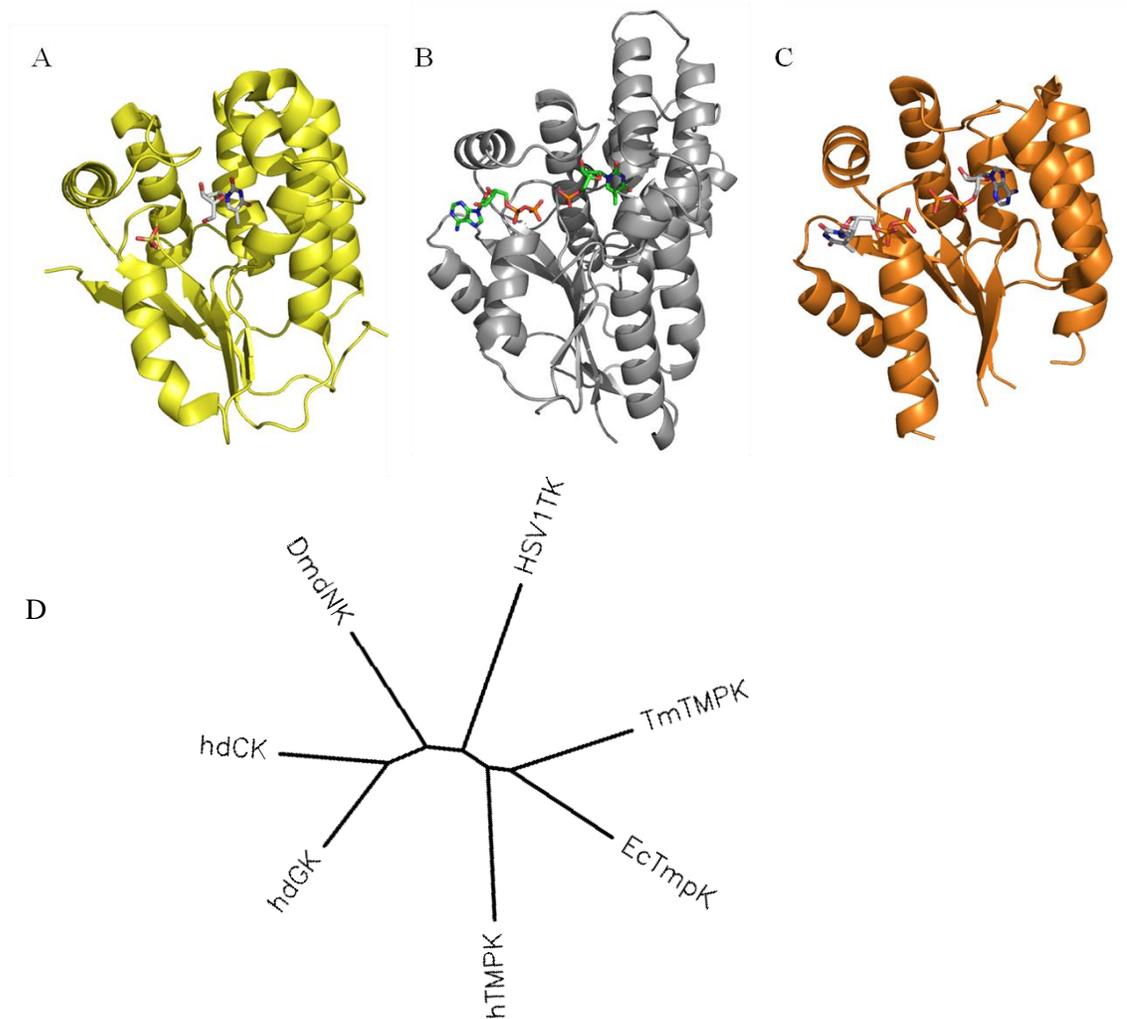


Figure 5.1 The structure of three enzymes with their substrates reveals high homology. (A) *DmdnK* with thymidine and sulfate (PDB code: 1OT3; (Mikkelsen et al., 2003)). (B) HSV1-TK with ADP and TMP (PDB code: 1VTK; (Wild et al., 1997)). (C) *TmTMPK* with TDP and ADP (PDB code 3HJN; (Yoshikawa, 2009)). (D) Phylogenetic relationship among dNKs, TMPKs and HSV1-TK calculated by ClustalW multiple sequence alignment.

5.2 Results and discussion

5.2.1 Site-directed mutagenesis study

Structure based alignment was applied to three dNKs, three TMPKs, HSV1-TK and EHV-TK to find the non-conserved residues in conserved regions that are relevant to catalysis (Figure 5.2). In theory, residues that are conserved in one family but not in the other are more likely to contribute to the different substrate specificity. Based on this principle, four highly conserved motifs around the active site were identified: the P-loop, the DRH motif, the X(P/R)EP motif and the Lid. Highlighted in red are the residues of interest.

5.2.1.1 Mutating the P-loop

The most important role of the P-loop is to accommodate a phosphoryl donor (ATP or GTP) by binding and positioning the α - and β -phosphate groups through backbone contacts involving conserved glycine residues and by the hydroxyl oxygen of a conserved serine or threonine. Aligning residues in this region reveals the difference between TK and TMPK. TMPKs have an additional hydrogen bond with the 3'-hydroxyl of the deoxyribose of the phosphate acceptor provided by a conserved glutamate /aspartate residue. An Ile residue occupies this position in dNKs, and the 3'-hydroxyl group hydrogen bonds with two residues from other regions - a glutamate from the lid region and a tyrosine residue from $\alpha 3$. In HSV1-TK, this position has a histidine residue. A hydrogen bond with the 3'-hydroxyl group of the phosphate acceptor is presumably formed with only a glutamate from the lid region (Fig 5.3).

Structure based sequence alignment

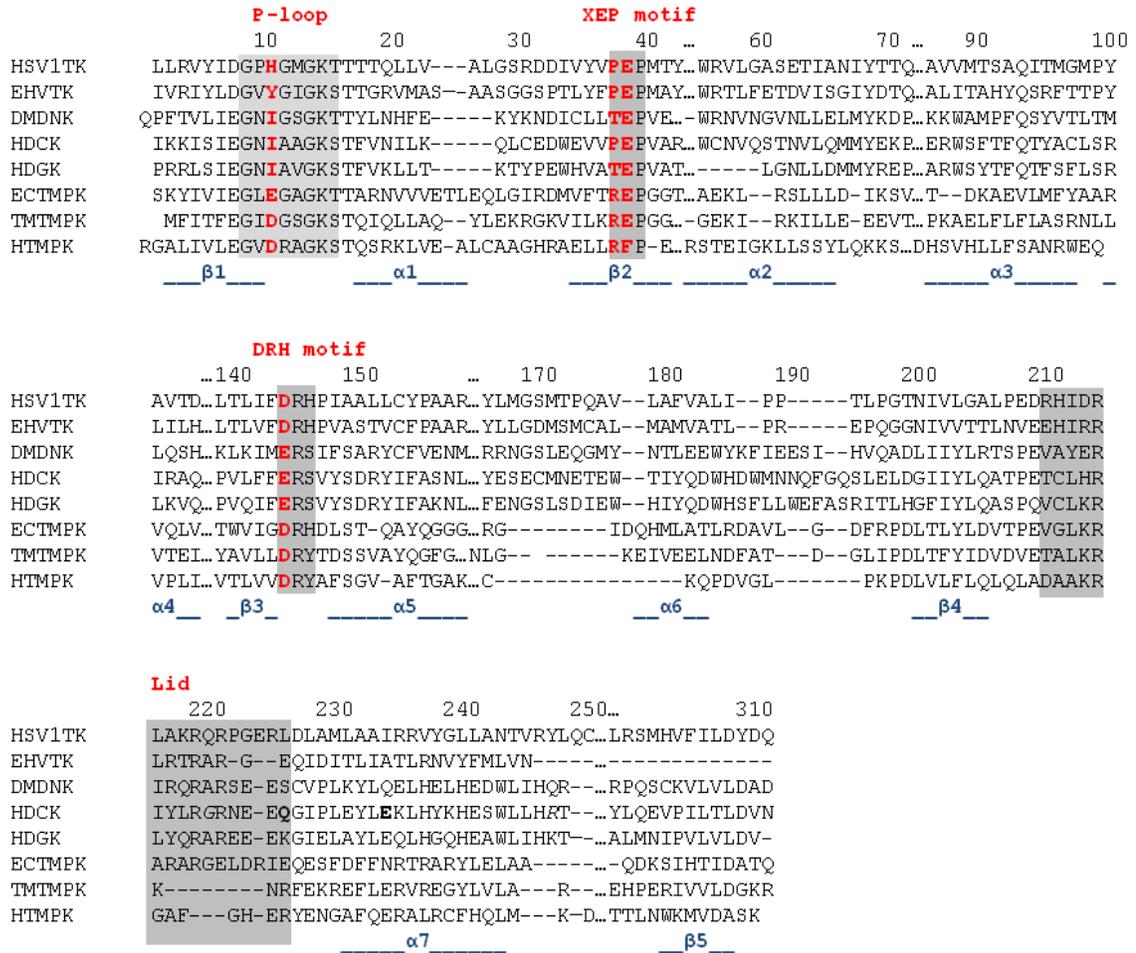


Figure 5.2 Structure based sequence alignment of dNKs, TMPKs, HSV1-TK and EHV-1 TK using the Combinatorial Extension (CE) Method (Shindyalov and Bourne, 1998). The catalytic regions are colored in grey, and the non-conserved residues are highlighted in red.

In a mutagenesis study, this histidine residue was mutated to Leucine, which caused a large increase of K_M for thymidine (~600 fold) and a decrease of k_{cat} (~60 fold) (Pilger et al., 1999). We hypothesize that the different hydrogen bonding pattern of the 3'-OH group in these enzymes may be responsible for the different substrate specificity because positioning of the deoxyribose directly influences the position of the reacting phosphate group and its distance to catalytic residues. Therefore, an I29D mutation was introduced in *DmdNK* to add an aspartate residue to the P-loop so that both thymidine and TMP can form a hydrogen bond through their 3'-OH group either with the glutamate-tyrosine pair or the aspartate in P-loop. However, kinetic measurements show that this mutation is lethal to the enzyme. It completely destroyed the thymidine activity of *DmdNK* and did not introduce any TMP activity to the enzyme (Table 5.1). One possible explanation is that the repulsion between aspartate and glutamate residues caused them to be in the wrong orientation and hydrogen bonds with the substrate cannot form.

