STRUCTURAL STUDIES OF THE S-ADENOSYLMETHIONINE-DEPENDENT METHYLTRANSFERASES

by

YI PENG

Submitted in partial fulfillment of the requirements

For the degree of Doctor of Philosophy

Dissertation Advisor: Dr.Vivien C. Yee

Department of Biochemistry

CASE WESTERN RESERVE UNIVERSITY

January, 2009

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Yi Peng

candidate for the Ph.D. degree *.

(signed) Menachem Shoham

(chair of the committee)

Vivien Yee

Jun Qin

Thomas Gerken

Paul Carey

(date) 11/19/2008

*We also certify that written approval has been obtained for any proprietary material contained therein.

To my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

It would not have been possible to write this doctoral thesis without the help and support of the nice people around me in the past five years, to only some of whom it is possible to give particular mention here.

First, I am extremely grateful to my advisor, Vivien Yee, for her invaluable help, support and advice on both an academic and a personal level. She taught me how to approach the research questions and write academic papers, brought out the good advice in me by her unsurpassed knowledge of crystallography, encouraged me to improve the ability to communicate in English, and had confidence in me. She is not only a great mentor, but also a friend. She made the lab a comfortable workplace and home by the interesting discussions and being friendly to be with the people in the lab. In the past five years with her, I went through the most important period of my life full of joy and sadness: getting married, reuniting with my husband after two-year separation, becoming a new mother, etc. Fortunately, I have been to her student because she shared my happiness and helped me get over the most difficult time since I came to the USA from China.

Besides my advisor, I would like to thank the rest of my thesis committee: Drs Menachem Shoham, Thomas Gerken and Jun Qin for their helpful discussion and insightful comments about my thesis as well as their friendship and encouragement. Thanks also go to Dr. Focco van den Akker for his support, encouragement and good questions.

Let me also say ‘thank you’ to the following people at Mayo Clinic College of Medicine-Mayo Clinic: Dr. Richard M. Weinshilboum for the collaboration. Qiping Feng

and Araba A. Adjei from the Weinshilboum lab for their hard work on Thiopurine S-methyltransferase activity assays. I also want to thank Dr. Monica Emanuelli from Institute of Biochemical Biotechnologies, Italy for the collaboration on the study of nicotinamide N-methyltransferase.

A special thank goes to Dennis Wilk, who made great effort to purify and provide me with the protein for crystallization. Many thanks also go to former and current colleagues in Dr. Yee and Dr. van den Akker's lab: Laura Fromondi and Wei Ke for the enjoyable lunch talks. Seungjoo Lee for distracting me joyfully. Pamela Hall, Allison Miketa, Rune Hartmann and Xiaolei Ma for helping me start my study and life in the lab when I joined the lab. Priyaranjan Pattanaik, Pius Padayatti, Kaustubha Qanungo, Rajaganapathi Jagannathan, Jared Sampson, Elena Filaretova, Elizabeth Rodkey and Derek Veith for their kindness and support.

I also wish to thank those good friends I come across in Cleveland for the talk and play we enjoy together. You really cheered me up. I am greatly indebted to Dr. Qing-xin Hua and his wife Wenhua Jia for their warm-hearted help.

Last, but not least, I thank my family: my parents for their unconditional support to pursue my PhD study, even when the study went beyond boundaries of language and geography. My brother and sister for their encouragements and having confidences in my ability to succeed in my study. Especially, my husband for his support, great patience and love at all times and my adorable son for the joy and love he brought to me.

LIST OF ABBREVIATIONS

MTase, methyltransferase

SAM or AdoMet, S-adenosylmethionine

TPMT, thiopurine S-methyltransferase

NNMT, nicotinamide N-methyltransferase

AS3MT, arsenic methyltransferase

AdoMet-dependent MTase, S-adenosylmethionine-dependent methyltransferases

SAH or AdoHcy, S-adenosylhomocysteine

COMT, catechol O-methyltransferase

MetH, cobalamin(vitamin B_{12})-dependent methionine synthase

CbiF, cobalt-precorrin-4 methyltransferase

ChOMT, chalcone O-methyltransferase

IOMT, isoflavone O-methyltransferase

GNMT, glycine N-methyltransferase

6MP, 6-mercaptopurine

6TG, 6-thioguanine

AZA, azathioprine

TGN, thioguanine nucleotide

HGPRT, hypoxanthine-guanine phosphoribosyl transferase

XO, xanthine oxidase

PDNS, purine *de novo* synthesis

6-Me-Thio-IMP, S-methyl-thioinosine 5'-monophosphate

SNP, single nucleotide polymorphism

PEG, polyethylene glycol

APS, Advanced Photon Source

ALS, Advanced Light Source

NSLS, National Synchrotron Light Source

MAD, multiple anomalous dispersion

MR, molecular replacement

SeMet, selenomethionine

SAD, single anomalous dispersion

PDB, Protein Data Bank

mTPMT, murine TPMT

hTPMT, human TPMT

SAMT, salicylic acid carboxyl methyltransferase

PrmC, N5-glutamine methyltransferase, the product of the protein release factor methylation gene

PAPT, putrescine aminopropyltransferase

TIMP, thioinosine monophosphate

dcAdoMet, decarboxylated S-adenosyl-L-methionine

AdoDATO, S-adenosyl-1,8-diamino-3-thiooctane

wt, wild-type

RMSF, root-mean-square fluctuation

TIB, 3,4,5,-triiodobenzoic acid

4IB, 4-iodobenzoic acid

TP, thiophenol

INMT, indolethylamine N-methyltransferase

PNMT, phenylethanolamine N-methyltransferase

DXMT, *Coffea canephora* 3,7-dimethylxanthine methyltransferase

XMT, *Coffea canephora* xanthosine methyltransferase

hAS3MT, human AS3MT

mAS3MT, mouse AS3MT

Structural Studies of the S-adenosylmethionine-dependent Methyltransferases

Abstract

by

YI PENG

S-adenosylmethionine-dependent methyltransferases (AdoMet-dependent MTases) are a main subfamily of MTases, which play critical roles in diverse methylation reactions in many significant biological processes. AdoMet-dependent MTases catalyze methylation reactions utilizing the methyl donor AdoMet. This thesis describes structure-function studies of several members of this enzyme family which are biomedically important. By combining experimental (X-ray crystallography) and theoretical (molecular dynamics simulation calculations) structural biology techniques with molecular biology and functional studies, the work presented here provides molecular insight into mechanisms of enzyme function and drug response.

The first enzyme studied, thiopurine S-methyltransferase (TPMT), modulates the cytotoxic effects of thiopurine prodrugs such as 6-mercaptopurine (6MP) by methylating them in a reaction using AdoMet as the donor. Patients with TPMT variant allozymes exhibit diminished levels of protein and/or enzyme activity and are at risk for thiopurine drug-induced toxicity. We have determined two crystal structures of wild-type murine TPMT, as a binary complex with the product S-adenosyl-L-homocysteine (AdoHcy) and

as a ternary complex with AdoHcy and the substrate 6MP, to 1.8 Å and 2.0 Å resolution, respectively. Comparison of the structures reveals that an active site loop becomes ordered upon 6MP binding. The positions of the two ligands are consistent with the expected S_N2 reaction mechanism. Arg147 and Arg221, the only polar amino acids near 6MP, are highlighted as possible participants in substrate deprotonation. Structure-based mutagenesis and enzyme activity assays suggest that either Arg152 or Arg226 may participate in some fashion in the TPMT reaction, with one residue compensating when the other is altered, and that Arg152 may interact with substrate more directly than Arg226, consistent with observations in the murine TPMT crystal structure.

In addition, we have compared the catalytic activity of wild-type and *5 variant TPMTs, and found that the variant's binding affinity for its methyl acceptor and donor substrates are reduced 10-fold and 2-fold, contributing to decreased enzyme activity of murine TPMT*5. We have determined two crystal structures of murine TPMT*5, as a binary complex with AdoHcy and as a ternary complex with AdoHcy and 6MP, respectively. The TPMT*5 crystal structures together with molecular dynamics simulation calculations reveal that the active site loop is more flexible in TPMT*5, which affects the AdoMet and 6MP substrate affinity and results in loss of the enzyme activity. In addition, these TPMT*5 crystal structures and the computational modeling of other TPMT variants using wild-type murine TPMT structures aid our understanding of the molecular consequences of TPMT polymorphisms. Furthermore, crystal structures of TPMT complexes with benzoic acid inhibitors and thiophenol substrate reveal that TPMT possesses a flexible active site which can accommodate both a smaller acceptor substrate

such as thiophenol and larger benzoic acid inhibitors. These structures provide insights into the connection between the subtle variation in binding of different acceptor substrate site ligands to TPMT and the different degree of inhibition by these benzoic acid inhibitors. The structural features of the acceptor binding site characterized by the ensemble of TPMT structures reported here may be useful in identifying new small molecule modulators for optimization of thiopurine-based therapy.

Nicotinamide N-methyltransferase (NNMT) catalyzes the N-methylation of nicotinamide, pyridines and other structural analogs using AdoMet as methyl donor. The crystal structure of human NNMT, which plays a significant role in the regulation of metabolic pathways and cancers, was solved as the ternary complex bound to both AdoHcy and nicotinamide. The structure reveals the structural basis for nicotinamide binding, highlights several residues in the active site, which may play roles in nicotinamide recognition and NNMT catalysis, and provides a structural basis for the design of NNMT mutants to further investigate the enzyme's catalytic mechanism. The structure-based mutagenesis of NNMT is being pursued in ongoing studies.

Arsenic methyltransferase (AS3MT) is the third important AdoMet-dependent MTase included in our studies. It is involved in methylation of inorganic arsenic and relevant to public health. To obtain the crystal structure of AS3MT for elucidation of the mechanism of arsenic methylation and to probe the relationship between AS3MT polymorphisms and individual variation in arsenic metabolism, a number of AS3MT constructs have been prepared and characterized, and efforts to crystallize AS3MT are ongoing.

CHAPTER I

Background and Significance

1.1 Introduction

Methylation plays a wide variety of roles in biological processes, such as biosynthesis, metabolism, detoxification, signal transduction, protein sorting and repair, and nucleic acid processing. The diversity in the roles of methylation is owed to the tremendous number of methyltransferases (MTases). MTases are widely distributed in living organisms to catalyze the methylation of DNA, RNA, proteins, phospholipids, and even small molecules like hormones, neurotransmitters, and therapeutic drugs (Martin and McMillan 2002). Among the several classes of MTases, S-adenosylmethionine-dependent MTases catalyze most of the methylation reactions utilizing the methyl donor S-adenosylmethionine (SAM or AdoMet) (Cheng and Blumenthal, 1999). The thiopurine S-methyltransferase (TPMT), nicotinamide N-methyltransferase (NNMT), and arsenic methyltransferase (AS3MT) studied in this work all belong to AdoMet-dependent MTases and play important clinical, pathological and biological roles respectively.

TPMT is an enzyme catalyzes the methylation of thiopurine drugs used for cancer therapy. Moreover, TPMT genetic polymorphisms were discovered that alter the level of TPMT activity in humans and therefore control the interindividual variation in the metabolism, therapeutic efficacy and toxicity of thiopurine drugs (Watters and McLeod, 2003). Due to its clinical significance, TPMT has been extensively studied and become one of the most

classic paradigms of pharmacogenetics, which is the study of the influence of genetic differences on individual drug response (Weinshilboum, 2001). Elucidation of the molecular mechanism and biochemical consequences of TPMT polymorphisms can provide invaluable knowledge to achieve the therapeutic goals of thiopurine drugs and avoid toxic side effect in patients. As we know that the structures of the enzymes can provide us with better understanding of enzyme function and catalytic mechanism, the determination of TPMT structures is essential to understand TPMT polymorphisms, and provide a possible strategy to design small molecule modulators or customizing drugs for optimization of thiopurine-based therapy.

The second target of our study is NNMT, which catalyzes the N-methylation of pyridines, particularly nicotinamide. The nicotinamide comprises part of NAD(H) and NADP(H), which are crucial for many metabolic pathways (Williams *et al.*, 2005). Therefore, NNMT could contribute to the regulation of the metabolic pathways involving NAD(H) and NADP(H). An abnormal expression of NNMT was detected in glioblastoma, stomach adenocarcinoma, papillary thyroid cancers, renal carcinoma, colorectal cancer (Sartini *et al.*, 2006). The function of NNMT in regulation of metabolic pathways and cancers remains unknown. In addition, NNMT may play a crucial role in the biotransformation and detoxification of drugs and xenobiotic compounds like many other drug metabolizing methyltransferases (Aksoy *et al.*, 1994). Thus the structural study of NNMT will contribute to the elucidation of its catalysis mechanism and function in cancer progression, and to the design of new NNMT inhibitors as new drugs.

AS3MT is another important MTase associated with public health application. As a carcinogen, arsenic is widely spread in drinking water (Hughes, 2002). The oxidative methylation of inorganic arsenic in humans is mediated by AS3MT. AS3MT genetic polymorphism is also related to interindividual variation in the biomethylation of arsenic (Wood *et al.*, 2006). The structure of AS3MT will be invaluable in order to understand the biological and pathological consequences of arsenic toxicity and carcinogenesis.

To provide a general context, I will begin with a brief introduction of AdoMet-dependent MTases along with a discussion of AdoMet, the methyl donor common to all three enzymes in my studies. Then, the background of TPMT, NNMT and AS3MT research will be reviewed in great detail. Finally, a brief overview of the X-ray crystallography methods and molecular dynamics simulations utilized for the studies reported here will be provided.

1.2 S-adenosylmethionine-dependent Methyltransferases

1.2.1 S-adenosylmethionine and Methylation

Since the structure of AdoMet was first characterized in 1951, AdoMet was found to be involved in many biological processes, such as methylation, polyamine synthesis, and radical-based catalysis (Grillo and Colombatto, 2008). AdoMet is the second most widely used enzyme substrate after ATP (Cantoni, 1975; Chiang *et al.*, 1996). Although AdoMet can be employed as a source of a variety of chemical groups, it is mainly recognized as a methyl donor, the only role we will focus on in our studies (Fontecave *et al.*, 2004). Considering the wide involvement of AdoMet in methylation and the critical role of

methylation mentioned above, AdoMet has profound effects on various cellular processes from gene expression to membrane fluidity. Additionally, its significance in tissue function results in that AdoMet has been studied as a possible therapeutic agent for the treatment of various clinical disorders (Mato *et al.*, 1997).

A series of reactions form the metabolic cycle of AdoMet (**Figure I-1**) (Fontecave *et al.*, 2004). AdoMet is synthesized from methionine and ATP in a reaction catalyzed by SAM synthetase. A high energy sulfonium ion is formed when the adenosyl moiety of ATP is transferred to methionine (Lu, 2000). The sulfonium group can activate each of the attached carbon atoms toward nucleophilic attack so that the methyl group of AdoMet can easily transfer to acceptor substrates. Thus, following AdoMet synthesis, the methyl group of AdoMet can be transferred to a methyl acceptor, leaving the product S-adenosylhomocysteine (SAH or AdoHcy). This is the methylation reaction, which can be catalyzed by various AdoMet-dependent MTases. It is notable that the increase of AdoHcy level and the decrease in the AdoMet level or AdoMet to AdoHcy ratio can inhibit the methylation reaction. To adjust the intracellular AdoHcy level, AdoHcy is sequentially hydrolyzed to form adenosine and homocysteine by SAH hydrolase. Finally, homocysteine is modified by methionine synthase to regenerate methionine. Given the strong electrophilic property of its methyl group, AdoMet becomes the favorable methyl donor over others like folate (Fontecave *et al.*, 2004). AdoMet-dependent MTases exploit the intrinsic character of AdoMet to bring it together with nucleophilic groups of substrates and carry out most of the methylation reactions in living organisms. Based on the substrate specificity (from arsenite to DNA and proteins) and the diversity of atomic targets (carbon,

oxygen, nitrogen, sulfur or even halides (Ohsawa *et al.*, 2001; Attieh *et al.*, 1995)), roughly 120 members of AdoMet-dependent MTases have been classified (Martin and McMillan 2002).

1.2.2 The Structures of S-adenosylmethionine-dependent Methyltransferases

After the initial characterization of AdoMet, the first AdoMet-dependent MTase crystal structure was determined over four decades later (Cheng *et al.*, 1993). For several years thereafter, several structures from AdoMet-dependent MTases were determined and found to share a common core structure, which is known now as the Class I AdoMet-dependent MTase core fold (Cheng and Blumenthal, 1999). More recently however, four more structurally distinct classes of AdoMet-dependent MTases (Classes II-V) have been reported (Kozbial and Mushegian, 2005). The structural features of the five known classes of AdoMet-dependent MTases will be compared in detail in the following paragraph.

Class I AdoMet-dependent MTases: The first reported M.Hhal DNA-MTase (Cheng *et al.*, 1993) and catechol O-MTase (COMT) structures (Vidgren *et al.*, 1994) are representatives of this group. These structures are similarly comprised of alternating β stands and α helices, forming a seven-stranded β sheet flanked by three helices on each side **(Figure I-2a)**. Strand 7 is antiparallel to the other six strands, forming a β hairpin at the C-terminus of the sheet (Schubert *et al.*, 2003a). It is also noted that the structure is strikingly similar to that of NAD(P)-binding Rossmann fold proteins (Cheng and Blumenthal, 1999). Both of them form a deep cleft in the center for AdoMet or NAD(P) binding. However, despite the high conservation of their core structures, there are some variations to the core structure. One of

the smallest Class I AdoMet-dependent MTases is COMT with only two minor modifications to the core fold. In contrast, most Class I AdoMet-dependent MTases contain various insertion patterns throughout the entire structure. These specific modifications to the core fold appear to play roles in substrate recognition (Martin and McMillan, 2002).

Additionally, Class I AdoMet-dependent MTases share limited sequence identity, especially at the AdoMet binding site. AdoMet binds to Class I AdoMet-dependent MTases at the equivalent position of the fold and adopts a similar extended conformation. But the chemistry of AdoMet binding varies considerably (Martin and McMillan, 2002; Schubert *et al.*, 2003a). The sequence alignment based on the structures of Class I AdoMet-dependent MTases indicates only one strongly conserved region E/DXGXGXG between the first α helix and β strand (motif I), which interacts with the amino acid part of AdoMet. In addition, an acidic loop between the second α helix and β strand (motif II) interacts with ribose hydroxyls of AdoMet. A hydrophobic residue from the linker between the fourth α helix and β strand interacts with the adenine ring of AdoMet. The residue has been identified as either phenylalanine, isoleucine, valine, cysteine, tryptophan, proline or methionine. In summary, the Class I AdoMet-dependent MTases maintain a conserved core fold and the same AdoMet binding region for the common role of AdoMet-dependent MTases. The differences in structural modifications and the interactions explain the tremendous diversity of substrates and function (Martin and McMillan, 2002; Schubert *et al.*, 2003a). It is interesting to note that both the TPMT and NNMT structures determined in this study show the Class I AdoMet-dependent MTase core fold.

Class II AdoMet-dependent MTases: The structure determination of the *Escherichia coli* cobalamin(vitamin B_{12})-dependent methionine synthase (MetH) reactivation domain first indicated that not all AdoMet-dependent MTases exhibited the same core fold (Dixon *et al.*, 1996). The MetH reactivation domain (Class II MTase) contains a long antiparallel β-sheet flanked by helices at either end (**Figure I-2b**). It is dramatically different from the Class I MTases in both overall architecture and AdoMet binding. AdoMet is bound to a shallow groove on the surface of the MetH reaction domain in an extended conformation. Hydrogen bonds are formed between AdoMet and a conserved RXXXGY motif (Schubert *et al.*, 2003a).

Class III AdoMet-dependent MTases: In 1998, the homodimeric structure of cobalt-precorrin-4 methyltransferase (CbiF), an important MTase in cobalamin biosynthesis, was first reported (Schubert *et al.*, 1998). In Class III AdoMet-dependent MTases the active site is located on a groove in the N-terminal αβα domain composed of five strands and four helices (**Figure I-2c**). Like the Class I AdoMet-dependent MTases, it has a GXGXG motif at the C-terminus of the first β strand but it does not interact with AdoMet in the absence of the precorrin substrate. On the contrary, AdoMet binds between the two αβα domains in a tightly folded conformation. Based on sequence analysis, the Class III AdoMet-dependent MTase structure is predicted to apply to other enzymes. This might suggest a diversity of substrates for Class III AdoMet-dependent MTases (Schubert *et al.*, 2003a).

Class IV AdoMet-dependent MTases: More recently, the SPOUT family of RNA MTases exhibited the Class IV AdoMet-dependent MTase structure (Anantharaman *et al.*, 2002; Michel *et al.*, 2002; Nureki *et al.*, 2002). This family consists of a six-stranded parallel β sheet flanked by seven helices (**Figure I-2d**). The active site is close to the subunit interface of a homodimer. The most unique feature is that a significant portion of the C terminus (~30 residues) is tucked back into the structure to form a `knot´ (Schubert *et al.*, 2003a). The structure of YibK MTase from *Haemophilus influenzae* (HI0766) also shows that AdoHcy is bound in a tightly folded conformation (Lim *et al.*, 2003).

Class V AdoMet-dependent MTases: The SET-domain proteins represent the newest structural family of AdoMet-dependent MTases, Class V (Zhang *et al.*, 2002; Min *et al.*, 2002; Trievel *et al.*, 2002; Wilson, 2002; Jacobs *et al.*, 2002). These Class V AdoMet-dependent MTases mainly include eight curved β strands with the combination of three small sheets (**Figure I-2e**). The C terminus forms a substructure similar to the knot of the Class IV AdoMet-dependent MTases. AdoHcy is bound to a concave surface of the SET domain in a conformation similar to that of the Class III AdoMet-dependent MTases (Schubert *et al.*, 2003a).

1.2.3 The Mechanisms of S-adenosylmethionine-dependent Methyltransferases

To date the determined structures for substrate-bound complexes of AdoMet-dependent MTases are mainly Class I AdoMet-dependent MTases. These structures demonstrate that nucleophilic attack happens exclusively at the methyl group of AdoMet. This led to the conclusion that the methylation catalyzed by AdoMet-dependent MTases adopts an

S_N2-like mechanism with direct transfer of methyl group to substrate (Schubert *et al.*, 2003a). This reaction also requires that a proton be removed before, concurrent with, or after methyl transfer. Class I AdoMet-dependent MTases bind AdoHcy/AdoMet at almost the same site with an extended conformation, reflecting their shared AdoMet dependence. Nevertheless they have diverse active sites to accommodate their wide variety of substrates. Taking account of the property of atomic targets, a wide variety of mechanisms have evolved to activate the catalytic nucleophile, such as O-methylation by chalcone O-methyltransferase (ChOMT), isoflavone O-methyltransferase (IOMT) (Zubieta *et al.*, 2001) and COMT (Vidgren *et al.*, 1994), N-methylation by glycine N-methyltransferase (GNMT) (Huang *et al.*, 2000), and C-methylation by M.Hhal DNA-MTase (Klimasauskas *et al.*, 1994). In particular, the structures of TPMT complexes determined in our study will illuminate S-methylation.

From the evidence cited above, it is indicated that AdoMet-dependent MTases play a significant role in all living organisms. Although most AdoMet-dependent MTases share a common Class I core fold, more structural classes have been and likely will be identified. So far an S_N2-like mechanism is thought to apply to AdoMet-dependent MTases. However, even for the structurally conserved Class I AdoMet-dependent MTases, the mechanism for the catalytic activation is expected to be distinct for each AdoMet-dependent MTase. Overall, to explore the specific mechanisms and significant roles of TPMT, NNMT and AS3MT related to drug metabolism, cancer pathology and public health, we carried out the structural studies dwelled on in the following chapters.

1.3 Thiopurine S-Methyltransferase

The studies of thiopurine S-methyltransferase (TPMT) can be tracked back to 1951 when thiopurines were originally synthesized by Elion and Hitchings. Soon after that, the success of thiopurines in the treatment of childhood acute lymphoblastic leukemia resulted in the Food and Drug Administration (USA) to approve them for use in 1953 (Coulthard and Hogarth, 2005). Since then, thiopurine drugs have been commonly used to treat a variety of conditions, such as acute leukemia, inflammatory bowel disease (IBD) and certain auto-immune diseases, as well as organ transplant rejection. In 1960s, while studying the metabolism of thiopurines, Remy first described specific TPMT activity in rat (Remy, 1963). However, it was not until 1978 that TPMT activity was first assayed in human tissue (Weinshilboum *et al*., 1978). Soon the purified TPMT protein from human kidney was characterized (Woodson and Weinshilboum, 1983). Human TPMT is a 28kDa protein with 245 amino acids. Currently, TPMT has been found in most human tissues and in both prokaryotes and eukaryotes (Krynetski and Evans, 2003). The important role of TPMT in thiopurine metabolism stimulated great interest of researchers in TPMT studies. The inherited variations of TPMT activity in red blood cell were first described in 1980 (Weinshilboum and Sladek, 1980). After the human TPMT cDNA and gene had been cloned and characterized (Honchel *et al.*, 1993; Szumlanski *et al.*, 1996), the first inactive allele of *TPMT* was identified in 1995 (Krynetski *et al.*, 1995a). To date, at least 28 *TPMT* genetic polymorphisms have been identified. Currently, TPMT genetic polymorphism has become a striking example of pharmacogenetics, which is the study of inherited interindividual variation in drug response (Relling and Dervieux, 2001; Eichelbaum *et al.*, 2006). The subsequent discussion will briefly review the development of TPMT

pharmacogenetics and our understanding of its evolutionary conservation and biological role.

1.3.1 TPMT Pharmacogenetics

Thiopurine metabolism: Thiopurines, including 6-mercaptopurine (6MP), 6-thioguanine (6TG) and azathioprine (AZA) (**Figure I-3a**), are a family of inactive prodrugs used for acute lymphoblastic leukemia, autoimmune disorders, IBD as well as organ transplant recipients (Weinshilboum 2001; Krynetski and Evans, 1999). In order to exert their cytotoxic effects, they require metabolism to thioguanine nucleotides (TGNs) (**Figure I-3b**). AZA can be reduced to 6MP. Both 6MP and 6TG are first activated by hypoxanthine-guanine phosphoribosyl transferase (HGPRT) and then metabolized by multiple enzymatic reactions to TGNs, before incorporation into DNA and RNA (Weinshilboum 2001; Watters and McLeod, 2003; Krynetski and Evans, 1999). The incorporation of TGNs into DNA can lead to DNA-protein cross-links, single strand breaks, interstrand cross-links, sister chromatid exchange, even inhibition of nucleotide and protein synthesis and finally inhibition of lymphocyte proliferation (Coulthard and Hogarth, 2005; Sahasranaman *et al.*, 2008). Besides the activation of thiopurines by the HGPRT pathway, two major metabolic reactions inactivate them *in vivo* (**Figure I-3b**), oxidation by xanthine oxidase (XO) and methylation by TPMT (Wang and Weinshilboum, 2006). As the level of XO activity in the bone marrow is low, the methylation by TPMT represents the predominant inactivation pathway (Coulthard and Hogarth, 2005; Wang and Weinshilboum, 2006). Therefore, TPMT activity is inversely correlated with the concentration of TGNs. In addition, an intermediate of the AZA/6MP pathway, thioinosine monophosphate, can also

be methylated by TPMT to generate S-methyl-thioinosine 5'-monophosphate (6-Me-Thio-IMP) (**Figure I-3b**). This is a strong inhibitor of purine *de novo* synthesis (PDNS), which also contributes to the cytotoxic action of 6MP through the inhibition of lymphocyte proliferation (Coulthard and Hogarth, 2005; Sahasranaman *et al.*, 2008). However, thiopurine drugs were also found to have a relatively narrow therapeutic index and be capable of causing life-threatening toxicity, most often resulting in myelosuppression (Sahasranaman *et al.*, 2008). Individual variation of TGN concentration after thiopurine medication is correlated with this toxicity (Krynetski and Evans, 2003; Sahasranaman *et al.*, 2008). The clinical significance, genomic studies and molecular mechanisms of TPMT polymorphisms will be addressed to illuminate the individual variation of drug response.

Clinical significance of TPMT polymorphism: In 1980, individual differences in human tissue TPMT enzyme activity were reported (Weinshilboum and Sladek, 1980). The family studies demonstrated the differences are inherited in an autosomal co-dominant fashion (Krynetski and Evans, 1999). The inherited phenotypic variations were primarily associated with TPMT genetic polymorphisms (Wang and Weinshilboum, 2006). Approximately 90% of individuals exhibited high activity resulting from being homozygous for a gene with high activity; 10% were heterozygous and exhibited intermediate activity; 0.3% were homozygous for a gene with low TPMT activity and displayed low or no activity (Weinshilboum, 2001; Krynetski and Evans, 2003). When the TPMT-deficient patients were treated with standard doses of thiopurine drugs, they had enhanced concentrations of TGNs and were at high risk of severe toxicity. Even

those with a heterozygous phenotype still suffered severe toxicity and some adverse side effects. To avoid the toxicity, patients were required to be treated based on their genotype (**Figure I-4**). In contrast, the patients with high TPMT activity may not respond to thiopurine therapy (Lennard *et al.*, 1990) and need to be given higher doses of thiopurine drugs. In the study of the relative significance of the TPMT polymorphism for thioguanine and mercaptopurine, it was indicated that mercaptopurine is more affected by TPMT polymorphism than thioguanine (Hartford *et al.*, 2007). This also provides a preclinical system to optimize the dosages of thiopurine medications.

Genomic studies of TPMT polymorphisms: To understand consequences of clinically important TPMT polymorphisms, the initial step was to identify the inactivating mutations in the human TPMT cDNA and gene which alter encoded amino acid sequences of human TPMT (Weinshilboum, 2001). Since the characterization of the first inactive allele of human TPMT (TPMT*2) which contains a single nucleotide polymorphism (SNP) (G238C) leading to the substitution of Ala80Pro (Krynetski *et al.*, 1995a), at least 28 TPMT polymorphisms have been reported to date (Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Lindqvist *et al.*, 2007; Garat *et al.*, 2008; Tamm *et al.*, 2008). While several polymorphisms influence transcription or mRNA splicing (Krynetski and Evans, 2003; Wang and Weinshilboum, 2006), 24 are nonsynonymous coding SNPs (cSNPs) which code for TPMT variants with varying degrees of reduced enzyme activity. A few allelic variants need further functional analysis to prove their implication in TPMT deficiency (Sasaki *et al.*, 2006). The studies also showed that the frequency of TPMT polymorphisms is related to ethnicity. The most predominant variant

allele for low TPMT activity in Caucasians is TPMT*3A, in which two SNPs (G460A and A719G) located respectively in exons 7 and 10 lead to the alterations of two amino acids (Ala154Thr and Tyr240Cys) (Szumlanski *et al.*, 1996). TPMT*3C, which is the major variant allele in East Asians, has a single A719G mutation (Wang and Weinshilboum, 2006). The three predominant alleles (TPMT*3C, *3A and *2) account for over 95% of inherited TPMT deficiency while the other variant alleles are rare (McLeod *et al.*, 2000) (**Figure I-4**). In addition, polymorphisms in promoter and introns can also lead to the TPMT activity variation by influencing transcript and mRNA splicing (Krynetski and Evans, 2003; Wang and Weinshilboum, 2006). More recently, TPMT activity in leukemia cells was found to be determined by chromosomal number and the property of the acquired chromosomes with a wild-type or variant TPMT allele. The accumulation of TGNs can be decreased by an additional acquired chromosome with a wild-type TPMT allele. This study suggested that the difference in allele-specific chromosomal number may be another genetic feature correlated with the individual variation in thiopurine drug response (Cheng *et al.*, 2005).

Molecular mechanisms of TPMT polymorphism: As a result of its clinical importance, the functional effects of TPMT polymorphisms were further studied, which opened a new fields for pharmacogenetics research. In the studies of TPMT polymorphism, it was first indicated that the striking decrease in TPMT*3A allozyme activity resulted from decreased quantity of TPMT protein (Tai *et al.*, 1997 & 1999). Evidence was obtained that the transcription/translation rate of variant TPMT proteins was similar to the wild-type TPMT while variant TPMT proteins were significantly unstable relative to wild-type.

The accelerated proteolysis of the variant proteins was revealed. It also turned out that the chaperone proteins might highly target the TPMT*3A allozyme for degradation compared to the wild-type enzyme (Tai *et al.*, 1997 & 1999; Wang *et al.*, 2003). Subsequently, thirteen of the known TPMT polymorphisms were further investigated by Salavaggione *et al.* (Salavaggione *et al.*, 2005). It was found that the decreased TPMT allozyme activity can be significantly correlated with their decreased protein level with the exception of TPMT*5 (**Figure I-5a**). But the change of substrate kinetics can not sufficiently explain the obvious alteration in TPMT enzyme activity. Instead, from comparing the rates of degradation for recombinant human TPMT allozymes, it was suggested that accelerated degradation might lead to the decrease in protein quantity of TPMT enzyme. It is notable that of all the studied TPMT variants, TPMT *5 is the only variant with a stable protein level. This indicated that TPMT*5 might change the ability of the protein to catalyze S-methylation of its substrate. Structures of TPMT*5 are expected to be very interesting and useful to help us understand the molecular cause of diminished TPMT variant function. In Chapter III, the structural study of TPMT*5 will be addressed in detail.

Furthermore, recent findings indicated that the TPMT*3A polymorphism may have resulted in aggregate formation as well as accelerated degradation of misfolded TPMT*3A allozymes (Wang *et al.*, 2005). Specifically, TPMT*3A can form aggresomes in COS-1 cells treated with the proteasome inhibitor (MG132) compared with wild-type TPMT (**Figure I-5b**). In addition, TPMT*3A aggresome formation can be reduced by not only microtubule destabilizing agent (vinblastin) but also HDAC inhibitor (scriptaid). This

suggested that TPMT*3A aggresome formation may involve the microtubule network. Both size exclusion chromatography and CD spectroscopy further confirmed that the TPMT*3A polymorphism lead to protein aggregation. Together, these mechanistic studies of TPMT pharmacogenetics demonstrated that the decreased activity of TPMT variant with alteration of encoded amino acid is the consequence of the reduction of protein quantity (Hernandez *et al.*, 1991; Weinshilboum and Wang, 2004b), which is caused by the enhanced degradation and aggregate formation of misfolded proteins (Weinshilboum and Wang, 2006).

Overall, TPMT pharmacogenetics has been studied so extensively to benefit thiopurine therapy. As a result, it broadens our understanding of mechanisms by which genetic polymorphisms can influence protein function. Nevertheless, many important questions about TPMT remain unclear, particularly its biological role since its endogenous substrate is still not known. In the following paragraphs, the evolutionary sequence conservation will be discussed in relation to the biological role of TPMT.