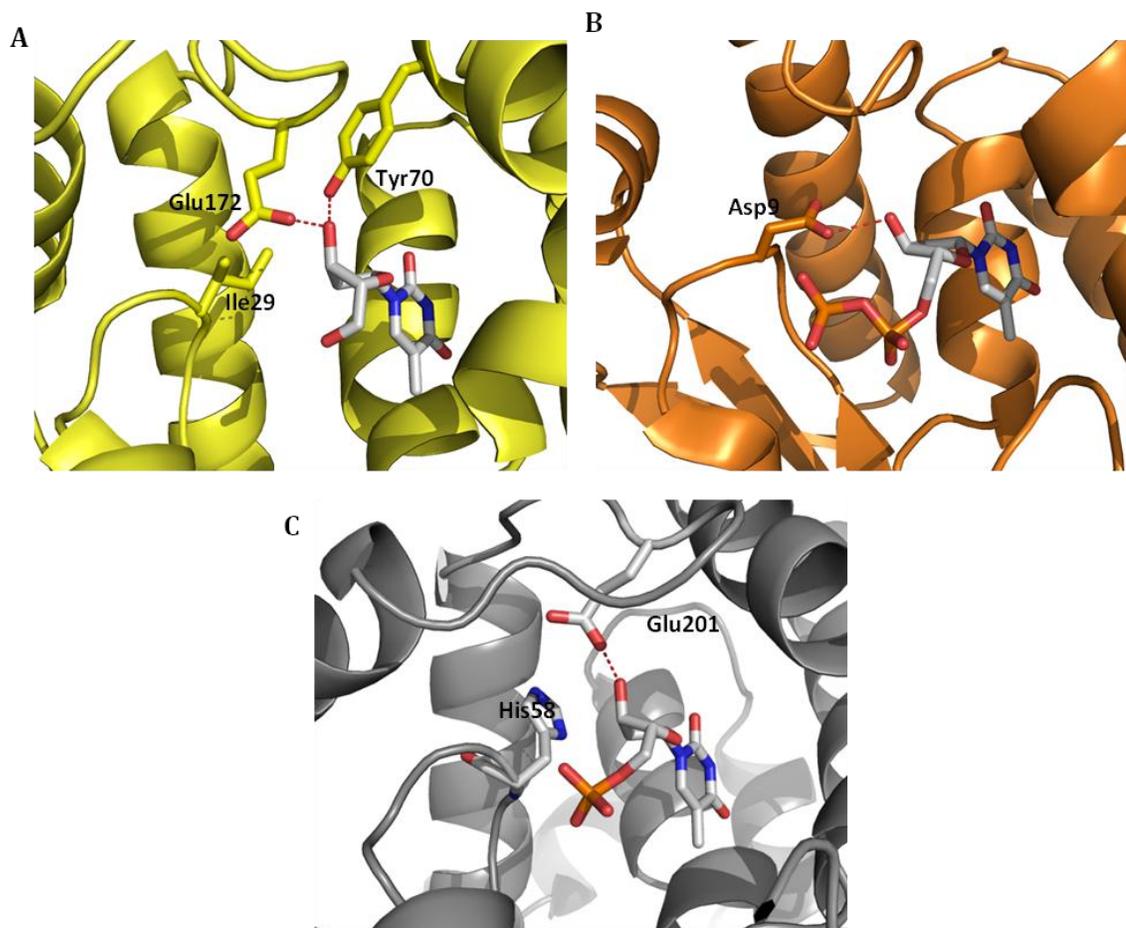


Figure.5.3 Comparison of nucleoside 3'-hydroxyl group hydrogen bonding with active site residues in (A) *DmdNK*, (B) *TmTMPK* and (C) HSV1-TK.

5.2.1.2 Mutating the DRH/ERS motif — *DmdNK_E104D* mutation

Another region of interest is the DRH motif, which is located adjacent to $\beta 3$ in the protein core. Sequence alignment shows that in TMPKs, HSV1-TK and EHV-TK (enzymes with TMP function) the sequence is DRH; but in dNKs, the sequence is ERS. The conserved arginine residue directly interacts with the phosphate group from the monophosphate in the phosphate acceptor site (Figure 5.4). The glutamate/aspartate residue is proposed to interact with a phosphate group of ATP through a Mg^{2+} ion. The different residue choice by dNK and TMPKs implies that the difference between glutamate and aspartate (a different side chain length by one carbon) might be the key to the different substrate specificity. Therefore, this glutamate residue was replaced with an aspartate in *DmdNK* to test if this mutation has any effect on substrate specificity. Kinetic assay shows that this mutant has very similar thymidine kinetic properties as the wild type *DmdNK* but no TMPK activity (Table 5.1). Overall this residue change is not a key player in determining thymidine and thymidylate activity.

Table 5.1 Kinetic properties of *DmdNK* mutants

	Thymidine			TMP
	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\times 10^6 \text{M}^{-1}\text{s}^{-1}$)	
<i>DmdNK</i>	1.9 \pm 0.3	17.5 \pm 0.6	9.1	n.d.
<i>DmdNK_I29D</i>	n.d	n.d	n.d	n.d.
<i>DmdNK_E104D</i>	0.86 \pm 0.05	6.9 \pm 0.04	7.9	n.d.

Thymidine activity was measured by a spectrophotometric assay, and the TMP activity was measured by a HPLC assay.

n.d.- not detected

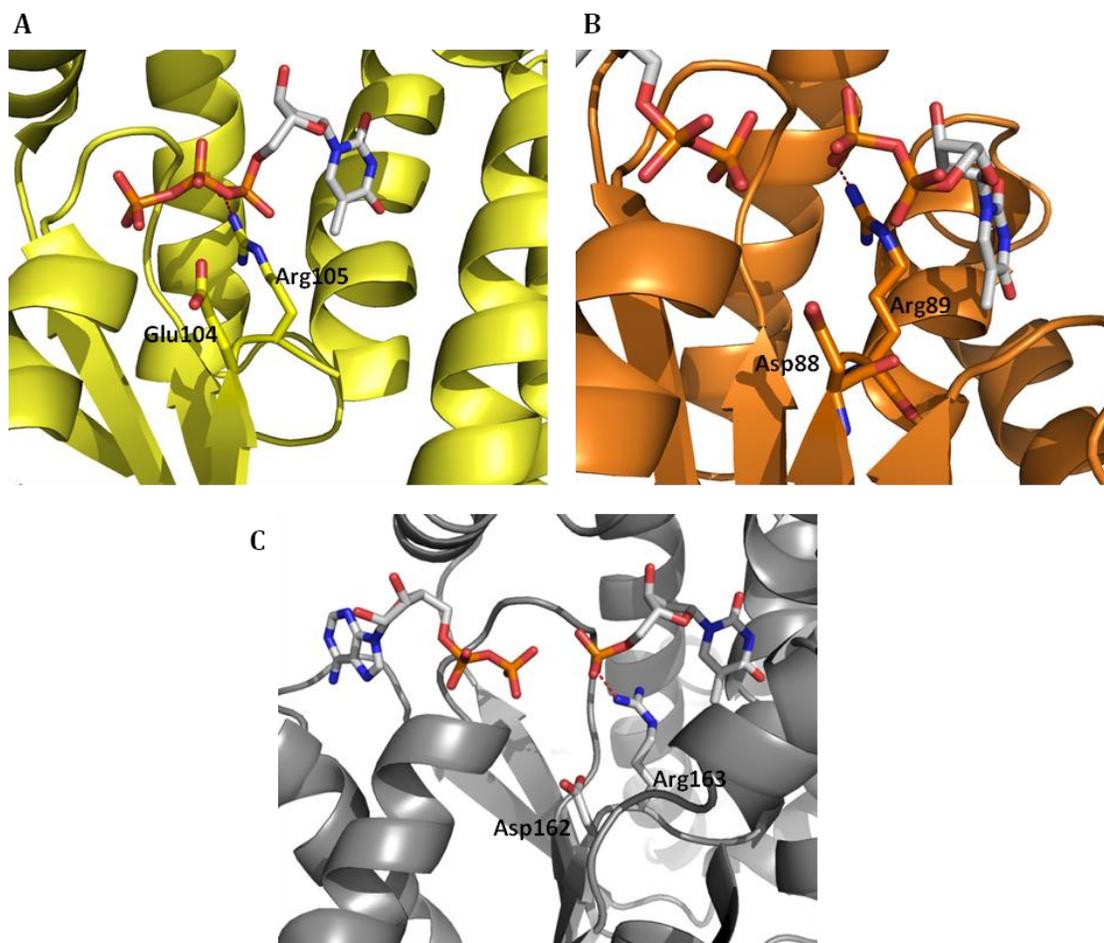


Figure 5.4 Comparison of the DRH/ERS motif interactions with substrates. (A) ERS motif in *DmdNK*, (B) DRH motif in *TmTMPK*, (C) DRH motif in HSV1-TK.

5.2.1.3 Mutating the XEP motif.

The XEP motif contains an important glutamate residue, which is conserved among all enzymes; however, the residue adjacent to it varies in the sequence alignment. The sequence is T/PEP in dNKs and HSV1-TK but REP in TMPKs. Revealed in the crystal structure, the conserved Glu residue forms a hydrogen bond with the 5'-hydroxyl group of the deoxyribose. It has been proposed that this residue acts as a general base to deprotonate the 5'-hydroxyl in the ester forming reaction catalyzed by TK. However, in the second step reaction (formation of an anhydride), the role of this Glu is unclear. In the crystal structures of TMPKs, the Glu residue rotates 90 degree away from the active site, which indicates that it is not required for the reaction (Fig 5.5). Therefore, in HSV1-TK, the same Glu residue must adopt a different orientation upon binding the monophosphate in the second step reaction. However, this prediction is not congruent with the crystal structures in complex with thymidine and TMP. The Glu residue adopts the same orientation.

To investigate the role of this residue, three different mutations were introduced to HSV1-TK at this position — E83A, E83D and E83S. If the Glu residue is only important for the first step reaction, then the mutants should maintain the ability to phosphorylate TMP. Protein expression and purification results show that all these mutants seem to be stable and have similar yield compared to wild type enzyme. However, the result from kinetic measurements is contrary to our prediction; all three mutants completely lost the activity for both thymidine and TMP indicating that this residue is important for both reactions (Table 5.2).

Another interesting observation from the sequence alignment is that at the position adjacent to Glu, TMPKs have a conserved Arg residue. From the crystal structures, this Arg residue is very important for binding the phosphate group of TMP. Combined with another Arg from the DRH motif, it forms a hydrogen bond with the phosphate group of TMP to bring the phosphate acceptor and donor in close proximity thus facilitating the phosphate transfer reaction. Therefore, we also introduced additional basic residues into HSV1-TK at this position to evaluate its impact on HSV1-TK activity. From the kinetic assay (Table 5.2), the P82A, P82K and P82R mutants have decreased TK activity mainly due to a ten-fold increase in K_M (Table 5.2). As the size of the mutated residue increases (Ala<Lys<Arg), the K_M of all three mutants also increases. The bulkier residue probably blocks substrate entrance to the active site. The kinetics of TMP activity also shows a decrease for the mutants. This change mainly comes from a four-fold decrease in k_{cat} although the K_M of all three mutants is about two fold higher than the wild type enzyme. These results suggest that the charged residues actually have a negative impact on the phosphate transfer reaction rather than facilitating the reaction as shown in TMPKs. Another explanation is that the mutants may have lower stability, which can also result in decreased of enzyme activity.

5.2.1.4 Conclusion of mutagenesis studies

In summary, the targeted residues identified by comparing the structure alignment of the two enzyme families are important for their distinct function. Mutation in *DmdNK* and HSV1-TK to the corresponding residues in *TmTMPK* results in a decrease or even abolishment of the TK activity. However, none of these mutations were able to introduce

or improve the TMPK activity. One possible explanation is that there are other regions in the enzymes that dictate substrate specificity for TMPK, for example, the lid region. Another explanation is that the determining factor exists in the global structure rather than specific residues. In HSV1-TK, the active site is more spacious than *DmdNK* and mimics TMPK, as shown by its ability to accommodate both TP₄A and TP₅A inhibitors.

Overall, site-directed mutagenesis in this case is difficult to correlate the structural perturbations to functional changes. Without rigorous biochemical and biophysical characterizations, it is hard to predict the exact role of each mutation, since a minor perturbation around the active site can cause a dramatic change in protein structure. Pre-steady state kinetic studies, substrate binding and computational models are some of the approaches that could guide the iterative structure-function study in the future.

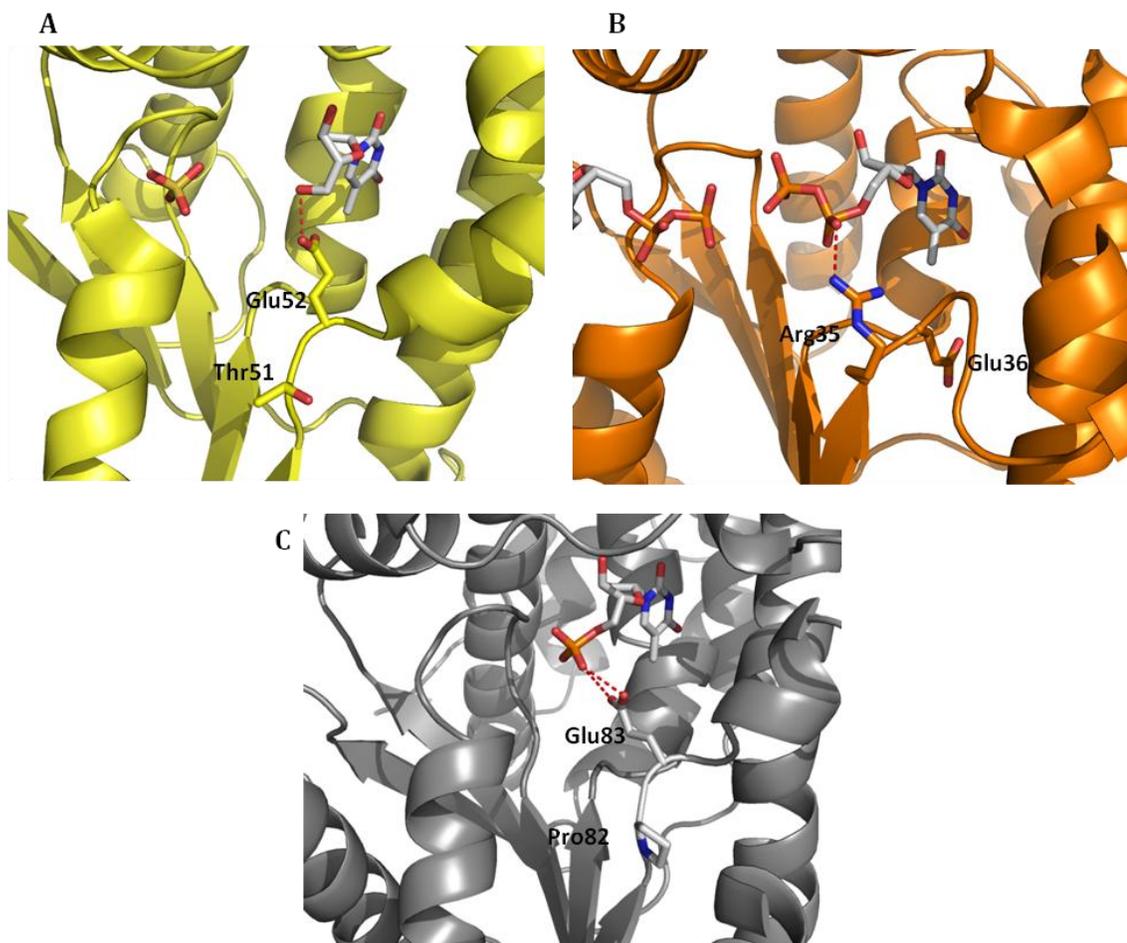


Figure 5.5 Comparison of XEP motif interactions with substrates in (A) *DmdNK*, (B) *TmTMPK* and (C) HSV1-TK.

Table 5.2 Kinetic properties of HSV1-TK mutants

	thymidine			TMP		
	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$)
wt	0.15±0.02	0.22±0.002	1.5×10^6	120±34	1.2±0.15	1.0×10^4
E83A	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E83D	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
E83S	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P82A	1.5±0.4	0.26±0.012	1.7×10^5	243±56	0.29±0.02	1.2×10^3
P82K	2.5±0.54	0.22±0.01	8.8×10^4	206±46	0.28±0.02	1.4×10^3
P82R	3.7±0.72	0.26±0.01	7.0×10^4	142±29	0.28±0.02	1.3×10^3

Thymidine activity was measured by a spectrophotometric assay, and TMP activity was measured by a HPLC assay.

n.d.-not detected

5.2.2 Calorimetric study of HSV1-TK, *DmdNK* and *TmTMPK*

5.2.2.1 *The model of enzyme-substrate ternary complex formation.*

In HSV1-TK, *DmdNK* and *TmTMPK*, the phosphate donor (ATP) and phosphate acceptor (thymidine or TMP) are located in two distinct binding sites. The formation of the ternary enzyme-substrate complex can either follow a sequential binding pathway or a random binding pathway (Figure 5.6). Perozzo *et. al.* proposed that HSV1-TK follows a sequential pathway, in which thymidine binds to the enzyme first followed by ATP (Perozzo *et al.*, 2000). In contrast, thymidylate kinase from *Streptococcus pneumonia* adopts a reversed sequential binding pathway whereby ATP binds to the enzyme first and TMP binds second (Petit and Koretke, 2002). A comprehensive thermodynamic study of ligand (substrate, product or inhibitor) binding and a detailed analysis of thermodynamics with structural insights can help decipher the substrate recognition of HSV1-TK, *DmdNK* and *TmTMPK*. Because substrate binding to the enzyme has a significant exothermic heat change, ITC was used to determine the thermodynamic parameters of binding for different ligands and to distinguish between sequential binding and random binding pathways in these three enzymes.

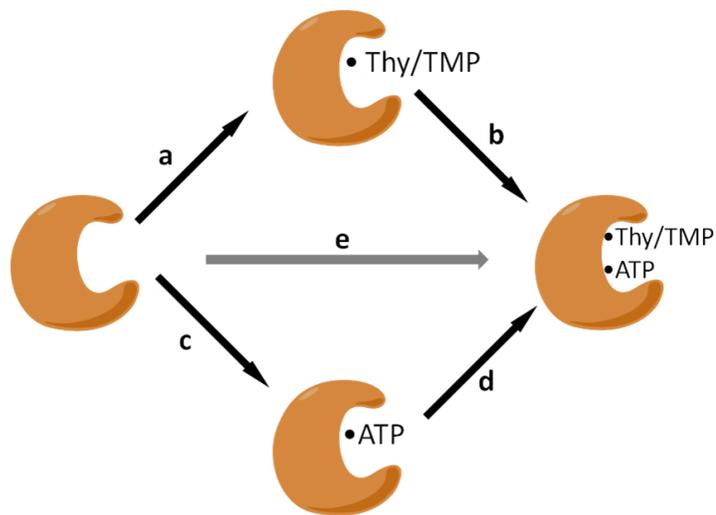


Figure 5.6 Mechanism of ternary enzyme-substrate complex formation. An ordered sequential pathway can either follow (a) plus (b) or (c) plus (d). A random binding mechanism is represented by (e), where the enzyme can bind substrates in any combination.

5.2.2.2 Determination of the thermodynamic parameters of thymidine and ATP binding

To characterize substrate recognition and probe the reaction mechanism, ITC studies of all three enzymes titrated with the substrates thymidine and ATP were performed. In order to prevent enzymatic turnover during ITC measurements, EDTA was added to chelate Mg^{2+} and a non-hydrolysable substrate analog adenosine 5'- γ -thiotriphosphate (ATP- γ S) was used instead of ATP. Figure 5.7 shows representative titrations of substrate to enzyme. The integrated heat of each injection is plotted against the molar ratio of substrate to enzyme binding site to calculate the thermodynamic parameters. All the parameters are summarized in Table 5.3.

It is interesting that HSV1-TK, *DmdNK* and *TmTMPK* all bind thymidine similarly. The binding is dominated by large favorable enthalpy and unfavorable entropy changes, which suggest significant hydrogen bond formation and increased conformational changes. The fact that *TmTMPK* also binds thymidine with a favorable enthalpy indicates that its active site residues probably can form the same hydrogen bonding interactions to the thymine base as the other enzymes.

For ATP, apo HSV1-TK does not show a typical binding isotherm, but both *DmdNK* and *TmTMPK* bind it. However, the binding signature is different when compared to thymidine binding. It has a less favorable change in enthalpy, which indicates less hydrogen bond formation, and favorable entropy change, which suggests hydrophobic interactions and a more ordered protein conformation forms. The different binding pattern suggests that ATP binds to an unfavorable binding pocket in the enzymes. The broad substrate specificity of *DmdNK* may explain how the enzyme can accommodate ATP, probably by binding it into its phosphate acceptor pocket.

For both HSV1-TK and *Dmd*NK, pre-incubation with thymidine increases the binding affinity for ATP, which is accompanied by a favorable enthalpy change favoring hydrogen bond formation, and an unfavorable entropy contribution corresponding to a conformational change. This behavior is consistent with the ordered sequential binding model reported in the literature (Perozzo et al., 2000). For *Tm*TMPK, the experiments did not show typical heat effects of binding for ATP. One possibility is that the conformational change induced by binding of thymidine is not suitable for ATP binding if *Tm*TMPK also follows a sequential binding mechanism. Another possibility is that *Tm*TMPK uses a different binding mechanism.

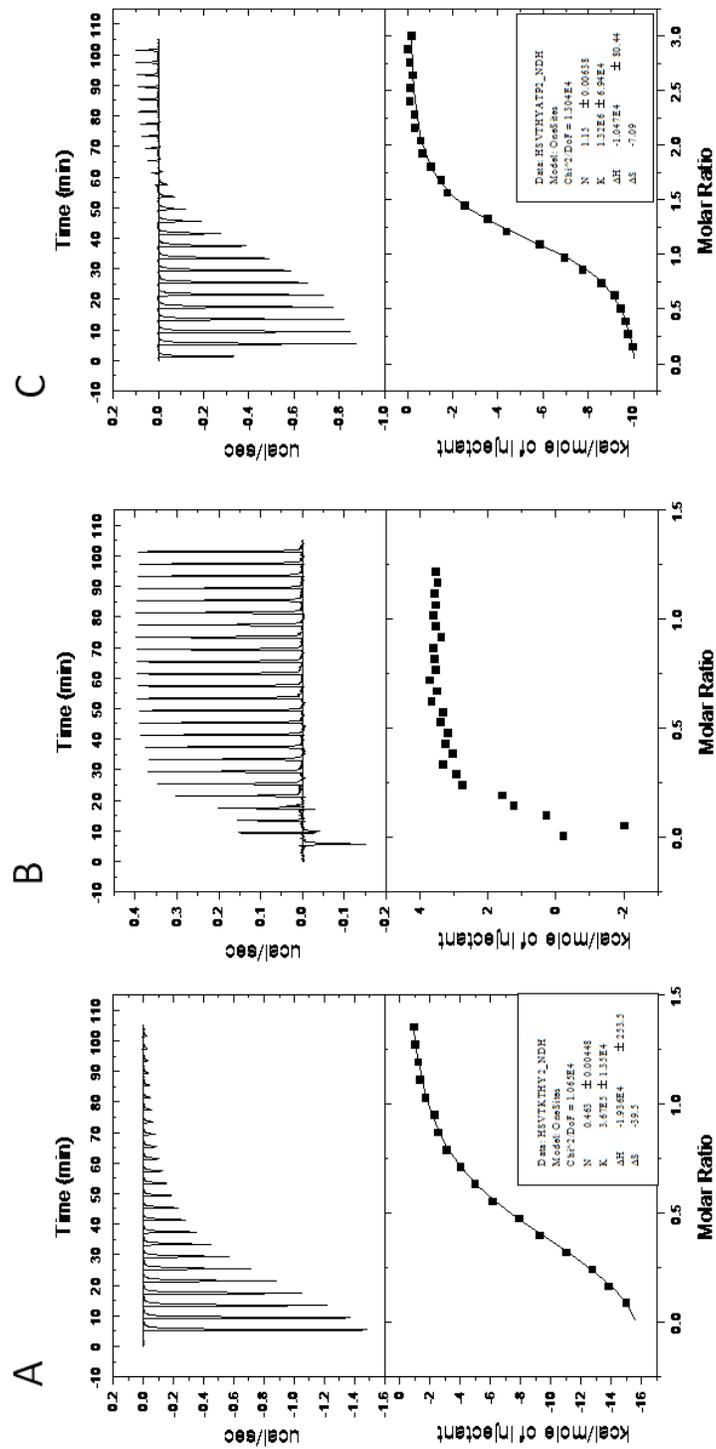


Figure 5.7 Representative isothermal titration calorimetry measurements of HSV1-TK without Mg^{2+} . HSV1-TK was titrated with 1 mM thymidine (A) or 1.0 mM ATP- γ S (B). HSV1-TK in the presence of 1 mM thymidine was titrated with 1 mM ATP- γ S (C). The upper panels are untreated data shown as differential power signals (μ cal /s). The lower panels are binding isotherms.