1.3.2 Evolutionary Conservation and Biological Role of TPMT

Although TPMT has been comprehensively studied for its importance in thiopurine cancer therapy, the biological role and natural substrate of TPMT still remain unknown. Phenotypes of TPMT-deficient and TPMT-proficient individuals appear to be indistinguishable except when they are treated with thiopurines (Krynetski and Evans, 2003). The evolution of TPMT may provide the basis to elucidate the biological function of TPMT. The significant homology among genetic and amino-acid sequences of TPMT

in primates, rodents and bacteria reveals the ancient origin of TPMT (**Figure I-6**). Furthermore, the residues which are altered by the cSNPs of human TPMT are conserved (**Figure I-6**) (Krynetski and Evans, 2003). TPMT activity was widely found in mammalian, avian and amphibian species and several bacteria (Remy 1963). In particular, the homolog from *P. syringae* has TPMT activity and resists the bactericidal agent tellurite (Cournoyer *et al.*, 1998). Furthermore, the expression of this gene in *E. coli* demonstrated that the bacterial TPMT can catalyze methylation of selenite and methylselenocysteine as well (Ranjard *et al.*, 2002). Another important finding is that human TPMT also displays the ability to catalyze methylation of selenium-containing aromatic compounds (Deininger *et al.*, 1994). Comparison of bacterial and mammalian TPMT protein sequences thus identified a conserved family of AdoMet-dependent MTases. Together, these observations suggest that TPMT may play a role of methylating sulfur-, selenium, or tellurium-containing organic and inorganic compounds (Krynetski and Evans, 2003).

In summary, the significant progress of TPMT studies provides a comprehensive insight of how pharmacogenetics can help achieve therapeutic goals. X-ray crystallographic studies of TPMT are expected to help us further understand the possible structural consequences of nonsynonymous cSNPs and the molecular mechanism of TPMT polymorphisms. Although the crystal structure of human TPMT bound to AdoHcy (Wu *et al.*, 2007) and the NMR structure of a bacterial TPMT orthologue from *Pseudomonas syringae* (Scheuermann *et al.*, 2003) have been reported ahead of our TPMT structures, these structures only characterized the TPMT fold and the AdoMet binding site and did not uncover the details

of methyl acceptor recognition. We have determined the crystal structures of wild-type murine TPMT binary and ternary complexes with AdoHcy and 6MP (see Chapter II). Our structures reveal a flexible active site loop which plays an important role of substrate recognition in TPMT. Together with the structure-based mutagenesis we carried out in the human enzyme, it is suggested that Arg147 and Arg221 (Arg152 or Arg226 in human TPMT) close to the 6MP acceptor in the structure of the murine TPMT bound to AdoHcy and 6MP may participate in the TPMT reaction by contributing to 6MP binding. In addition to the wild-type studies, the crystal structures of murine TPMT*5 complexes with AdoHcy and 6MP were also solved to reveal the possible mechanism by which human TPMT*5 leads to a marked decrease in enzyme activity (see Chapter III). Our structures together with molecular dynamics simulation calculations indicate that the hydrophobic leucine residue mutation to serine residue in TPMT*5 leads to active site conformational differences that can affect both AdoMet and 6MP binding and result in loss of enzyme activity. Finally, the crystal structures of murine TPMT complexes with several TPMT inhibitors were obtained to further characterize substrate recognition and perhaps aid the discovery of new small molecule modulators for optimization of thiopurine-based therapy (see Chapter IV).

1. 4 Nicotinamide N-Methyltransferase

Nicotinamide N-methyltransferase (NNMT) methylates nicotinamide, pyridines and other structural analogs using AdoMet as methyl donor. After Wilhelm His first reported the methyl conjugation of pyridine in 1884, the enzyme NNMT catalyzing that reaction was characterized in 1951 (Aksoy *et al.*, 1994). Like other methylation reactions, it may play

a crucial role in the biotransformation and detoxification of xenobiotic compounds (Aksoy *et al.*, 1994). NNMT is mainly expressed in the liver, where its activity has a bimodal frequency distribution with 25% of the general population having a high enzyme activity (Aksoy *et al.*, 1994). This finding suggests its enzyme activity may be regulated by a genetic polymorphism, which may have functional implications for individual differences in drug metabolism and the therapeutic effect (Aksoy *et al.*, 1994). However, the enhanced NNMT activity also appears to produce toxic N-methylpyridinium compounds in Parkinson's disease, which have been identified as possible neurotoxins (Williams *et al.*, 2005; Williams and Ramsden, 2005a).

NNMT is also the only known enzyme to catalyze N-methylation of nicotinamide, in the metabolic pathway for nicotinamide excretion (Williams and Ramsden, 2005b). Therefore NNMT can regulate intracellular levels of nicotinamides. As already known, nicotinamide can form NAD(H) and NADP(H), which are associated with a large number of biological metabolisms, such as energy production, cellular resistance to stress or injury, and longevity (Sartini *et al.*, 2006). Consequently, NNMT might participate in the regulation of these cellular processes.

Although NNMT is predominantly expressed in the liver, lower expression is also found in kidney, lung, placenta, heart, brain and muscle. The abnormal expression of NNMT has been identified in glioblastoma, stomach adenocarcinoma, papillary thyroid cancers, renal carcinoma, colorectal cancer and so on (Sartini *et al.*, 2007). However the function of NNMT in cancer progression is still unclear. NNMT crystal structures will provide a

structural basis to elucidate the enzymatic catalysis and function. Although the crystal structures of human and mouse NNMT bound to AdoHcy has been deposited into Protein Data Bank (PDB accession code: 2IIP and 2I62), they didn't reveal details of the nicotinamide binding. In the study of NNMT reported in Chapter V, the crystal structure of human NNMT bound to AdoHcy and nicotinamide has been achieved. This structure provides a structural scaffold for the design of NNMT mutants to probe mechanism, and the future design of new NNMT inhibitors as possible leads for drug discovery.

1.5 Arsenic Methyltransferase

The metalloid arsenic is an important carcinogen widely found in ground water and natural environments (Hughes, 2002). Its exposure can lead to neurotoxicity, carcinogenesis, hepatic injury, cardiac failure, leucopenia, and even death (Yoshida *et al.*, 2004; Simeonova and Luster, 2004). Arsenic toxicity is a significant concern of public health. Moreover, As_2O_3 is an active drug in therapy for promyelocytic leukemia (Sanz *et al.*, 2005). The mechanism of arsenic toxicity remains unclear. It was proposed (1) arsenic in the trivalent state might inhibit protein activity by reacting with free thiols in protein residues; (2) arsenic in the pentavalent state might compete with phosphate and bind to polyphosphates; (3) arsenic can induce cancer by genotoxicity, cell proliferation, altered DNA repair and DNA methylated oxidative stress, co-carcinogenesis and tumor promotion (Hughes, 2002).

The metabolism of inorganic arsenic in mammals can yield many reactive and toxic products, which also have an important influence on arsenic toxicity. The AdoMet-

dependent methylation of inorganic arsenic is an important metabolic reaction in humans (Thomas *et al.*, 2004; Vasken Aposhian *et al.*, 2004). For many years, the methylation of arsenic was considered to be a major mechanism of detoxication. However, it was demonstrated that the methylated metabolites of arsenic in the trivalent state are more cytotoxic and genotoxic than arsenate and arsenite (Styblo *et al.*, 2000; Mass *et al.*, 2001; Styblo *et al.*, 2002).

Arsenic methyltransferase (AS3MT) was first characterized in rat as the enzyme catalyzing the methylation of arsenic using AdoMet as the methyl donor (Lin *et al.*, 2002). Since the rat AS3MT gene was cloned, the characterization of homologous genes of AS3MT was performed across a variety of species. This identified proteins encoded in genomes of purple sea urchin (*Strongylocentrotus purpuratus*), sea squirt (*Ciona intestinalis)*, rainbow trout (*Oncorhynchus mykiss)*, chicken (*Gallus gallus*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), cow (*Bos taurus*), chimpanzee (*Pan troglodytes*), and human (*Homo sapiens*). All sequences of AS3MT homologs have several common features: 1) five fully conserved cysteines and 2) three sequence motifs (I, II, III) common to AdoMet-dependent MTases (Thomas *et al.*, 2007). These suggested that the conserved function of theses AS3MT homologs might be to catalyze arsenic methylation. A species variation is also identified in AS3MT activity. A few species are not able to methylate arsenic, such as marmoset monkey and chimpanzee. In *in vitro* tests, the livers of rabbit, rat, mouse, hamster, pigeon and rhesus monkey exhibited AS3MT activity, while no activity was measured in the liver of marmoset monkey, chimpanzee and others (Vasken Aposhian, 1997; Vahter, 2000; Li *et al.*, 2005).

Human AS3MT (~43kD) is encoded by the *cyt19* gene. With the growing studies of human AS3MT, it was also indicated that human AS3MT catalyzes the formation of the major urinary arsenic metabolites, methylated and dimethylated arsenic (Drobna *et al.*, 2006). In order to determine the contribution of human AS3MT to the overall capacity to methylate arsenic, AS3MT expression was knocked down by short hairpin RNA (shRNA) in human hepatocellular carcinoma (HepG2) cells (Drobna *et al.*, 2006). The study indicated that reduced human AS3MT expression resulted in a significant decreased methylation of arsenic. It suggests that human AS3MT is an important enzyme in the pathway to methylate arsenic in humans (Drobna *et al.*, 2006). Currently, two conceptual models can describe the catalytic function of AS3MT (**Figure I-7**) (Thomas *et al.*, 2007). In Model A, oxidative methylation of arsenicals alternates with the reduction of pentavalent arsenicals to trivalency. In Model B, the methyl groups are sequentially transferred to thiol-containing complexes of trivalent arsenicals. Both models utilize AdoMet as the methyl donor. Moreover, the studies suggested that the presence of GSH and other physiological reductants can stimulate rates of reactions catalyzed by AS3MT.

On the other hand, genetic studies of human AS3MT provide new data on association of human AS3MT polymorphisms and metabolic phenotypes (Meza *et al.*, 2005; Wood *et al.*, 2006). Pharmacogenetics study of human AS3MT identified 26 SNPs, three of which were nonsynonymous cSNPs (Arg173Trp, Met287Thr, and Thr306Ile). The enzyme activity and immunoreactive protein levels of two allozymes (Trp^{173} and Ile^{306}) were obviously decreased relative to wild-type allozyme while the Thr^{287} allozyme displayed increased activity and protein level. However, the differences in protein levels of AS3MT

allozymes did not appear to be due to accelerated degradation (Wood *et al.*, 2006). These indicated that the human AS3MT polymorphisms might contribute to individual variation in biology and pathology of chronic arsenic exposure as well as the response to arsenic-containing therapeutic agents (Vahter, 2000; Drobna *et al.*, 2004; Wood *et al.*, 2006). The determination of a crystal structure for AS3MT may provide more useful information to elucidate the enzymatic mechanism for arsenic metabolism and understand AS3MT polymorphisms. Therefore, the efforts continue to crystallize AS3MT (see Chapter VI).

1.6 Protein X-ray Crystallography: a brief introduction

The three-dimensional structural knowledge of proteins is vital to understanding how proteins function inside living organisms. Especially, the information of protein interactions and enzyme catalysis acquired from the structures can be conducive for rational drug design and the development of effective therapeutic agents. Since the crystal structure of myoglobin successfully solved by Perutz and Kendrew (Kendrew *et al.*, 1958), X-ray crystallography has been widely used for three-dimensional structure determination of macromolecules. As yet, X-ray crystallography is the most commonly used technique to elucidate macromolecular structures. In the structural studies of AdoMet-dependent methyltransferases reported here, X-ray crystallography is the main tool utilized to obtain a detailed description of protein folds and to reveal the structural features of these enzyme active sites. The basic methods of protein crystallography used for the structural studies of AdoMet-dependent MTases (**Figure I-8**) are briefly described here in order to better understand the structural studies reported in the following chapters.

1.6.1 Protein Preparation

Obtaining crystals is the prerequisite for protein structure determination by X-ray crystallography. Aiming to produce well-ordered, large single protein crystals to diffract X-ray beam first requires multi-milligram quantities of pure protein. Large amounts of proteins can be isolated from the natural sources or can be produced from recombinant protein expression in a microorganism. After purification by chromatographic methods, the high quality of purified protein suitable to carry out crystallization experiment requires that the protein is soluble, properly folded, active, homogenous and stable over the period of crystal growth (Bergfors, 1999).

1.6.2 Protein Crystallization

Crystallization techniques: Growth of diffraction-quality single crystals is crucial for solving protein structures. Protein crystallization is characterized normally by two steps: nucleation and growth, ideally to yield a diffraction-quality crystal (Chernov, 2003). Currently, there are many crystallization techniques, all of which attempt to bring the protein solution to a supersaturation state and consequently form crystals. They include sitting drop vapor diffusion, hanging drop vapor diffusion, sandwich drop, microbatch under oil, microdialysis, and free interface diffusion. Among them, vapor diffusion techniques are most widely used. In the present work only vapor diffusion methods will be discussed as it was the main method used in the work reported here.

In sitting drop and hanging drop vapor diffusion (**Figure I-9**), a small droplet of protein solution is mixed with crystallization reagent and equilibrated against a reservoir

containing the crystallization reagent. Since the initial concentration of crystallization reagent in the reservoir is higher than that in the droplet, the reservoir will pull water molecules from the droplet in a vapor phase until the reagent in the reservoir is approximately the same as that in the drop. Eventually, this process increases the relative supersaturation of the sample in the drop. Both sitting and hanging drop methods are easy to perform and only demand a small amount of protein sample. However, sitting drop is preferable to hanging drop in speed and simplicity. On the other hand, the advantage of hanging drop is 1) to be capable of observing the drop without optical interference from plastic and 2) to reduce the chance of crystals sticking to the well surface that makes removal difficult. (http:// www.hamptonresearch.com/).

Crystallization experiments: The goal of crystallization experiments is to identify solution conditions that favor the development of a well diffracting crystal by the crystallization techniques. However, the extreme difficulty of predicting protein crystallization conditions (Rupp and Wang, 2004) results in crystallization experiments being a trial-and-error process. Furthermore, because the factors influencing crystal growth are too numerous to allow an exhaustive search, the initial crystallization conditions only can be identified by screening over a restricted representative sampling of combinations of these variables, such as buffer type and pH, type of salts, salt concentration, precipitants type, precipitant concentration, protein concentration, and temperature (Chayen, 2005). Therefore, commercially available screens are a rapid and economical means in the search for the preliminary crystallization conditions. The solution conditions in these screens are empirically derived, based on those conditions

proven successful in published protein crystallization reports. The screening can often provide the initial conditions that direct the subsequent attempt to determine appropriate conditions to yield single and good diffracting crystals (Jancarik and Kim, 1991; Cudney *et al.*, 1994). Moreover, in order to optimize the size and diffracting power of crystals, the initial conditions identified in the screens need to be optimized in the composition of buffer pH, additive, precipitating solution, temperature, drop volume, protein concentration and protein:well drop volume ratio as well as seeding.

Unfortunately, in many cases proteins may fail to crystallize using the screens. At that time, the use of fusion protein expression systems, mutation and truncation of the protein, addition of ligands, and improvement of purification protocol are attempted to aid in solving the difficulty of protein crystallization. Although modifications of crystallization experiments and strategies, together with the application of new crystallization techniques, have improved the success rate in growing crystals, often ingenuity and luck are critical for obtaining diffraction-quality crystals (Gilliland and Ladner, 1996).

1.6.3 Data Collection

Crystal mounting: After crystal growth, the crystals are mounted for X-ray diffraction experiment. Although the crystals can still be mounted using glass capillaries for room temperature studies as previously, today most crystals are mounted in a tiny loop, made of plastic or nylon and attached to a solid metal rod base, and then flash-frozen with liquid nitrogen. The advantage of the cryo cooling technique is the reduction of the radiation damage of the X-rays. However, crystals often crack or are surrounded by

diffracting ice crystals if flash-frozen without appropriate cryoprotectant (http://en.wikipedia.org/wiki/X-ray_crystallography#cite_note-Powell-63). Therefore, it is necessary to identify a suitable cryoprotectant for the crystals by extensive trial and error (Jeruzalmi, 2006; Helliwell, 2005). Cryo-solutions usually consist of both crystallization solution and an anti-freeze agent (glycerol, polyethylene glycol (PEG), etc.). The crystallization condition and cryoprotectants for murine TPMT and human NNMT crystals are described in the following chapters respectively.

X-ray sources: The first required hardware for the X-ray diffraction experiment is an X-ray source. Considering the higher intensity and tunability of synchrotrons relative to in-house sources, synchrotron sources have become the more useful and powerful X-ray sources for protein structure determination. In-house sources generally use copper radiation with a fixed wavelength of 1.54 Å, which limits the resolution and phasing options. Therefore, in the structural studies of AdoMet-dependent MTases, in-house sources are primarily used to check the quality of crystals and screen proper cryo-solutions so that the best quality crystals can be identified for further diffraction studies at synchrotron beamlines. In contrast to in-house sources, X-rays generated at synchrotrons are several orders of magnitude more intense, which allows for determination of higher resolution structures and quicker data collection. In addition, a tunable monochromator makes it possible to change the wavelength of synchrotron radiation, which allows for single or multi-wavelength anomalous dispersion (SAD/MAD) phasing experiments described below (Drenth, 1994). This method was used to solve the structure of the wild-type murine TPMT binary complex with AdoHcy. Moreover, all data reported here were

collected at several synchrotron facilities: beamlines 19ID, 19BM, Advanced Photon Source (APS), beamline 4.2.2, Advanced Light Source (ALS), and beamlines X25, X29, National Synchrotron Light Source (NSLS) at Brookhaven.

X-ray diffraction theory: X-ray diffraction by crystals is made possible by the periodicity of crystal architecture. In order to better understand the structural studies of AdoMet-dependent MTases here, X-ray diffraction theory will be introduced (Drenth, 1994; Glusker *et al.*, 1994; Hall, 2004).

The unit cell and crystal lattice: Crystals can be described as infinite arrays in which the building blocks (the unit cell) are repeated in a regular manner in each direction. Generally, a unit cell is defined by three vectors, a, b, and c, and the angles between them α (between b and c), β (between a and c), and γ (between a and b). The linear periodic arrangement of these unit cells forms a crystal lattice in real space. In addition to the crystal lattice in real space, there is another lattice in diffraction space, the reciprocal lattice, which is reciprocal to the crystal lattice and corresponds to reflection positions. The size or intensity of a reflection is a function of the contents of the entire unit cell. Within the crystal lattice, there are an infinite number of lattice planes passing through crystal lattice points. These lattice points, which correspond to diffraction from a given set of lattice planes in the crystal, are designated by a set of three numbers called the Miller indices (reciprocal lattice vector, *hkl*).

Bragg's Law: Bragg showed that diffraction from single crystals can be mathematically treated as a reflection from sets of equivalent parallel lattice planes in the crystals. Constructive interference between the scattered X-rays from adjacent planes will happen when the angle θ between the plane and rays results in a path-length difference between these rays that is equal to a whole number n of wavelengths λ. This is expressed in an equation known as Bragg's Law (also see **Figure I-10**):

$$n\lambda = 2d\sin(\theta)$$

where d is the spacing between planes of atoms. When *n* is a whole number, the reflected waves are in phase and a diffraction spot, with a given *hkl* designation, is generated.

Ewald's sphere: Ewald proposed a geometrical description of diffraction (the Ewald construction), which allows us to calculate which Bragg reflections will be observed when we know the orientation of the crystal relative to the incident beam. The Ewald's sphere is an important method to construct the diffraction pattern and widely used to determine space group, obtain crystal orientation, index reflections and evaluate diffraction intensity. The Ewald's sphere is centered on the crystal M with a radius proportional to $1/\lambda$ (**Figure I-11**). The primary beam (s_0) is scattered by the crystal (M). The reciprocal lattice, drawn on the same scale as that of the Ewald's sphere, is then placed with its origin centered at O. A reflection *hkl* occurs in the direction MP (s) when reciprocal lattice point P comes in contact with the Ewald's sphere. Based on Bragg's law, reflection *hkl* in the direction MP corresponds to both the reciprocal lattice point and diffraction from the crystal lattice plane with the same values for the Miller indices *hkl*.

During the rotation of the crystal new reciprocal lattice points intersect with Ewald's sphere so as to produce new reflections.

Waves and scattering factors: A periodic wave, such as X-rays, can be described by its amplitude, wavelength and relative phase angle. In X-ray crystallography, the wave is normally represented as a vector in an Argand diagram (**Figure I-12a**). The vector length (F) is proportional to the amplitude of the diffracted wave. The angle of the vector to the horizontal axis equals its relative phase angle (α). A periodic wave can be considered as a sum of the vectors of individual waves. This helps interpret X-ray waves scattered by a protein crystal as the sum of scattering by individual atoms (**Figure I-12b**). The ability of an atom to scatter X-rays is determined by the number and position of electrons. The information of atomic scattering factor for a given atom is available. Since there are *n* atoms at different positions in a unit cell relative to the unit cell origin, the total scattering from the unit cell is

$$\mathbf{F(S)} = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j * \mathrm{S}]$$

where f_j are the atomic scattering factors for each of the *n* atoms at positions **r**, with respect to the unit cell origin. The scattering vector S is perpendicular to the reflecting plane. **F**(**S**) is dependent on the positions and structure of all the atoms in the unit cell, and is termed the structure factor. In sum, the wave scattered by a crystal is depicted by a summation of the atomic scattering factors over all unit cells relative to a common origin. Its amplitude and phase are dependent on the arrangement of atoms with respect to the incident X-ray beam and the orientation of the scattering vector and plane.

Data processing: A crystal scattering X-rays generates a pattern of spots (also known as reflections) that is recorded by a two-dimensional detector, which is typically either a phosphorescent imaging plate or a charge-coupled device (CCD) image sensor. Based on X-ray diffraction theory, data processing will convert the diffraction pattern of spots to a crystallographic dataset consisting of structure factor amplitudes. The relative intensities of these reflections provide information to determine the arrangement of molecules in the crystal. The diffraction pattern of the multiple images recorded at the beginning can be indexed to identify the unit cell dimensions, space group, and even the resolution limit. Then the data are integrated to calculate and sort the intensities of many thousands of reflections on hundreds of images according to their Miller indices and output the data as structure factor indices, intensities, and estimated errors, into a single file. The last step of data processing is scaling. During the scaling, these reflections are merged and scaled to calculate average (for multiply measured reflections) decay-corrected intensities for the reflections. The relative intensities of the reflections are crucial information for structure determination. Moreover, many symmetry-equivalent reflections are also recorded multiple times, which leads to merging or symmetry related R-factor (R_{sym} or $R_{merge)}$ to be calculated, on the basis of how similar the measured intensities of symmetry equivalent reflections are, with lower values corresponding to more internally consistent and reliable datasets. Thus, the R-factor is a standard to evaluate the quality of the data. (http://en.wikipedia.org/wiki/X-ray_crystallography#cite_note-Powell-63). HKL2000 (Otwinowski and Minor, 1997) and d*TREK (Pflugrath, 1999) are the programs which were mainly used for data processing of murine TPMT and human NNMT data collected at synchrotrons mentioned before.

1.6.4 Phasing

As discussed in X-ray Diffraction Theory, the structure factor is a complex number containing information regarding the amplitude as well as phase of a wave. In order to convert the diffraction pattern to electron density through the mathematical technique of Fourier transformation, both amplitude and phase information of each individual reflection are required. Although the intensities (from which amplitudes can be calculated) of the diffracted waves can be recorded during data collection, phases cannot be directly measured. This results in the phase problem in X-ray crystallography (http://en.wikipedia.org/wiki/X-ray_crystallography#cite_note-Powell-63). Currently there are many methods to obtain phase information, such as direct methods (Hauptman, 1997; Usón, 1999), anomalous dispersion (Ealick, 2000), isomorphous replacement and molecular replacement (Taylor, 2003). In order to better understand the phase problem, Fourier transformation will be briefly described. Subsequently, multiple anomalous dispersion (MAD) and molecular replacement (MR) mainly used in this structural study of murine TPMT and human NNMT will be discussed below.

Fourier transformation: The intensity of a reflection **I**(*hkl*) is proportional to the square of the amplitude of its corresponding structure factor **F**(*hkl*). As previously described, the structure factor **F(S)** depends on the distribution of electrons within the unit cell. The diffraction from all atoms can be summed up, with all electrons within the unit cell integrated for their diffraction at a given position **r** within the unit cell represented by ρ(**r**):

$$\mathbf{F(S)} = \int_{\text{cell}} \rho(\mathbf{r}) \exp[2\pi i \mathbf{r} * \mathrm{S}] dv$$

If x, y and z are given in fractional coordinates of the unit cell ($0 \leq x < 1$), V is the volume of the unit cell, and the product of the vectors describing the relative positions of the diffracting electrons (**r**) and diffracting plane (**S**) can be defined as $\mathbf{r} * \mathbf{S} = hx + ky + lz$, then:

$$\mathbf{F}(\mathbf{S}) = \mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i (hx + ky + lz)] dx dy dz$$

The Fourier transform of $\mathbf{F}(hkl)$ aims at summing the diffraction from the electron density at all positions within the crystal:

$$\rho(xyz) = 1/V \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(hkl) \exp[-2\pi i (hx + ky + lz)]$$

or

$$\rho(xyz) = 1/V \sum_{h} \sum_{k} \sum_{l} |\mathbf{F}(hkl)| \exp[-2\pi i (hx + ky + lz) + i\alpha(hkl)]$$

The measured intensities ***I(hkl)*** will provide the amplitudes $|\boldsymbol{F(hkl)}|$, but the phase angles $\alpha(hkl)$ need to be determined separately by some phasing method. (Drenth, 1994; Hall, 2004)

Multiple anomalous dispersion (MAD): As mentioned before, MAD was used to solve the structure of the wild-type murine TPMT binary complex with AdoMet, using selenomethionine-(SeMet-) substituted protein. Friedel's law requires that a reflection ***F(hkl)*** has the equal amplitude and opposite phase of its "opposite" reflection, ***F(-h-k-l)***. This law is broken in the presence of anomalous scattering. Exactly based on this principle, MAD was developed as a purely experimental approach for phasing. As shown in **Figure I-13,** a significant enhancement of anomalous scattering factors can be observed by sharp spectral features when the wavelength of the incident x-ray approaches the absorption

maximum of an anomalous scatterer. At this wavelength, the scattered X-ray beam of free electrons is different in phase from the incident beam by exactly 180° while the difference is not actually 180° as a real result of the inner shell electrons binding to the nucleus more tightly than outer shell electrons. These anomalous differences in data collected near the absorption edge of an anomalous scattering atom can be used to identify the position of the anomalous scatterers. This allows the phases of the scattering by the anomalous scatterer to be calculated, which in turn restricts the possible phase angles for the scattering by the protein atoms, and allows an indirect path for determination of phases for the measured structure factors. (Drenth, 1994; Hall, 2004)

Generally, the strong anomalous scatterers can be soaked into the crystal using a crystallization solution supplemented with the heavy atom, or can be incorporated by substituting the methionine of the protein with selenomethionine (SeMet) by the protein expression techniques. Since 1) SeMet can be easily incorporated into proteins for crystallization, 2) the structure of SeMet substituted proteins are remarkably similar to those of their wild-type counterparts, and 3) the incorporation of SeMet is uniform and complete, SeMet has become the most successfully used anomalous scatterer for MAD phasing to date (http://www.ccp4.ac.uk/courses/proceedings/1997/j_smith/main.html).

To take advantage of the differences in the anomalous signal at different wavelengths, MAD is usually done at three wavelengths, which correspond to the inflection point of the absorption edge, the absorption maximum above the edge, and a remote point at higher energy (**Figure I-13**) (Walsh *et al.*, 1999; Hendrickson and Ogata, 1997). It is synchrotron

sources that provide intense and tunable X-ray beams and the instruments to boost the wide application of MAD. Moreover, anomalous dispersion phasing can be done at a single wavelength (single anomalous dispersion, SAD), which can minimize the time in the X-ray beam by the crystal, thus reducing potential radiation damage during data collection. The programs used for MAD phasing in this work are Solve and Resolve (Terwilliger and Berendzen, 1999; Terwilliger, 2000&2003).

Molecular replacement (MR): MR is the method used to solve the structures of murine TPMT and human NNMT. MR requires a previously known structure of a homologous protein, or even the same protein from a different crystal form. As the level of similarity of two protein structures correlates well with the level of sequence identity, MR only can be used to solve an unknown structure when one has a good model which is fairly complete and shares at least ~40% sequence identity with the unknown structure. However, with the increase of known structures, MR has become a very useful method for structure determination (Hall, 2004).

In this method, the rotation and translation searches are performed to achieve the maximum overlap of the model's calculated diffraction pattern with the observed diffraction pattern for the new crystal. As a result, the orientation and position of the search model (known structure) in the unit cell of the target molecule (unknown structure) can be determined. Thus, this method provides initial phase information for the new crystal. In the rotation step, the rotation of the model which optimizes the agreement of calculated and measured diffraction data is determined. In the translation step, the

correctly oriented model is positioned optimally to maximize calculated and measured data agreement. As phases are not experimentally acquired by MR, a potential for model bias exists. This demands more careful refinement and analysis (Drenth, 1994). MR programs utilized in the studies here include EPMR (Kissinger *et al.*, 1999) and MOLREP (Vagin and Teplyakov, 1997).

1.6.5 Model Building, Refinement and Structure Validation

Having obtained structure factors and initial phases by one of the procedures mentioned above, an electron density map can be generated by Fourier transform calculations and used for an initial model building. High resolution maps with accurate phases are easily interpreted by automatic fitting programs, whereas low resolution maps with less unambiguous density need interactive computer programs for manual model building (Hall, 2004). When the initial model is built, the model will be refined against the diffraction data. After several rounds of manual model rebuilding and refinement calculations, the ideal result is to converge on a final model such that its calculated structure factors (F_{calc}) agree with the observed (experimental) structure factors (F_{obs}) as much as possible. During the iterative cycles of model building and refinement, the parameters of the model, namely the positional coordinates (*x, y, z)* and temperature factors (*B*), will be optimized. Through the refinement, the *R*-factor will be calculated as below to measure the agreement between the two sets of structure factors (F_{obs} and F_{calc}).

$$R = \frac{\sum_{hkl} || F_{obs}| - k | F_{\text{calc}} ||}{\sum_{hkl} | F_{obs}|} * 100\%$$

Low R factors correspond to structures which agree well with their measured diffraction datasets. It is possible for a low R factor to result from an incorrect model when the refinement uses too many model parameters, i.e. due to over-refinement. Thus, a randomly selected fraction of the data (5-10%) is kept aside and used to calculate an R-factor unbiased by the refinement process (R_{free}) by cross-validation (Brunger, 1992&1993). During the model refinement, both R_{work} (from the 90%-95% of reflections used in refinement) and R_{free} (from the 5%-10% of reflections omitted from refinement) are calculated and compared. Generally, the R-factor for a good model is roughly no more than the resolution limit in Ångstrom divided by 10. If the difference between R_{work} and R_{free} is over 10%, it indicates possible problems with the model or errors in the refinement process. With the further refinement of the model, both R factors should decrease while R_{free} remains a few percent higher than R_{work}. The other statistics to gauge the refinement include root-mean-square deviation (rmsd) values from expected bond lengths and bond angles, which should be below about 0.015 (Å) and 1.5 (deg) respectively. The programs for model building and refinement of murine TPMT and human NNMT include COOT (Emsley and Cowtan, 2004), CNS (Brunger *et al.*, 1998), and Refmac (Murshudov *et al.*, 1997).

The structure validation is necessary before deposition of the final structural coordinates in the Protein Data Bank (PDB). The distribution of the mainchain dihedral angles (ϕ and ψ) shown in a Ramachandran plot is examined to ensure no residues fall in the disallowed region. A number of software packages are also used to identify outliers of the bond lengths, bond angles, and close contacts in the structural studies here, such as

PROCHECK (Laskowski *et al.*, 1993), SFCHECK (Vaguine *et al.*, 1999) and MOLPROBITY (Davis *et al.*, 2004).

1.7 Molecular Dynamics Simulations in Biology: a brief introduction

Biological macromolecules are inherently dynamic systems in which the internal motions and conformational changes have a profound effect on their function (Karplus and Petsko, 1990). Molecular dynamics simulations are powerful tools for computationally exploring the dynamic nature of the structure and function of biological macromolecules (Karplus and McCammon, 2002; Karplus and Kuriyan, 2005), which is experimentally difficult to probe. The first molecular dynamics simulation of a biological macromolecule was the molecular dynamics simulation of bovine pancreatic trypsin inhibitor (BPTI) (McCammon *et al.*, 1977). Since then, molecular dynamics simulations have been widely applied in not only molecular motion and the interpretation of experiments but also simulated annealing methods for X-ray structure refinement (Brunger *et al.*, 1987) and NMR structure determination (Nilsson *et al.*, 1986). Nowadays the advances in computer power as well as the continued development of new parameters for the potential function and new techniques for running a molecular dynamics simulation have accelerated the development of successful molecular dynamics simulations of biological macromolecules (Dodson *et al.*, 2008). The applications of molecular dynamics simulations in biology demonstrated the significance of molecular internal motions in biological function, including the flexibility of tRNA (Harvey *et al.*, 1984), the fluctuations of heme proteins for ligand entrance and exit (Case and Karplus, 1979), enzyme catalysis (Joseph *et al.*, 1990; Rod *et al.*, 2003; Watney *et al.*, 2003), understanding of signal pathways (Lee *et al.*,

2007; Somani *et al.*, 2007), and even bimolecular assemblies (Kazmirski *et al.*, 2005; Hamacher *et al.*, 2006). In the following paragraphs, we begin with an overview of molecular dynamics, continue with the significance of molecular dynamics simulations, and close with a discussion of the application of molecular dynamics simulations in the studies reported here.

1.7.1 The Methodology of Molecular Dynamics Simulations

Molecular dynamics: Molecular dynamics is a specialized discipline of simulating the motions of a system of particles (Karplus and Petsko, 1990). A molecular dynamics simulation requires a definition of the interaction potential for the particles and the equations of motion describing the dynamics of the particles. Although the interaction potential may vary from the simple gravitational interaction to the complicated interaction between biological macromolecules, classical Newton's equations of motion are enough for many systems, including the biological macromolecules of primary concern here (Karplus and Petsko, 1990). However, hydrogen atoms are light and quantum mechanical character is sometimes essential for the motion of protons (van der Spoel *et al.*, 2005). Considering the unique property of molecular dynamics simulations in probing the space and timescales simultaneously, molecular dynamics simulations can help better understand the detailed dynamic process of biology, which the experiments alone can not characterize (Tama and Brooks, 2006; Dodson *et al.*, 2008). In addition, the approximate potential used in a simulation can be user-modified as the need arises or knowledge improves (van der Spoel *et al.*, 2005). The property of potential also confers

molecular dynamics simulation capable of calculating free energy differences (Karplus and Petsko, 1990).

Potential energy function: The total energy of an atom arrangement can be represented as a potential energy function with atomic positions as variables. In molecular dynamics simulations the forces derived from a potential energy function can be used to calculate the dynamic behavior of the system by solving Newton's equations of motion (Karplus and Petsko, 1990). Considering computational efficiency in large biomacromolecular systems, the potential energy function should be simple, and since force calculations are required in molecular dynamics simulations, the function must also have to be analytically differentiable (Norberg and Nilsson, 2003). The empirical energy functions used for biomacromolecules consist of bonding terms representing bond lengths, bond angles, and torsional angles and non-bonding terms composed of van der Waals interactions and electrostatic contributions (**Figure I-14**) (Karplus and Petsko, 1990; Levitt, 2001). Additionally, the accuracy of the energy function directly affects the accuracy and stability of molecular dynamics simulations of biomacromolecules. Therefore, *ab initio* quantum chemical calculations on model compounds are essential in the development of these empirical energy functions (Norberg and Nilsson, 2003).