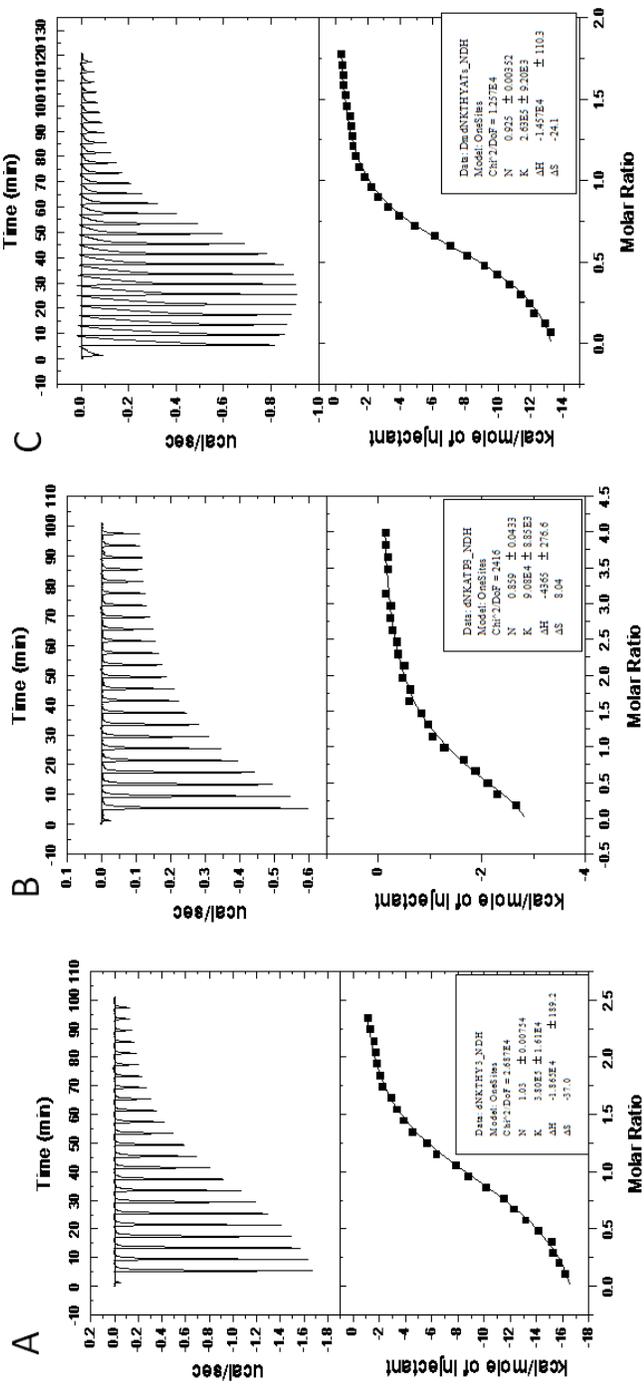


Figure 5.8 Representative isothermal titration calorimetry measurements of *DmdNK* without Mg^{2+} . *DmdNK* was titrated with 1 mM thymidine (A) or 1.0 mM ATP- γ S (B). *DmdNK* in the presence of 1 mM thymidine was titrated with 1 mM ATP- γ S (C). The upper panels are untreated data shown as differential power signals (μ cal/s). The lower panels are binding isotherms.

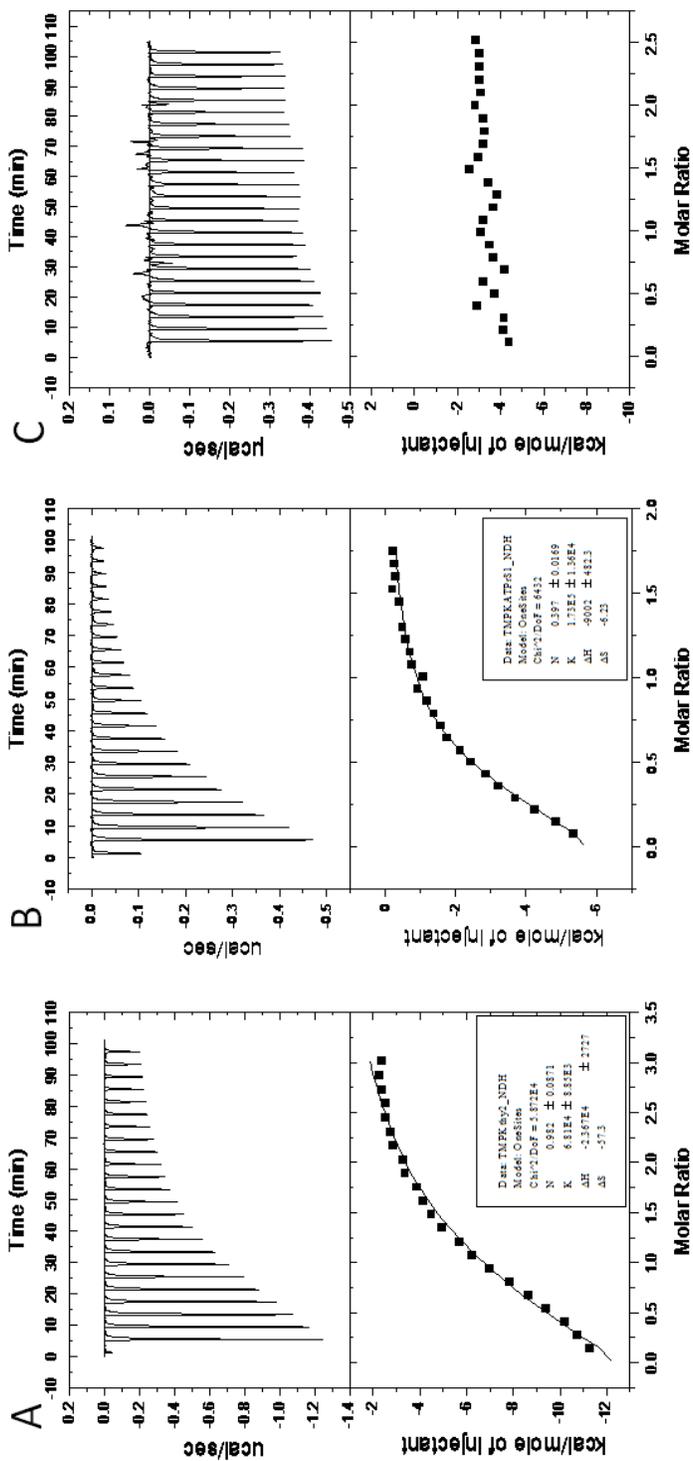


Figure 5.9 Representative isothermal titration calorimetry measurements of *Tm*TMPK without Mg^{2+} . *Tm*TMPK was titrated with 1 mM thymidine (A) or 1.0 mM ATP- γ S (B). *Tm*TMPK in the presence of 1 mM thymidine was titrated with 1 mM ATP- γ S (C). The upper panels are untreated data shown as differential power signals ($\mu\text{cal/s}$). The lower panels are binding isotherms.

Table 5.3 Thermodynamic parameters for the binding of thymidine and ATP to *DmdNK*, *HSV1-TK* and *TmTMPK*.

	thymidine					ATP					E:Thy+ATP					
	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)
<i>DmdNK</i>	2.7 \pm 0.1	-18.7 \pm 0.4	-9.9	-8.8	10.5 \pm 0.7	-4.1 \pm 0.2	-7.9	-4.1	3.3 \pm 0.4	-17.7 \pm 2.9	-9.1	-8.7				
<i>HSV1-TK</i>	2.9 \pm 0.2	-21.6 \pm 2.0	-12.9	-8.7	n.d.	n.d.	n.d.	n.d.	0.84 \pm 0.11	-10.6 \pm 0.10	-0.94	-9.6				
<i>TmTMPK</i>	15.9 \pm 1.3	-24.1 \pm 1.9	-16.5	-7.6	7.3 \pm 0.1	-11.0 \pm 2.2	-2.5	-8.1	n.d.	n.d.	n.d.	n.d.				

Values represent the mean of three experiments.

n.d. - not detectable

weak binding - the binding curve does not show sigmoid shape, thus can not fit to the Origin software.

5.2.2.3 Determination of the thermodynamic parameters of TMP binding.

TMP binds to all three enzymes. *Tm*TMPK binds TMP with favorable enthalpy and unfavorable entropy changes indicating considerable hydrogen bonding and conformational changes (Table 5.4). The signature is similar to those for thymidine binding. HSV1-TK and *Dmd*NK both bind TMP with moderate hydrogen bonding and hydrophobic interactions; the binding does not produce large conformational changes as shown by favorable entropy contributions. For thymidine kinases, TMP is the product of the reaction; therefore, this binding signature might be common for all enzyme product binding. It is not as favorable as the thymidine binding, which facilitates product release.

When all three enzymes were incubated with TMP, none of them show strong binding for ATP. HSV1-TK shows very weak binding by ITC experimentation; the curve does not show any sigmoidal shape, which prevents Origin from fitting the data correctly. *Dmd*NK does not display any binding characteristics. Surprisingly, *Tm*TMPK also does not show binding of ATP when the enzyme was incubated with TMP. The result is contrary to previous ITC studies of TMPK from *Streptococcus pneumonia*, which showed that TMPK adopts an ordered binding pathway in which ATP binds to the enzyme first and TMP second (Petit and Koretke, 2002).

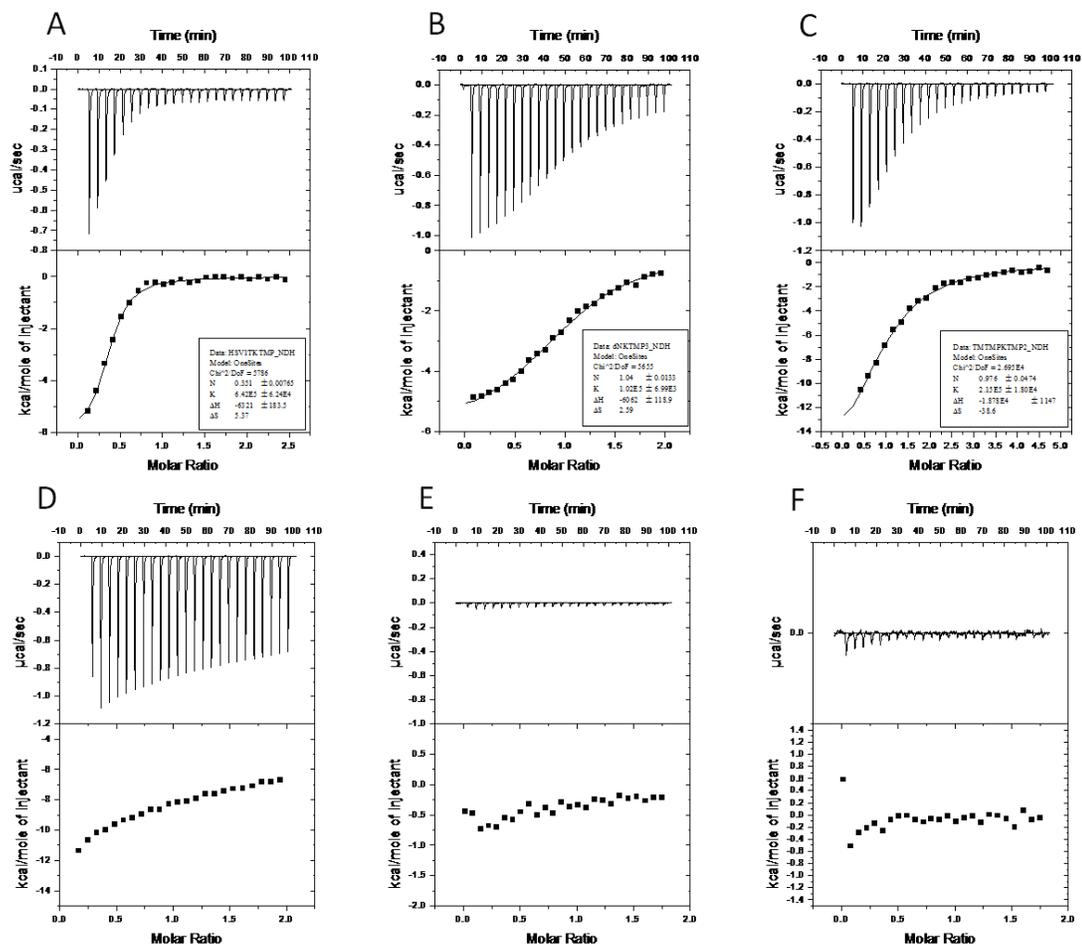


Figure 5.10 ITC experiments with TMP. (A) HSV1-TK titrated with TMP (B) *DmdNK* titrated with TMP (C) *TmTMPK* titrated with TMP; (D) HSV1-TK preincubated with TMP and titrated with ATP, (E) *DmdNK* preincubated with TMP and titrated with ATP, (F) *TmTMPK* preincubated with TMP and titrated with ATP.