Today, a number of different programs for molecular dynamics simulations of biomacromolecules have been developed, including the energy functions. Most of them adopt the same form of the empirical energy functions mentioned above. The major difference between them is the way that the parameters of the energy function are

obtained. Overall, these energy functions have similar performance (Norberg and Nilsson, 2003). With the increased computer power, the development of newer energy functions will focus on the balance between solute-solute, solute-solvent and solvent-solvent interaction energies as the result of more biological macromolecular simulations performed in solution (Norberg and Nilsson, 2003).

Simulation methods: A molecular dynamics simulation requires as input the initial conditions including the coordinates and velocities of all atoms in the system. The coordinates can be obtained from the structures determined by X-ray crystallography, NMR or homology modeling (Karplus and Petsko, 1990). When a starting structure is far from equilibrium, the excessively large forces may result in the failure of the molecular dynamics simulation (van der Spoel *et al.*, 2005). Thus, an energy minimization is required to relieve the stresses of the initial structure due to overlaps of non-bonded atoms, bond-length distortions and so on. Sequentially, atoms are assigned velocities taken randomly from a Maxwell–Boltzmann distribution for a low temperature. The movement of the atoms is simulated by numerically solving Newton's equations of motion (Karplus and Petsko, 1990). Modern molecular dynamics simulations of biomacromolecules are generally performed in the nanosecond range.

Currently, the commonly used program packages for molecular dynamics simulations of biomolecules are mainly AMBER (Cornell *et al.*, 1995), CHARMM (Brooks *et al.*, 1983; MacKerell *et al.*, 1998), GROMOS (Scott *et al.*, 1999), NAMD (Phillips *et al.*, 2005), ENCAD (Levitt *et al.*, 1995) and GROMACS (Bekker *et al.*, 1993; Berendsen *et al.*,

1995; Lindahl *et al.*, 2001; van der Spoel *et al.*, 2005). GROMACS was applied in the molecular dynamics simulations carried out in Chapter III.

1.7.2 The Significance of Molecular Dynamics Simulations in Biology

Since simulations can describe the individual particle motions as a function of time, they can be often used to investigate dynamic processes more easily than experiments. The comparison of experimental and calculated results can confirm the accuracy of the calculated results and improve methodology (Karplus and McCammon, 2002). With the general availability of computer hardware and software, molecular dynamics simulation has become quite useful in elucidating possible features of biological macromolecules at an atomic level (Norberg and Nilsson, 2003). The molecular dynamics simulations have been applied to address a variety of motional phenomena in biology since the first molecular dynamics simulation of a biological macromolecule reported in 1977 (McCammon *et al.*, 1977). Molecular dynamics simulations combined with experiments have been applied to investigate possible features in mechanisms of protein folding and enzyme catalysis, conformational change in the functional mechanism of chaperones, dynamic coupling of protein modules in protein kinases, and the mechanism of signal pathways and bimolecular assemblies (Karplus and McCammon, 2002; Karplus and Kuriyan, 2005; Dodson *et al.*, 2008). With the explosive growth of experimentally-determined structures which involved in various biological processes, the increasing use of molecular dynamics simulations plays an increasingly significant role in expanding understanding of comprehensive biological issues.

1.7.3 The Application of Molecular Dynamics Simulations in the Structural Studies of TPMT

As mentioned earlier, the decreased activity of TPMT variants with alterations of encoded amino acids is typically the consequence of a reduction of protein quantity (Hernandez *et al.*, 1991; Weinshilboum and Wang, 2004b), which is caused by enhanced degradation and aggregate formation of misfolded proteins (Weinshilboum and Wang, 2006). Molecular dynamics simulation calculations can be used to investigate dynamic processes more easily than by experimental methods. To better understand the molecular and structural consequences of TPMT variant proteins which are prone to misfold and are subject to enhanced degradation and aggregate formation, molecular dynamics simulation calculations were employed to complement the experimentally-determined TPMT crystal structures (see Chapter III).

In the studies presented here, GROMACS (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; van der Spoel *et al.*, 2005) was used for molecular dynamics simulation calculations. Three sets of calculations were carried out, using as starting models the wild-type murine TPMT crystal structure (PDB accession code 3BGD, Peng *et al.*, 2008), the TPMT*5 crystal structure (PDB accession code 3BKO), and a computationally modeled murine TPMT*5 (Leu44Ser) generated by computationally modeling the mutation into the wild-type murine TPMT crystal structure coordinates (Chapter III). The molecular dynamics simulation calculations confirmed the observations from the crystal structures that the hydrophobic leucine residue mutation to serine residue in TPMT*5 is incompatible with the native wild-type protein conformation, and leads to a

more flexible active site conformation that can affect both AdoMet and 6MP binding and result in loss of enzyme activity (see Chapter III).

In addition, a new TPMT variant (Tyr107Asp) has recently been identified in an organ transplant patient in Thailand, who was having toxicity issues with 6MP therapy. This variant protein was expressed and displayed low levels of both protein quantity and enzyme activity (personal communication, Dr. Weinshilboum). Structural analysis of human TPMT structures (PDB accession code 2BZG, Wu *et al.*, 2007) showed that Tyr107 is in a water-mediated surface hydrogen-bonding network, but given its location on the protein surface it is not obvious that Tyr107Asp would destabilize the native structure. To investigate the molecular basis for low levels of both protein quantity and enzyme activity of this TPMT variant, two sets of molecular dynamics simulation calculations were performed, using the wild-type human TPMT crystal structure and a computationally modeled Tyr107Asp variant generated by computationally "mutating" the wild-type human TPMT crystal structure as starting points. These calculations suggest that the tyrosine residue substitution by aspartic acid is not compatible with the wild-type protein conformation and may destabilize the native structure (Chapter III).

In summary, the combination of molecular dynamics simulations and crystal structures in the structural studies of TPMT reported here improve our understanding of the molecular basis of TPMT polymorphisms. Given the continuing improvements of computer power and molecular dynamics simulations as well as the availability of molecular dynamics

programs, we look forward to more comprehensive applications of molecular dynamics simulations in the future.

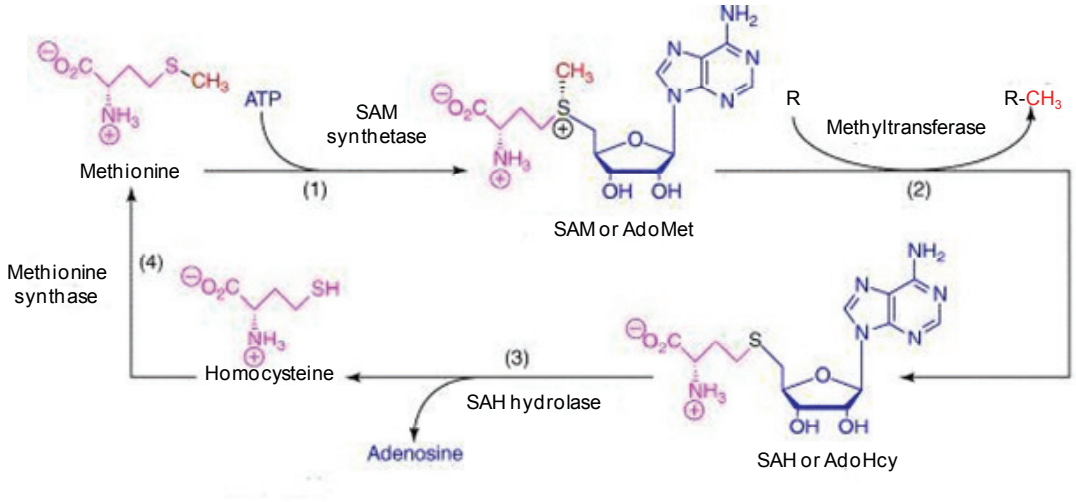


Figure I-1 The metabolic cycle of S-adenosylmethionine (SAM or AdoMet).

(1) AdoMet is synthesized by SAM synthetase using methionine and ATP;

(2) Methyltransferases catalyze the transfer of the methyl group of AdoMet to nucleophilic substrates (R) and AdoMet is converted to S-adenosylhomocysteine (SAH or AdoHcy);

(3) AdoHcy is hydrolyzed by SAH hydrolase to generate homocysteine and adenosine;

(4) Homocysteine is methylated by methionine synthase to regenerate methionine. (Adapted from Fontecave *et al.*, 2004).

Figure I-2

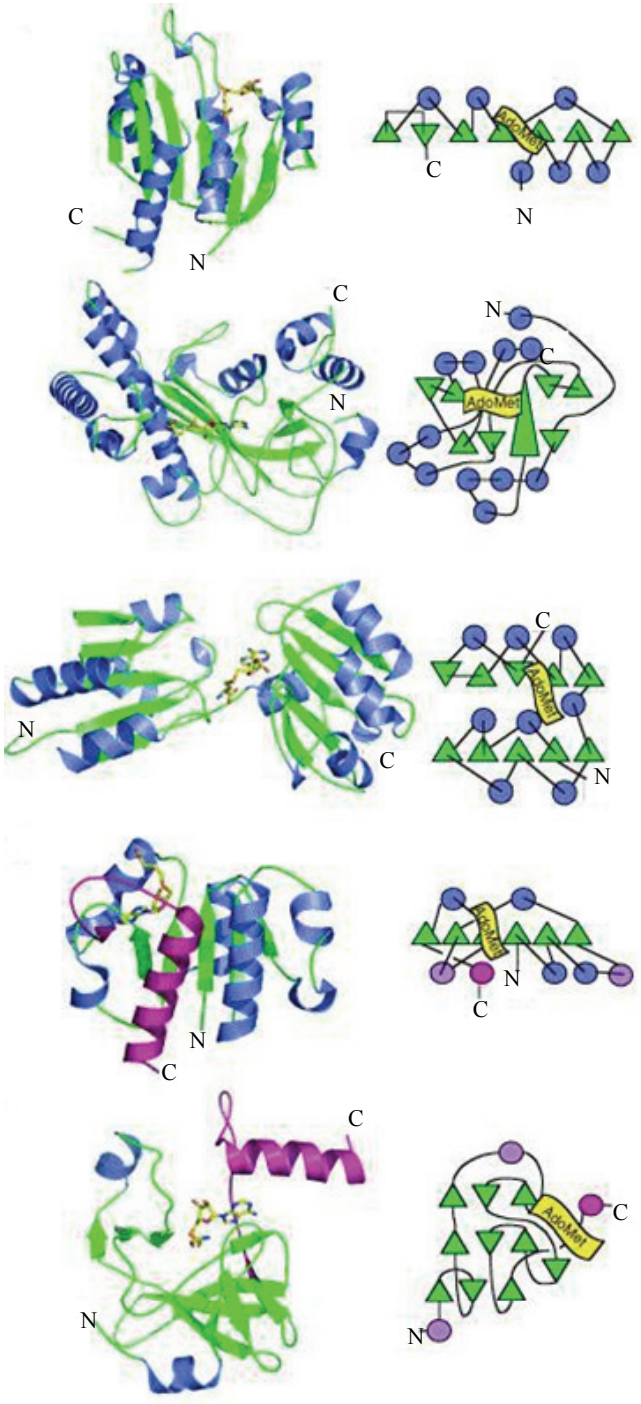

(a)
C
N
AdoMet
C
N
(b)
C
N
N
C
AdoMet
(c)
N
C
C
AdoMet
N
(d)
N
C
AdoMet
N
C
(e)
C
N
AdoMet
C
N

Figure I-2 Structures of the five classes of S-adenosylmethionine-dependent methyltransferases (AdoMet-dependent MTases). In each case a representative structure (left) and topology diagram (right) is displayed.

(a) Class I: the structures of over 33 family members have been solved. All of them include a seven-stranded β sheet flanked by α helices. (enzyme:PDB code, M.HhaI:6MHT).

(b) Class II: the reactivation domain of methionine synthase comprises a series of long β strands and AdoMet is bound to a shallow groove on the surface of the domain (MetH:1MSK).

(c) Class III: the bilobal structure of CbiF has an AdoMet binding site between the two αβα domains (CbiF:1CBF).

(d) Class IV: the SPOUT family of RNA MTases displays a unique knot structure (magenta) at the C terminus (YibK:1MXI).

(e) Class V: the SET-domain formed by the combination of three small β sheets. AdoMet is bound to a shallow groove of the protein and the substrate active site is close to the C-terminal tail (Set7/9:1O9S). (Adapted from Schubert *et al.*, 2003a)

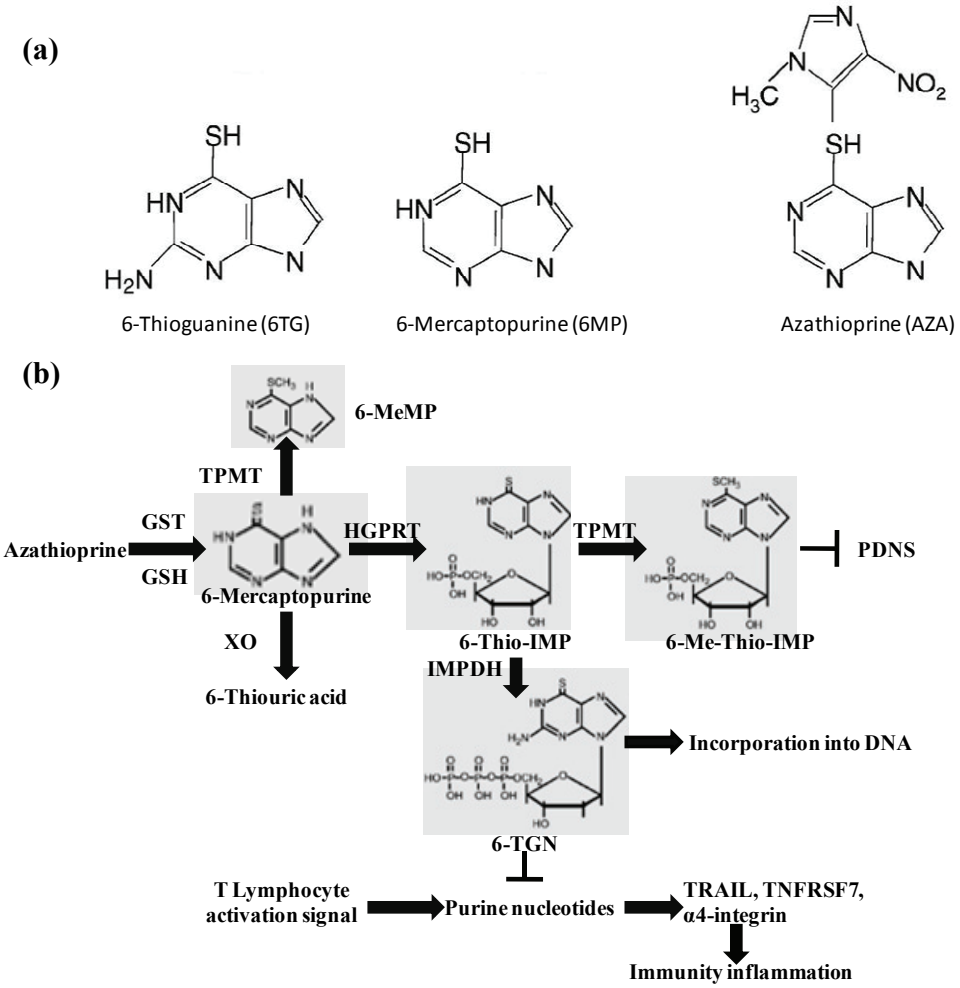


Figure I-3 Thiopurines.

(a) Structures of thiopurines.

(b) Schematic representation of thiopurine metabolism. HGPRT, Hyproxanthine-guanine-phosphoribosyl-transferase; IMPDH, inosine monophosphate dehydrogenase; TPMT, thiopurine S-methyltransferase; XO, xanthine oxidase; 6-MeMP, 6-methylmercaptopurine; 6-Thio-IMP, thioinosine 5'-monophosphate; 6-Me-Thio-IMP, S-methyl-thioinosine 5'-monophosphate; PDNS, purine *de novo* synthesis; 6-TGN, thioguanine nucleotide. ((a) Adapted from Coulthard and Hogarth, 2005 and (b) adapted from Sahasranaman *et al.*, 2008).

Figure I-4

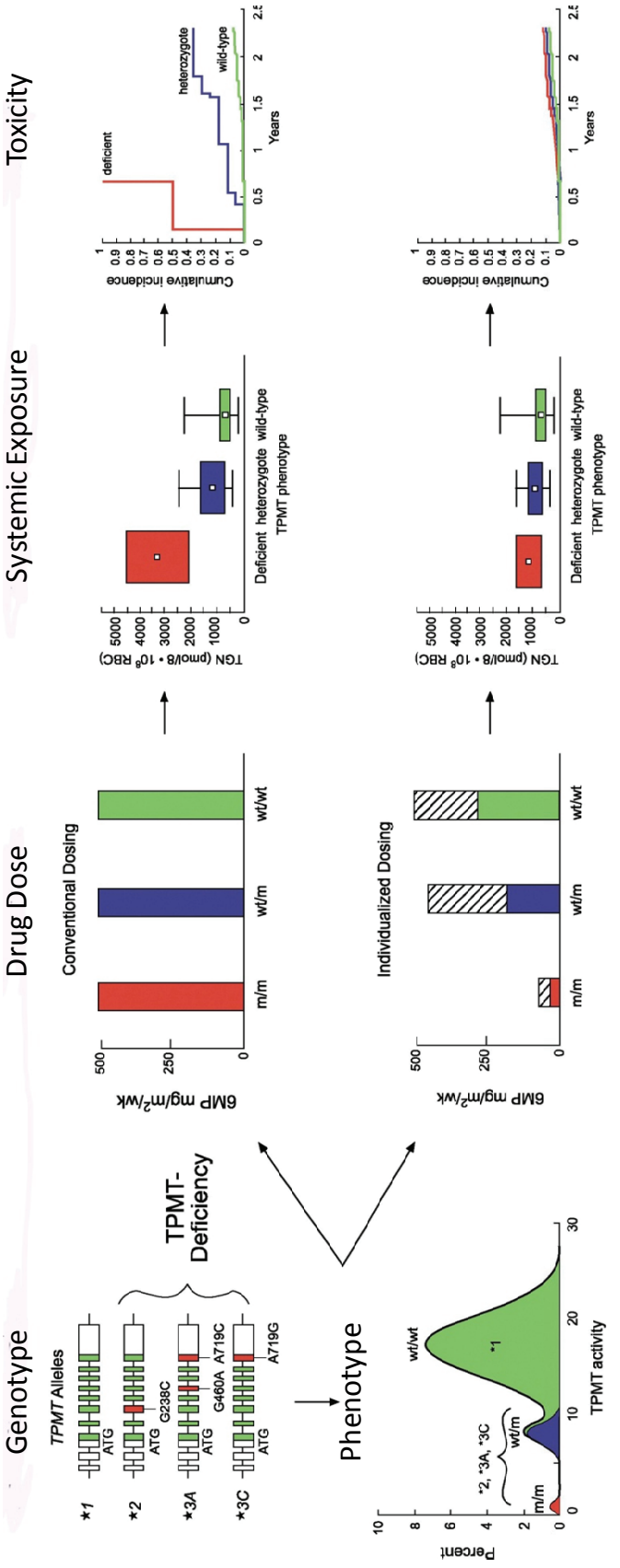
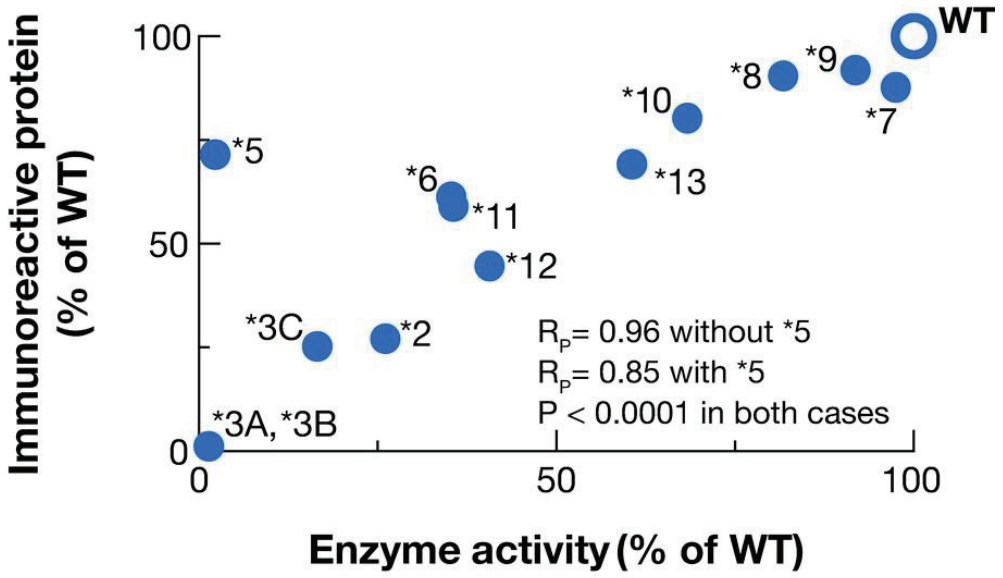

Genotype
TPMT Alleles
*1
*2
*3A
*3C
ATG
G238C
G460A
A719C
A719G
TPMT-Deficiency
Phenotype
Percent
TPMT activity
wt/wt
wt/m
m/m
*1
*2, *3A, *3C
Drug Dose
Conventional Dosing
Individualized Dosing
6MP mg/m²/wk
Systemic Exposure
TGN (pmol/8 • 10⁸ RBC)
Deficient heterozygote wild-type
TPMT phenotype
Toxicity
Cumulative Incidence
deficient
heterozygote
wild-type
Years

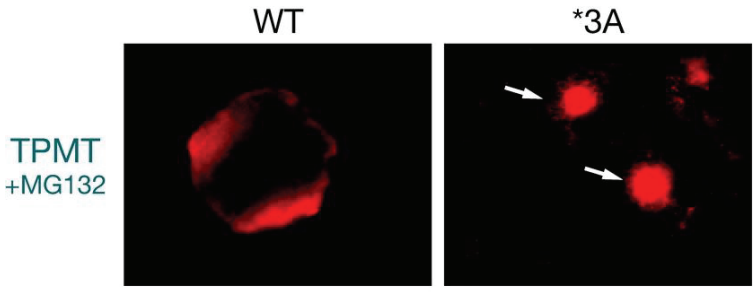
Figure I-4 Thiopurine S-methyltransferase (TPMT) polymorphism and its clinical role in drug response. The left panels show the predominant TPMT variant alleles (top), which cause the inherited individual difference of TPMT activity in humans (bottom). As depicted in the adjacent top panels, when conventional dosages of thiopurine are administered to all patients (left), TPMT-deficient patients accumulate remarkably higher cellular TGN concentrations (middle), which can result in a higher frequency of toxicity (right). As depicted in the bottom three panels, when individual dosage is used according to the genotype of patients, similar cellular TGN concentrations are obtained and all patients have no risk of acute toxicity. (Adapted from Eichelbaum *et al.*, 2006).

Figure I-5

(a)


Immunoreactive protein (% of WT)
100
50
0
0
50
100
WT
*9
*8
*7
*10
*13
*5
*6
*11
*12
*3C
*2
*3A, *3B
R_P= 0.96 without *5
R_P= 0.85 with *5
P < 0.0001 in both cases
Enzyme activity (% of WT)

(b)


WT
*3A
TPMT
+MG132

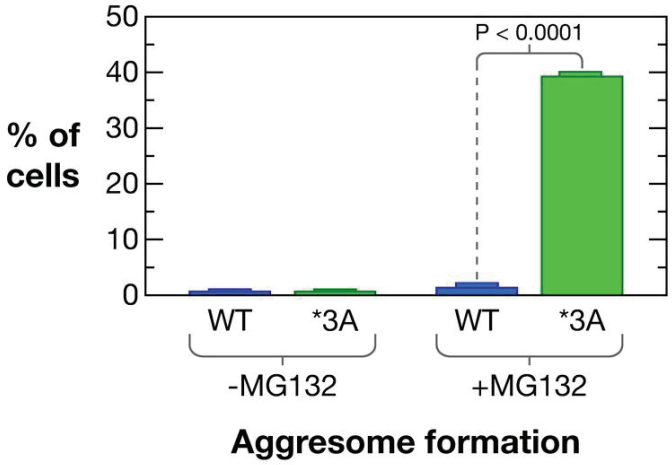

% of cells
50
40
30
20
10
0
P < 0.0001
WT
*3A
WT
*3A
-MG132
+MG132
Aggresome formation

Figure I-5 Thiopurine S-methyltransferase (TPMT) variant allozyme functional genomics and aggresome formation.

(a) TPMT variant allozyme functional genomics. The figure shows correlation of human TPMT recombinant allozyme activity and level of immunoreactive protein expressed in COS-1 cells. Correlation coefficients are shown as Rp.

(b) TPMT*3A aggresome formation. COS-1 cells transfected with WT and TPMT*3A was treated with MG132 and then observed by fluorescence microscopy. Aggresomes were shown by the arrows. The bar graph shows comparison of TPMT*3A aggresome formation with WT TPMT. (Adapted from Weinshilboum and Wang, 2006).

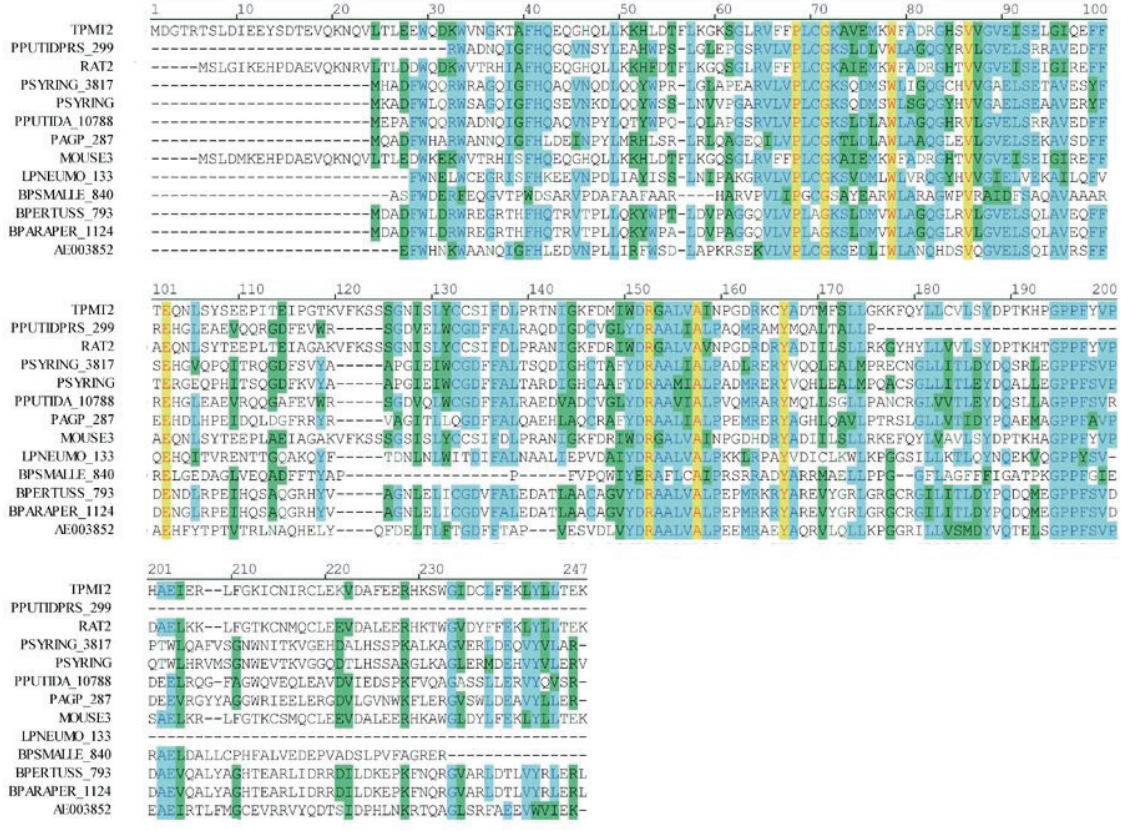


Figure I-6 Protein sequence comparison of Thiopurine S-methyltransferase (TPMT). Human (TPMT2), *P. putida* (PPUTIDPRS_299), rat (RAT2), *P. syringae* (PSYRING_3817), *P. syringae p.v.* (PSYRING), *P. putida* (PPUTIDA_10788), *P. aeruginosa* (PAGP_287), mouse (MOUSE3), *L. pneumophila* (LPNEUMO_133), *B. pseudomallei* (BPSMALLE_840), *B. pertussis* (BPERTUSS_793), *B. parapertussis* (BPARAPER_1124), and *V. cholerae* (AE003852). (Adapted from Krynetski and Evans, 2003)

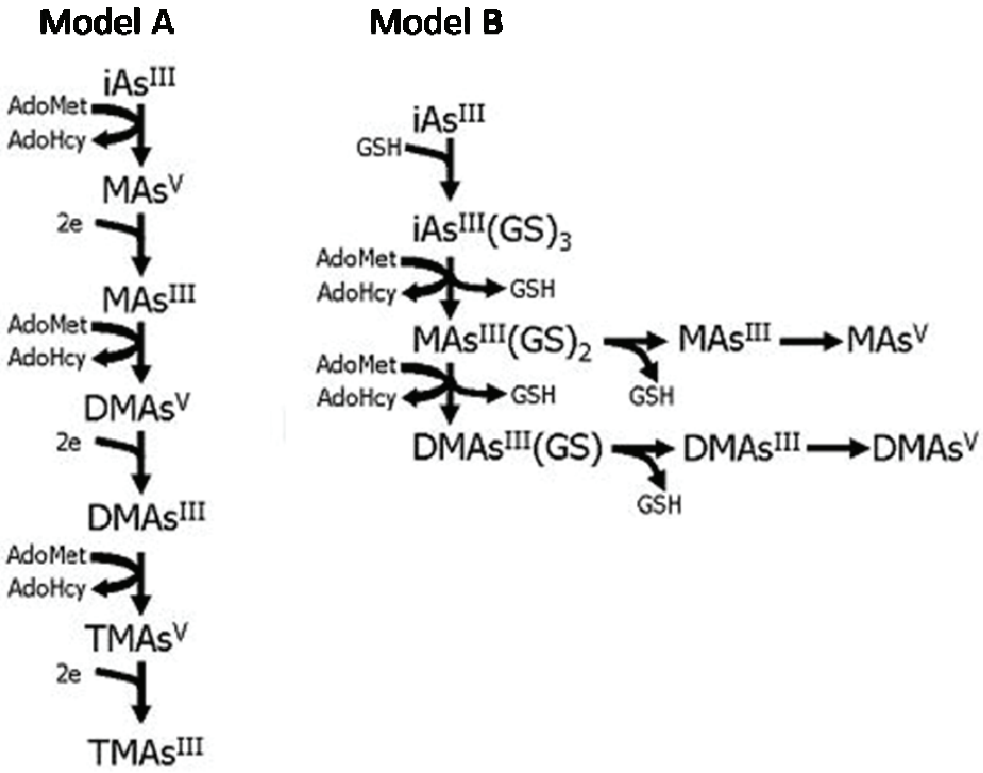


Figure I-7 Two conceptual models for methylation of inorganic arsenic by arsenic methyltransferase (AS3MT). In Model A, oxidative methylation of trivalent arsenicals alternates with reduction of pentavalent arsenicals. In this pathway, arsenite (iAs^{III}) is converted to methylarsonic acid (MAs^{V}), which is then reduced to methylarsonous acid (MAs^{III}). Subsequently, MAs^{III} is converted to dimethylarsinic acid ($DMAs^{V}$), which is reduced to dimethylarsinous acid ($DMAs^{III}$). Finally, $DMAs^{III}$ is converted to trimethylarsine oxide ($TMAs^{V}$), which is reduced to trimethylarsine ($TMAs^{III}$). In Model B, the methyl groups are sequentially transferred to thiol-containing complexes of trivalent arsenicals. Both models utilize the AdoMet as the methyl donor. (Adapted from Thomas *et al.*, 2007)

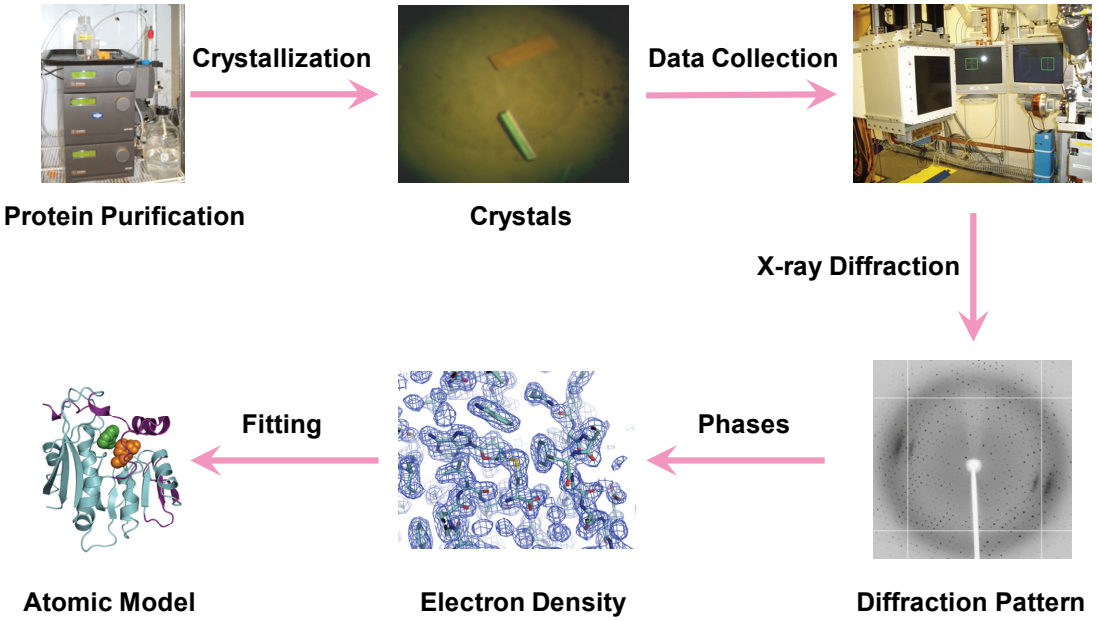


Figure I-8 The major steps in structure determination by protein X-ray crystallography: protein preparation, crystallization, data collection, phasing, model building and refinement.

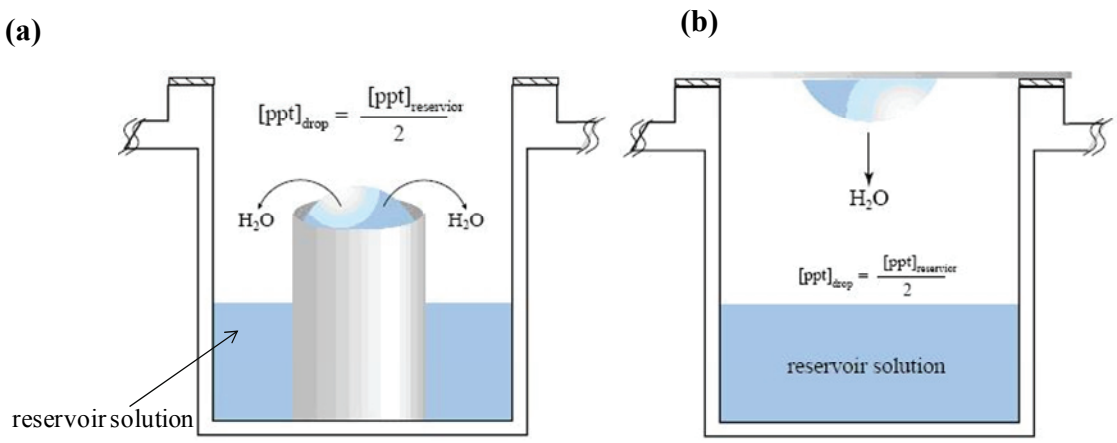


Figure I-9 Crystallization techniques.

(a) Sitting drop vapor diffusion.

(b) Hanging drop vapor diffusion. ((a) & (b) Adapted from http://www.hamptonresearch.com/).

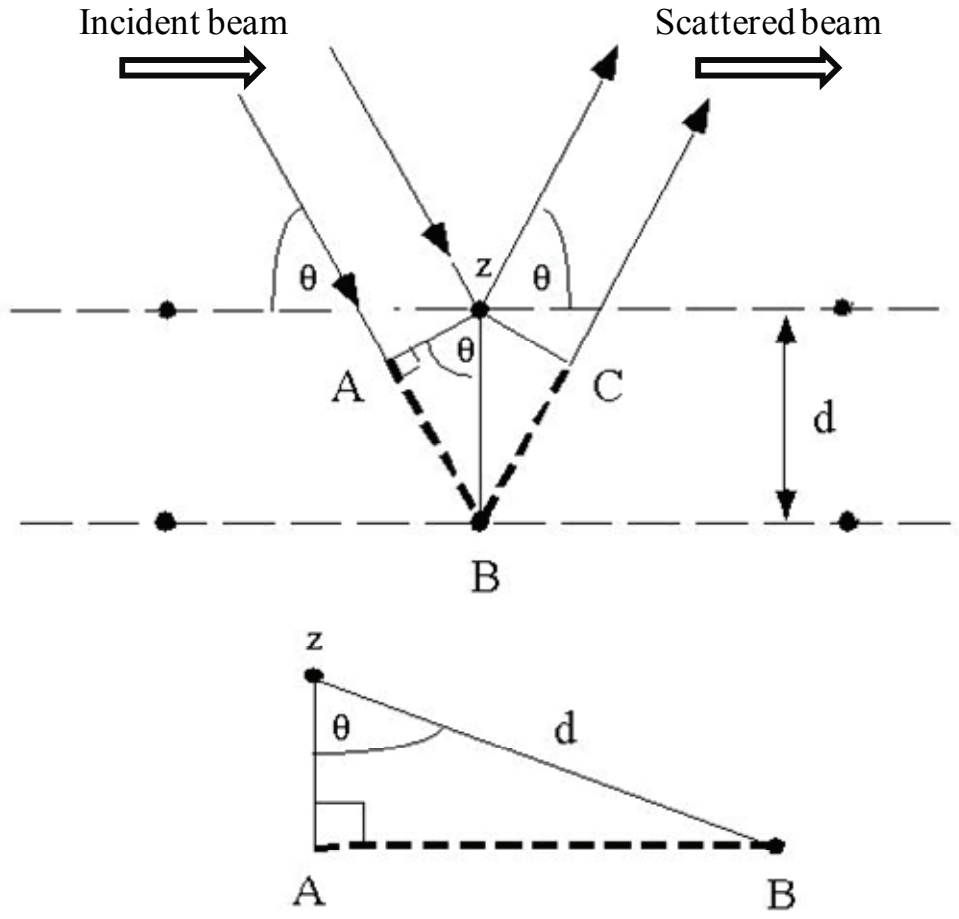


Figure I-10 Bragg model of diffraction. Bragg's law is derived using the reflection geometry. The lower beam must travel the extra distance (AB+BC) to continue traveling parallel and adjacent to the upper beam. (Adapted from http://www.eserc.stonybrook.edu/ProjectJava/Bragg/index.html).

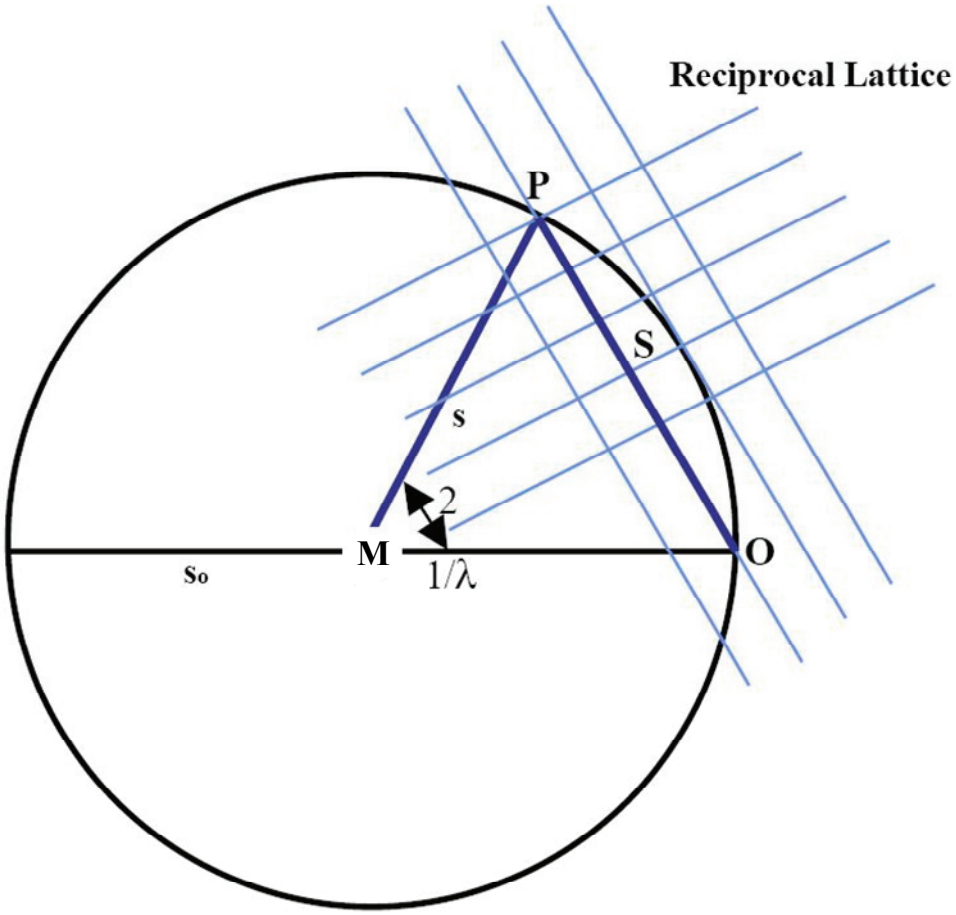


Figure I-11 Diffraction occurs when a reciprocal lattice point intersects the Ewald's sphere. (Adapted from http://www.tdx.cbuc.es/TESIS_UAB/AVAILABLE/TDX-1209102-142528//eep2de4.pdf)

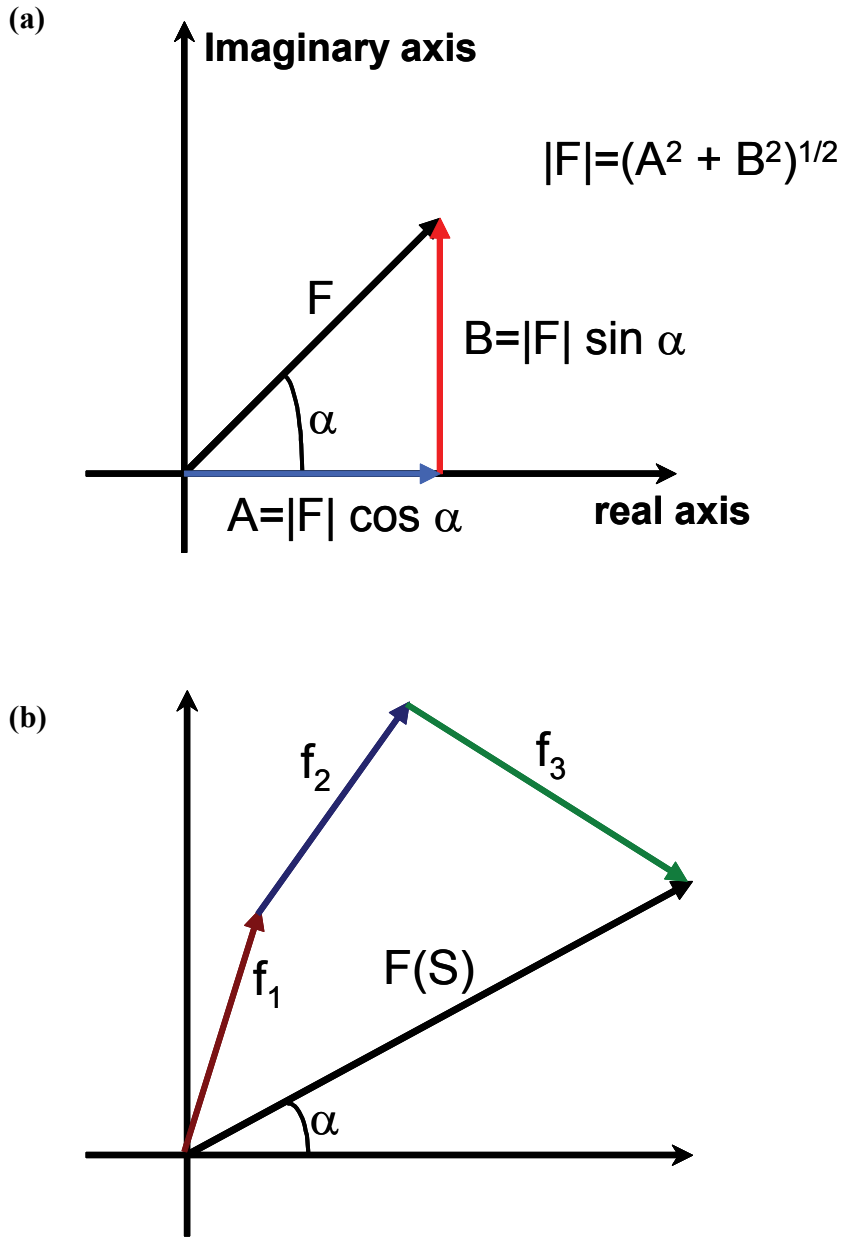


Figure I-12 Descriptions of waves.

(a) Representation of a wave by a vector F.

(b) Scattering by the individual atoms in the unit cell can be summed to obtain the structure factor F(S). ((a) & (b) Adapted from Hall, 2004)

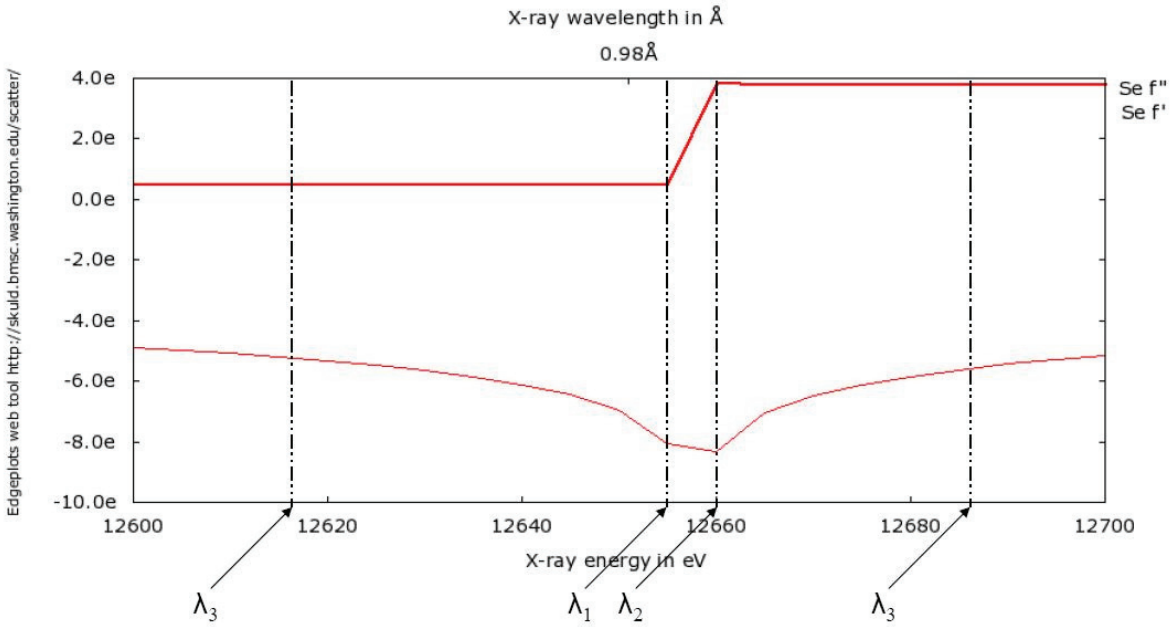


Figure I-13 Plot of the theoretical absorption edge of selenium from 12600eV to 12700eV with the edge, peak and short/long remote labeled as λ_1, λ_2 and λ_3 (Generated by http://skuld.bmsc.washington.edu/scatter/). In this plot, the top line Se f'' (the imaginary anomalous scattering factor), represents the change in absorption of fluorescence as a function of X-ray energy. The bottom line Se f' (the real anomalous scattering factor), is calculated by numerical integration from f''.

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2} K_b (b-b_0)^2 + \sum_{\text{All Angles}} \tfrac{1}{2} K_\theta (\theta-\theta_0)^2$$

$$+ \sum_{\text{All Torsion Angles}} K_\Phi [1-\cos(n\Phi+\delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \varepsilon[(r_0/r)^{12} - 2(r_0/r)^6]$$

$$+ \sum_{\text{All partial charges}} 332\, q_i q_j / r$$

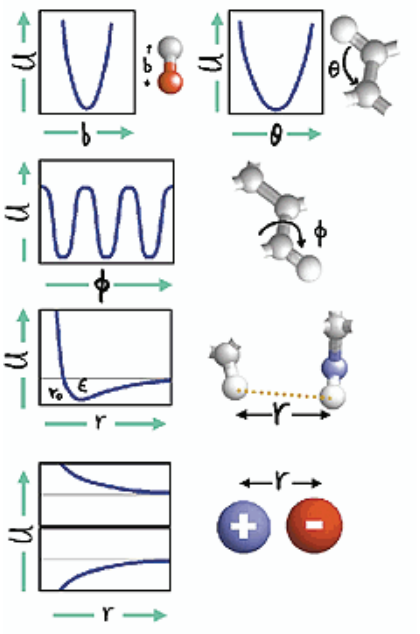


Figure I-14 A description of the empirical potential energy functions for simulations of biomolecules. The total potential energy of any molecule is the sum of bonding terms describing bonds lengths, bond angles and torsional angles, and non-bonding terms composed of van der Waals interactions and electrostatics. (Adapted from Levitt, 2001)

CHAPTER II

Crystal Structures of Wild-type TPMT:

Structural Basis of Substrate Recognition in Thiopurine S-Methyltransferase

(Peng *et al.*, 2008)

2.1 Introduction

Thiopurines such as 6-mercaptopurine (6MP), 6-thioguanine, and azathioprine are cytotoxic and are immunosuppressant compounds which are used to treat childhood acute lymphoblastic leukemia, inflammatory bowel disease, and transplant rejection (Paterson and Tidd, 1975; Lennard, 1992; Coulthard and Hogarth, 2005). They are prodrugs which are extensively metabolized to give thioguanine nucleotides that exert their cytotoxic effects by incorporation into DNA or inhibiting purine *de novo* synthesis. Thiopurine S-methyltransferase (TPMT) is a cytosolic enzyme which modulates the cytotoxicity of thiopurines by taking them as substrates in an S-methylation reaction. TPMT is a classic example of pharmacogenetics, in which allelic variation in patients results in a range of drug response (Cheok and Evans, 2006; Wang and Weinshilboum, 2006). Variant alleles result in TPMT enzymes whose activities range from barely detectable to near wild-type level, and as a consequence a 10- to 15-fold range in thiopurine dosage is necessary in order to minimize toxicity and other complications in patients.

The TPMT reaction takes two substrates, the methyl donor S-adenosyl-L-methionine (SAM or AdoMet) and a methyl acceptor, and generates S-adenosyl-L-homocysteine (SAH or AdoHcy) and a methylated product. Although 6MP and other thiopurines are

well characterized as TPMT methyl acceptors, a natural substrate for this enzyme has yet to be identified. The crystal structure of human TPMT bound to AdoHcy, and the NMR structure of a bacterial TPMT ortholog have been reported (Scheuermann *et al.*, 2003; Wu *et al.*, 2007). While these structures characterized the TPMT fold and the AdoMet binding site, they did not reveal the details of methyl acceptor recognition. We report here the crystal structures of murine TPMT bound to AdoHcy, with and without 6MP additionally present in the methyl acceptor binding site. These structures demonstrate structural and conformational features of 6MP binding and highlight two nearby arginine residues which were subsequently tested by mutagenesis for their role in substrate binding and TPMT activity.

2.2 Materials and Methods

2.2.1 Expression and Purification of TPMT

The full-length murine TPMT (mTPMT) and human TPMT (hTPMT) coding regions were amplified by PCR from eukaryotic expression plasmids (Salavaggione *et al.*, 2005) and cloned into the *NdeI* and *EcoRI* sites of pET-28a (Novagen) to give expression constructs coding for TPMTs with an N-terminally fused His_6-tag. Transformed *E. coli* BL21(DE3) cells were grown at 37 ^{o}C to an OD_{600} of ~0.8 before induction with 0.4 mM IPTG and were incubated for an additional 8 h at 37 ^{o}C. Cells were then harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.9, 500 mM NaCl, and 5 mM imidazole) with the protease inhibitor cocktail set II (Calbiochem) for sonication. After centrifugation, the supernatant was loaded onto Ni^{2+}-NTA resin (Qiagen), washed with the lysis buffer, and the expressed TPMT protein eluted with 50

mM Tris-HCl, pH 7.9, 500 mM NaCl, and 150 mM imidazole. The pooled eluted fractions were dialyzed against 50 mM Tris-HCl, pH 8.2, and 2 mM DTT. Final purification was by gel filtration on a Superdex75 10/300 GL column (Amersham Biosciences) equilibrated with 20 mM Tris-HCl, pH 8.0. MALDI-MS analysis of the purified mTPMT revealed that both AdoMet and AdoHcy had co-purified with mTPMT, with AdoHcy the major species of the two (data not shown, Mass Spectrometry Facility, Institute of Pathology, Case Western Reserve University). Selenomethionine (SeMet) substituted mTPMT was obtained with the same expression system by inhibition of methionine biosynthesis (van Duyne *et al.*, 1993) and purified as above except that 5 mM DTT was added in dialysis and gel filtration. Point mutations in the hTPMT expression plasmid were prepared using the QuikChange site-directed mutagenesis kit (Stratagene), and the hTPMT mutants were expressed and purified using the wild-type protocol.

2.2.2 TPMT Enzyme Activity Assays

TPMT enzyme activity was measured with a radiochemical assay based on the methylation of 6MP with [^{14}C-methyl]-AdoMet as the methyl donor (Weinshilboum *et al.*, 1978). The reaction mixture contained 7.5mM 6MP (Sigma), 24.2μM AdoMet (7.2μCi/μmol, Perkin-Elmer), 1mM DTT, 50μM allopurinol, and 150 mM KH_2PO_4 buffer (pH 6.5). For substrate kinetic experiments, 6MP concentrations varied from 30μM to 7.5mM and AdoMet concentrations ranged from 0.09μM to 48.4 μM. One unit of TPMT activity represented the formation of 1 nmol of 6-methyl-mercaptopurine per second of incubation at 37 °C. The kinetic experiments were carried out in triplicate and

standard deviations were calculated for the average values of Km and Vmax determined (**Table II-2**).

2.2.3 Crystallization and Data Collection of TPMT

All crystals were grown by sitting drop vapor diffusion at 20 °C using a protein concentration of ~6 mg/ml and a 1:1 protein:precipitant volume ratio. Four different mTPMT crystal forms were grown from a variety of conditions which fell into four types: condition 1 with 22%-32% (w/v) PEG3350, 0.2 M lithium sulfate/ammonium sulfate and 0.1 M imidazole/Mes/bis-tris pH 5.2-6.5; condition 2 with 20-22% PEG3350 and 0.15 malic acid (pH 5.5); condition 3 with 30-35% pentaerythritol ethoxylate, 0.05 M ammonium sulfate, and 0.05 M imidazole/Mes/bis-tris (pH 6.0-6.5); and condition 4 with 20% PEG3350 and 0.2 M potassium sulfate. Crystals of mTPMTwtSeMet-AdoHcy were grown from condition 1 with 25%-32% (w/v) PEG3350 and 0.1 M imidazole (pH 5.2-6.0). mTPMTwt complex crystals were grown by mixing proteins with AdoHcy (2mM), AdoMet (5mM), and 6MP (10mM) respectively. Crystals of mTPMTwt-AdoHcy-6MP were grown from condition 1 with 25-27% (w/v) PEG3350 and 0.1 M imidazole (pH 5.5), and from condition 2. Crystals of mTPMTwt-AdoMet were obtained from condition 1 with 22-27% (w/v) PEG3350 and 0.1 M imidazole/Mes (pH 6.0 or 6.5), and from conditions 2 and 3. mTPMTwt-AdoHcy crystals grew from condition 4. Crystals were cryoprotected in artificial mother liquor containing 35% (w/v) PEG3350 or pentaerythritol ethoxylate and flash cooled in liquid nitrogen. All data were collected at APS beamlines 19ID (mTPMTwt-AdoHcy-6MP, mTPMTwt-AdoHcy,) and 19BM (SeMet-mTPMT-AdoHcy). The data were processed with HKL (Otwinowski and Minor,

1997) (**Table II-1**). Crystals belong to space group $P2_1$ and contain two molecules in each asymmetric unit.

2.2.4 TPMT Structure Determination and Refinement

SeMet MAD data were measured at the peak and edge wavelengths (0.97877 Å and 0.97888 Å) at APS 19BM from a mTPMTwtSeMet-AdoHcy crystal. Initial heavy atom sites were found with SOLVE (Terwilliger and Berendzen, 1999) and improved by RESOLVE (Terwilliger, 2000&2003), such that the mean figure of merit increased from 0.29 to 0.59. During the model-building process, an unpublished structure of truncated hTPMT with AdoHcy was deposited in the Protein Data Bank (PDB accession code 2BZG) and was used to aid the interpretation of electron density maps. The initial mTPMTwt-AdoHcy model was in turn used to solve the mTPMTwt-AdoHcy-6MP structures. Interactive model building was carried out with COOT (Emsley and Cowtan, 2004) and refinement calculations were performed in CNS (Brunger *et al.*, 1998) and REFMAC (Murshudov *et al.*, 1997). The quality of all models was assessed by PROCHECK (Laskowski *et al.*, 1993). The majority of the residues are located in the most favored region of the Ramachandran plot. No clear electron density for the N-terminal His_6-tag and 8-9 residues of mTPMT was observed, and these residues were not included in the final models. A summary of the final refinement statistics is provided in **Table II-1**. Molecular figures were generated using PyMol (DeLano, 2004). Coordinates and structure factors for mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP have been deposited in the Protein Data Bank with accession codes 3BGI and 3BGD, respectively.

2.3 Results

2.3.1 Overall Structures of TPMT Complexes

Although all possible binary and ternary complexes of mTPMTwt with AdoMet, AdoHcy, and 6MP were crystallized, the electron density maps revealed that AdoHcy was bound in all crystal forms, even in those which were grown from mTPMT incubated with AdoMet. This is presumably due to the predominance of AdoHcy, which is proved to copurify with mTPMT by MALDI-MS analysis mentioned above. Thus, the many crystal forms grown finally resulted in two unique crystal structures of two different complexes: mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP (**Table II-1**). These structures of mTPMTwt possesses the expected class I AdoMet-dependent MTase core fold (Martin and McMillan, 2002; Schubert *et al.*, 2003a; Kozbial and Mushegian, 2005), described in Chapter I (**Figure II-1a**). Moreover, the mTPMT structure has several modifications to the MTase core: an additional 40 residues at the N-terminus, a reversed β hairpin inserted from residues Tyr102-Ser122, and a small helix (Glu220-Trp225) inserted between the last two β strands.

A DALI structural database search with the mTPMT coordinates identified over fifty similarly folded proteins, all with <20% shared sequence identity (Holm and Sander, 1993), but limited information on acceptor substrate binding. Only five structural matches are of complexes with both the methyl donor and acceptor binding sites occupied. Of these, the two most similar structural homologs are phenylethanolamine N-methyltransferase (Martin *et al.*, 2001) (3.1 Å rmsd for 189 superimposed Cα atoms, 12% sequence identity) and histamine N-methyltransferase (Horton *et al.*, 2005) (3.6 Å rmsd

for 178 superimposed Cα atoms, 12% sequence identity), each bound to AdoHcy and an inhibitor. Both ternary complexes revealed inhibitors bound in positions or orientations too far away from the AdoHcy (> 7 Å) to represent productive acceptor substrate binding modes. The remaining three complex structures reveal ligands bound productively in the acceptor binding site and are of *Clarkia breweri* salicylic acid carboxyl methyltransferase (SAMT, 3.4 Å rmsd for 191 superimposed Cα atoms, 8% sequence identity) (Zubieta *et al.*, 2003), *Thermotoga maritima* N^5-glutamine methyltransferase (product of the protein release factor methylation gene, PrmC, 3.2 Å rmsd for 155 superimposed Cα atoms, 10% sequence identity) (Schubert *et al.*, 2003b), and *Thermotoga maritime* putrescine aminopropyltransferase (PAPT, 3.5 Å rmsd for 166 superimposed Cα atoms, 7% sequence identity) (Korolev *et al.*, 2002).

For each complex crystal structure we determined, there are two copies of mTPMT in the asymmetric unit. Superpositions of the Cα atoms for molecules in the same crystal and across crystals give overall rmsd of 0.2-0.8 Å and 0.8-1.0 Å, respectively. Unless otherwise stated, structural details provided in the following sections are for the mTPMTwt-AdoHcy-6MP structure and are similar in the mTPMTwt-AdoHcy complex.

2.3.2 AdoHcy Binding to TPMT

In all mTPMT molecules in both structures, the electron density for AdoHcy is well-defined and clearly shows that AdoHcy is present since there is no density corresponding to the sulfur-bound methyl group characteristic of AdoMet (**Figure II-1b**). As seen in other class I MTases, AdoHcy binds to mTPMT in an extended conformation at the N-

terminal half of the central β-sheet, with the N-terminal extension and inserted helix covering the occupied active site (**Figures II-1a, b**). Hydrogen bonding interactions between AdoHcy and mTPMT are similar to those in other class I MTase structures as described in Chapter I. The AdoHcy terminal amine group forms hydrogen bonds with a water molecule and the carbonyl oxygen atoms of Leu64 and Arg147 in all four complex molecules. The ligand's carboxylate is surrounded by a more flexible region of the protein and participates in different sets of interactions in different structures, interacting with one or two solvent molecules as well as with the side chain of Arg147 (only in mTPMTwt-AdoHcy-6MP). The AdoHcy ribose hydroxyl groups form hydrogen bonds with the Trp28 and Glu85 side chains while the adenine's ring nitrogen atoms interact with the main chain amines of Ile86 and Ile130 and a water molecule. Among the other contacts that stabilize AdoHcy binding is the sandwiching of the adenine ring between the Leu64 and Ile86 side chains. AdoHcy binding to mTPMTwt, when comparing the two molecules within each crystal and between the two crystal structures, is with the ligand in essentially the same position in identical conformations (rmsd of ~0.4 Å for pairwise superpositions). Of the AdoHcy binding residues (**Figure II-2a, b**), those whose side chains interact with AdoHcy (Trp28, Leu64, Glu85, Ile86) are conserved in not only human and murine, but also other mammalian TPMTs (**Figure II-3**); Ile130, which is involved in only a main chain interaction, is conservatively substituted with a leucine in some of the other mammalian sequences.

2.3.3 6MP Binding to TPMT

As documented before, Class I MTases bind AdoMet (or the demethylated product AdoHcy) at the same methyl donor site in their structures, reflecting their shared AdoMet dependence. But they must also have diverse active sites to accommodate their wide variety of acceptor substrates (Martin and McMillan, 2002). In the mTPMT complex structures reported here, the electron density for the 6MP acceptor substrate is weaker and less defined than for AdoHcy (**Figure II-1c, d**), which is consistent with the observation that the 6MP binding affinity is weaker than that of AdoMet by 2-3 orders of magnitude (**Table II-2**). For one molecule in the mTPMTwt-AdoHcy-6MP, the electron density had a flattened shape suitable for the bicyclic purine structure, but the location of the protruding thiol was ambiguous (**Figure II-1c**). Therefore, we modeled 6MP in each of the two possible positions, related by ~180º rotation perpendicular to the plane of the rings. However, upon each possible 6MP position was superimposed the structure of thioinosine monophosphate (TIMP), a larger TPMT substrate which is an intermediate in thiopurine metabolism (Krynetski *et al.*, 1995b), TIMP was found to fit in only one orientation so that no severe clashes happen with the TPMT polypeptide backbone. Since the two TPMT substrates are expected to bind similarly in the active site due to their high degree of chemical similarity, this orientation was chosen for final refinement of 6MP. The density for the second molecule in mTPMTwt-AdoHcy-6MP was defined enough to allow identification of the sulfur position and thus of the 6MP orientation (**Figure II-1d**).

Moreover, the mTPMTwt-AdoHcy-6MP structure shows the 6MP bound to the two independent molecules similarly, in essentially the same position in the active site but in different orientations (**Figure II-2c, d and II-4a**). The inter-sulfur distances of 3.4 and

4.1 Å for the two molecules are about double the 1.85 Å length of a sulfur-carbon single bond, consistent with the presumption mentioned before that AdoMet-dependent MTases adopt a conserved S_N2-like mechanism in which the methyl group is transferred directly from AdoMet to the acceptor. Participating in no direct hydrogen bonding interactions with the protein, 6MP is bound in the active site by a number of van der Waals interactions (**Figure II-2a, b**). It is noticeable that Pro191 and Pro190 are conserved in all TPMTs (**Figure II-3**), and their *cis*-peptide bonds define one face of the 6MP binding site. In addition, the side chains of Pro191 and Phe35, conserved in all mammalian TPMTs, form a hydrophobic clamp that orients the 6MP purine ring. Arg147, which forms a salt bridge with the AdoHcy carboxylate, and Arg221 are the only hydrophilic residues within 4 Å of 6MP and thus are the only candidates that may possibly be involved in the deprotonation of 6MP which is expected before the subsequent methylation by AdoMet. Arg147 is about 3Å from 6MP in both mTPMT-AdoHcy-6MP complexes, while Arg221 is 3.2 Å and 5.4 Å from 6MP in the two molecules.

2.3.4 A Flexible Active Site Loop in TPMT

By comparing the two molecules in each of the mTPMTwt-AdoMet and mTPMTwt-AdoHcy-6MP structures, the regions which change conformation upon 6MP binding are revealed. The two molecules in mTPMTwt-AdoHcy superimpose to give an rmsd of 0.8 Å for all Cα atoms. The regions whose conformations differ the most between the two molecules are residues Ser34-His47 which form an extended loop over the active site, (the "active site loop", **Figure II-2c**), and residues Glu220-Trp225 which is a short helix in an adjacent loop that contains Arg221, one of the two arginine residues which is near

the bound 6MP in mTPMTwt-AdoHcy-6MP. It's interesting to note that these two segments are not part of the core protein fold of the class I AdoMet-dependent MTases, and are characteristic of TPMT structure. When these two regions are omitted from the superposition of the two molecules in mTPMTwt-AdoHcy, the Cα rmsd drops to 0.2 Å. In addition, these two regions also have a much higher average B factor (70 $Å^2$) than for the rest of the molecule (21$Å^2$), characteristic of high flexibility. In contrast, the two molecules in mTPMTwt-AdoHcy-6MP superimpose to give a low 0.2 Å r. m. s. d. for all Cα atoms, and the average B factor for the two loop regions (36 $Å^2$) is much closer to that of the rest of the protein (23 $Å^2$), suggesting that upon 6MP binding the active site loop and the adjacent Arg221-containing helix become ordered (**Figure II-2d)**. A similar superposition of the mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP structures highlights the extended active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix as the segments which change the most upon 6MP binding: the pairwise Cα rmsd values drop from 0.8-1.0 Å to 0.3-0.5 Å when the two regions are omitted.

2.3.5 Structure-based Mutagenesis of TPMT

To investigate whether either of the arginine residues near the bound 6MP is important for enzyme activity, single and double mutants were prepared in the more biomedically relevant hTPMT. In the single mutants, Arg152 in hTPMT (Arg147 in mTPMT) or Arg226 (Arg221 in mTPMT) was substituted with either alanine, glutamic acid, or histidine. The double mutant hTPMT Arg152Ala/Arg226Ala was also generated. Enzyme activity assays were carried out using radiolabeled AdoMet, and substrate kinetic experiments were performed by varying AdoMet and 6MP concentrations separately

(**Table II-2**). Most of the mutants exhibit Km and Vmax values which are similar to those of wild-type enzyme. The only mutants which deviate substantially from wild-type activity are Arg152Glu and Arg152Ala/Arg226Ala. When Arg152 is replaced with an oppositely charged glutamic acid, the Km for the 6MP acceptor, but not for the AdoMet donor, increases significantly (by ~7- fold), and the Vmax decreases modestly by ~2- to 5-fold. Similarly, when both arginine residues are substituted with alanines, a small decrease in Vmax (~3-fold) accompanies a substantial increase in Km for 6MP (~9- fold) while the Km for AdoMet is not significantly altered.

2.4 Discussion

While TPMT's ability to methylate thiopurine prodrug substrates is well characterized, natural substrates for this enzyme have yet to be identified. To obtain some structural understanding of TPMT substrate recognition, the crystal structures of mTPMT with and without its 6MP acceptor substrate bound were determined. Consistent with the 80% sequence conservation, the mTPMT protein fold is very similar to that of hTPMT, for which co-crystallization efforts with acceptor substrate were unsuccessful (Wu *et al.*, 2007) and to an unpublished crystal structure of mTPMT deposited in the PDB during the course of this work that also did not contain bound acceptor substrate (accession code 2GB4). When the mTPMT-AdoHcy and mTPMT-AdoHcy-6MP crystal structures are superimposed with that of hTPMT-AdoHcy (Wu *et al.*, 2007), the rmsd for equivalent Cα atoms are 1.1 and 0.7 Å respectively and AdoHcy is bound in the same conformation in the active sites of the two enzymes. Both mTPMTwt and hTPMTwt have higher affinity for AdoMet (low micromolar) than for 6MP (low millimolar) (**Table II-2**). This

difference is consistent with the larger accessible surface area buried and greater number of ligand-protein hydrogen bonds formed upon AdoHcy binding to mTPMTwt (~400 $Å^2$, 6-7 hydrogen bonds) compared to 6MP binding to the enzyme (~200 $Å^2$, 1 hydrogen bond). The modest binding affinity of 6MP is also consistent with the observations that 6MP interacts with the enzyme through hydrophobic and van der Waals interactions, in a relatively loose manner reflected by its two different orientations in the two independent molecules in the mTPMTwt-AdoHcy-6MP crystal structure (**Figure II-2a, b**).