Table 5.4 Thermodynamic parameters for the binding of TMP and ATP to *DmdNK*, *HSV1-TK* and *TmTMPK*.

	TMP				ATP				E:TMP+ATP			
	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)	Kd (μ M)	Δ H (kcal/mol)	T Δ S (kcal/mol)	Δ G (kcal/mol)
<i>DmdNK</i>	9.5 ± 1.7	-6.2 ± 0.4	1.7	-7.9	10.5 ± 0.7	-4.1 ± 0.2	-7.9	-4.1	n.d.	n.d.	n.d.	n.d.
<i>HSV1-TK</i>	1.7 ± 0.1	-5.5 ± 0.7	3.6	-9.1	n.d.	n.d.	n.d.	n.d.	n.d.	weak binding		
<i>TmTMPK</i>	4.5 ± 0.3	-16.8 ± 1.7	-8.4	-8.5	7.3 ± 0.1	-11.0 ± 2.2	-2.5	-8.1	n.d.	n.d.	n.d.	n.d.

Values represent the mean of three experiments.

n.d.- not detectable

weak binding - the binding curve does not show sigmoid shape, thus cannot be fit to the Origin software

5.2.2.4 Role of the Magnesium ion.

Based on the important role of the magnesium ion in both reactions catalyzed by TK and TMPK, it may also be imperative for substrate binding. It could influence the substrate binding by either direct participation in the bonding network in the binding pocket or by indirect structural rearrangements. For this reason all the experiments were repeated in HEPES buffer with 10 mM Mg^{2+} for comparison. The data is summarized in Table 5.5.

For *DmdNK*, binding of thymidine and TMP in the presence of Mg^{2+} do not show much difference when compared to previous experiments in the absence of Mg^{2+} . However when Mg^{2+} is present, the apo enzyme does not bind ATP, which is consistent with the sequential binding mechanism observed in HSV1-TK. This suggests that Mg^{2+} might facilitate the process by directing substrate binding to the correct enzyme site.

For HSV1-TK, thymidine and TMP binding in the presence of Mg^{2+} also show similar results when compared to previous experiments. There is a slight increase in favorable enthalpy contributions in TMP binding (from -5.5 kcal/mol to -9.2 kcal/mol), which suggests that in the presence of Mg^{2+} TMP forms more or stronger hydrogen bonds with residues in the active site. In comparing the two thymidine pre-incubation ATP binding experiments (with or without Mg^{2+}), fewer hydrogen bonds form when Mg^{2+} is present. One explanation could be that ATP forms interactions with Mg^{2+} as shown in the crystal structure instead of directly hydrogen bonding to residues in the enzyme.

For *TmTMPK*, binding of thymidine and TMP when Mg^{2+} is present shows nearly the same result, which suggests that Mg^{2+} does not play an important role in phosphate acceptor binding. It is also interesting to see that in the presence of Mg^{2+} , the apo enzyme

does not bind ATP anymore. This suggests that Mg^{2+} helps direct phosphate donor binding to the correct pocket, similar to the case of *DmdNK*.

5.2.2.5 Binding of thymidine triphosphate (TTP)

TTP was reported to function as a feedback inhibitor for a majority of thymidine kinases. To investigate the hypothesis that the reason HSV1-TK cannot bind ATP is due to its inability to accommodate ATP in the phosphate acceptor site, TTP titration experiments were carried out on all three enzymes. As expected, HSV1-TK binds TTP with a high affinity. When Mg^{2+} is present, the binding is accompanied with more favorable enthalpy changes thus more hydrogen bond formation. For *DmdNK*, it is interesting that without Mg^{2+} , the enzyme does not bind TTP. When Mg^{2+} is present, *DmdNK* binds TTP with a high affinity of 8.4 nM. For *TmTMPK*, the binding of TTP also shows higher affinity when Mg^{2+} is present. It is clear that Mg^{2+} has an important role in substrate binding.

Table 5.5 ITC measurements in the presence of Mg²⁺

		Kd (μM)	ΔH(kcal/mol)	TΔS(kcal/mol)	ΔG(kcal/mol)
<i>DmdNK</i>	Thymidine	8.5±0.3	-20.3±0.4	-12.3	-8.0
	TMP	11.0±1.9	-6.4±0.4	1.4	-7.8
	ATP	n.d.			
	E•Thy +ATP	2.3±0.3	-4.7±0.7	4.2	-8.9
<i>HSV1-TK</i>	Thymidine	2.7±0.2	-21.8±2.0	-13.0	-8.8
	TMP	1.2±0.1	-9.2±0.4	0.16	-9.3
	ATP	n.d.	n.d.	n.d.	n.d.
	E•Thy +ATP	1.1±0.1	-3.3±0.3	6.1	-9.4
	E•TMP +ATP		weak	binding	
<i>TmTMPK</i>	Thymidine	2.4	-11.3	-2.4	-8.9
	TMP	3.3±0.2	-10.2±0.9	-1.5	-8.7
	ATP	n.d.			
	E•TMP+ATP		weak	binding	

Values represent the mean of three measurements.

n.d.- not detectable

weak binding - the binding curve does not show a sigmoidal shape, thus cannot fit using the Origin software.

Table 5.6 ITC measurements of TTP

		Kd (μ M)	Δ H(kcal/mol)	T Δ S(kcal/mol)	Δ G(kcal/mol)
no Mg ²⁺	HSV1-TK	0.75 \pm 0.1	-2.5 \pm 0.3	7.2	-9.7
	<i>DmdNK</i>	n.d.	n.d.	n.d.	n.d.
	<i>TmTMPK</i>	0.49 \pm 0.04	-10.2 \pm 0.9	-0.23	-9.97
Mg ²⁺	HSV1-TK	0.81 \pm 0.1	-6.8 \pm 0.7	2.8	-9.6
	<i>DmdNK</i>	0.0084 \pm 0.00007	-4.6 \pm 0.5	8.2	-12.8
	<i>TmTMPK</i>	0.14 \pm 0.01	-5.1 \pm 0.2	5.7	-10.8

Values represent the mean of three measurements.

n.d.- not detectable

weak binding - the binding curve does not show a sigmoidal shape, thus cannot fit by the

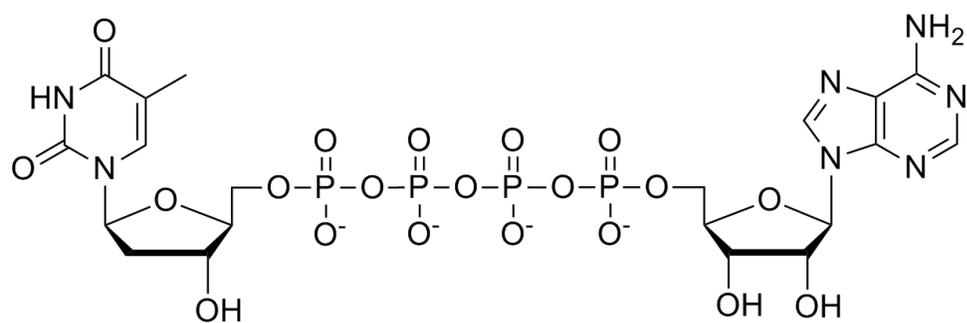
Origin software.

5.2.2.6 Determination of the thermodynamic parameters of the bi-substrate inhibitors

TP₄A and TP₅A

To investigate if the dual function of HSV1-TK is because it possesses a larger active site that can accommodate an extra phosphate group, ITC experiments using the bi-substrate inhibitors TP₄A and TP₅A were carried out for all three enzymes (Figure 5.8). These two inhibitors have been used to crystallize HSV1-TK; however, their binding affinity to HSV1-TK has not been reported (Gardberg et al., 2003a). If the space in the active site has a causal relationship with the size of substrate it can accommodate, then *DmdNK* would not show favorable bonding towards TP₅A, *TmTMPK* would not show favorable binding of TP₄A and HSV1-TK would favorably bind both. This hypothesis is partially supported by the ITC titration experimental results. *DmdNK* prefers TP₄A over TP₅A, binding TP₄A with a higher affinity and better hydrogen bonding indicated by a more favorable enthalpy contribution. HSV1-TK binds both inhibitors similarly, which supports the hypothesis that the dual function comes from the ability to accommodate the extra phosphate group. *TmTMPK* shows preference for TP₅A, as the binding of TP₅A produced more favorable enthalpy (better hydrogen bond formation) and a positive entropy indicating conformational changes.

A



B

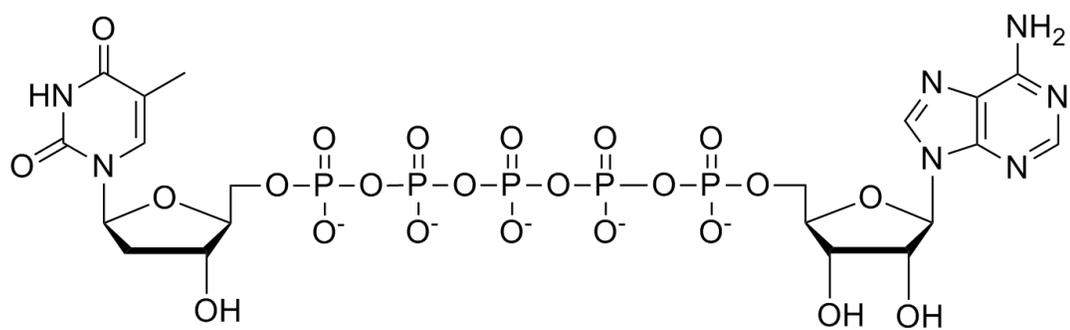


Figure 5.11 Bi-substrate inhibitor structures. (A) TP₄A and (B) TP₅A.

Table 5.7 ITC measurements of TP₄A and TP₅A

		Kd (μ M)	Δ H(kcal/mol)	T Δ S(kcal/mol)	Δ G(kcal/mol)
TP ₄ A	HSV1-TK	0.89	-6.4	3.2	-9.6
	<i>DmdNK</i>	1.0	-7.5	1.9	-9.4
	<i>Tm</i> TMPK	0.53	-9.4	0.52	-9.9
TP ₅ A	HSV1-TK	0.23	-8.4	2.1	-10.5
	<i>DmdNK</i>	2.2	-3.8	5.2	-9.0
	<i>Tm</i> TMPK	0.61	-12.0	-2.2	-9.8

Values represent one measurement due to limited ligand quantity.

5.2.2.7 Conclusion of ITC experiments

In summary, the ITC experiments suggest that *DmdNK*, HSV1-TK and *TmTMPK* follow the same sequential binding mechanism. The enzymes bind the phosphate acceptor (thymidine or TMP) first and the phosphate donor (ATP) second in the presence of Mg^{2+} . This finding is consistent with the previous studies of HSV1-TK in the literature (Perozzo et al., 2000). Comparing thymidine and TMP binding among all three enzymes shows that a favorable substrate binding is always accompanied with a large favorable enthalpy contribution, which indicates good hydrogen bond formation, and unfavorable entropy, which probably corresponds to a conformational change upon binding. Ligand binding without a favorable enthalpy seems to be indication of reaction product binding or binding to the wrong pocket in enzyme. Comparing the same titration experiments with or without Mg^{2+} suggests that Mg^{2+} plays an integral role in substrate binding. It probably helps the enzyme arrange the active site residues in a favorable orientation for substrate binding and directs the substrate especially the phosphate donor (ATP) to the correct binding pocket. Therefore, in the future, ITC titration experiments of TKs and TMPKs should always be carried out in the presence of Mg^{2+} to obtain correct enzyme binding information.

More importantly, comparing ITC titration experiments of enzymes with bi-substrate inhibitors TP₄A and TP₅A reveals that the determinant of HSV1-TK's dual function might be the size of its active site. It can accommodate both inhibitors with similar thermodynamics, but *DmdNK* and *TmTMPK* clearly show a preference for one of the two inhibitors when comparing the enthalpy contribution.

A major drawback of the ITC experiment is that it has a limited detection range for binding. It is only feasible for high affinity binding events. In several experiments, the binding is too weak (not displaying a sigmoidal curve) to be fit using the Origin software to generate the thermodynamic parameters. Other techniques measuring binding affinity (*i.e.*, surface plasmon resonance (SPR), fluorescence) could be helpful to measure the weaker binding events in these scenarios in the future.

5.3 Materials and methods

5.3.1 Materials

Enzymes were purchased from New England Biolabs (Beverly, MA) unless otherwise indicated. *Pfu* Turbo DNA polymerase (Stratagene, La Jolla, CA) was used for cloning. Pyruvate kinase and lactate dehydrogenase were purchased from Roche Biochemicals (Indianapolis, IN). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless noted.

5.3.2 Cloning and mutagenesis

The genes encoding *DmdNK* were isolated previously (Gerth et al., 2004). The HSV1-TK gene was kindly provided by Dr. Margret Black (Washington State University, Pullman, WA). Primers (HSV1TK_for and HSV1TK_rev) were used to amplify the gene and to clone it into the *Nde*I and *Spe*I restriction sites of pET14b (Novagen, Madison, WI) for overexpression and purification. All the mutations are introduced by overlap extension PCR, and the primers are listed in Table 5.8. All constructs were confirmed by DNA sequencing.

Table. 5.8 Primers used for plasmid construction and mutagenesis

Name	Sequence
t-dNK_for	5'-CGCCATATGAAGTACGCCGAGGGCACCC-3'
t-dNK_rev	5'-CGCACTAGTTCAGGGCTGTTGGTACTTGA-3'
dNKI29D	5'-ATCGAGGGCAACGACGGCAGCGGGAAGACC -3'
dNKE104D	5'-CTAAAAATAATGGACCGCTCCATTTTTAGC-3'
HSV1TK_for	5'-CGCCATATGGCTTCGTACCCCGGC-3'
HSV1TK_rev	5'-CGCACTAGTTTAGTTAGCCTCCCCATCTC-3'
HSVE83A	5'-GTCTACGTACCCGCGCCGATGACTTACTGG-3'
HSVE83D	5'-GTCTACGTACCCGACCCGATGACTTACTGG-3'
HSVE83S	5'-GTCTACGTACCCTCGCCGATGACTTACTGG-3'
HSVP82A	5'-ATCGTCTACGTAGCGGAGCCGATGACTTAC-3'
HSVP82K	5'-ATCGTCTACGTAAAAGAGCCGATGACTTAC-3'
HSVP82R	5'-ATCGTCTACGTACGCGAGCCGATGACTTAC-3'

5.3.3 Protein expression and purification

Escherichia coli BL21(DE3)pLysS cells (Novagen, Madison, WI) were transformed with pET14b vector containing mutated genes. Cells were grown at 37°C in LB broth containing ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml). At OD₆₀₀ ~ 0.6, isopropyl-b-D-thiogalactopyranoside (IPTG) was added to 0.4 mM (final concentration) to induce protein expression. The protein expression induction condition for each enzyme are: *DmdNK*, 30°C, 6h; HSV1-TK, 25°C, 8h; *TmTMPK* 37°C, 4h. The cells were harvested by centrifugation at 3000g at 4°C. The cell pellets were stored at -20°C.

For protein purification, cell pellets were suspended in the lysis buffer (50 mM KH₂PO₄, 150 mM sodium chloride, 10mM imidazole) The buffer at pH 7.5 was used for *DmdNK* and HSV1-TK while buffer at pH 8.0 was used for *TmTMPK*. Protease inhibitor cocktail, rLysozyme and Benzonase (Novagen, Madison, WI) were added according to the manufactures' protocols. Cells were lysed by sonication on ice, and the lysate was separated by centrifugation (10,000g, 4°C, 30 min). Lysate with 5–10 mg of protein was incubated with 1 ml of Ni-NTA resin at 4°C for 1 h, and the resin was separated by centrifugation and loaded to a polypropylene column (Bio-Rad, Hercules, CA). After three wash steps with 5 column volumes of lysis buffer containing 2 mM MgCl₂ and 10 mM ATP, 10 mM and 15 mM imidazole, the target proteins were eluted with lysis buffer containing 250 mM imidazole. Amicon-Ultra centrifugal filters (Amicon Bioseparations, Billerica, MA) were used to exchange the purified protein into storage buffer (50 mM KH₂PO₄, 150 mM NaCl, 5 mM MgCl₂ (pH 7.5 for *DmdNK* and HSV1-TK; pH 8.0 for *TmTMPK*)). Protein concentrations were quantified by measuring A₂₈₀ (*DmdNK* ε =

38,360 M⁻¹ cm⁻¹; HSV1-TK $\epsilon = 38,420 \text{ M}^{-1} \text{ cm}^{-1}$, *Tm*TMPK, $\epsilon = 14,650 \text{ M}^{-1} \cdot \text{cm}^{-1}$, calculated according to Gill and von Hippel (S.C. Gill, 1989)).

5.3.4 Thymidine kinase kinetic assay

Spectrophotometric assays to measure thymidine phosphorylation were performed as described previously (Materials and methods of chapter 2). All experiments were performed in triplicate. Data were fit to the Michaelis-Mention equation using Origin (OriginLab, Northhampton, MA).

5.3.5 Thymidylate kinase kinetic assay

The TMP kinase activity was measured by HPLC assay based on separation of nucleoside substrates and products. The main reason for using this assay instead of the coupled-enzyme assay is that the recycling of TDP by pyruvate kinase during the assay makes it difficult to determine the true reaction velocity (Blondin et al., 1994). Therefore, a direct measurement of reactant consumption and product formation would be more reliable to measure TMPK activity. The HPLC assay reaction was carried out in a final volume of 500 μl solution containing 1 mM DTT, 2.5 mM MgCl₂, 1 mM ATP, 50 mM Tris-HCl, and 50 mM KCl. Tryptophan at 200 mM concentration was used as internal standard. Substrate ranges from 10 to 1000 μM . To follow the enzymatic kinetics, 100 μl of reaction mixture is taken every minute after enzyme addition, and the reaction was quenched by adding 2% trichloroacetic acid (final concentration) and then neutralized to pH 7.0 by Tris-base. The concentration of nucleotide is measured via absorbance at 267 nm (a mean value between the maximum of absorption of purine nucleotides and that of

pyrimidine nucleotides) by HPLC using an Eclipse plus C18, 4.6×250 mm, 5 μm particle size column (Agilent Technologies, Santa Clara CA). A gradient of buffer A (10 mM tetrabutylammonium hydroxide, 10 mM KH₂PO₄ 0.25% methanol, pH 7.0) and buffer B (2.8 mM tetrabutylammonium hydroxide, 100 mM KH₂PO₄, 30% methanol, pH 5.5) was used for elution of nucleosides. Gradient was formed as follows: 0-5min from 0%-55% buffer B; 5-10 min from 55%-100% buffer B. The flow rate was 1.5 ml/min. The reaction velocities were calculated by monitoring reduction of TMP in terms of optical absorbance per minute at 267 nm (Di Pierro et al., 1995). Data were fit to the Michaelis-Menten equation using Origin (OriginLab, Northampton, MA).

5.3.6 Isothermal Titration Calorimetry

ITC experiments were carried out using the VP-ITC MicroCalorimeter (Microcal Inc., Northampton, MA). Experiments were performed at 25 °C in buffer A (50 mM Tris-HCl, 4 mM EDTA or buffer B (50 mM HEPES, 10 mM MgCl₂) at pH 7.5 for *DmdNK* and HSV1-TK and pH 8.0 for *TmTMPK*. Solutions of the protein were filled in the sample cell and titrated with substrate in the same buffer to minimize heat of dilution. The procedure of the titration experiment is: a first control injection of 1 μl, followed by 24 injections, each 5.5 μl of 15 s duration, with a 4 min interval in between. Raw data were collected and integrated using Origin (OriginLab, Northampton, MA). The data were fitted to a single-site binding model by a non-linear regression analysis to yield binding constants (K_b) and enthalpy of binding (ΔH), which are used to calculate free energy (ΔG) and entropy (ΔS).

Chapter 6 Conclusions and perspectives

6.1 Summary

In this dissertation I have described the development of a new selection system for TK and TMPK activity, the attempt to engineer enzymes with dual TK and TMPK capability, and the investigation of the mechanism for the dual function of HSV1-TK. First, to provide a good model for protein engineering, a complete biophysical and kinetic characterization of the TMPK from *Thermotoga maritima* (*Tm*TMPK) is described in chapter 2. *Tm*TMPK has robust activity for TMP and is similar to the activity of *E. coli* TMPK. It also has high thermostability ($T_m = 99^\circ\text{C}$), which makes it a good candidate for protein engineering since mutagenesis tends to destabilize enzymes.

Another requirement of enzyme engineering is the establishment of a suitable selection method. In chapter 3, I developed a novel selection system, a conditional auxotroph *E. coli* strain, for the selection of enzymes with dual TK and TMPK function. This is the first system reported to be able to select for consecutive reactions in the same cellular pathway. By placing TMPK under the arabinose promoter, the selection pressure for the second step phosphorylation can be easily tuned by changing the arabinose concentration. This selection system will be widely applicable to future engineering works focusing on dual TK and TMPK function.

Two model enzymes, *Dmd*NK and *Tm*TMPK, were chosen for directed evolution experiments in chapter 4. Given their high structural homology and low sequence identity, a non-homologous recombination technique SCRATCHY was applied to generate protein chimeras of the two enzymes. To improve the existing SCRATCHY protocol, I introduced SCHEMA, a computational program, to guide primer design in the multiple-crossover generation step. In the previous protocol, this step is completely

arbitrary, which renders the chimera library vulnerable of improper protein folding. SCHEMA calculations guide primer design to locations that produce minimal disruption in the protein structure upon recombination. Combining SCRATCHY and SCHEMA, several hybrid enzyme libraries with different crossovers were generated and submitted to selection. The results showed that although *Dmd*NK and *Tm*TMPK are structurally similar to the dual function enzyme HSV1-TK, chimeras made of these two enzymes do not possess dual activity.