As addressed above, the structural database search identified crystal structures of SAMT (Zubieta *et al.*, 2003), PrmC (Schubert *et al.*, 2003b), and PAPT (Korolev *et al.*, 2002) as the most similar ternary complexes for comparison with TPMT. Of these enzymes, TPMT, SAMT, and PrmC use AdoMet as a methyl donor while PAPT uses decarboxylated AdoMet (dcAdoMet) as an aminopropyl donor. All four enzymes bind their donor substrates in similar positions and orientations (**Figure II-4**). Class I AdoMet-dependent MTases are expected to bind donor and acceptor molecules in a manner consistent with an S_N2-like reaction mechanism involving direct transfer between substrates (Woodard *et al.*, 1980; Schubert *et al.*, 2003a&b), and their active sites are expected to be chemically and structurally diverse to accommodate a variety of acceptor substrates (Martin and McMillan, 2002). The mTPMTwt-AdoHcy-6MP structure reveals 6MP-AdoHcy sulfur-sulfur distances of 3.4-4.1 Å consistent with direct methyl transfer (**Figure II-4a**). Superposition of the mTPMT ligands with the AdoHcy and salicylic acid in the SAMT active site (salicylic acid-AdoHcy oxygen-sulfur distances of 4.3-4.7 Å), and the AdoMet and N^5-methylglutamine bound to PrmC (glutamine-AdoMet nitrogen-

sulfur distances of 3.3-3.8 Å) show that the acceptor molecules are bound in similar positions and orientations relative to their corresponding donor molecule **(Figure II-4b,c**). The similar juxtaposition of donor and acceptor molecules is striking considering that these three enzymes transfer methyl groups to different types of atoms (sulfur, oxygen, and nitrogen), in chemically diverse acceptor substrates. The bisubstrate adduct inhibitor S-adenosyl-1,8-diamino-3-thiooctane (AdoDATO) bound to the aminopropyltransferase PAPT identifies donor and acceptor binding sites which coincide with those of the three methyltransferases, although the conformation of the dcAdoMet donor moiety of the inhibitor differs from that seen for AdoHcy/AdoMet in the methyltransferase structures, reflecting the absence of the carboxylate group in the donor, the different chemical group transferred, and the covalent attachment of donor and acceptor substrate groups (**Figure II-4d**).

In addition, the MTase reaction mechanism also requires a deprotonation step before, during, or after the methyl transfer (Schubert *et al.*, 2003a&b). Inspection of the hTPMT-AdoHcy crystal structure led to speculation that Lys32 and Arg226 may be candidates for substrate deprotonation (Wu *et al.*, 2007). The mTPMTwt-AdoHcy-6MP structure we present here provides a first view of acceptor substrate bound to the TPMT active site. The closest approach of mTPMT Lys27 (homologous to hTPMT Lys32) to 6MP is ~6-7 Å, too long for this residue to participate in substrate deprotonation. Superposition of the human and murine structures shows that hTPMT Lys32 and mTPMT Lys27 are in very similar conformations, and that Lys32 is also ~6 Å from 6MP. Only two hydrophilic residues are found within 4 Å of the 6MP bound in the mTPMT active site: Arg147 and

Arg221 (Arg152 and Arg226 in hTPMT). Arg147 in both molecules in the mTPMT structure, and Arg152 in the hTPMT structure, are in nearly identical positions and conformations, and all are about 3Å from 6MP (for Arg152, when the hTPMT and mTPMT structures are superimposed). In contrast, the mTPMT Arg221 and hTPMT Arg226 conformations are less conserved, and the closest approaches to 6MP are 3.0 Å and 5.3 Å (two molecules in mTPMT), and 2.7 Å and 3.5 Å (hTPMT superimposed onto the two mTPMT molecules). Substitution of either Arg152 or Arg226 residues in hTPMT with alanine, glutamic acid, or histidine resulted in modestly diminished enzyme activity for only Arg152Glu (**Table II-2**). Tandem mutations of both residues gave Arg152Ala/Arg226Ala, which also exhibited modestly compromised activity. The average decreases in Vmax are ~ 3-fold for both Arg152Glu and Arg152Ala/Arg226Ala, while Km increased 7- and 9- fold for the single and double mutants, respectively. These observations are consistent with the proximity of the two arginine residues to the 6MP acceptor and suggest that these two residues, while not essential for TPMT activity, nonetheless may play a role in the reaction by contributing to 6MP binding. That the double alanine mutant has altered Vmax and Km for 6MP when the single alanine substitutions have wild-type behavior suggests that the two arginines have overlapping roles such that one residue may compensate when the other is altered. That the only single mutant with altered behavior is Arg152Glu indicates that this residue may contribute more than Arg226, consistent with the more direct interaction of the homologous Arg147 with 6MP that is observed in the mTPMTwt-AdoHcy-6MP structure. The greater importance of Arg152 is also consistent with this residue (mTPMT Arg147) being conserved among all TPMTs while Arg226 (mTPMT Arg221) can vary (**Figure II-**

3). The modest effect on Vmax for these designed TPMT mutants suggests that neither arginine is responsible for substrate deprotonation. In addition, the mTPMT ternary structure does not highlight any solvent molecule or other moiety near the bound 6MP that can be speculated to prepare the 6MP thiol group for the methyltransferase reaction by general acid/base catalysis. The absence of a general base or solvent molecule near the acceptor substrate in mTPMT is a shared observation with the SAMT and PrmC ternary complexes; for the latter the deprotonation step is speculated to occur during or following product release (Schubert *et al.*, 2003b).

Comparison of the mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP structures reveals that a loop containing residues Arg31-Gln55 that extends over the active site, and an adjacent helical turn (Glu220-Trp225) which includes Arg221, are the two regions which change conformation most upon 6MP binding. The active site loop is very flexible in the absence of 6MP, as evidenced by weak density, high B factors, and high rmsd between the two mTPMT molecules (Figure II-2c,d). However, upon 6MP binding, this segment is characterized by well-defined density, lower B factors which are similar to those for the rest of the protein, and a low rmsd between the two molecules. This pronounced conformational alteration of the mTPMT active site loop is reminiscent of a 10-residue "gatekeeping" loop in PAPT which changes conformation and becomes ordered upon AdoDATA binding. Interestingly, the PAPT loop corresponds to mTPMT residues Arg147-Gly156; while this loop does not change conformation in mTPMT upon 6MP binding, it contains the Arg147 which interacts with the acceptor substrate. Acceptor substrate diffusion through a solvent channel to the active site has been proposed upon

surface analysis of the hTPMT crystal structure without acceptor ligand bound (Wu *et al.*, 2007). The characterization of a flexible active site loop in mTPMT identifies an alternative or additional structural feature which plays a role in substrate access and recognition.

The two mTPMT crystal structures reported here identify molecular features that are important for the enzyme's catalytic function. The structures reveal an active site which binds 6MP in two different, overlapping orientations. Other TPMT substrates in thiopurine metabolism include 6-thioguanine and the nucleotides 6-thioinosine-5'-monophosphate and 6-thioguanosine-5'-monophosphate (Krynetski *et al.*, 1995b), which result from coupling of thiopurines to phosphoribosylpyrophosphate by hypoxanthine phosphoribosyltransferase. Structural modeling confirms that these larger acceptor substrates can be accommodated in both bound 6MP orientations since the phosphoribosyl groups would be directed away from the binding site toward solvent. The other and more effective known TPMT acceptor substrates are thiophenol derivatives (Woodson *et al.*, 1983; Ames *et al.*, 1986); their Km values of two to three orders of magnitude less than that of 6MP illustrate that the presence of a heterocyclic aromatic ring is not necessary for substrate recognition, consistent with the observation of few hydrophilic residues near the bound 6MP. Although natural TPMT substrates have yet to be identified, the flexibility and partial solvent accessibility of the 6MP binding site suggest that TPMT may be capable of binding molecules which are substantially larger than 6MP and the related thiopurine metabolites. In addition to aiding speculation on the nature of natural TPMT substrates, the structural features of the acceptor binding site

might also be exploited in designing small molecule modulators or customizing drugs which may allow optimization of thiopurine-based therapy.

Table II-1 Data collection, phasing, and refinement statistics for wild-type mTPMT

	SeMet-wt-AdoHcy		wt-AdoHcy	wt-AdoHcy-6MP
	Peak	Inflection		
Data collection				
cell dimensions	62.60	62.57	62.91	62.89
a, b, c (Å)	66.09	66.12	65.51	69.78
	72.86	72.92	72.90	72.23
β (deg)	115.48	115.53	115.29	115.78
wavelength (Å)	0.97877	0.9788	1.0332	0.97912
resolution (Å)	50-2.5 (2.59-2.5)	50-2.5 (2.59-2.5)	50-1.8 (1.86-1.8)	30-2.0 (2.07-2.0)
R_{sym}	12.6 (38.2)	13.4 (43.2)	4.6 (34.1)	7.3 (26.3)
$I / \sigma I$	10.8 (3.8)	9.7 (2.5)	31.4 (4.7)	13.0 (3.5)
completeness (%)	99.4 (99.1)	99.1 (95.7)	98.5 (96.6)	96.9 (83.6)
redundancy	4.1 (3.9)	3.9 (3.1)	4.9 (4.9)	2.9 (2.4)
Refinement				
resolution (Å)			50-1.8 (1.86-1.8)	30-2.0 (2.07-2.0)
no. of reflections			46591	35011
R_{work} / R_{free}			18.74 (22.07)	20.89 (26.21)
no. of atoms				
protein, ligand, water			3746, 52, 456	3760, 92, 307
average *B*-factors				
protein, AdoHcy, 6MP, water			25.8, 17.3, -, 30.9	24.3, 14.7, 40.9, 29.6
rmsd*				
bond lengths (Å)			0.013	0.013
bond angles (deg)			1.440	1.461
Ramachandran				
most favored (%)			89.1	92.8
additional allowed (%)			10.7	7.2
disallowed (%)			0.2	0.0

Numbers in parentheses refer to the highest resolution shell.
*rmsd, root mean square deviation.

Table II-2 Enzyme activity of TPMT wild-type and mutant proteins

		AdoMet		6MP	
mTPMT	hTPMT	K_m (μM)	V_{max} (units/mg protein)	K_m (mM)	V_{max} (units/mg protein)
wt		5.6 ± 0.5	0.96 ± 0.05	0.350 ± 0.014	0.604 ± 0.016
	wt	18.5 ± 1.7	1.59 ± 0.16	0.68 ± 0.17	1.23 ± 0.03
	Arg152Ala	25.4 ± 1.4	1.18 ± 0.06	2.73 ± 0.23	1.49 ± 0.07
	Arg152Glu	44.5 ± 3.6	**0.32** ± 0.05	**4.63** ± 1.75	**0.51** ± 0.16
	Arg152His	23.7 ± 2.2	1.25 ± 0.10	1.43 ± 0.39	0.92 ± 0.17
	Arg226Ala	19.7 ± 2.4	1.24 ± 0.14	1.61 ± 0.11	1.59 ± 0.04
	Arg226Glu	14.8 ± 1.4	1.16 ± 0.05	1.2 ± 0.15	1.37 ± 0.06
	Arg226His	19.9 ± 2.2	1.55 ± 0.09	0.91 ± 0.13	1.26 ± 0.06
	Arg152Ala/ Arg226Ala	46.3 ± 12.5	**0.50** ± 0.13	**6.32** ± 0.37	**0.426** ± 0.014

Figure II-1

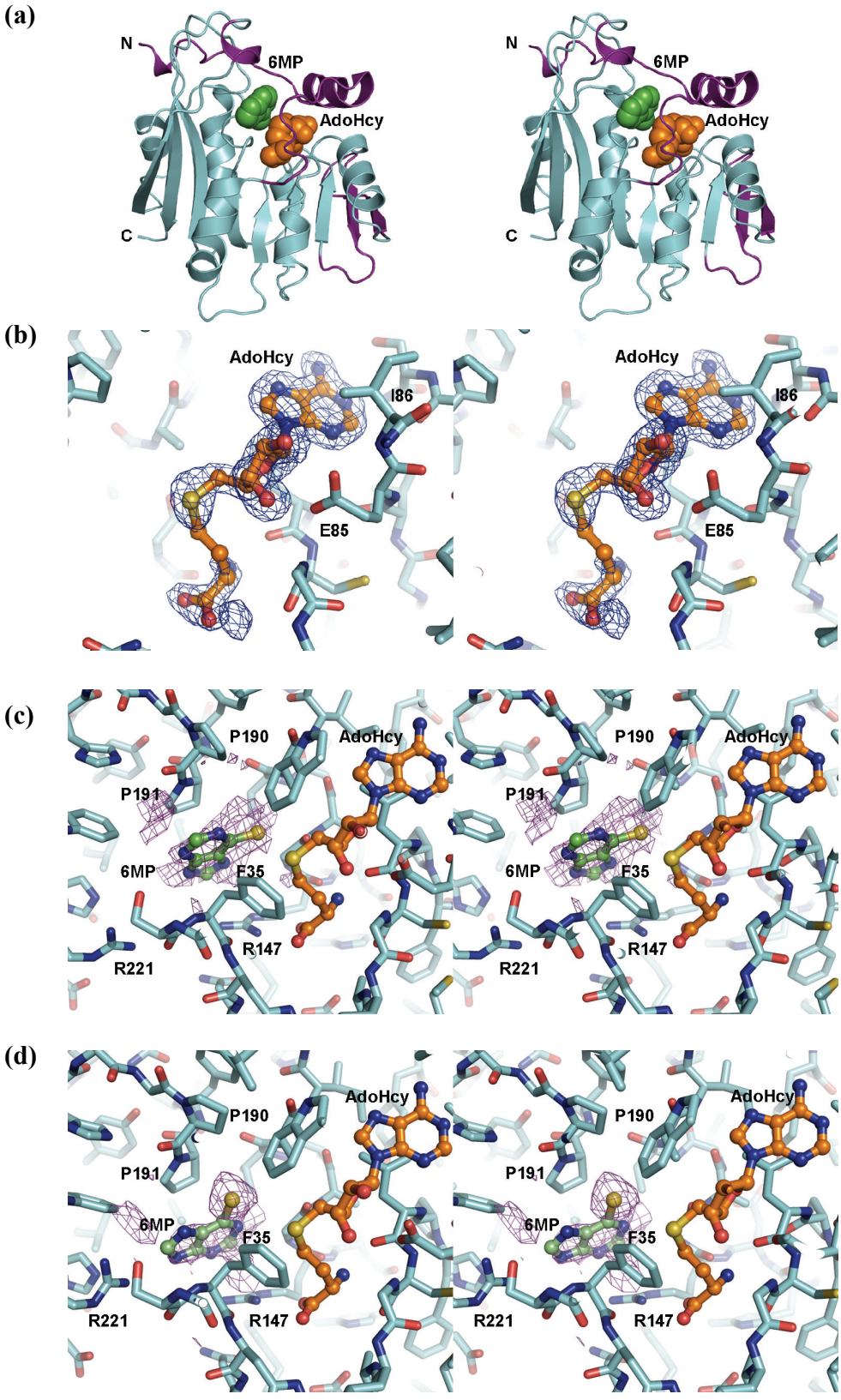

(a)
N
6MP
AdoHcy
C
N
6MP
AdoHcy
C
(b)
AdoHcy
I86
E85
AdoHcy
I86
E85
(c)
P190
AdoHcy
P191
6MP
F35
R147
R221
P190
AdoHcy
P191
6MP
F35
R147
R221
(d)
AdoHcy
P190
P191
6MP
F35
R221
R147
AdoHcy
P190
P191
6MP
F35
R221
R147

Figure II-1 Crystal structures of mTPMT.

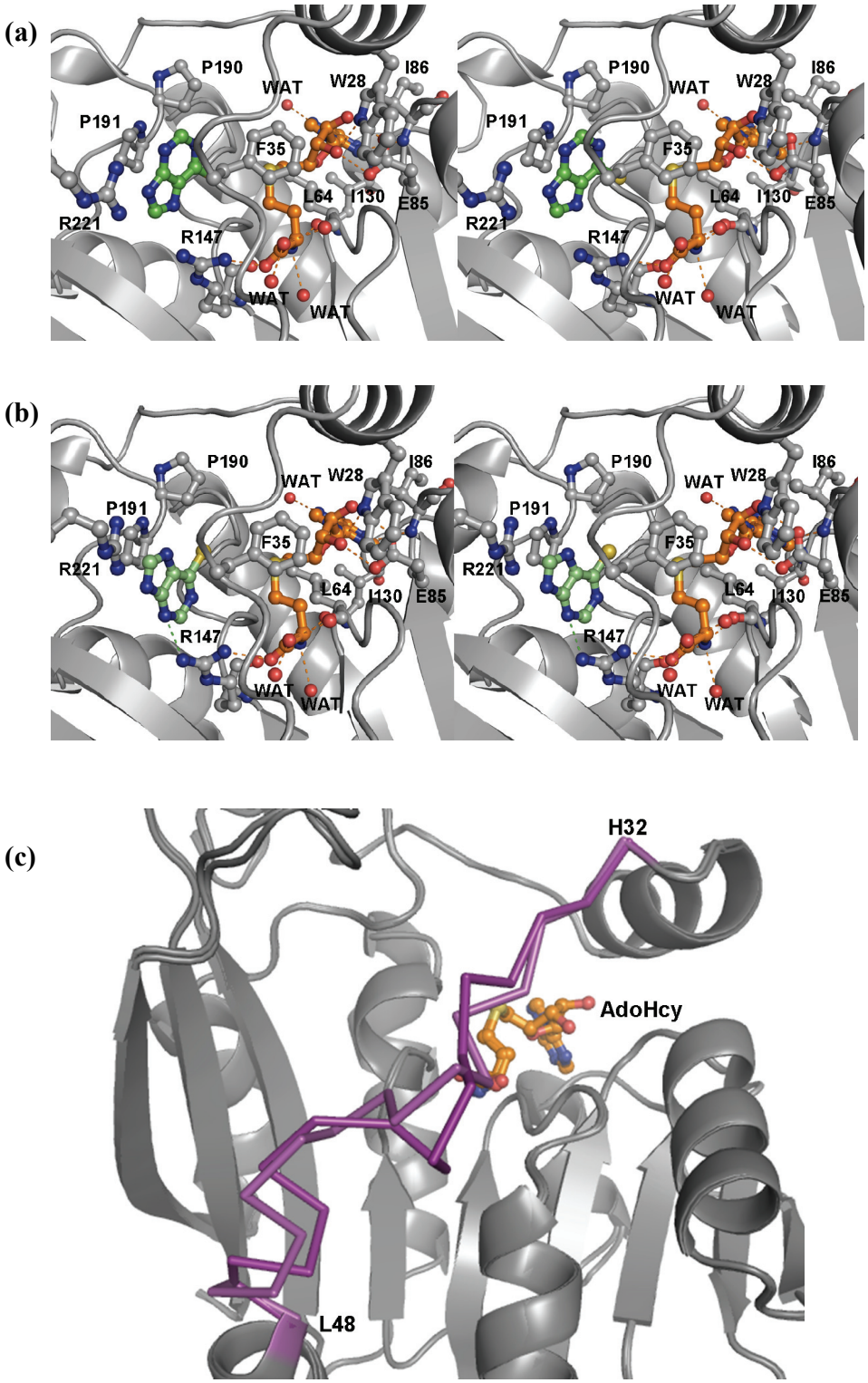
(a) Stereo ribbon diagram of mTPMT. The conserved class I MTase core is shown in cyan, and the inserted regions characteristic of the TPMT structure are purple. AdoHcy (orange) and 6MP (green) are shown as space-filling spheres.

(b) AdoHcy bound to the active site of molecule A in mTPMTwt-AdoHcy. Stereo view of 1.8 Å resolution omit $|F_o|-|F_c|$ density contoured at 3 σ is shown; similar density is present for molecule B, and for the two AdoHcy molecules in mTPMTwt-AdoHcy-6MP.

(c) 6MP bound to the active site of mTPMTwt-AdoHcy-6MP. Stereo view of 2.0 Å resolution omit $|F_o|-|F_c|$ density contoured at 1.5 σ is shown for the 6MP bound to molecule A.

(d) Similar omit density for 6MP bound to molecule B.

Figure II-2


(a)
P190
WAT
W28
I86
P191
F35
L64
I130
E85
R221
R147
WAT
WAT
(b)
P190
WAT
W28
I86
P191
F35
R221
L64
I130
E85
R147
WAT
WAT
(c)
H32
AdoHcy
L48

(d)

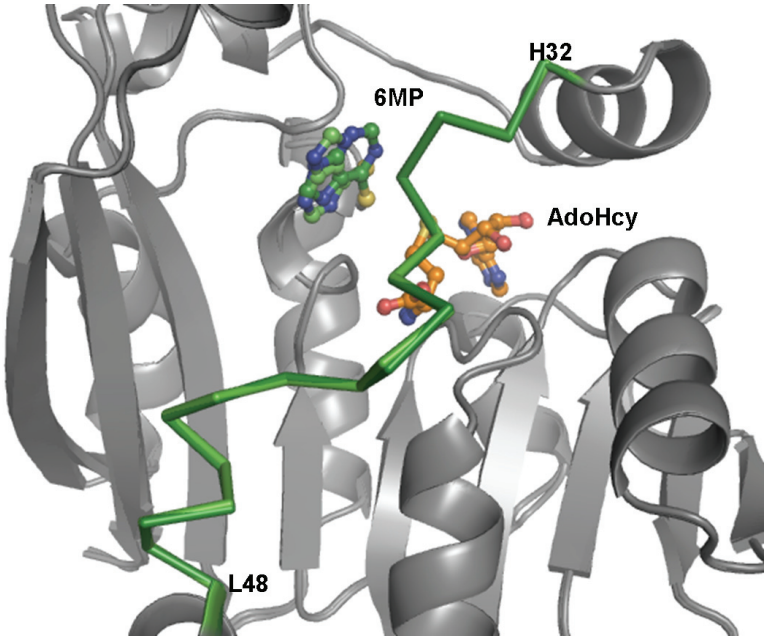


Figure II-2 Ligand binding to mTPMT.

(a) and (b) Stereo views of the active sites of molecules A and B in the mTPMTwt-AdoHcy-6MP structure. AdoHcy, 6MP, and side chains of nearby mTPMT residues are shown as ball-and-stick structures.

(c) Active site loop flexibility in mTPMTwt-AdoHcy. Superposition of molecules A and B shows that the active site loop (purple) is flexible and adopts two different conformations.

(d) Active site loop conformation in mTPMTwt-AdoHcy-6MP. When both AdoHcy and 6MP are bound, the superimposed active site loops (green) become better ordered and adopt the same conformation.

```
mTPMT    -----MSLDMKEHPDAEVQKNQVLTLEDWKEKWVTRHISFHQEQGHQLLKKHLDTFLKGQSGLRVFFPLCGKAIE 70
hTPMT    MDGTRTSLDIEEYSDTEVQKNQVLTLEEWQDKWVNGKTAFHQEQGHQLLKKHLDTFLKGKSGLRVFFPLCGKAVE 75
gTPMT    MDGTRTSLDIEEYSDTEVQKNQVLTLEEWQDKWVNGKTAFHQEQGHQLLKKHLDTFLKGKSGLRVFFPLCGKAVE 75
chTPMT   MDGTRTSLDIEEYSDTEVQKNQVLTLEEWQDKWVNGKTAFHQEQGHQLLKKHLDTFLKGKSGLRVFFPLCGKAVE 75
dTPMT    MDKTRTFLDVKEYPDTEVQKNRVLTLEEWQEKWVSRRIGFHQEQGHKLLKKHLDTFLKGENGLRVFFPLCGKAVE 75
cTPMT    MDDTSTLIDVKEYPDTEVQKNRVLTLEEWREKWVDGKIGFHQEQGHQLLKKHLDTFLKGENVLRVFFPLCGKAVE 75
psTPMT   ------------------------------------GSHQSEVNKDLQQYWSSLN-VVPGARVLVPLCGKSQD 36

mTPMT    MKWFADRGHTVVGVEISEIGIREFFAEQNLSYTEEPLAEIAGAKVFKSSSGSISLYCCSIFDLPRANIGKFDRIW 145
hTPMT    MKWFADRGHSVVGVEISELGIQEFFTEQNLSYSEEPITEIPGTKVFKSSSGNISLYCCSIFDLPRTNIGKFDMIW 150
gTPMT    MKWFADRGHSVVGVEISELGIQEFFTEQNLSYSEEPITEIPGTKVFKSSSGNISLYCCSIFDLPRTNIGKFDMIW 150
chTPMT   MKWFADRGHSVVGVEISELGIREFFTEQNLSYSEEPITEIPGTKVFKSSSGNISLYCCSIFDLPRTNIGKFDMIW 150
dTPMT    MKWFADRGHSVVGVEISELGIREFFAEQNLSYTEEPIVEIPGGKIFKSSSGNISLYCCSLFDLPRANIGKFDRIW 150
cTPMT    MKWFADRGHCVVGVEISELGIREFFTEQNLSYSEEPIMEIPGAKVFKSSSGNISLYCCNLFDLPRVNIGKFDRIW 150
psTPMT   MSWLSGQGYHVVGAELSEAAVERYFTER----GEQPHITSQGDFKVYAAPG-IEIWCGDFFALTARDIGHCAAFY 106

mTPMT    DRGALVAINPGDHDRYADIILSLLRKEFQYLVAVLSYDPTKHAGPPFYVPSAELKRLFGTKCSMQCLEEVDALEE 220
hTPMT    DRGALVAINPGDRKCYADTMFSLLGKKFQYLLCVLSYDPTKHPGPPFYVPHAEIERLFGKICNIRCLEKVDAFEE 225
gTPMT    DRGALVAINPGDRKCYADTMLSLLGKKFQYLLCVLSYDPTKHPGPPFYVPHAEIERLFGKICNIRCLEKVDAFEE 225
chTPMT   DRGALVAINPGDRKCYADTMLSLLGKKFQYLLCVLSYDPTKHPGPPFYVPHAEIERLFGKICNIRCLEKVDAFEE 225
dTPMT    DRGALVAINPGDRERYADIMLSLTRKGFHYLLAVLCYDPTKHAGPPFYVPEAEIKKLFGSICNIHCLEKVDVFEE 225
cTPMT    DRGALVAVNPGDRKCYTDIMLSLTRKGFRYLLAVLSYDPTKHPGPPFYVPDAEIKNLFGSTCNIHCLEKVDVFEE 225
psTPMT   DRAAMIALPADMRERYVQHLEALMPQACSGLLITLEYDQALLEGPPFSVPQTWLHRVMSGNWEVTKVGGQDTLHS 181

mTPMT    RHK--AWGLDYLFEKLYLLTEK 240
hTPMT    RHK--SWGIDCLFEKLYLLTEK 245
gTPMT    RHK--SWGIDCLFEKLYLLTEK 245
chTPMT   RHK--SWGIDCLFEKLYLLTEK 245
dTPMT    QHK--SWGIDYIIEKLYLFTEK 245
cTPMT    RHK--SWGIDYIVEKLYLLTEK 245
psTPMT   SARGLKAGLERMDEHVYVLERV 203
```

Figure II-3 Alignment of TPMT and related sequences. Sequences of six representative mammalian and one bacterial TPMTs are included (NCBI accession numbers): mouse (mTPMT, AAC25919), human (hTPMT, AAB27277), gorilla (gTPMT, AAX37643), chimpanzee (chTPMT, AAX37639), dog (dTPMT, AAL18006), cat (cTPMT, Q6EIC1), and Pseudomonas syringae (psTPMT, PDB 1pjz). Residues involved in AdoHcy binding (orange), 6MP binding (green), and hTPMT polymorphisms (blue) are highlighted.

Figure II-4

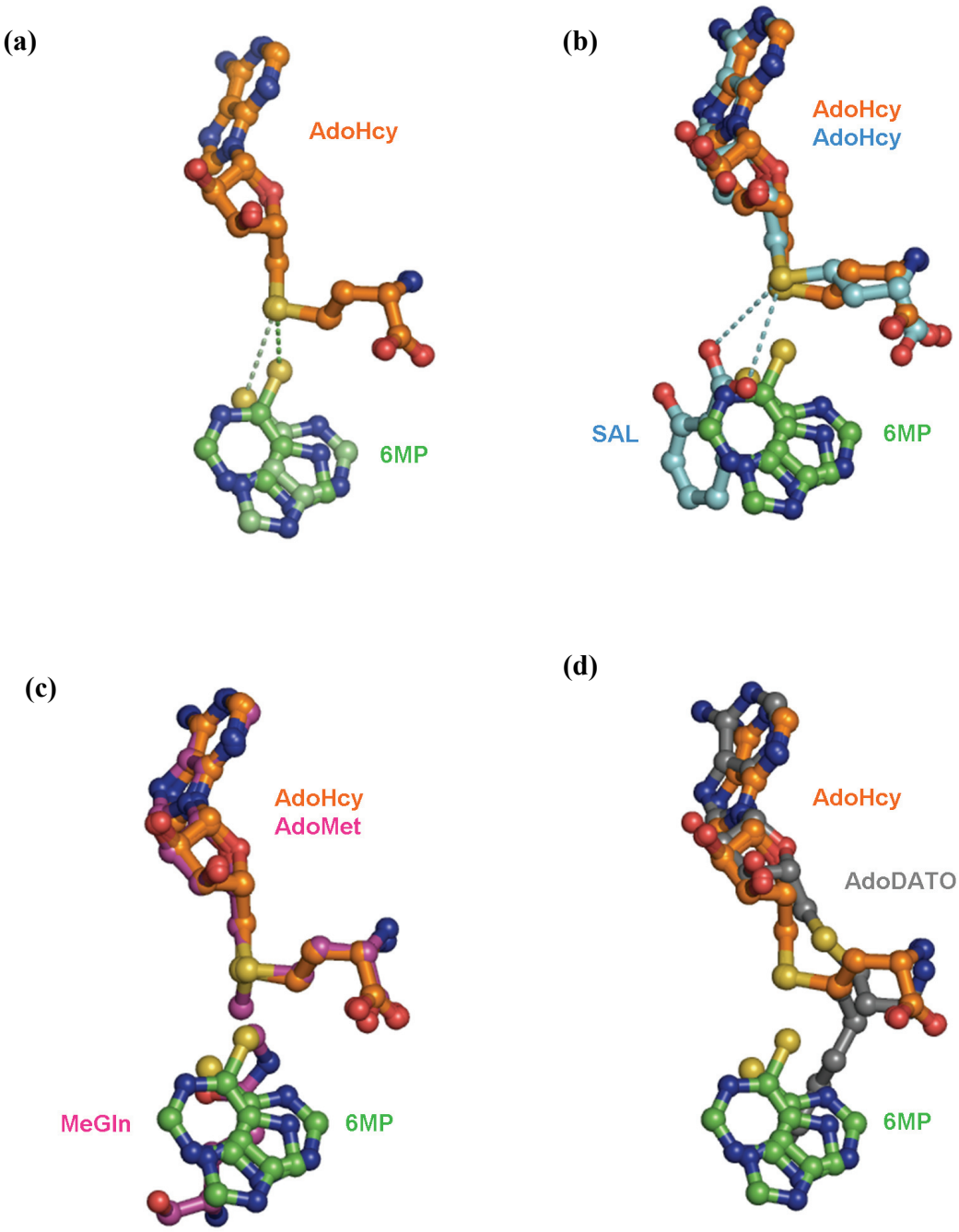

(a)
AdoHcy
6MP
(b)
AdoHcy
AdoHcy
SAL
6MP
(c)
AdoHcy
AdoMet
MeGln
6MP
(d)
AdoHcy
AdoDATO
6MP

Figure II-4 Donor and acceptor substrate binding to MTase fold enzymes.

(a) AdoHcy demethylated donor and 6MP acceptor bound to mTPMTwt-AdoHcy-6MP. Superposition of the AdoHcy bound to molecules A and B shows that the two 6MP are in the same position but different orientations. Dashed lines represent the possible path of methyl transfer between the AdoHcy and 6MP sulfur atoms.

(b) AdoHcy demethylated donor and salicylic acid acceptor bound to SAMT (Zubieta *et al.*, 2003). Superposition of the SAMT and mTPMT structures shows that salicylic acid and 6MP approach their respective AdoHcy from the same general direction. Dashed lines represent possible paths of methyl transfer between the salicylate carboxylate oxygen atoms and the AdoHcy sulfur in SAMT.

(c) AdoMet donor and N5-methylglutamine methylated acceptor bound to PrmC (Schubert *et al.*, 2003b). Both methylated PrmC ligands represent mixtures of (de)-methylated donor (AdoMet and AdoHcy) and acceptor (methylglutamine and glutamine) molecules. Superposition of the PrmC and mTPMT structures shows that N5-methylglutamine and 6MP approach their respective AdoMet/AdoHcy in similar fashion.

(d) The bisubstrate adduct inhibitor AdoDATO bound to PAPT (Korolev *et al.*, 2002). The adenosyl and aminopropyl groups at opposite ends of AdoDATA superimpose well with the adenosyl group of AdoHcy and 6MP bound to mTPMT. The central region of AdoDATA is in a different conformation from the homocysteine moiety of AdoHcy, probably reflecting the differences in the aminopropyl and methyl groups transferred and in the (de)carboxylated states of the donor molecules.

CHAPTER III

Crystal Structures of TPMT*5

3.1 Introduction

The wide variability in how individuals respond to drugs can lead to complications such as adverse drug reactions and differences in therapeutic effect. Characteristics of the patient, disease, and drug contribute to the many factors that influence the variation in drug response. The field of pharmacogenetics aims to study the role of inheritance, in particular DNA sequence variations in patients, in the heterogeneity of individual drug response (Weinshilboum and Wang, 2006; Eichelbaum *et al.*, 2006). Single-gene traits that affect drug metabolism were the first examples identified of human genetic polymorphisms that influence drug response. One of the best-studied examples of these is the thiopurine S-methyltransferase (TPMT).

TPMT is a cytosolic AdoMet-dependent MTase that metabolizes thiopurines such as 6-mercaptopurine (6MP), azathioprine and 6-thioguanine (6TG), which are used in treatments for childhood acute lymphoblastic leukemia, inflammatory bowel disease, autoimmune disorders, and organ transplant patients (Wang and Weinshilboum, 2006; Cheok and Evans, 2006). The thiopurines are prodrugs which are converted to thioguanine nucleotides and then exert the cytotoxicity by incorporation into DNA (Watters and McLeod, 2003). The inactivation of thiopurine prodrugs and some of their intermediate metabolites by TPMT methylation influences the concentration of cytotoxic

thioguanine nucleotides. Patients with variant TPMTs are thus at risk for thiopurine drug-induced toxicity.

Analysis of TPMT enzyme activity levels in human tissues revealed a large variation that followed a pattern of inheritance (Weinshilboum and Sladek, 1980). Upon cloning and characterization of the human TPMT cDNA and gene, the range of TPMT activities could be explained by genetic sequence variations (Honchel *et al.*, 1993; Szumlanski *et al.*, 1996, Tai *et al.*, 1996). To date, at least 28 TPMT polymorphisms have been reported, most of which are associated with decreased enzyme activity and/or thiopurine drug-induced toxicity (Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Lindqvist *et al.*, 2007; Garat *et al.*, 2008; Tamm *et al.*, 2008). While several polymorphisms influence transcription or mRNA splicing, 24 are nonsynonymous coding single nucleotide polymorphisms (cSNPs) which code for TPMT variants with varying degrees of reduced enzyme activity. Expression of 13 of the variants in COS-1 cells showed that nearly all had an enzyme activity which correlated with the amount of immunoreactive protein. The most functionally compromised variants, *3A, *3B, and *3C, are the result of some of the earliest identified and most commonly observed variant alleles. One variant, *5 (Leu49Ser), is notable due to the observation that it exhibited very low enzyme activity but only modestly decreased protein level (Salavaggione *et al.*, 2005). To investigate the possible structural basis for the diminished enzyme activity, we have determined the crystal structures of murine TPMT*5 (mTPMT*5), as binary complexes bound to the reaction product S-adenosylhomocysteine (AdoHcy) and as ternary complexes bound to both AdoHcy and 6MP. By comparison with the two wild-type (wt) TPMT complex

structures documented in Chapter II (Peng *et al.*, 2008), these two TPMT structures reveal conformational changes resulting from the leucine to serine substitution in the *5 variant, and differences between the wt and *5 structures which can explain the diminished enzyme activity of the *5 variant.