To investigate the mechanism of the dual function of HSV1-TK, I carried out site-directed mutagenesis experiments that target non-conserved residues in conserved catalytic regions, which are identified from the sequence alignment of dNKs, TMPKs and HSV1-TK. Although I was unable to introduce TMPK function into *Dmd*NK or to improve the existing TMPK function in HSV1-TK, the identified residues are critical to their natural function in that most mutations are lethal to the enzyme. In addition, Isothermal Titration Calorimetry (ITC) was used to investigate the thermodynamics of substrate recognition and inhibitor binding in *Dmd*NK, *Tm*TMPK and HSV1-TK. Comparing thymidine and TMP binding among all three enzyme shows that a favorable substrate binding is always accompanied by a large favorable enthalpy contribution, which indicates good hydrogen bond formation, and unfavorable entropy, which probably corresponds to the conformational change upon binding the substrate. HSV1-TK's superior ability to bind both bi-substrate inhibitors, TP4A and TP5A, compared to *Dmd*NK and *Tm*TMPK suggests that a more flexible and spacious binding pocket might be the key to its dual functionality.

6.2 New selection system enables future protein engineering

One of the most important discoveries of this dissertation is the development of the novel selection system for TK and TMPK function either separately or concurrently, which opens the door to future engineering of multifunctional kinases. Positive genetic selection provides a rapid method to identify mutants with the desired function from combinatorial libraries. Although our SCRATCHY libraries did not produce a dual functional variant, other techniques can be applied to engineer TK and TMPK and the libraries can be easily assessed by the YL-1 strain under the dual selection condition.

Another aspect that has not been explored is the enhancement of the TMPK function of HSV1-TK to make it more robust towards both phosphorylation steps. In the past, studies on engineering HSV1-TK have primarily focused on the initial phosphorylation of nucleoside analogs such as ganciclovir and acyclovir (Black et al., 2001; Munir et al., 1993). Only a small number of these studies have focused on the impact of mutations on the second reaction, conversion of TMP to TDP. Although the conditional auxotroph YL-1 is mainly used as a dual functional selection tool in this dissertation, it also has the ability to select for TMPK function and can be used to assess the impact of specific amino acids on HSV1-TK's TMPK activity. With the controllable expression of chromosomal TMPK under the arabinose promoter, the selection pressure for TMPK function can be easily tuned as shown in chapter 3. Since HSV1-TK already has TMPK activity, directed evolution of an existing but residual function seems to be more promising than creating a completely new enzyme with dual activity. Applying directed evolution or rational approaches on HSV1-TK has a great potential to create a more robust enzyme for both phosphorylation reactions.

6.3 Future investigation of HSV1-TK's dual function

The real challenge for creating dual functional enzymes lies in a better understanding of the mechanism for the dual functionality of HSV1-TK. HSV1-TK was first discovered over 30 years ago, and investigation of its dual functionality has only been approached by crystallography. Both the studies by Wild *et. al* and Lavie *et.al* show that no substantial changes occurred in proteins crystallized with thymidine, TMP or bi-substrate inhibitors that mimic the two step reactions. The similar crystal structures could be an artifact of crystal packing. Envisioning the process of two reactions, there must be conformational changes in HSV1-TK to favor different substrate binding. HSV1-TK may also use different residues for each reaction since the reaction mechanisms are different; the first phosphorylation requires a general base, but the second reaction does not.

Site-directed mutagenesis studies described in this dissertation provide a starting point for future investigations. From ITC, we can now assess the impact of each mutation on substrate binding, which will help to explain the steady-state kinetic data. In addition, mutations in other catalytic regions such as the Lid region should also be evaluated. For the wild type HSV1-TK, combining ITC studies with protein dynamic techniques such as NMR and FT-IR will also help identify the residues that change conformation upon binding of thymidine or TMP, as well as inhibitors.

References:

- Abramoff, M.D., Magelhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International 11*, 36-42.
- Arigoni, F., Talabot, F., Peitsch, M., Edgerton, M.D., Meldrum, E., Allet, E., Fish, R., Jamotte, T., Curchod, M.L., and Loferer, H. (1998). A genome-based approach for the identification of essential bacterial genes. *Nat Biotechnol 16*, 851-856.
- Arner, E.S., and Eriksson, S. (1995). Mammalian deoxyribonucleoside kinases. *Pharmacol Ther 67*, 155-186.
- Arnon Lavie, M.K. (2004). Structural Requirements for Efficient Phosphorylation of Nucleotide Analogs by Human Thymidylate Kinase. *Mini-Reviews in Medicinal chemistry 4*, 351-359.
- Baldwin, S.A., Mackey, J.R., Cass, C.E., and Young, J.D. (1999). Nucleoside transporters: molecular biology and implications for therapeutic development. *Mol Med Today 5*, 216-224.
- Balzarini, J., Herdewijn, P., and De Clercq, E. (1989). Differential patterns of intracellular metabolism of 2',3'-didehydro- 2',3'-dideoxythymidine and 3'-azido-2',3'-dideoxythymidine, two potent anti-human immunodeficiency virus compounds. *J Biol Chem 264*, 6127-6133.
- Bellinzoni, M., Haouz, A., Grana, M., Munier-Lehmann, H., Shepard, W., and Alzari, P.M. (2006). The crystal structure of *Mycobacterium tuberculosis* adenylate kinase in complex with two molecules of ADP and Mg^{2+} supports an associative mechanism for phosphoryl transfer. *Protein Sci 15*, 1489-1493.

Bertrand, T., Briozzo, P., Assairi, L., Ofiteru, A., Bucurenci, N., Munier-Lehmann, H., Golinelli-Pimpaneau, B., Barzu, O., and Gilles, A.M. (2002). Sugar specificity of bacterial CMP kinases as revealed by crystal structures and mutagenesis of *Escherichia coli* enzyme. *J Mol Biol* 315, 1099-1110.

Black, M.E., Kokoris, M.S., and Sabo, P. (2001). Herpes Simplex Virus-1 Thymidine Kinase Mutants Created by Semi-Random Sequence Mutagenesis Improve Prodrug-mediated Tumor Cell Killing. *Cancer Res* 61, 3022-3026.

Black, M.E., and Loeb, L.A. (1993). Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants in vitro. *Biochemistry* 32, 11618-11626.

Black, M.E., Newcomb, T.G., Wilson, H.M., and Loeb, L.A. (1996). Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc Natl Acad Sci U S A* 93, 3525-3529.

Blondin, C., Serina, L., Wiesmuller, L., Gilles, A.M., and Barzu, O. (1994). Improved spectrophotometric assay of nucleoside monophosphate kinase activity using the pyruvate kinase/lactate dehydrogenase coupling system. *Anal Biochem* 220, 219-221.

Brown, S.H., Sjöholm, C., and Kelly, R.M. (1993). Purification and characterization of a highly thermostable glucose isomerase produced by the extremely thermophilic eubacterium, *Thermotoga maritima*. *Biotechnol Bioeng* 41, 878-886.

Brundiars, R., Lavie, A., Veit, T., Reinstein, J., Schlichting, I., Ostermann, N., Goody, R.S., and Konrad, M. (1999). Modifying Human Thymidylate Kinase to Potentiate Azidothymidine Activation. *J Biol Chem* 274, 35289-35292.

Chaperon, D.N. (2006). Construction and complementation of in-frame deletions of the essential *Escherichia coli* thymidylate kinase gene. *Appl Environ Microbiol* 72, 1288-1294.

Chen, M.S., and Prusoff, W.H. (1978). Association of thymidylate kinase activity with pyrimidine deoxyribonucleoside kinase induced by herpes simplex virus. *J Biol Chem* 253, 1325-1327.

Cheng, Y.C., and Prusoff, W.H. (1973). Mouse ascites sarcoma 180 thymidylate kinase. General properties, kinetic analysis, and inhibition studies. *Biochemistry* 12, 2612-2619.

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

Coulthard, S.A., Hogarth, L.A., Little, M., Matheson, E.C., Redfern, C.P., Minto, L., and Hall, A.G. (2002). The effect of thiopurine methyltransferase expression on sensitivity to thiopurine drugs. *Mol Pharmacol* 62, 102-109.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97, 6640-6645.

De Clercq, E. (2005). (E)-5-(2-bromovinyl)-2'-deoxyuridine (BVDU). *Med Res Rev* 25, 1-20.

Di Pierro, D., Tavazzi, B., Perno, C.F., Bartolini, M., Balestra, E., Calio, R., Giardina, B., and Lazzarino, G. (1995). An ion-pairing high-performance liquid chromatographic method for the direct simultaneous determination of nucleotides, deoxynucleotides, nicotinic coenzymes, oxypurines, nucleosides, and bases in perchloric acid cell extracts. *Anal Biochem* 231, 407-412.

Dipak K. Dube, J.D.P., David C. French, David S. Cahill, Syamalima Dube, Marshall S. Z. Horwitz, Khan M. Munir, Lawrence A. Loeb (1991). Artificial Mutants Generated by the Insertion of Random Oligonucleotides into the Putative Nucleoside Binding Site of the HSV-1 Thymidine Kinase Gene. *Biochemistry* 30, 11760-11767.

Eriksson, S., Munch-Petersen, B., Johansson, K., and Ecklund, H. (2002). Structure and function of cellular deoxyribonucleoside kinases. *Cellular and Molecular Life Sciences (CMLS)* 59, 1327-1346.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., *et al.* (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403.

Furman, P.A., Fyfe, J.A., St Clair, M.H., Weinhold, K., Rideout, J.L., Freeman, G.A., Lehrman, S.N., Bolognesi, D.P., Broder, S., Mitsuya, H., *et al.* (1986). Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proc Natl Acad Sci U S A* 83, 8333-8337.

Galmarini, C.M., Mackey, J.R., and Dumontet, C. (2002). Nucleoside analogues and nucleobases in cancer treatment. *Lancet Oncol* 3, 415-424.

Gardberg, A., Shuvalova, L., Monnerjahn, C., Konrad, M., and Lavie, A. (2003a). Structural Basis for the Dual Thymidine and Thymidylate Kinase Activity of Herpes Thymidine Kinases. *Structure* 11, 1265.

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

- Gerth, M.L., Patrick, W.M., and Lutz, S. (2004). A second-generation system for unbiased reading frame selection. *Protein Engineering, Design and Selection* 17, 595-602.
- Godsey, M.H., Ort, S., Sabini, E., Konrad, M., and Lavie, A. (2006). Structural basis for the preference of UTP over ATP in human deoxycytidine kinase: illuminating the role of main-chain reorganization. *Biochemistry* 45, 452-461.
- Guettari, N., Loubiere, L., Brisson, E., and Klatzmann, D. (1997). Use of herpes simplex virus thymidine kinase to improve the antiviral activity of zidovudine. *Virology* 235, 398-405.
- Hamzeh, F.M., and Lietman, P.S. (1991). Intranuclear accumulation of subgenomic noninfectious human cytomegalovirus DNA in infected cells in the presence of ganciclovir. *Antimicrob Agents Chemother* 35, 1818-1823.
- Hasegawa, Y., Emi, N., and Shimokata, K. (1995). Retroviral transfer of HSV1-TK gene into human lung cancer cell line. *J Mol Med* 73, 107-112.
- Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C., and Arnold, F. (2010). Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng Des Sel* 23, 871-880.
- Hible, G., Daalova, P., Gilles, A.M., and Cherfils, J. (2006). Crystal structures of GMP kinase in complex with ganciclovir monophosphate and Ap5G. *Biochimie* 88, 1157-1164.
- Hiraga, S., Igarashi, K., and Yura, T. (1967). A Deoxythymidine Kinase-Deficient Mutant of *Escherichia Coli*. *Biochimica et Biophysica Acta* 145, 41-51.