3.2 Materials and Methods

3.2.1 Expression, Purification, Functional Analysis, Crystallization and Data Collection

A construct for expression of the mTPMT*5 mutant (Leu44Ser) protein was prepared by the QuikChange Site-Directed Mutagenesis Kit (Stratagene). The mutant was expressed and purified as described for wild-type protein in Chapter II (Peng *et al.*, 2008). Enzyme activity was measured as described previously in Chapter II. All crystals were grown by sitting drop vapor diffusion at 20 °C, using a protein concentration of ~6 mg/ml. Crystals of mTPMT*5-AdoHcy were obtained from 25-27% (w/v) PEG3350, 0.2 M lithium sulfate/ammonium sulfate and 0.1 M imidazole (pH 5.5 or 6.0). Crystals of mTPMT*5-AdoHcy-6MP were grown from two conditions: (1) 25-27% (w/v) PEG335, 0.2M lithium sulfate/ammonium sulfate and 0.1M MES/bis-tris (pH 6.0), and (2) 20-22% PEG3350 and 0.15 malic acid (pH 5.5). Crystals were cryoprotected in artificial mother liquor containing 35% (w/v) PEG3350 and flash cooled in liquid nitrogen. All data were collected at APS beamlines 19ID. The data were processed with HKL (Otwinowski 1997) (**Table III-1**). Crystals belong to space group $P2_1$ and contain two molecules in each asymmetric unit.

3.2.2 Structure Determination and Refinement

The structures of mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP were determined by molecular replacement using mTPMTwt-AdoHcy as the search model. Interactive model building was carried out with COOT (Emsley and Cowtan, 2004) and refinement calculations were performed in CNS (Brunger *et al.*, 1998) and REFMAC (Murshudov *et al.*, 1997). The quality of all models was evaluated by PROCHECK (Laskowski *et al.*, 1993). The only outliers in the Ramachandran plot, His36 and Lys45 of molecule A in the mTPMT*5-AdoHcy-6MP structure, are in flexible regions with poor density. The electron density for the AdoHcy in both structures is clear and well-defined (**Figure III-1b**) while the electron density for 6MP is only detectable in molecule A of mTPMT*5-AdoHcy-6MP (**Figure III-1c**). The electron density for 6MP in molecule A of mTPMT*5-AdoHcy-6MP was defined enough to allow identification of the sulfur position and thus of the 6MP orientation. Therefore, 6MP is only modeled in molecule A of mTPMT*5-AdoHcy-6MP. A summary of the final refinement statistics is provided in **Table III-1**. Molecular figures were generated using PyMol (DeLano 2002). Coordinates and structure factors for mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP have been deposited in the Protein Data Bank with accession codes 3BKE and 3BKO, respectively.

3.2.3 Molecular Dynamics Simulations of mTPMT and hTPMT

The starting structure for the simulation of wt mTPMT was the crystal structure of mTPMT bound to AdoHcy and 6MP (PDB accession code 3BGD, residues 9-240, Chain A). The molecule A of the mTPMT*5-AdoHcy-6MP crystal structure is the starting

structure for the simulation of mTPMT*5 (PDB accession code 3BKO, residues 9-240, Chain A). In addition, a computationally modeled mTPMT*5 (Leu44Ser) structure was generated by substituting Leu with Ser in the wt mTPMT crystal structure. Since molecule B of mTPMTwt-AdoHcy-6MP is very similar to molecule A, only molecule A was used for the simulation. All 6MP and water molecules were removed before the simulations. The starting structure for the simulation of wt hTPMT was the crystal structure of hTPMT bound to AdoHcy (PDB accession code 2BZG, residues 17-245). AdoHcy and B3P were removed and selenomethionine residues (76, 148, and 170) were replaced with methionine. A computationally modeled TPMT variant (Tyr107Asp) structure was generated by substituting Tyr with Asp in the wt hTPMT crystal structure. Two water molecules which are involved the hydrogen network of Tyr107 were kept for the simulations.

All molecular dynamics simulations of wt and *5 mTPMTs, modeled mTPMT*5, wt hTPMT, and modeled hTPMT variant (Tyr107Asp) were performed using GROMACS3.3.3 (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; van der Spoel *et al.*, 2005) software package, and the GROMOS-87 (van Gunsteren and Berendsen, 1987) force field with modified interaction parameters between water-oxygens and carbon atoms (van Buuren *et al.*, 1993; Mark *et al.*, 1994) as well as 10 extra atom types (van Buuren *et al.*, 1993; Mark *et al.*, 1994; Jorgensen *et al.*, 1983; van Buuren and Berendsen, 1993; Liu *et al.*, 1995). The water model used in all the simulations is the SPC model (Berendsen *et al.*, 1981). Proteins were solvated in a cubic periodic box with walls extending at least 9 Å from the protein molecules. The energy

was minimized in 500 steps using a steepest descents energy minimization procedure. In order to soak the water into the protein without distorting the protein, 30ps of position-restrained molecular dynamics was calculated. Finally, the molecular dynamics simulation was run for 2ns with coordinates saved every 1ps for later analysis. A time step of 2fs was used. Temperature was coupled to a reference bath of 300K, with a coupling time constant of 0.1ps. Cα RMSF values were calculated relative to the average structure. The backbone of each structure from a trajectory is compared to that of the starting structure to generate RMSD values over the time frame.

3.3 Results

3.3.1 TPMT*5 Variant Has Reduced Substrate Affinity and Enzyme Activity

The expression studies suggested that most of the nonsynonymous cSNPs encode amino acid substitutions which decrease enzyme activity by compromising TPMT folding, stability or trafficking, with the exception of *5, in which the substitution alters specific enzyme activity (Salavaggione *et al.*, 2005). As part of a comprehensive functional genomic approach to TPMT pharmacogenetics, we have purified bacterially-expressed *5 variant of human and murine TPMTs and show that K_m of the variant increases ~10-fold for the 6MP methyl acceptor substrate and ~2-fold for the AdoMet methyl donor; V_{max} of the variant decreases about 3-10% that of the wt enzyme (**Table III-2**). TPMT has a much higher affinity for AdoMet (low micromolar) than for 6MP (low millimolar). The lower TPMT*5 enzyme activity compared to wt enzyme results from a decrease in substrate affinity, primarily for 6MP, and to a lesser extent for AdoMet.

3.3.2 Overall Structure of TPMT*5

Both mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP show that mTPMT*5 contains the expected class I AdoMet-dependent MTase core fold described previously (Martin and McMillan, 2002; Schubert *et al.*, 2003; Kozbial and Mushegian, 2005) (**Figure III-1a**). For each complex crystal structure, there are two copies of mTPMT*5 in the asymmetric unit. Superpositions of the Cα atoms for molecules in the same crystal and across crystals of mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP give overall rmsd of 0.7-0.8 Å and 0.6-0.9 Å, respectively. Unless otherwise stated, structural details provided in the following sections are for molecule A in the mTPMT*5-AdoHcy-6MP structure and are similar in the other complexes.

3.3.3 AdoHcy and 6MP Binding to TPMT*5

AdoHcy binding is conserved in wt and *5 variant TPMT. As seen in wt mTPMT structures, AdoHcy binds to mTPMT*5 in an extended conformation at the N-terminal half of the central β-sheet (**Figure III-1b, d**). The AdoHcy terminal amine group still forms hydrogen bonds with the carbonyl oxygen atoms of Leu64 and Arg147. The AdoHcy's carboxylate interacts with several water molecules. As in wt mTPMT structures, the AdoHcy's ribose hydroxyl groups interact with the side chains of Trp28 and Glu85 and the adenine's ring nitrogen atoms form hydrogen bonds with the main chain amines of Ile86 and Ile130. AdoHcy's adenine ring is stacked between the Ile86 and Leu64 side chains. AdoHcy binding to mTPMTwt and to mTPMT*5, when comparing the two molecules within each crystal and between crystal structures, is with the ligand in essentially the same position in identical conformations.

The mTPMT*5-AdoHcy-6MP structure shows that 6MP binds to wt and *5 variant TPMT in different modes. In the mTPMT*5-AdoHcy-6MP structure, the electron density of 6MP is only found in molecule A (**Figure III-1c**). Although 6MP is bound to mTPMT*5 in the same general area as it is found in mTPMTwt, the 6MP ligand approaches its donor substrate AdoMet from different direction, about 90° away (**Figure III-1d**) from the 6MP-AdoHcy approach when bound to mTPMTwt. The AdoHcy and 6MP are further apart from each other when bound to mTPMT*5 compared to wt mTPMT: the sulfur-sulfur distance of 4.9 Å in mTPMT*5-AdoHcy-6MP is significantly longer than the 3.4-4.1 Å observed in the wt complex structure.

3.3.4 A Flexible Active Site Loop in TPMT*5

The observation that TPMT*5 retains a high protein level makes it a variant uniquely amenable to structural studies. The mTPMT*5 crystal structures show that the variant binds AdoHcy in exactly the same manner as wt enzyme. In contrast, 6MP binding to the variant is different from that observed in the wt mTPMT structures, providing an explanation for the alterations in substrate affinity and enzyme activity. The cause of the difference in 6MP binding is apparent upon a detailed comparison of the Leu44 (Leu49 in hTPMT, wt) and Leu44Ser (*5) environments in the ensemble of mTPMT crystal structures. In the wt mTPMT structures, the Leu44 side chain is buried in a hydrophobic environment and the residue is located in the first turn of an α-helix (His41-Lys53) following an extended loop (Ser34-Gly40) which covers the active site (**Figure III-2a**). In the absence of 6MP, the active site loop (Ser34-His47) of wt mTPMT structures is flexible and becomes more ordered when 6MP is bound (Peng *et al.*, 2008, **Figure III-**

3a). In the mTPMT*5 structures, the hydrophilic Leu44Ser side chain is solvent-exposed, and there is a partial unwinding of the helix residues (**Figure III-2b**). As a result, the Ser34-His47 loop is significantly more flexible both in the absence and presence of 6MP (poor density, rmsd of 1.9-2.6 Å for superpositions of the Cα atoms for the active site loop in the same crystal and across crystals of mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP, **Figure III-3b**). Superpositions of the Cα atoms for molecules in the same crystal and across crystals of mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP give rmsd of 0.6-0.9 Å. When the active site loop region is omitted from the superpositions, the Cα rmsd drops to 0.2-0.5 Å. In the mTPMT*5 structures, the active site loop also have a higher average B factor (67-73 $Å^2$) than for the rest of molecule (~26 $Å^2$), which is consistent with the high flexibility of the active site loop. In the presence of 6MP, the active site loop of mTPMT*5 also is more flexible than that of wt mTPMT (**Figure III-3c**). The unwinding of the helical residues His41-His47 and increase in Ser34-His47 loop flexibility provide an explanation for the decreased affinity for 6MP, and to a lesser degree for AdoMet. The increase in active site flexibility extends to Phe35, whose side chain in mTPMTwt-AdoHcy-6MP packs against 6MP (Peng *et al.*, 2008) but in mTPMT*5-AdoHcy-6MP is directed away from the substrate (**Figure III-1d)**. Similarly, the Arg147 side chain in mTPMTwt-AdoHcy-6MP interacts with AdoHcy, is directed toward 6MP, and was confirmed by mutagenesis to play a role in 6MP binding (Peng *et al.*, 2008), but in mTPMT*5-AdoHcy-6MP it is directed away from both AdoHcy and 6MP (**Figure III-1d**).

In addition, molecular dynamics simulations were performed to investigate the flexibility of the active site loop. The rmsd over the time frame of the simulations relative to the starting structures (**Figure III-4a**) and the Cα root-mean-square fluctuations (RMSF, **Figure III-4b**) are shown. Simulations using either the wt mTPMT or mTPMT*5 crystal structures as starting points were stable, as evidenced by the plateau of rmsd values. In contrast, the simulation using the computationally modeled mTPMT*5 structure, in which the leucine to serine substitution was made in the wt mTPMT crystal structure, was not stable and generated rmsd values which continued to increase throughout the course of the simulation. This suggests that the substitution of Leu44 with Ser44 is incompatible with, and destabilizes, the wt mTPMT crystal structure conformation. The RMSF value from the wt mTPMT simulation is in a similar pattern to the B-factors of crystal structure. The RMSF values from the mTPMT*5 and modeled mTPMT*5 simulations show that most variations occurred in the loop regions with the central β-sheet remaining stable. The N-terminus (residues 9-18) had the largest RMSF values for both simulations, and the active site loop was also quite mobile. Large fluctuations were also seen in several additional regions for mTPMT*5 simulation (residues 26-32, 37-41, 53-58) and only seen in one region for the modeled mTPMT*5 simulation (residue 20-35). These observations suggest that the substitution of Leu44 with Ser44 has an effect on the local structure of mTPMT, especially for α helix (residues 20-33) and part of the active site loop. Together, the observations indicate that the loss of the Leu44 hydrophobic interactions leads to active site conformational differences in mTPMT*5 that can affect both AdoMet and 6MP binding and result in loss of enzyme activity.

3.3.5 Destabilization of hTPMT Variant (Tyr107Asp)

Recently, a new TPMT variant (Tyr107Asp) was identified in an organ transplant patient in Thailand, who was having toxicity issues with 6MP therapy. When expressed in a mammalian cell line, this variant protein displayed low levels of both protein quantity and enzyme activity. In addition, the variant was found to undergo accelerated degradation when expressed in rabbit reticulocyte lysate. Finally, treatment with the proteasome inhibitor (MG132) increased the amount of detectable Tyr107Asp protein. Together, these observations suggested that the decreased Tyr107Asp variant enzyme activity may be due to the accelerated degradation of misfolded proteins (personal communication, Dr. Weinshilboum). Structural analysis of hTPMT structures (PDB accession code 2BZG, Wu *et al.*, 2007) showed that Tyr107 is in a water-mediated surface hydrogen-bonding network and it is not obvious that Tyr107Asp would destabilize native structure (**Figure III-5**). Molecular dynamics simulation calculations were performed to investigate the molecular basis for low levels of both protein quantity and enzyme activity of TPMT variant (Tyr107Asp). The rmsd over the time frame of the simulations relative to the starting structures (**Figure III-6a**) and the Cα root-mean-square fluctuations (RMSF, **Figure III-6b**) are shown. Compared to the simulation using the wt hTPMT crystal structure as starting points, the simulation using the modeled hTPMT Tyr107Asp structure, in which the tyrosine to aspartic acid substitution was made in the wt hTPMT crystal structure, generated rmsd values which continued to increase throughout the course of the simulation. This suggests that the substitution of Tyr107 with Asp107 is incompatible with, and destabilizes, the wt hTPMT crystal structure conformation. Comparison of the RMSF value from wt hTPMT and the

modeled hTPMT Tyr107Asp simulations shows that most variations occurred in the loop regions. However, large fluctuations relative to wt hTPMT were seen in several regions for hTPMT Tyr107Asp simulation (residues 26-40, 52-62, 89-98, 100-104, and 137-141). These observations suggest that the substitution of Tyr107 with Asp107 has an effect on the native structure of hTPMT, especially for several α helices (residues 25-36, 46-57, 93-102, and 137-141). Together, these calculations suggest that the tyrosine residue mutation to aspartic acid residue destabilizes the native structure of wt hTPMT.

3.4 Discussion

Of the 28 hTPMT nonsynonymous cSNPs, 23 alter residues conserved in mTPMT and the remaining four alter amino acids with similar chemistry except TPMT*20(b) whose activity phenotype is unknown (Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Sasaki *et al.*, 2006; Lindqvist *et al.*, 2007; Garat *et al.*, 2008; **Table III-3**). There is no hotspot for the substituted residues, which are distributed throughout the TPMT protein fold (**Figure III-7**). By using the wt mTPMT and hTPMT crystal structures (Peng *et al.*, 2008; Wu *et al.*, 2007) as starting scaffolds, computational modeling of the amino acid alterations leads to the observation that the protein level and enzyme activity of the hTPMT polymorphisms appear to correlate with whether the introduced residue is compatible with the native protein structure (**Table III-3**). The significantly decreased levels of enzyme activity for TPMT*2, *3A-C, *5, *6, *10-13, *17, *18 and *23 (Salavaggione *et al.*, 2005; Lindqvist *et al.*, 2007) can be explained by the incompatibility of their associated amino acid substitutions with their local environment and thus the native wt protein fold. The relatively modest effect on TPMT activity observed for

TPMT*7-9, *16, *19-25 (Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Garat *et al.*, 2008) may be explained by the observation that the associated substitutions are with hydrophilic amino acids in solvent-exposed positions which are more easily structurally accommodated.

Of particular interest are the hTPMT*3A-C variants, which were the first characterized and are the most prevalent after the wt enzyme (Szumlanski *et al.*, 1996; Tai *et al.*, 1996), and hTPMT*5, which exhibits low enzyme activity despite a near wild-type protein level (Salavaggione *et al.*, 2005). Analysis of the wt TPMT crystal structures reveals that the Ala154Thr (Ala149Thr in mTPMT) substitution associated with hTPMT*3A and *3B replaces Ala154, a small residue buried in a hydrophobic environment near the active site, with the larger Thr154 which introduces sterically unfavorable short clashes that cannot be accommodated without altering at least the local protein conformation. The Tyr240Cys (mTPMT Tyr235Cys) substitution associated with hTPMT*3A and *3C replaces Tyr240, which is a buried residue far from the active site whose side chain forms a hydrogen bond with His201 (mTPMT Ser196). Replacement with the smaller Cys240 removes the hydrogen bonding interaction and introduces a void in the structure, both structural consequences which may compromise native protein folding and/or structural stability. It is interesting to note that when overexpressed in bacterial systems, hTPMT*3B is inactive even when a portion of the protein remains monomeric while hTPMT*3C retains a small amount of enzyme activity with the presence of some monomeric protein (Wang *et al.*, 2005). This difference in behavior may be attributed to the location of the hTPMT*3B (Ala154Thr) substitution near the enzyme's active site.

Bacterially expressed hTPMT*3A variant, which contains both the *3B (Ala154Thr) and *3C (Tyr240Cys) substitutions, is entirely aggregated and exhibits no enzyme activity (Wang *et al.*, 2005), suggesting that the structural consequences of the two amino acid replacements are additive. In addition, the molecular dynamics simulations of hTPMT*3A-C variants (Rutherford and Daggett, 2008) also led to similar speculations on the destabilizing effects of hTPMT*3A-C on wt TPMT protein structure.

The mTPMT*5 crystal structures reported here reveal conformational details that explain the compromised substrate binding and enzyme activity for the TPMT*5 variant. Both wt mTPMT and wt hTPMT have higher affinity for AdoMet (low micromolar) than for 6MP (low millimolar) (Peng *et al.*, 2008). The 6MP binding becomes even weaker in mTPMT*5-AdoHcy-6MP, due to the increase in flexibility of the active site loop. The side chain of Phe35 in mTPMT*5-AdoHcy-6MP is directed away from the substrate (**Figure III-1d),** which destabilizes the hydrophobic clamp formed by Phe35 and Pro191 to position 6MP in wt mTPMT structures. In the mTPMT*5-AdoHcy-6MP structure, the observation that 6MP is bound further away from AdoHcy may also help explain the low enzyme activity of the variant enzyme (**Figure III-1d**). The increased flexibility of the active site loop also weakens AdoMet binding; in the mTPMT*5 structures this involves the Arg147 swinging away from AdoHcy rather than forming a hydrogen bond with the methyl donor's carboxylate as it does in the wt mTPMT structure. The resulting decrease in AdoMet binding affinity is a modest 2-fold (**Table III-2**), since the many other hydrogen bonds and additional interactions appear to be enough to keep AdoMet bound to mTPMT*5 in exactly the same conformation and position as to wt mTPMT. These

active site structural differences between wt mTPMT and mTPMT*5 provide an explanation for the decreased substrate binding affinities that contribute to the low V_{max} observed for the TPMT*5 which is ~3-10% that of wt TPMT (**Table III-2**). Comparison of the local conformations for the Leu44 in wt mTPMT and the Ser44 in mTPMT*5 structures (**Figure III-2**) reveals that the replacement of the buried Leu44 side chain in a hydrophobic environment with the hydrophilic Ser44 which prefers to be solvent-exposed results in unwinding of a helical turn that in turn leads to the increased flexibility of the active site loop, thus affecting the substrate binding affinities and enzyme activity (**Figure III-3b, c**). The molecular dynamics simulations also suggest that the substitution of Leu44 with Ser44 in mTPMT*5 has a destabilizing effect on the wt mTPMT structure and increases the flexibility of the N-terminus, especially an α helix (residues 20-33) and part of the active site loop (**Figure III-4**).

More recently, a new TPMT variant (Tyr107Asp) was identified. When expressed in cell culture or cell-free systems, low levels of both protein quantity and enzyme activity were observed, presumably as a result of accelerated degradation of misfolded proteins (personal communication, Dr. Weinshilboum). In the wt hTPMT crystal structure, Tyr107 is in a water-mediated surface hydrogen-bonding network and it is not obvious that Asp107 would destabilize the native structure (**Figure III-5**). Thus molecular dynamics simulation calculations were very useful since they indicated that the substitution of Tyr107 with Asp107 has a destabilizing effect on the wt hTPMT structure and increases the flexibility of the N-terminus and several α helices (residues 25-36, 46-57, 93-102, and 137-141) (**Figure III-6**).

The structural implications of TPMT polymorphisms discussed here help us understand the molecular mechanism of TPMT polymorphisms. Furthermore, the mTPMT*5 crystal structures together with molecular dynamics simulation calculations also explains the exception of TPMT*5 variant, which exhibits low enzyme activity with a near wild-type protein level (Salavaggione *et al.*, 2005). The replacement of the hydrophobic Leu44 with the hydrophilic Ser44 leads to the more flexible active site loop in TPMT*5, which affects the AdoMet and 6MP substrate affinity and results in loss of the enzyme activity. Molecular dynamics simulation calculations also confirmed that the substitution of Tyr107 with Asp107 has a destabilizing effect on the wt hTPMT structure.

Table III-1 Data collection and refinement statistics for mTPMT*5

	*5-AdoHcy	*5-AdoHcy-6MP
Data collection		
cell dimensions		
a, *b*, *c* (Å)	62.84 65.57 72.70	62.92 65.74 72.74
(deg)	115.76	115.68
wavelength (Å)	0.97912	0.97912
resolution (Å)	30-1.85 (1.92-1.85)	30-1.73 (1.79-1.73)
R_{sym}	4.1 (22.4)	3.7 (20.4)
I / I	20.8 (3.3)	23.2 (3.7)
completeness (%)	97.4 (89.3)	91.6 (77.1)
redundancy	2.5 (2.3)	2.5 (1.9)
Refinement		
resolution (Å)	30-1.85 (1.92-1.85)	30-1.73 (1.79-1.73)
no. of reflections	42152	48653
R_{work} / R_{free}	18.06(22.34)	20.36(24.43)
no. of atoms		
protein, ligand, water	3756, 52, 461	3756, 62, 347
average *B*-factors		
protein, AdoHcy, 6MP, water	29.1, 17.6 , -, 35.0	28.8, 19.3, 43.2, 32.5
rmsd*		
bond lengths (Å)	0.014	0.013
bond angles (deg)	1.462	1.468
Ramachandran		
most favored (%)	90.6	89.4
additional allowed (%)	9.4	10.1
disallowed (%)	0.0	0.5

Numbers in parentheses refer to the highest resolution shell.
*rmsd, root mean square deviation.

Table III-2 Enzyme activity of wild-type and variant TPMT

			AdoMet		6MP	
			K_m (μM)	V_{max} (units/mg protein)	K_m (mM)	V_{max} (units/mg protein)
mTPMT	wt	--	5.6 ± 0.5	0.96 ± 0.05	0.350 ± 0.014	0.604 ± 0.016
	*5	L44S	8.8±0.98	0.031±0.002	4.98±2.02	0.025±0.008
hTPMT	wt	--	18.5 ± 1.7	1.59 ± 0.16	0.68 ± 0.17	1.23 ± 0.03
	*5	L49S	33.6±0.8	0.178±0.006	6.79±0.96	0.11±0.01

*The data for wild-type mTPMT and hTPMT are adapted from Peng *et al.*, 2008. The data for *5 are prepared by Qiping Feng from Dr. Weinshilboum's lab at Mayo Clinic. One unit of TPMT activity represented the formation of 1 nmol of 6-methyl-mercaptopurine per second of incubation at 37 °C. The kinetic experiments were carried out in triplicate and standard deviations were calculated for the average values of K_m and V_{max} determined.

Table III-3 TPMT polymorphisms structural speculation based on mTPMT structures

TPMT polymorphism	Amino acid change in hTPMT	Amino acid in mTPMT	hTPMT activity phenotype[b]	Structural speculation[c]
*1(WT)	---	---	high	-------
*2	Ala80Pro	Ala75	low	destabilize helix
*3A	Ala154Thr /Tyr240Cys	Ala149 Tyr235	low	*3B and *3C
*3B	Ala154Thr	Ala149	low	alter local structure of active site and destabilize Arg152 necessary for deprotonation
*3C	Tyr240Cys	Tyr235	low	remove stabilizing hydrogen bonds to destabilize the C-terminal β hairpin
*3D	Glu98STOP/ Ala154Thr/Tyr240Cys		low	
*5[a]	Leu49Ser	Leu44	low	destroy local hydrophobic interaction to destabilize helix to flexible loop and active site
*6	Tyr180Phe	Tyr175	low	remove hydrogen bond with Asp151 to destabilize active site
*7	His227Gln	His222	high	solvent-exposed, easily accommodated
***8**	**Arg215His**	**Gln210**	**intermediate**	solvent-exposed, easily accommodated
*9	Lys119Thr	Lys114	intermediate	solvent-exposed, easily accommodated
*10	Gly144Arg	Gly139	low	solvent-exposed, the larger Arg mutation can be repulsed by Arg64 to destabilize the local structure
*11	Cys132Tyr	Cys127	low	the larger Tyr side chain can affect AdoHcy/AdoMet binding site in order to accommodate with local conformation adjustment
*12	Ser125Leu	Ser120	low	solvent-exposed, exposed Leu hydrophobic surface can destabilize the protein
***13**	**Glu28Val**	**Asp23**	**low**	solvent-exposed and at the helix, exposed Val hydrophobic surface can destabilize the helix
*14	Met1Val		low	not present in the current structure
*15	Loss of 25	Loss of 25		

	residues 140-165	residues 135-160	low	
***16**	**Arg163His**	**His158**	**intermediate**	partly solvent-exposed, easily accommodated
*17	Gln42Glu	Gln37	intermediate	solvent-exposed, easily accommodated; acidic side chain of Glu can attract basic side chain of Arg152 to affect the catalysis
*18	Gly71Arg	Gly66	intermediate	destabilize AdoHcy/AdoMet binding site
*19	Lys122Thr	Lys117	high	solvent-exposed, easily accommodated
*20(a)	Lys238Glu	Lys233	intermediate	solvent-exposed, easily accommodated
*21	Leu69Val	Leu64	intermediate	destabilize AdoHcy/AdoMet binding site
***22**	**Arg163Pro**	**His158**	**intermediate**	solvent-exposed, and at the helix
***20(b)**	**Gly36Ser**	**Arg31**	**None**	solvent-exposed
*23	Ala167Gly	Ala162	low	solvent-exposed, and at the helix, Gly makes the helix flexible
*24	Gln179His	Gln174	high	solvent-exposed, easily accommodated
*25	Cys212Arg	Cys207	intermediate	destroy local hydrophobic interaction

TPMT*8, 13, 16, 22, and 20(b) are not conserved between human and mouse and are in bold.

[a] From observations in the mTPMT*5 crystal structures.

[b] References for hTPMT activity phenotype: Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Sasaki *et al.*, 2006; Lindqvist *et al.*, 2007; Garat *et al.*, 2008.

[c] hTPMT residue numbers are used.

Figure III-1

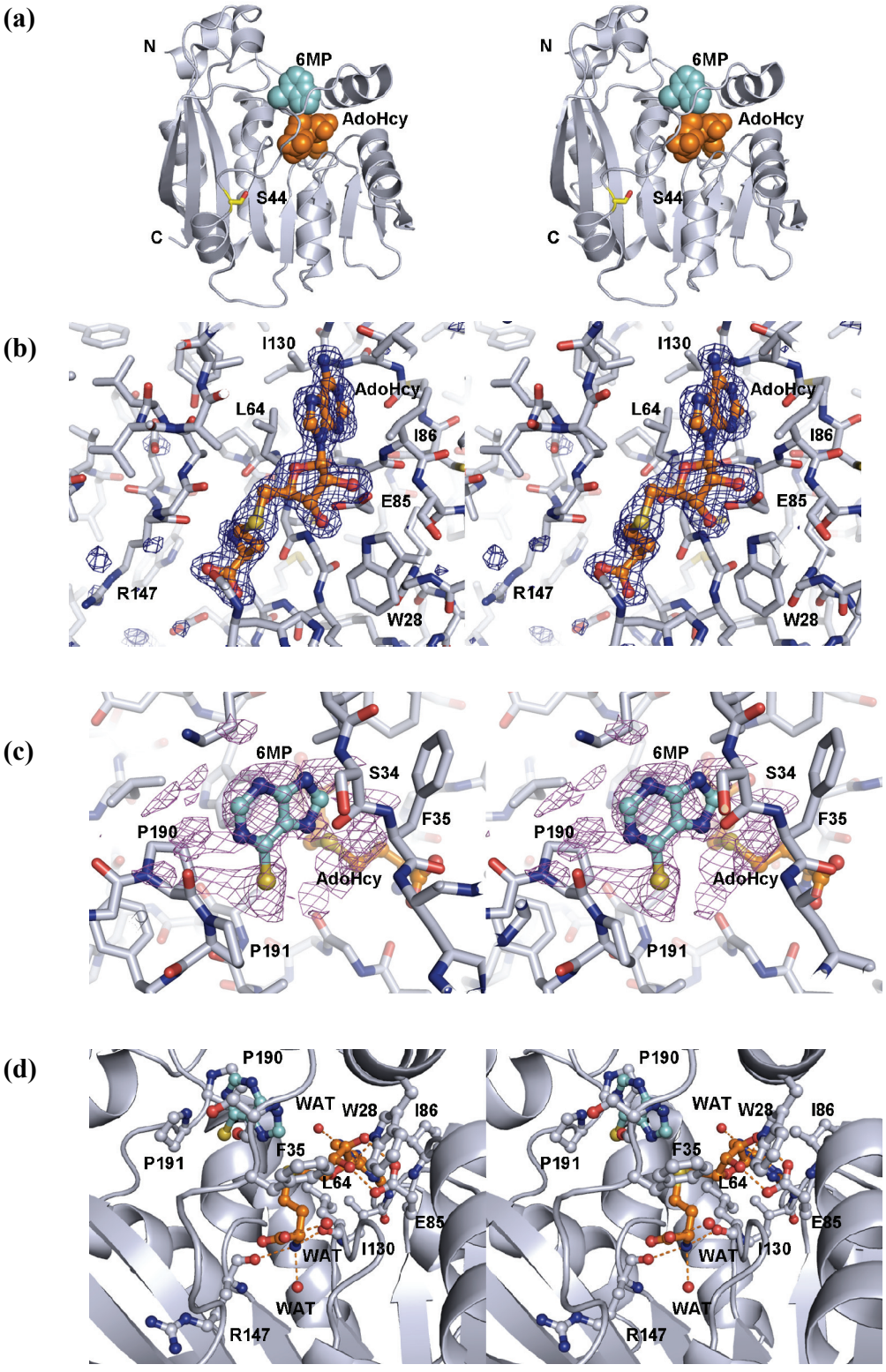

(a)
N
6MP
AdoHcy
S44
C
N
6MP
AdoHcy
S44
C
(b)
I130
AdoHcy
L64
I86
E85
R147
W28
I130
AdoHcy
L64
I86
E85
R147
W28
(c)
6MP
S34
F35
P190
AdoHcy
P191
6MP
S34
F35
P190
AdoHcy
P191
(d)
P190
WAT
W28
I86
P191
F35
L64
E85
WAT
I130
WAT
R147
P190
WAT
W28
I86
P191
F35
L64
E85
WAT
I130
WAT
R147

Figure III-1 Crystal structures of mTPMT*5.

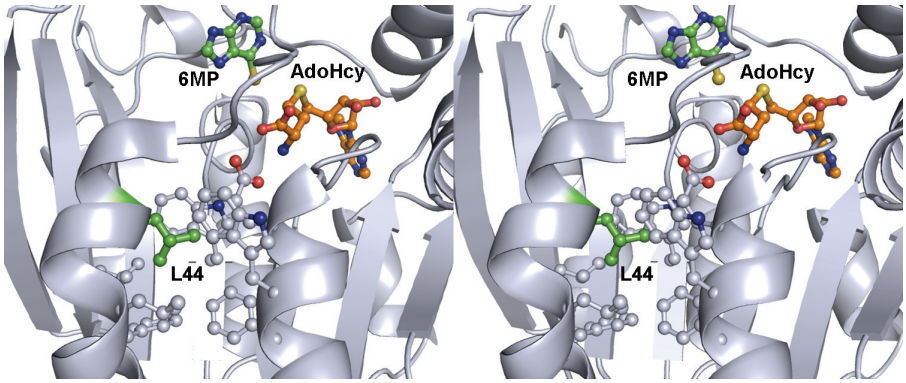
(a) Stereo ribbon diagram of mTPMT*5. AdoHcy (orange) and 6MP (cyan) are shown as space filling spheres. The side chain of Ser44 (yellow) is shown as a ball-and-stick structure.

(b) AdoHcy bound to the active site of molecule A in mTPMT*5-AdoHcy. Stereoview of 1.85 Å resolution omit $|F_o|$ - $|F_c|$ density contoured at 3σ is shown; similar density is present for molecule B and for two AdoHcy molecules in mTPMT*5-AdoHcy-6MP. The nearby residues are labeled.

(c) 6MP bound to the active site of molecule A in mTPMT*5-AdoHcy-6MP. Stereoview of 1.73 Å resolution omit $|F_o|$ - $|F_c|$ density contoured at 1σ is shown. The nearby residues are labeled.

(d) Ligand binding to mTPMT*5. Stereoview of the active site of molecule A in the mTPMT*5-AdoHcy -6MP structure is shown. AdoHcy (orange), 6MP (cyan) and side chains of nearby mTPMT*5 residues (light gray) are shown as ball-and-stick structures.