Huang, S.H., Tang, A., Drisco, B., Zhang, S.Q., Seeger, R., Li, C., and Jong, A. (1994). Human dTMP kinase: gene expression and enzymatic activity coinciding with cell cycle progression and cell growth. *DNA Cell Biol* 13, 461-471.

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

Jong, A.Y., and Campbell, J.L. (1984). Characterization of *Saccharomyces cerevisiae* thymidylate kinase, the CDC8 gene product. General properties, kinetic analysis, and subcellular localization. *J Biol Chem* 259, 14394-14398.

Jordheim, L.P., Galmarini, C.M., and Dumontet, C. (2006). Gemcitabine resistance due to deoxycytidine kinase deficiency can be reverted by fruitfly deoxynucleoside kinase, DmdNK, in human uterine sarcoma cells. *Cancer Chemother Pharmacol* 58, 547-554.

Kawarasaki, Y., Griswold, K.E., Stevenson, J.D., Selzer, T., Benkovic, S.J., Iverson, B.L., and Georgiou, G. (2003). Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. *Nucleic Acids Res* 31, e126.

Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L., and Keasling, J.D. (2001). Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* 147, 3241-3247.

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

Kong, W., Engel, K., and Wang, J. (2004). Mammalian nucleoside transporters. *Curr Drug Metab* 5, 63-84.

Lavie, A., Konrad, M., Brundiers, R., Goody, R.S., Schlichting, I., and Reinstein, J. (1998a). Crystal Structure of Yeast Thymidylate Kinase Complexed with the Bisubstrate Inhibitor P1-(5'-Adenosyl) P5-(5'-Thymidyl) Pentaphosphate (TP5A) at 2.0 Angstrom Resolution: Implications for Catalysis and AZT Activation. *Biochemistry* 37, 3677-3686.

Lavie, A., Ostermann, N., Brundiers, R., Goody, R.S., Reinstein, J., Konrad, M., and Schlichting, I. (1998b). Structural basis for efficient phosphorylation of 3'-azidothymidine monophosphate by *Escherichia coli* thymidylate kinase. *PNAS* 95, 14045-14050.

Lavie, A., Schlichting, I., Vetter, I.R., Konrads, M., Reinstein, J., and Goody, R.S. (1997). The bottleneck in AZT activation. *Nat Med* 3, 922-924.

Lee, L.S., and Cheng, Y. (1977). Human thymidylate kinase. Purification, characterization, and kinetic behavior of the thymidylate kinase derived from chronic myelocytic leukemia. *J Biol Chem* 252, 5686-5691.

Li, Y., Zhang, Y., and Yan, H. (1996). Kinetic and thermodynamic characterizations of yeast guanylate kinase. *J Biol Chem* 271, 28038-28044.

Liang, P., Averboukh, L., Zhu, W., Haley, T., and Pardee, A.B. (1995). Molecular characterization of the murine thymidylate kinase gene. *Cell Growth Differ* 6, 1333-1338.

Link, A.J., Phillips, D., and Church, G.M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J Bacteriol* 179, 6228-6237.

- Liu, L., Li, Y., Liotta, D., and Lutz, S. (2009). Directed evolution of an orthogonal nucleoside analog kinase via fluorescence-activated cell sorting. *Nucleic Acids Res* 37, 4472-4481.
- Lutz, S., and Benkovic, S.J. (2000). Homology-independent protein engineering. *Curr Opin Biotechnol* 11, 319-324.
- Lutz, S., Liu, L., and Liu, Y. (2009). Engineering Kinases to Phosphorylate Nucleoside Analogs for Antiviral and Cancer Therapy. *Chimia (Aarau)* 63, 737-744.
- Lutz, S., Ostermeier, M., and Benkovic, S.J. (2001a). Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res* 29, E16.
- Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. (2001b). Creating multiple-crossover DNA libraries independent of sequence identity. *Proc Natl Acad Sci U S A* 98, 11248-11253.
- Macchi, B., and Mastino, A. (2002). Pharmacological and biological aspects of basic research on nucleoside-based reverse transcriptase inhibitors. *Pharmacol Res* 46, 473-482.
- Maga, G., Focher, F., Wright, G.E., Capobianco, M., Garbesi, A., Bendiscioli, A., and Spadari, S. (1994). Kinetic studies with N2-phenylguanines and with L-thymidine indicate that herpes simplex virus type-1 thymidine kinase and thymidylate kinase share a common active site. *Biochem J* 302 (Pt 1), 279-282.
- Meyer, M.M., Hochrein, L., and Arnold, F.H. (2006). Structure-guided SCHEMA recombination of distantly related beta-lactamases. *Protein Eng Des Sel* 19, 563-570.

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

Miller, W.H., Daluge, S.M., Garvey, E.P., Hopkins, S., Reardon, J.E., Boyd, F.L., and Miller, R.L. (1992). Phosphorylation of carbovir enantiomers by cellular enzymes determines the stereoselectivity of antiviral activity. *J Biol Chem* 267, 21220-21224.

Mohri, H., Singh, M.K., Ching, W.T., and Ho, D.D. (1993). Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the blood of treated and untreated patients. *Proc Natl Acad Sci U S A* 90, 25-29.

Monnerjahn, C., and Konrad, M. (2003). Modulated Nucleoside Kinases as Tools to Improve the Activation of Therapeutic Nucleoside Analogues. *ChemBioChem* 4, 143-146.

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

Munch-Petersen, B., Piskur, J., and Sondergaard, L. (1998). Four Deoxynucleoside Kinase Activities from *Drosophila melanogaster* Are Contained within a Single Monomeric Enzyme, a New Multifunctional Deoxynucleoside Kinase. *J Biol Chem* 273, 3926-3931.

Munir, K.M., French, D.C., and Loeb, L.A. (1993). Thymidine kinase mutants obtained by random sequence selection. *Proc Natl Acad Sci U S A* 90, 4012-4016.

Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007). A simple selection strategy for evolving highly efficient enzymes. *Nat Biotechnol* 25, 1145-1147.

Ostermann, N., Schlichting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T., Goody, R.S., and Lavie, A. (2000). Insights into the phosphoryltransfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure* 8, 629-642.

Ostermeier, M., Nixon, A.E., Shim, J.H., and Benkovic, S.J. (1999). Combinatorial protein engineering by incremental truncation. *Proc Natl Acad Sci U S A* 96, 3562-3567.

Otey, C.R., Landwehr, M., Endelman, J.B., Hiraga, K., Bloom, J.D., and Arnold, F.H. (2006). Structure-guided recombination creates an artificial family of cytochromes P450. *PLoS Biol* 4, e112.

Perozzo, R., Jelesarov, I., Bosshard, H.R., Folkers, G., and Scapozza, L. (2000). Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study. *J Biol Chem* 275, 16139-16145.

Petit, C.M., and Koretke, K.K. (2002). Characterization of *Streptococcus pneumoniae* thymidylate kinase: steady-state kinetics of the forward reaction and isothermal titration calorimetry. *Biochem J* 363, 825-831.

Pilger, B.D., Perozzo, R., Alber, F., Wurth, C., Folkers, G., and Scapozza, L. (1999). Substrate Diversity of Herpes Simplex Virus Thymidine Kinase. IMPACT OF THE KINEMATICS OF THE ENZYME. *J Biol Chem* 274, 31967-31973.

Rai, K.R. (1998). Cladribine for the treatment of hairy cell leukemia and chronic lymphocytic leukemia. *Semin Oncol* 25, 19-22.

Rhoads, D.G., and Lowenstein, J.M. (1968). Initial velocity and equilibrium kinetics of myokinase. *J Biol Chem* 243, 3963-3972.

Richard, J.P., Carr, M.C., Ives, D.H., and Frey, P.A. (1980). The stereochemical course of thiophosphoryl group transfer catalyzed by adenosine kinase. *Biochem Biophys Res Commun* 94, 1052-1056.

Russell, J.S.a.D.W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd ed.

S.C. Gill, P.H.v.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* 182, 319-326.

Sales, S.D., Hoggard, P.G., Sunderland, D., Khoo, S., Hart, C.A., and Back, D.J. (2001). Zidovudine phosphorylation and mitochondrial toxicity in vitro. *Toxicol Appl Pharmacol* 177, 54-58.

Samuels, D.C. (2006). Mitochondrial AZT metabolism. *IUBMB Life* 58, 403-408.

Schelling, P., Folkers, G., and Scapozza, L. (2001). A Spectrophotometric Assay for Quantitative Determination of *k_{cat}* of Herpes Simplex Virus Type 1 Thymidine Kinase Substrates. *Analytical Biochemistry* 295, 82-87.

Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* 16, 559-565.

Sclafani, R.A., and Fangman, W.L. (1984). Yeast gene CDC8 encodes thymidylate kinase and is complemented by herpes thymidine kinase gene TK. *Proc Natl Acad Sci U S A* 81, 5821-5825.

Sekulic, N., Shuvalova, L., Spangenberg, O., Konrad, M., and Lavie, A. (2002). Structural characterization of the closed conformation of mouse guanylate kinase. *J Biol Chem* 277, 30236-30243.

Shindyalov, I.N., and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11, 739-747.

Siegele, D.A., and Hu, J.C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U S A* 94, 8168-8172.

Squires, K.E. (2001). An introduction to nucleoside and nucleotide analogues. *Antivir Ther* 6 Suppl 3, 1-14.

Stolworthy, T.S., Krabbenhoft, E., and Black, M.E. (2003). A novel *Escherichia coli* strain allows functional analysis of guanylate kinase drug resistance and sensitivity. *Anal Biochem* 322, 40-47.

Su, J.Y., and Sclafani, R.A. (1991). Molecular cloning and expression of the human deoxythymidylate kinase gene in yeast. *Nucleic Acids Res* 19, 823-827.

Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A., and Andersson, S.G. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* 296, 2376-2379.

Tornevik, Y., Ullman, B., Balzarini, J., Wahren, B., and Eriksson, S. (1995). Cytotoxicity of 3'-azido-3'-deoxythymidine correlates with 3'-azidothymidine-5'-monophosphate (AZTMP) levels, whereas antihuman immunodeficiency virus (HIV) activity correlates with 3'-azidothymidine-5'-triphosphate (AZTTP) levels in cultured CEM T-lymphoblastoid cells. *Biochemical Pharmacology* 49, 829-837.

Turriziani, O., Schuetz, J.D., Foher, F., Scagnolari, C., Sampath, J., Adachi, M., Bambacioni, F., Riva, E., and Antonelli, G. (2002). Impaired 2',3'-dideoxy-3'-thiacytidine accumulation in T-lymphoblastoid cells as a mechanism of acquired resistance

independent of multidrug resistant protein 4 with a possible role for ATP-binding cassette C11. *Biochem J* 368, 325-332.

Van Rompay, A.R., Johansson, M., and Karlsson, A. (1999). Phosphorylation of deoxycytidine analog monophosphates by UMP-CMP kinase: molecular characterization of the human enzyme. *Mol Pharmacol* 56, 562-569.

Van Rompay, A.R., Johansson, M., and Karlsson, A. (2000). Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacol Ther* 87, 189-198.

Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L., and Arnold, F.H. (2002). Protein building blocks preserved by recombination. *Nat Struct Biol* 9, 553-558.

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

Wild, K., Bohner, T., Folkers, G., and Schulz, G.E. (1997). The structures of thymidine kinase from herpes simplex virus type 1 in complex with substrates and a substrate analogue. *Protein Sci* 6, 2097-2106.

Yan, H., and Tsai, M.D. (1999). Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Adv Enzymol Relat Areas Mol Biol* 73, 103-134, x.

Yoshikawa, S., Nakagawa, N., Shirouzu, M., Yokoyama, S., Kuramitsu, S. (2009). Crystal structure of thymidylate kinase in complex with dTDP and ADP from *Thermotoga maritima*