(a)



(b)

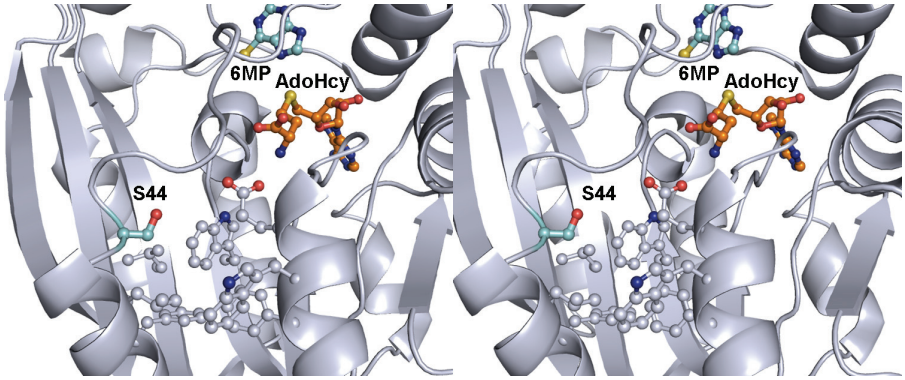


Figure III-2 The replacement of the hydrophobic Leu44 of wt mTPMT with the hydrophilic Ser44 of mTPMT*5.

(a) Leu44 located in the first turn of an α-helix (His41-Lys53) in wt mTPMT. The side chain of Leu44 is buried in a hydrophobic environment. The mTPMTwt-AdoHcy-6MP (PDB accession code 3BGD) is used for the figure preparation. AdoHcy (orange), 6MP (green), the side chain of Leu44 (green) and its nearby hydrophobic environment (light gray) are shown as ball-and-stick structures.

(b) Ser44 in mTPMT*5. The solvent-exposed side chain of Ser44 results from a partial unwinding of the helix residues. AdoHcy (orange), 6MP (cyan), the side chain of Ser44 (cyan) and its nearby environment (light gray) are shown as ball-and-stick structures in molecule A of mTPMT*5-AdoHcy-6MP.

Figure III-3

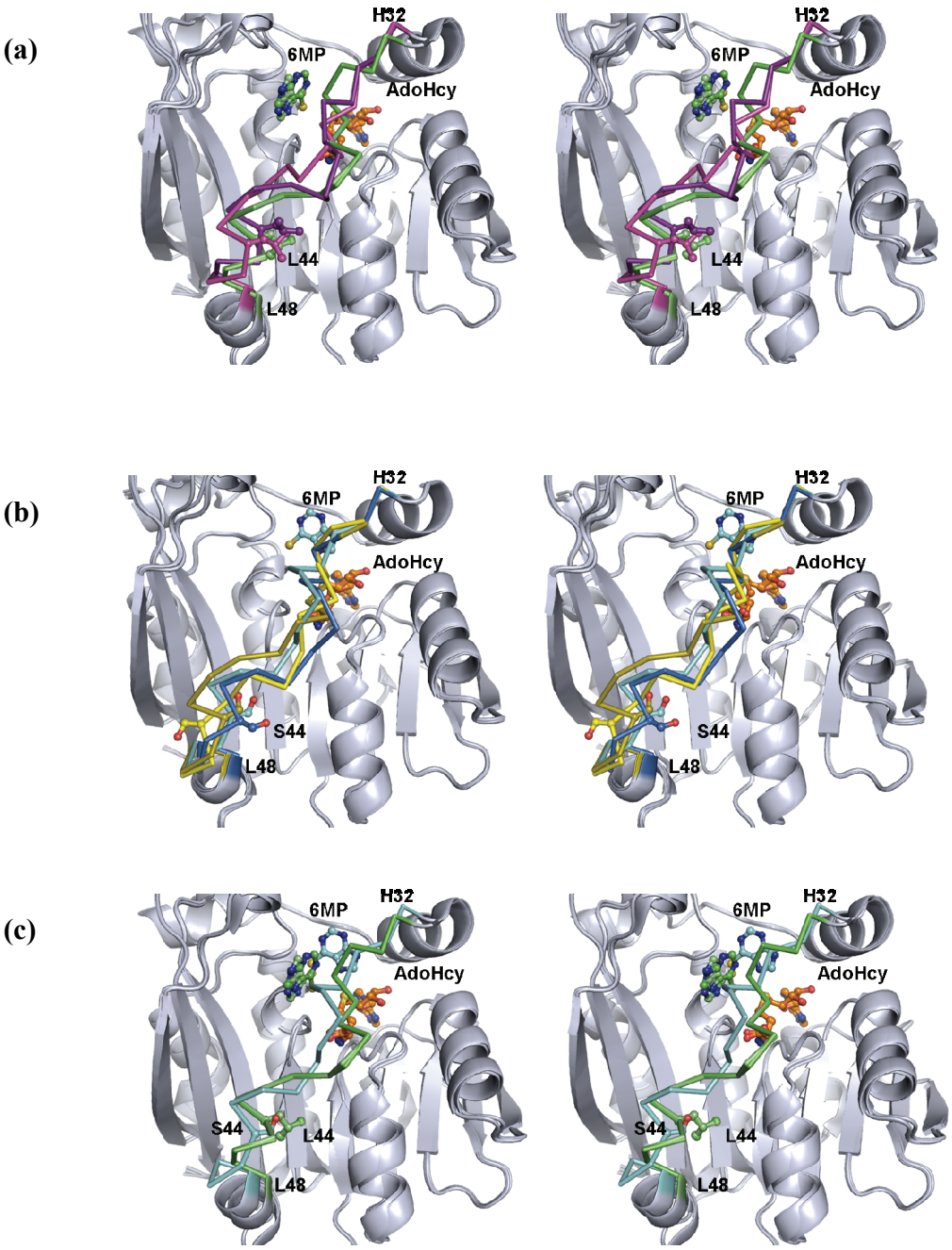

(a)
H32
6MP
AdoHcy
L44
L48
H32
6MP
AdoHcy
L44
L48
(b)
H32
6MP
AdoHcy
S44
L48
H32
6MP
AdoHcy
S44
L48
(c)
6MP
H32
AdoHcy
S44
L44
L48
6MP
H32
AdoHcy
S44
L44
L48

Figure III-3 mTPMT*5 active site loop flexibility in the crystal structures.

(a) Active site loop flexibility in wt mTPMT. Superpositions of mTPMTwt-AdoHcy (purple, PDB accession code 3BGI) and mTPMTwt-AdoHcy-6MP (green, PDB accession code 3BGD) show that the active site loop of wt mTPMT is flexible in the absence of 6MP and becomes better ordered when 6MP is bound. AdoHcy (orange), 6MP (green), and the side chains of Leu44 (purple and green) are shown as ball-and-stick structures.

(b) Active site loop flexibility in mTPMT*5. Superpositions of mTPMT*5-AdoHcy (yellow) and mTPMT*5-AdoHcy-6MP (molecule A, cyan; molecule B, blue) show that the active site loop of mTPMT*5 has different conformations. The Ser44 side chain is solvent-exposed due to the unwinding of a helical turn. AdoHcy (orange), 6MP (cyan), and the side chains of Ser44 are shown as ball-and-stick structures.

(c) Active site loop flexibility in the presence of 6MP. Superpositions of mTPMTwt-AdoHcy-6MP (green, PDB accession code 3BGD) and mTPMT*5-AdoHcy-6MP (molecule A, cyan) show that the active site loop of mTPMT*5 (cyan) remains flexible with both AdoHcy (organge) and 6MP (cyan) bound while those of wt mTPMT (green) become more ordered upon the binding of 6MP (green). The side chains of Leu44 in wt mTPMT and Ser44 in mTPMT*5 are shown as ball-and-stick structures.

Figure III-4

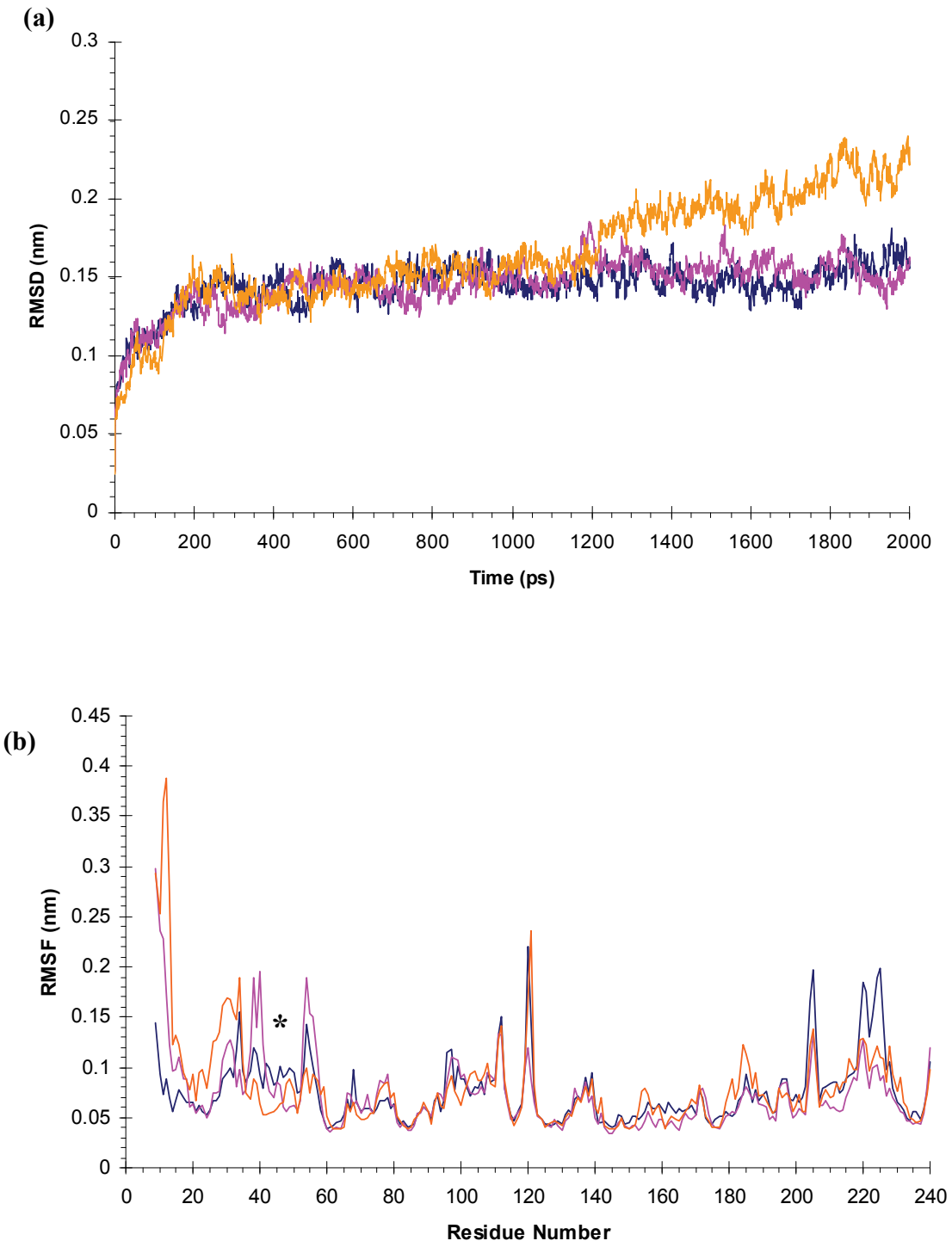

(a)
RMSD (nm)
0.3
0.25
0.2
0.15
0.1
0.05
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time (ps)
(b)
RMSF (nm)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
*
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue Number

Figure III-4 Flexibility of mTPMT*5 in molecular dynamics simulation calculations.

(a) Instantaneous backbone rmsd of trajectory structures from the starting structure. The backbone rmsd of residues 9-240 in wt mTPMT (blue), mTPMT*5 (purple) and modeled mTPMT*5 (orange) are overlaid.

(b) Atomic fluctuation of mTPMT*5. Cα RMSF in 200-2000ps trajectory is shown. Cα RMSF values for the wt mTPMT, mTPMT*5 and modeled mTPMT*5 are colored in blue, purple and orange, respectively. (*) Leu44Ser.

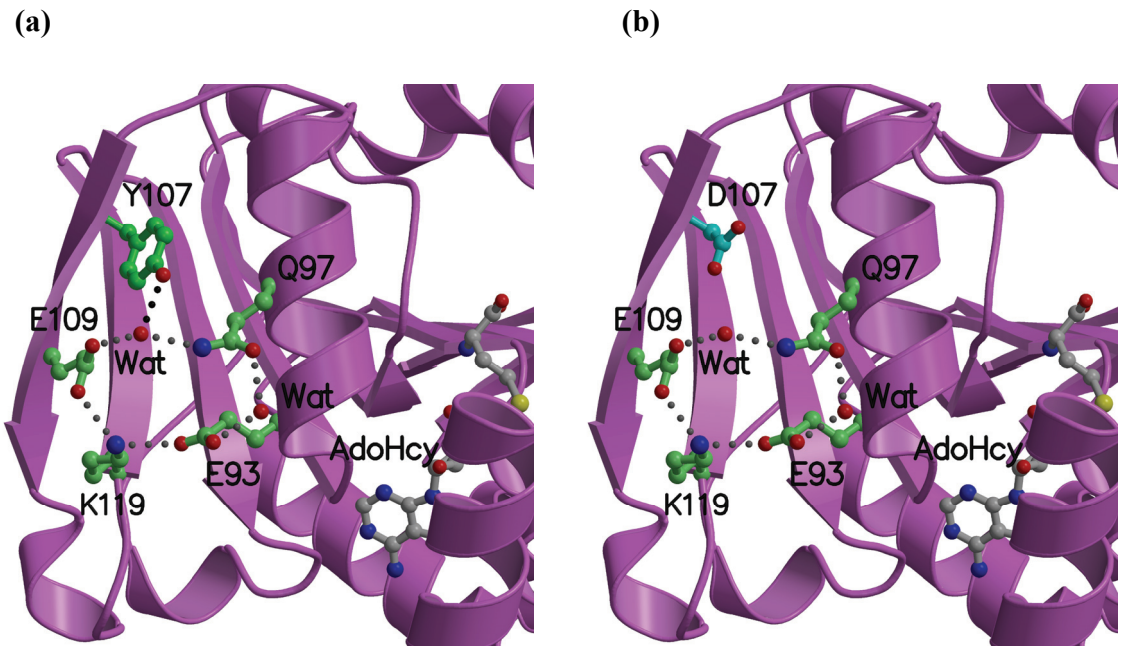


Figure III-5 The model of hTPMT variant (Tyr107Asp).

(a) Tyr107 is in a water-mediated surface hydrogen-bonding network. AdoHcy (gray) and the side chains of Tyr107 (bright green) and nearby residues (pale green) are shown as ball-and-stick structures. Water molecules are shown as isolated red spheres.

(b) Asp107 in the model of hTPMT variant (Tyr107Asp). AdoHcy (gray) and the side chains of Asp107 (light blue) and nearby residues (pale green) are shown as ball-and-stick structures. Water molecules are shown as isolated red spheres.

Figure III-6

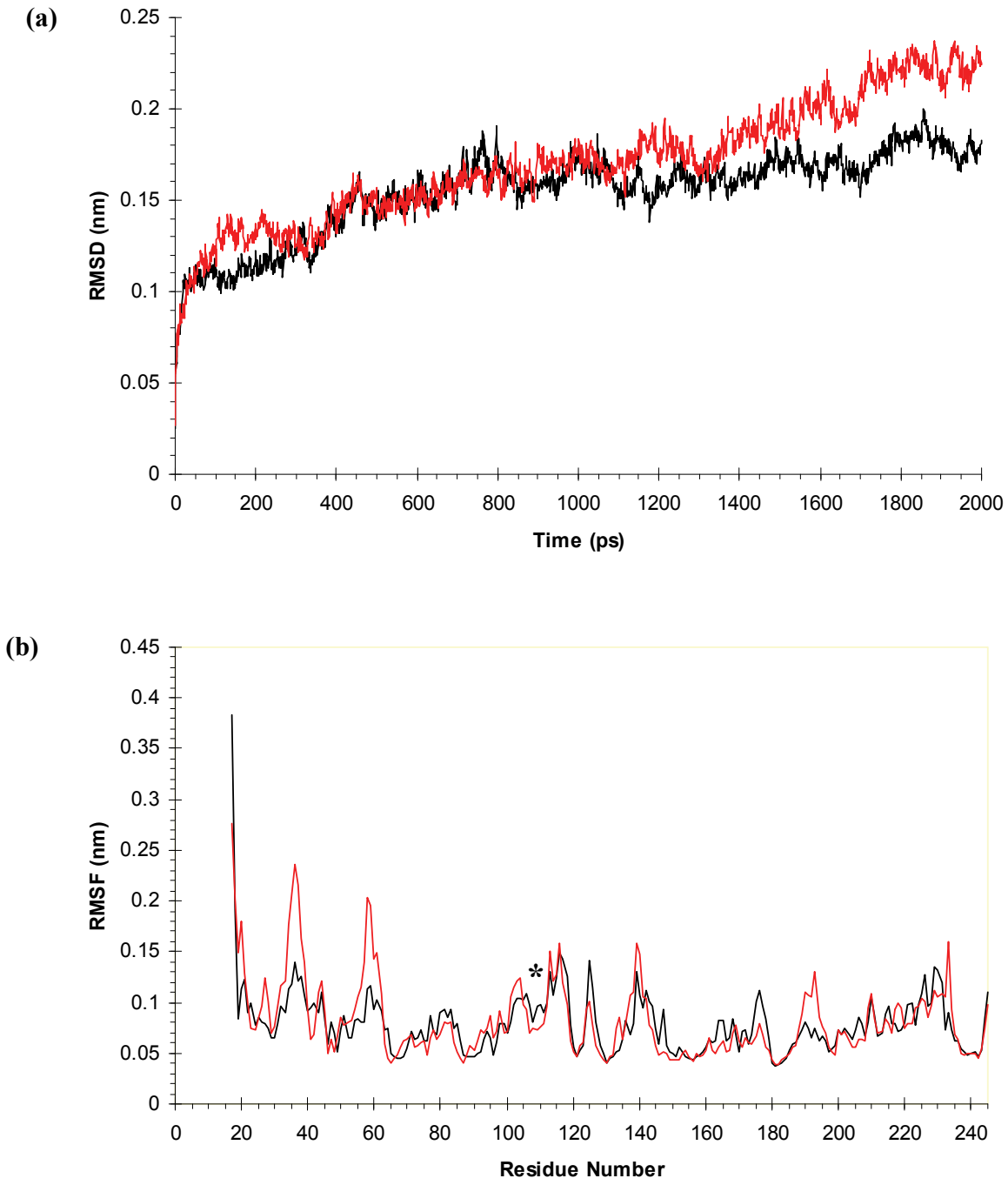

(a)
RMSD (nm)
0.25
0.2
0.15
0.1
0.05
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time (ps)
(b)
RMSF (nm)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
*
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue Number

Figure III-6 Destabilization of hTPMT variant (Tyr107Asp).

(a) Instantaneous backbone rmsd of trajectory structures from the starting structure. The backbone rmsd of residues 17-245 in wt hTPMT (black) and modeled hTPMT variant (Tyr107Asp, red) are overlaid.

(b) Atomic fluctuation of hTPMT variant (Tyr107Asp). Cα RMSF in 200-2000ps trajectory is shown. Cα RMSF values for the wt hTPMT and modeled hTPMT variant (Tyr107Asp) are colored in black and red, respectively. (*) Tyr107Asp.

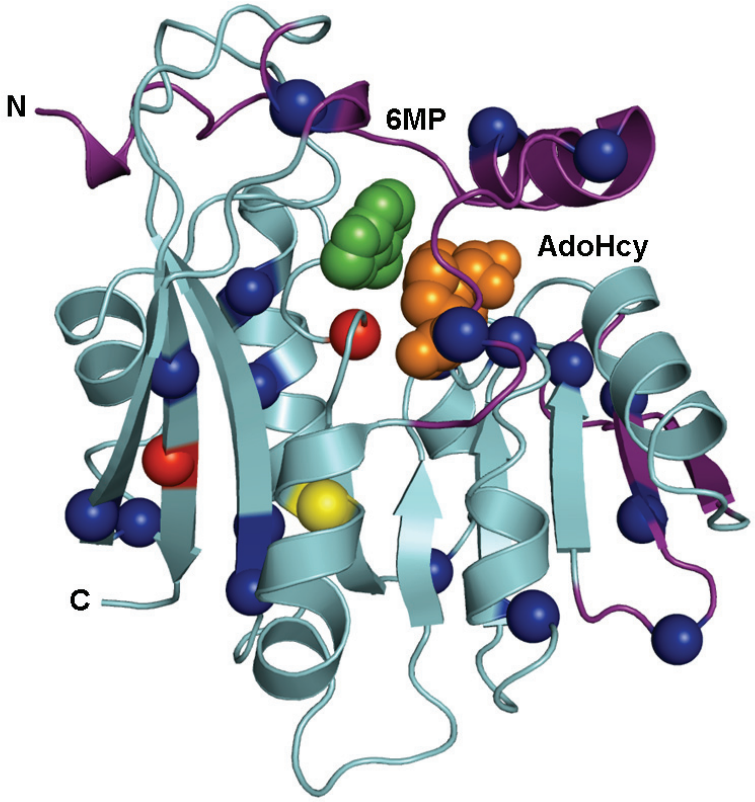


Figure III-7 Structural implication of TPMT polymorphisms. Distribution of residues altered in hTPMT variants, in the homologous mTPMTwt-AdoHcy-6MP structure (PDB accession code 3BGD; Peng *et al.*, 2008). The Cα atoms of amino acids altered in the most common variants *3A, *3B, and *3C are red, that substituted in *5 is yellow, and other variant residues are in blue. AdoHcy (organge) and 6MP (green) are shown as space filling spheres.

CHAPTER IV

Crystal Structures of TPMT Binding to Benzoic Acid Inhibitors and Thiophenol Substrates

4.1 Introduction

Thiopurine S-methyltransferase (TPMT) catalyzes the S-methylation of thiopurine and thiopyrimidine, an important metabolic pathway for thiopurine drugs. Thiopurines are prodrugs used to treat childhood acute lymphoblastic leukemia, inflammatory bowel disease, and transplant rejection (Coulthard and Hogarth, 2005). They exert their cytotoxic effects upon their metabolism to thioguanine nucleotides, which can be incorporated into DNA. TPMT genetic polymorphisms were discovered to control the level of TPMT activity in humans and therefore correlated with the interindividual variation in the metabolism, therapeutic efficacy and toxicity of thiopurine drugs (Watters and McLeod, 2003). Although TPMT has been well studied for its role in thiopurine metabolism, its natural substrate is still unknown. Discovery of potential inhibitors and new substrates for TPMT is important for experimental and clinical implications. Toward this end, benzoic acid compounds were identified as potential inhibitors of TPMT, and several nonheterocyclic aromatic thiols were identified as the substrates of TPMT (Woodson *et al.*, 1983; Ames *et al.*, 1986). Interestingly, K_m values of thiophenol derivatives are two to three orders of magnitude less than that of 6-mercaptopurine (6MP) (Woodson *et al.*, 1983; Ames *et al.*, 1986). In the crystal structure of the murine TPMT (mTPMT) ternary complex with S-adenosylhomocysteine (AdoHcy) and 6MP (see in Chapter II, Peng *et al.*, 2008), few hydrophilic residues were found near the bound 6MP.

This is consistent with the observation that a heterocyclic aromatic ring is not necessary for TPMT substrate recognition. To further aid discovery of small molecule modulators for optimization of thiopurine-based therapy, murine TPMT (mTPMT) was crystallized with several benzoic acid inhibitors, 3,4,5,-triiodobenzoic acid, 4-iodobenzoic acid, and 4-tert-butylbenzoic acid, and the substrate thiophenol. The resulting crystals extend our understanding of TPMT active site ligand recognition.

4.2 Materials and Methods

4.2.1 Crystallization and Data Collection

N-terminal His_6-tagged *wt* mTPMT was expressed and purified as described previously (see Chapter II, Peng *et al.*, 2008). Protein was concentrated to ~6 mg/ml and mixed with AdoHcy and inhibitors or thiophenol (final concentration 5-10mM). The crystals were grown from the conditions described in Chapter II. The crystals of mTPMTwt-AdoHcy were also soaked with these compounds at different final concentrations (1mM or 5mM) for varying time periods (4hrs or overnight). Native diffraction datasets up to 1.76 Å resolution have been measured at NSLS beamline X29 and APS beamline 19ID and processed with HKL (Otwinowski and Minor, 1997) (**Table IV-1**). All crystals belong to space group $P2_1$ and contain two molecules in each asymmetric unit.

4.2.2 Structure Determination and Refinement

The structures of mTPMT complexes with its inhibitors and thiophenol were determined by molecular replacement, built by COOT (Emsley and Cowtan, 2004), and refined by REFMAC (Murshudov *et al.*, 1997). The initial coordinates for the inhibitors and

thiophenol were generated using the PRODRG server (Schuettelkopf and van Aalten, 2004). The quality of all models was assessed by PROCHECK (Laskowski *et al.*, 1993). The majority of the residues are located in the most favored region of the Ramachandran plot. A summary of the final refinement statistics is provided in **Table IV-1**. Clear electron density for 3,4,5,-triiodobenzoic acid (**Figure IV-1a**) was observed for the crystal soaked overnight with 5mM of 3,4,5,-triiodobenzoic acid. Crystals of mTPMT were grown in the presence of 4-iodobenzoic acid, 4-ter-butylbenzoic acid and thiophenol. After structure determination, no density for 4-tert-butylbenzoic acid was observed. In contrast, electron density was present for 4-iodobenzoic acid (**Figure IV-1b**) and thiophenol (**Figure IV-1c**). Molecular figures were generated using PyMol (DeLano, 2004).

4.3 Results

The crystal structures of mTPMT bound to benzoic acid inhibitors, 3,4,5,-triiodobenzoic acid (mTPMTwt-AdoHcy-TIB) and 4-iodobenzoic acid (mTPMTwt-AdoHcy-4IB), and thiophenol substrate (mTPMTwt-AdoHcy-TP) were solved at a resolution of 1.76, 2.2 and 1.8Å, respectively. The crystal structures are isomorphous with the structures of mTPMTwt-AdoHcy-6MP discussed in Chapter II (Peng *et al.*, 2008). Superpositions of the Cα atoms for these structures with mTPMTwt-AdoHcy-6MP give overall rmsd values of 0.2-0.5Å. The secondary structure and most of the tertiary structure of mTPMT in the structures of these complexes as well as the binding mode and interactions formed for AdoHcy reported here are consistent with those described previously. The most

significant and interesting features of these structures relate to the binding of inhibitor/substrate to the active site, and are the focus of the following discussion.

4.3.1 Thiophenol Binding to TPMT

Other than thiopurine drugs, TPMT was also discovered to catalyze the methylation of nonheterocyclic aromatic thiols such as thiophenol derivatives (Woodson *et al.*, 1983). Thiophenol is a more effective TPMT acceptor substrate than 6MP. In the study of human TPMT enzyme, the *Km* value for thiophenol is ~0.36 μM (Woodson *et al.*, 1983) while *Km* for 6MP is ~0.68mM (Peng *et al.*, 2008). In the mTPMTwt-AdoHcy-TP structure, the electron density for the thiophenol substrate is strong enough to define the location of the thiol (**Figure IV-1c**). In addition, the electron density for the thiophenol substrate in the mTPMTwt-AdoHcy-TP structure is stronger than that for the 6MP substrate in the mTPMTwt-AdoHcy-6MP structure (Peng *et al.*, 2008), which is consistent with the observation that the 6MP binding affinity is weaker than that of thiophenol by 3 orders of magnitude.

The mTPMTwt-AdoHcy-TP structure reveals that thiophenol binds to the two independent molecules of the crystal, in the same position and orientation in the active site (**Figure IV-2a**). As seen in mTPMTwt-AdoHcy-6MP, the intersulfur distances of 4.0 and 3.8 Å for the two molecules are about double the 1.85 Å length of a sulfur-carbon single bond. Like 6MP, thiophenol has no direct hydrogen-bonding interactions with the enzyme. It is bound in the active site by essentially the same van der Waals interactions with 6MP described in Chapter II (**Figure IV-1c**). The hydrophobic clamp formed by the

side chains of Pro191 and Phe35 holds the thiophenol in the same position with 6MP (**Figure IV-2b**). Arg147 and Arg221 are the only hydrophilic residues within 4 Å of thiophenol. Therefore, Arg147 and Arg221 may be involved in the deprotonation of thiophenol, which is speculated to be a necessary step in a S_N2-like reaction mechanism of AdoMet-dependent MTases. Arg147 is about 3.2 and 3.4 Å from thiophenol in the two molecules of mTPMTwt-AdoHcy-TP while Arg221 is 5.5 and 3.3 Å from thiophenol in the two molecules. However, mutagenesis showed that neither Arg147 nor Arg221 is essential for 6MP methylation, indicating that neither is required for 6MP deprotonation (Peng *et al.*, 2008). Thus it is likely that neither arginine is involved in thiophenol deprotonation as well.

4.3.2 Benzoic Acid Inhibitors Binding to TPMT

4-tert-butylbenzoic acid has a bulky tert-butyl substituent in the 4-position. The addition of a tert-butyl is unfavorable for binding to TPMT and thus the pI_{50} value is much lower than 4-iodobenzoic acid, which has instead an iodine atom as the 4-substitutent (**Table IV-2**). The crystallization and crystal structure of mTPMTwt-AdoHcy-4IB supports this interpretation. Diffraction data measured from mTPMT co-crystallized with, or soaked in, 4-tert-butylbenzoic acid generated electron density maps which showed no density for the inhibitor bound in the active site. In contrast, co-crystallization with 4-iodobenzoic acid yielded diffraction data and electron density for bound 4-iodobenzoic acid (**Figure IV-1b**).

3,4,5,-triiodobenzoic acid has an iodine at the same position as in 4-iodobenzoic acid, along with two additional iodine atoms substituted *meta* to the carboxylate group. The additional iodo substituents are favorable for binding to TPMT, as evidenced by a pI_{50} value for 3,4,5,-triiodobenzoic acid that is substantially higher than that for 4-iodobenzoic acid (**Table IV-2).** The electron densities for 3,4,5,-triiodobenzoic acid in both mTPMTwt-AdoHcy-TIB molecules are well-defined and stronger than those for 4-iodobenzoic acid in the two mTPMTwt-AdoHcy-4IB molecules (**Figure IV-1a, b**). Superposition of the Cα atoms for two mTPMTwt-AdoHcy-TIB molecules gives overall rmsd of ~0.2 Å while that for the two mTPMTwt-AdoHcy-4IB molecules displays overall rmsd of ~0.4 Å. The binding mode and location of 3,4,5,-triiodobenzoic acid and 4-iodobenzoic acid are similar to those observed for 6MP and thiophenol. The 3,4,5,-triiodobenzoic acid and 4-iodobenzoic acid are also oriented by the hydrophobic clamp formed by the side chains of Pro191 and Phe35 (**Figure IV-1**). The significant difference is that the carboxylate of two benzoic acid inhibitors can form hydrogen bonds or salt bridge with the protein (**Figure IV-3**). 3,4,5,-triiodobenzoic acid is essentially identical in both mTPMTwt-AdoHcy-TIB molecules. The carboxylate of 3,4,5,-triiodobenzoic acid forms a salt bridge with Arg221 and hydrogen bonds with Ser34 and His222, while the carboxylate of 4-iodobenzoic acid only forms a salt bridge with Arg221 (**Figure IV-3**). Comparisons of all mTPMTwt-AdoHcy-TIB and mTPMTwt-AdoHcy-4IB molecules (**Figure IV-4**) reveal that Arg221, located in a flexible short helix (residues Glu220-Trp225, see Chapter II), is flexible and can adopt different conformations when it interacts with different inhibitors. The Cα atoms of Arg221 shift 0.4-0.9 Å between the different complexes. The side chains of Ser34 and His222 in the mTPMTwt-AdoHcy-4IB

complex also shift away from Arg221 when compared to their positions in the mTPMTwt-AdoHcy-TIB complex. The 4-iodobenzoic acid inhibitor bound in mTPMTwt-AdoHcy-4IB also is shifted from the bound position of 3,4,5,-triiodobenzoic acid, but such that their 4-iodo substituents are located in essentially identical positions. In effect, the 4-iodo atom acts as a pivot about which the inhibitors are rotated into their optimal binding orientation. Furthermore, the 4-iodo atoms of both 3,4,5,-triiodobenzoic acid and 4-iodobenzoic acid in these crystal structures form van der Waals interactions with Gly148 and Ala152. However, 3- and 5-iodo atoms of 3,4,5,-triiodobenzoic acid form a number of additional van der Waals interactions with Gly148, Val151 and Pro190, which optimize the binding orientation of 3,4,5,-triiodobenzoic acid compared to that 4-iodobenzoic acid. Together, these observations indicate that the 3- and 5-iodo substituents in 3,4,5,-triiodobenzoic acid influence the orientation of the inhibitor in order to optimize interactions between the inhibitor's carboxylate group and TPMT. Moreover, although the side chains of Arg147 don't have obvious conformational change in these complex structures, the shifts of benzoic acid ring orient the inhibitor about 4 Å from Arg147.

4.4 Discussion

Our results show that the methyl acceptor binding site in TPMT has some inherent plasticity. It is able to accommodate small substrates such as thiophenol, as well as larger substituted benzoic acid inhibitors. The plasticity for the binding of benzoic acid inhibitors is mainly due to the flexibility of the side chains of a few residues on the active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix (Peng *et al.*, 2008).

The crystal structure of the mTPMTwt-AdoHcy-TP complex shows thiophenol bound in the same position and orientation in the active sites of the two independent molecules. This is in contrast to 6MP, which was observed bound in the same position but different orientations in the two molecules of the mTPMTwt-AdoHcy-6MP crystal structure (**Figure IV-2**) (Peng *et al.*, 2008). Also, the electron density for thiophenol is stronger and better defined (**Figure IV-1c**) than that for 6MP. These observations are consistent with the higher affinity of thiophenol for TPMT, as evidenced by its lower *Km*. (Woodson *et al.*, 1983). The intersulfur distances of 4.0 and 3.8 Å for two molecules in the mTPMTwt-AdoHcy-TP structure are also consistent with the expectation that AdoMet-dependent MTases adopt a S_N2-like reaction mechanism. Similarly to that seen in the 6MP complex, the mTPMTwt-AdoHcy-TP structure shows that Arg147 and Arg221 may be the only residues that may potentially be involved in the deprotonation of thiophenol; by analogy with observations of Arg147 and Arg221 mutants, it is likely that Arg147 may interact with thiophenol more directly than Arg221, as was observed for 6MP.

Structural characterization of the mTPMTwt-AdoHcy-TIB and mTPMTwt-AdoHcy-4IB complexes allows identification of structural differences to explain the range of inhibition seen for 3,4,5,-triiodobenzoic acid, 4-iodobenzoic acid and 4-tert-butylbenzoic acid. The electron density for 3,4,5,-triiodobenzoic acid is stronger than that for 4-iodobenzoic acid (**Figure IV-1a, b**). The co-crystallization with 4-tert-butylbenzoic acid failed to give any detectable electron density for 4-tert-butylbenzoic acid. These results are consistent with the observations that 3,4,5,-triiodobenzoic acid is the most potent inhibitor among these

compounds, and 4-tert-butylbenzoic acid the weakest (**Table IV-2**) (Ames *et al.*, 1986). In the structures of complexes with benzoic acid inhibitors, all inhibitors are located in the active site of TPMT (**Figure IV-1a, b**). This observation suggests that inhibitions by these benzoic acid inhibitors are competitive with respect to 6MP, which contracts the original interpretation of the inhibition data in Woodson *et al.*. The carboxylate group of 3,4,5,-triiodobenzoic acid can form more interactions with TPMT than that of 4-iodobenzoic acid, due to that the influence of the 3- and 5-iodo substituents in orienting the inhibitor (**Figure IV-3**). Further analysis suggests that the side chains of Ser34, Arg221 and His222 on the active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix are flexible, since they interact differently with the 4-iodobenzoic acid and 3,4,5,-triiodobenzoic acid inhibitors (**Figure IV-4**). The structures of complexes with benzoic acid inhibitors also show that the aromatic ring of the inhibitor is about 4 Å from Arg147.

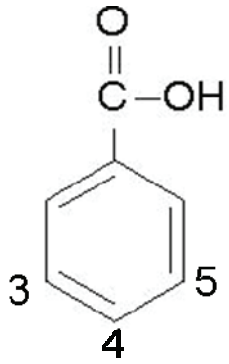
In summary, TPMT possesses a flexible active site which can accommodate the smaller acceptor substrate, thiophenol, or larger benzoic acid inhibitors. The crystal structures of TPMT complexes with benzoic acid inhibitors reveal that the flexibility of side chains of several residues on the active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix (Peng *et al.*, 2008) can aid the binding of benzoic acid inhibitors in slightly different orientations. These structural characteristics of the acceptor binding site may be useful in identifying new small molecule modulators for optimization of thiopurine-based therapy.

Table IV-1 Data collection and refinement statistics for mTPMT binding to benzoic acid inhibitors and thiophenol

	wt-AdoHcy-TIB	wt-AdoHcy-4IB	wt-AdoHcy-TP
Data collection			
cell dimensions	63.01	62.93	63.01
a, b, c (Å)	70.29	67.17	70.08
	72.16	72.50	72.28
. *β* (deg)	115.84	115.73	115.79
wavelength (Å)	1.0809	1.0809	0.97857
resolution (Å)a	50-1.76(1.82-1.76)	50-2.2(2.28-2.2)	50-1.8(1.86-1.8)
R_{sym}	8.2(31.4)	15.9(25.00)	6.0(36.0)
$I / \sigma I$	17.8(3.9)	11.1(6.2)	17.1(2.7)
completeness (%)	94.1(71.7)	88.5(61.5)	96.9(95.3)
redundancy	5.5(3.8)	6.5(4.5)	3.0(2.4)
Refinement			
resolution (Å)	50-1.76(1.82-1.76)	50-2.2(2.28-2.2)	50-1.8(1.86-1.8)
no. of reflections	50310	23366	48405
R_{work} / R_{free}	22.08(26.78)	21.72(27.79)	20.04(23.37)
no. of atoms			
protein, ligand, water	3760, 76, 386	3760, 72, 112	3760, 66, 414
average *B*-factors			
protein, AdoHcy, ligand, water	22.4, 14.3, 27.3, 30.6	41.0, 30.6, 51.9, 41.5	31.0, 20.2,57.5, 39.3
rmsd*			
bond lengths (Å)	0.012	0.013	0.013
bond angles (deg)	1.418	1.503	1.443
Ramachandran			
most favored (%)	93.5	90.3	94.0
additional allowed (%)	6.5	9.7	6.0
disallowed (%)	0.0	0.0	0.0

TIB, 3, 4, 5,-triiodobenzoic acid.
4IB, 4-iodobenzoic acid
TP, thiophenol
Numbers in parentheses refer to the highest resolution shell.
*rmsd, root mean square deviation.

Table IV-2 Inhibition of TPMT by benzoic acid compounds

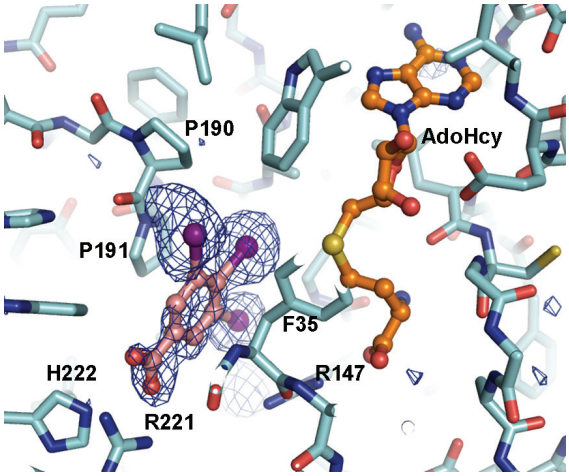


	Substituents			
	3	4	5	pI_{50}
3, 4, 5,-triiodobenzoic acid	**I**	**I**	**I**	**6.13**
4-iodobenzoic acid		**I**		**4.12**
4-ter-butylbenzoic acid		**$C(CH_3)_3$**		**2.68**

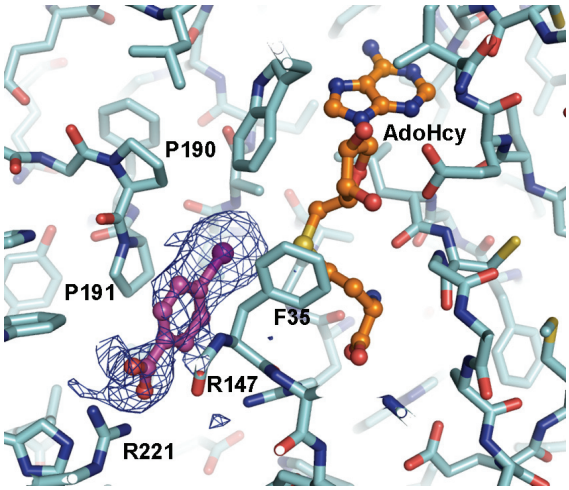
pI_{50} values are the –log of IC_{50} values. All values are cited from Ames *et al.*, 1986. IC_{50} refers to the concentration of compound that inhibits enzyme activity by 50%.

Figure IV-1

(a)



(b)



(c)

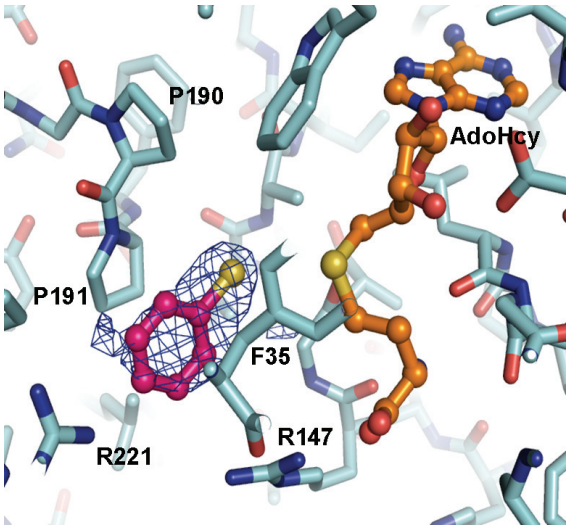


Figure IV-1 The omit $|F_o|-|F_c|$ density of benzoic acid inhibitors and thiophenol.

(a) The 1.76 Å resolution |Fo|-|Fc| omit electron density for 3,4,5,-triiodobenzoic acid (light pink), contoured at 3.0σ.

(b) The 2.2 Å resolution |Fo|-|Fc| omit electron density for 4-iodobenzoic acid (magenta), contoured at 2.0σ.

(c) The 1.8 Å resolution |Fo|-|Fc| omit electron density for thiophenol (pink), contoured at 3.0σ. The density shown is for molecules A in each crystal structure; similar density was observed for their corresponding molecules B. Active site residues near the bound inhibitors and thiophenol are labeled.

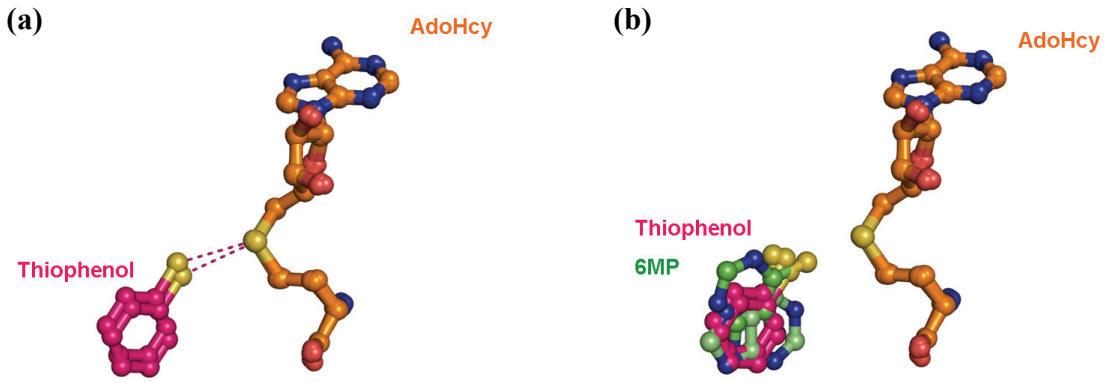


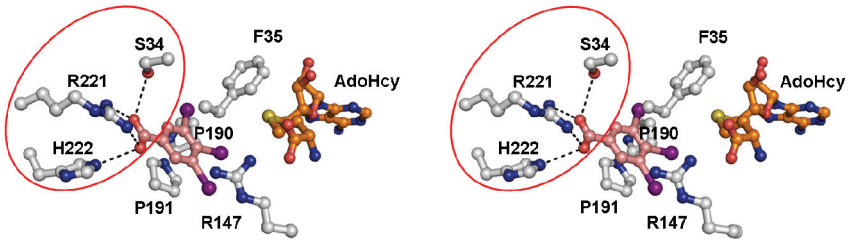
Figure IV-2 Thiophenol binding to mTPMT.

(a) Thiophenol and AdoHcy bound to mTPMTwt-AdoHcy-TP. Molecules A and B were superimposed and only the two molecules each of thiophenol (pink) and AdoHcy (orange) are shown for clarity. The two AdoHcy superimpose nearly perfectly since they are bound in the same position and conformation; the two thiophenols are also bound in similar positions and orientations. Dashed lines represent the possible path of methyl transfer between AdoHcy and thiophenol sulfur atoms.

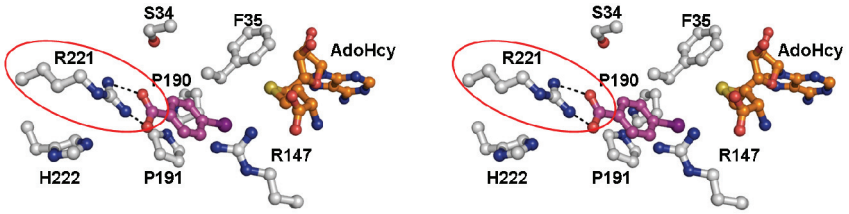
(b) 6MP and AdoHcy bound to mTPMTwt-AdoHcy-6MP (Peng *et al.*, 2008). Superposition of these mTPMT structures shows the thiophenols and 6MP (green) bound in similar positions.

Figure IV-3

(a)



(b)



(c)

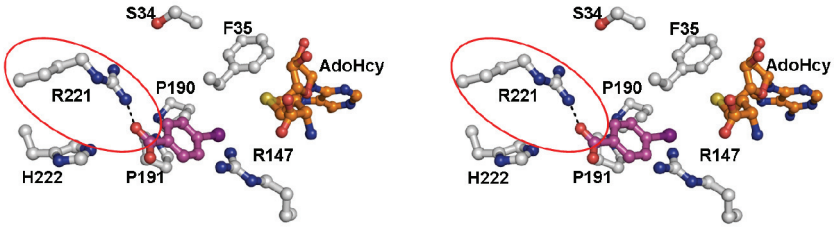


Figure IV-3 Enzyme-inhibitor interactions in the mTPMT active site with benzoic acid inhibitors bound. The side chains of residues that are involved in the binding of inhibitors are circled in red.

(a) Stereoview of 3,4,5,-triiodobenzoic acid (light pink) in mTPMTwt-AdoHcy-TIB. The inhibitor is bound similarly to both molecules A and B in the crystal structure.

(b) Stereoview of 4-iodobenzoic acid (magenta) binding to molecule A in mTPMTwt-AdoHcy-TIB.

(c) Stereoview of 4-iodobenzoic acid (magenta) binding to molecule B in mTPMTwt-AdoHcy-TIB.

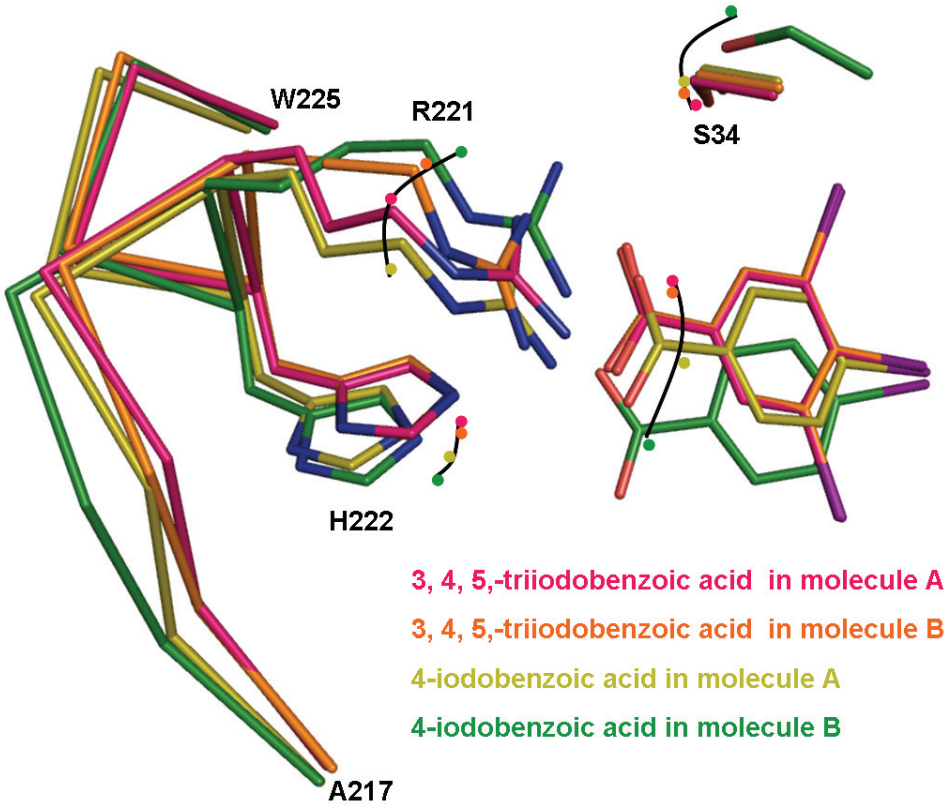


Figure IV-4 Flexibility of side chains of several residues on the active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix. The side chains of Ser34, Arg221 and His222 as well as the benzoic acid inhibitors are shown as ball-and-stick structures. Residues in Glu220-Trp225 near the end of a helix are labeled. The curved lines illustrate shifts in TPMT side chains and inhibitors when comparing the different structures.

CHAPTER V

Crystal Structure of Human Nicotinamide N-Methyltransferase

5.1 Introduction

Nicotinamide is known as a compound essential for the formation of NAD(H) and NADP(H), which are involved in many important biological processes (Sartini *et al.*, 2006). Nicotinamide N-methyltransferase (NNMT) is an enzyme that catalyzes the N-methylation of nicotinamide, pyridines and other structural analogs using S-adenosylmethionine as the methyl donor (Williams and Ramsden, 2005b). NNMT was characterized by Cantoni in 1951 and is mainly expressed in the liver, where its activity varies 5-fold among individuals and has a bimodal frequency distribution. NNMT is another case in which genetic polymorphism may be associated with its enzyme activity (Aksoy *et al.*, 1994). In addition, the enhanced NNMT activity also leads to the production of toxic N-methylpyridinium compounds in Parkinson's disease (Williams *et al.*, 2005; Williams and Ramsden, 2005a). Thus NNMT may play a crucial role in the biotransformation and detoxification of xenobiotic compounds.

Lower expression of NNMT is also found in kidney, lung, placenta, heart, brain and muscle. Interestingly, the abnormal expression of NNMT was identified in various cancers, such as glioblastoma, stomach adenocarcinoma, papillary thyroid cancers, renal carcinoma, colorectal cancer and so on (Sartini *et al.*, 2007). To elucidate the structural basis of NNMT catalysis and function for the design of new NNMT inhibitors and even the discovery of new drugs, crystal structures of NNMT bound to substrate or inhibitor are desirable.

Although the crystal structures of human and mouse NNMT bound to AdoHcy have been deposited in the Protein Data Bank (PDB accession code: 2IIP and 2I62), they didn't reveal nicotinamide binding. We reported here the crystal structure of human NNMT bound to both AdoHcy and nicotinamide. The structure highlights several residues in the active site which may play roles in nicotinamide recognition and NNMT catalysis, and provides a structural basis for the design of NNMT mutants to further investigate the enzyme's catalytic mechanism.

5.2 Materials and Methods

5.2.1 Expression and Purification of NNMT

Using the crystal structure of the human NNMT-AdoHcy complex as a guide (unpublished, PDB accession code: 2IIP), the full-length human NNMT triple mutant (Lys100Ala/Glu101Ala/Glu103Ala) coding region was obtained by the cleavage of the N*de*I and H*in*dIII sites of recombinant human NNMT triple mutant plasmid provided by Dr. Emanuelli (Institute of Biochemical Biotechnologies, Italy) and cloned into the N*de*I and H*in*dIII sites of pET-28a (Novagen) to express N-terminally His_6-tagged NNMT. The pET-28a-NNMT was transformed into *E.coli* BL21(DE3) cells that were grown at 37^{o}C to an OD_{600} of ~1.0. At that point, the cultures were induced with 1mM IPTG and incubated overnight at 25^{o}C. Cells were harvested by centrifugation and resuspended in lysis buffer (50mM Tris-HCl, pH8.0, 0.5M NaCl, 5mM imidazole, 2mM β-mercaptoethanol, and 5% glycerol) with protease inhibitor (1mM phenylmethyl sulfonyl fluoride, PMSF) before sonication. After centrifugation, the supernatant was loaded onto Ni^{2+}-NTA resin (Qiagen) and washed with 50mM Tris-HCl, pH8.0, 0.5M NaCl, 25mM

imidazole, and 5% glycerol. The expressed NNMT protein was then eluted with 50mM Tris-HCl, pH8.0, 0.5M NaCl, 250mM imidazole, and 5% glycerol. The combined eluted fractions were dialyzed against 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5% glycerol, and 1 mM DTT. Final purification was by gel filtration on a Superdex 75 10/300GL column (Amersham Biosciences) equilibrated with 20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5% glycerol, and 1 mM DTT.

5.2.2 Crystallization and Data Collection of NNMT

The purified NNMT was incubated with AdoHcy at 1:4 molar ratio before concentration to 10mg/ml. The concentrated NNMT complex with AdoHcy was then incubated with 10mM nicotinamide for ~30min on ice before crystallization. Crystals of NNMT bound to AdoHcy and nicotinamide were grown by sitting drop vapor diffusion at 20^{o}C using a 1:1 protein: precipitant volume ratio from two conditions: condition 1 with 0.1M bis-Tris (pH5.5) and 25% PEG3350, and condition 2 with 0.2M NaCl, 0.1M Bis-Tris (pH5.5) or imidazole (pH6.5), and 25-27% PEG3350. All crystals grew in clusters (**Figure V-1a**). Screening a variety of additives and employing microseeding failed to yield single crystals. Thus for data collection, single crystals were obtained by separation from the clusters using cryoloops. Cryoprotection in artificial mother liquor containing 35% PEG3350 was followed by flash cooling of the crystals in liquid nitrogen. One 2.7 Å resolution native diffraction dataset was measured at APS beamline 19ID (**Table V-1**). The data were processed with HKL (Otwinowski and Minor, 1997). The crystal belongs to space group P1 and contains four molecules in each asymmetric unit.

5.2.3 Structure Determination and Refinement of NNMT

The structure was determined by molecular replacement using the program MOLREP (Vagin and Teplyakov, 1997) and the protein coordinates from the crystal structure of human NNMT bound to AdoHcy (PDB accession code: 2IIP). Iterative rounds of model building were performed with COOT (Emsley and Cowtan, 2004), and refinement calculations were carried out in REFMAC (Murshudov *et al.*, 1997). A Ramachandran plot calculated with PROCHECK (Laskowski *et al.*, 1993) placed 81.7% of the residues in the most favored regions and 0.4% in disallowed regions. No density is observed for N- and C-terminal residues 1-2 and 262-264. The final refined structure includes four molecules of NNMT containing residues 3-261, AdoHcy and nicotinamide ligands bound to each, and 50 water molecules. A summary of the refinement statistics is provided in **Table V-1**. All figures were generated using PyMol (DeLano, 2004).

5.3 Results

5.3.1 Overall Structure of NNMT Ternary Complex

The crystal asymmetric unit contains four conformationally similar copies of NNMT. Pairwise superpositions of the Cα atoms give overall rmsd values ranging from 0.5-0.8Å. Thus for simplicity, structural details provided in the following sections are for molecule A unless otherwise noted.

NNMT possesses the class I AdoMet-dependent MTase core fold (see Chapter I) (Martin and McMillan, 2002; Schubert *et al.*, 2003a; Kozbial and Mushegian, 2005), which is comprised of a seven-stranded β sheet flanked by α helices on each side (**Figure V-1b**).

The NNMT structure has several modifications to the MTase core: two additional α helices at its N-terminus, two additional α helices inserted in Trp107-Ala134, a small inserted helix from Cys165-Cys170, and a reversed β hairpin inserted at Tyr203-Ser212 (shown in purple in **Figure V-1b**). A structural similarity search using DALI (Holm and Sander, 1993) identified two closest structural matches, which are both small molecule MTases: indolethylamine N-methyltransferase bound to AdoHcy (PDB accession code 2A14) (INMT, 1.2 Å rmsd for 257 superimposed Cα atoms, 52% sequence identity) and phenylethanolamine N-methyltransferase bound to AdoHcy and an inhibitor (PNMT, 1.5 Å rmsd for 251 superimposed Cα atoms, 39% sequence identity) (Gee *et al.*, 2007). Structural alignments with INMT and PNMT reveal similar patterns of insertions in the MTase core fold and highly similar site for AdoHcy/AdoMet binding. Both NNMT and PNMT display an enclosed active site with the "cover" formed by several insertion motifs (**Figure V-1c**). In the PNMT ternary complex, the inhibitor is positioned more than 6 Å from AdoHcy. This is in contrast to two other ternary complexes with both the methyl donor and acceptor binding sites occupied such that ligands are bound productively in the acceptor binding site. These structures are *Coffea canephora* 3,7-dimethylxanthine methyltransferase bound to AdoHcy and theobromine (DXMT, 2.8 Å rmsd for 196 superimposed Cα atoms, 14% sequence identity) (McCarthy and McCarthy, 2007), and *Coffea canephora* xanthosine methyltransferase bound to AdoHcy and xanthosine (XMT, 2.7 Å rmsd for 186 superimposed Cα atoms, 16% sequence identity) (McCarthy and McCarthy, 2007).

5.3.2 AdoHcy Binding to NNMT

Electron density corresponding to AdoHcy is very well-defined in all NNMT molecules (**Figure V-2a**). AdoHcy is bound in a similar position and conformation in NNMT as seen in other class I MTase structures (**Figure V-1b; Figure V-2a**), through extensive hydrogen-bonding and van der Waals interactions (see Chapter I). The adenine ring nitrogen atoms interact with the side chain of Asp142 and the main chain amine of Val143. Additionally, the adenine ring is sandwiched between the hydrophobic side chains of Tyr86 and Ala169 (**Figure V-2c**). The ligand's ribose hydroxyl groups form hydrogen bonds with the Asp85 and Asn90 side chains. At the other end of AdoHcy, the terminal amine group forms hydrogen bonds with the carbonyl oxygen atoms of Gly63 and Thr163, and the carboxylate group forms hydrogen-bonding interactions with the hydroxyls of Tyr20, Tyr25, Tyr69 and Thr163 side chains (**Figure V-2c**). Interestingly, one entire face of the AdoHcy is surrounded by a constellation of six tyrosines (Tyr11, Tyr20, Tyr25, Tyr69, Tyr86, and Tyr204) and one phenylalanine (Phe15) (**Figure V-2c**), similar to those seen in the PNMT structure, which have been confirmed to be important for PNMT function (Martin *et al.*, 2001). These residues involved in the binding site of AdoHcy are conserved among the NNMTs.

5.3.3 Nicotinamide Binding to NNMT

In the NNMT complex structure, the electron density for the nicotinamide acceptor substrate is weaker than for AdoHcy (**Figure V-2a,b**), which is consistent with the observation that human liver NNMT has higher affinity for AdoMet (K_m of 1.8 μM) than for nicotinamide (K_m of 0.43 mM) (Aksoy *et al.*, 1994). In the four molecules of our crystal structure, the reactive nitrogen atom in the nicotinamide aromatic ring is situated

3.7, 3.5, 3.7 and 4.5 Å from the sulfur atom of its corresponding AdoHcy. The nicotinamide nitrogen position is suitable for attack by the AdoMet methyl group, consistent with the presumption that catalysis by AdoMet-dependent MTases proceeds by a conserved S_N2-like mechanism in which the methyl group is transferred directly from AdoMet to the acceptor (see Chapter I). The nicotinamide is positioned in the active site through both hydrogen-bonding and van der Waals interactions similarly in all four molecules except for a small shift in molecule D **(Figure V-2c**). The amine of the nicotinamide forms hydrogen bonds with Asp197, Ser201, and Ser213. The Tyr20 hydroxyl group is also close enough to the nicotinamide ring to interact electrostatically. Tyr24 packs against the nicotinamide *via* hydrophobic interactions, and the side chains of Leu164 and Tyr204 form a hydrophobic clamp to orient the aromatic ring of the nicotinamide. In one NNMT molecule, Tyr242 is seen in a water-mediated interaction with the nicotinamide (**Figure V-2c**). Of these active site residues, Asp197, Ser201, Ser213, and Tyr20 may be involved in the deprotonation of nicotinamide and be important for NNMT function.

5.4 Discussion

The work reported here provides structural details of nicotinamide binding to NNMT and identifies a number of important residues that may be necessary for NNMT substrate recognition and catalysis. The NNMT protein fold in the ternary complex with AdoHcy and nicotinamide described here is very similar to that of the NNMT binary complex with AdoHcy (PDB accession code: 2IIP). The rmsd for superposition of two crystal structures

is ~0.4 Å, and AdoHcy is bound in the same conformation in the active sites of both complexes.

A structural database search identified DXMT and XMT (McCarthy and McCarthy, 2007) as the most similar ternary complexes containing bound AdoHcy and acceptor substrate. NNMT, DXMT and XMT bind AdoHcy in an extended conformation in similar positions and orientations (**Figure V-3**). Our NNMT complex structure reveals nicotinamide-AdoHcy nitrogen-sulfur distances of 3.5-4.5 Å that are consistent with an S_N2-like reaction mechanism involving direct methyl transfer (**Figure V-3a**). Superposition of the NNMT ligands with the AdoHcy and theobromine bound to DXMT (theobromine-AdoHcy nitrogen-sulfur distance of 3.6 Å) and the AdoHcy and xanthosine bound to XMT (xanthosine-AdoHcy nitrogen-sulfur distance of 4.0 Å) indicate that their acceptor substrates are bound in their active sites in similar positions and orientations relative to their methyl donor (**Figure V-3b,c**).

As mentioned before, the S_N2-like reaction mechanism adopted by AdoMet-dependent MTases also requires a deprotonation step before, during, or after the methyl transfer. Inspection of the NNMT ternary complex presented here provided a view of nicotinamide binding. Several residues found in the active site may be candidates for substrate deprotonation. Ser201, Ser213 and Asp197 form hydrogen bonds with the amide group of nicotinamide (**Figure V-2c**). The chemical character of Asp197 makes it a good base in a possible nicotinamide deprotonation step. In the four molecules of the NNMT crystal structure, at least one of the serines can hydrogen bond to a mainchain carbonyl oxygen

so that it may be activated as a base. In addition, Tyr20, Tyr25, and Tyr69 form hydrogen bonds to the AdoHcy carboxylate (**Figure V-2c**), which may stabilize the negative charge on the AdoHcy carboxylate so as to effectively activate the Tyr20 side chain hydroxyl group as a base to deprotonate nicotinamide (**Figure V-2c**). Furthermore, Tyr242 is seen close to a water-mediated interaction in one molecule. Therefore, it might also play a role in the catalytic mechanism of NNMT. To investigate the potential catalytic importance of these residues, a structure-based mutagenesis study is underway. First, a number of single mutants (Asp197Ala, Ser201Ala, Ser213Ala, Tyr20Ala and Tyr20Phe), double mutants (Ser201Ala/Ser213Ala, Ser201Ala/Asp197Ala, and Ser213Ala/Asp197Ala) and a triple mutant (Ser201Ala/Ser213Ala/Asp197Ala) have been designed to identify the possible residues in nicotinamide deprotonation step. As mentioned before, the amine of the nicotinamide forms hydrogen bonds with Asp197, Ser201, and Ser213. If any three residues are involved in the nicotinamide deprotonation, the single, double or triple mutants altering the deprotonating residue may lead to dramatically decreased enzyme activity, with or without decreased affinity for nicotinamide. If any or all of the residues are not involved in nicotinamide deprotonation, the single, double, or triple mutants may still affect the nicotinamide binding, but with little or no effect on enzyme activity. As seen in the crystal structure, the Tyr20 hydroxyl group is also close enough to the nicotinamide ring to interact electrostatically. Therefore, the Tyr20Phe and Tyr20Ala mutants may decrease the affinity for nicotinamide and decrease enzyme activity. If Tyr20 is important in nicotinamide deprotonation, both Tyr20Ala and Tyr20Phe mutants will result in dramatic loss of enzyme activity. Both alanine and phenylalanine substitutions are proposed in case the aromatic ring is important for protein

folding/stability. If Tyr20 is able to act as a base, it is likely due to activation by its interaction with the AdoMet carboxylate. Tyr25 and Tyr69 also hydrogen bond to the AdoMet carboxylate and could modulate the activation of Tyr20. Thus in order to test whether these two additional tyrosine residues may play a role in indirectly stabilizing Tyr20 as the catalytic base, the single mutants (Tyr25Ala, and Tyr69Ala) and a double mutant (Tyr25Ala/Tyr69Ala) have been designed. These mutants may decrease the enzyme activity with or without altering the affinity for AdoHcy/AdoMet. Asp197, Ser201, Ser213 and Tyr20 are the most likely candidates for the deprotonating base. If none are identified by functional analysis of the proposed mutants to be important, a second set of conserved active site tyrosine residues will be targeted for mutagenesis. Tyr242 is another candidate for the possible catalytic base, and Tyr24 and Tyr204 are likely to be important for nicotinamide binding. Overall, the structure of NNMT presented here identifies a number of potential residues involved in substrate recognition and catalysis. It provides a basis for the design of NNMT mutants for further mechanistic study, and for the future discovery of novel NNMT inhibitors as possible leads for drug discovery.

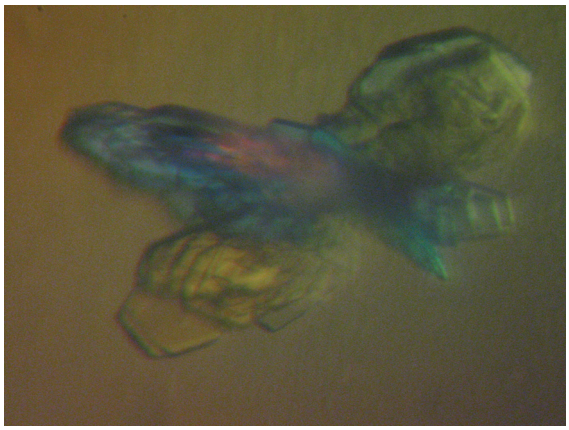
Table V-1 Data collection and refinement statistics for NNMT

Data collection	
cell dimensions	
a, b, c (Å)	60.66, 61.79, 74.34
α, β, γ (deg)	106.57, 103.98, 104.14
wavelength (Å)	0.97857
resolution (Å)	50-2.7 (2.8-2.7)
R_{sym}	9.1(47.6)
$I / \sigma I$	9.6(1.9)
completeness (%)	98.7(97.5)
redundancy	2.5(2.4)
Refinement	
resolution (Å)	50-2.7 (2.8-2.7)
no. of reflections	23684
R_{work} / R_{free}	20.88 (31.45)
no. of atoms	
protein, ligand, water	8092, 140, 50
average *B*-factors	
protein, AdoHcy, nicotinamide, water	35.6, 30.9, 43.4 , 24.4
rmsd*	
bond lengths (Å)	0.011
bond angles (deg)	1.528
Ramachandran	
most favored (%)	81.7
additional allowed (%)	17.9
disallowed (%)	0.4

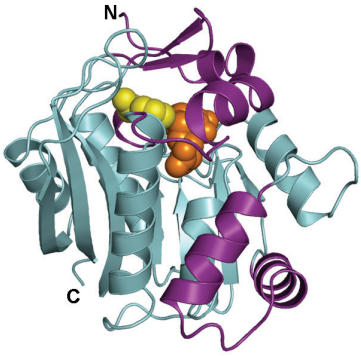
Numbers in parentheses refer to the highest resolution shell.
*rmsd, root mean square deviation.

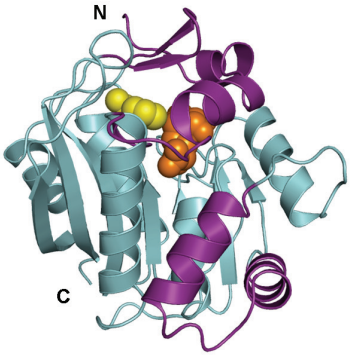
Figure V-1

(a)

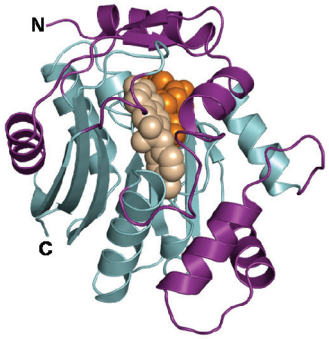


(b)





(c)



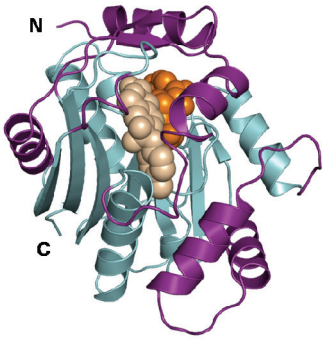


Figure V-1 Crystal and structure of NNMT.

(a) Cluster of NNMT crystals grown in the presence of AdoHcy and nicotinamide.

(b) Stereo ribbon diagram of NNMT. The conserved class I AdoMet-dependent MTase core fold is shown in cyan, and the inserted regions characteristic of the NNMT structure are in purple. AdoHcy (orange) and nicotinamide (yellow) are shown as space-filling spheres.

(c) Stereo ribbon diagram of PNMT (PDB accession code 2OBF; Gee *et al.*, 2007). The conserved class I AdoMet-dependent MTase core fold is shown in cyan, and the inserted regions characteristic of the PNMT structure are in purple. AdoHcy (orange) and inhibitor (light brown) of PNMT are shown as space-filling spheres.

Figure V-2

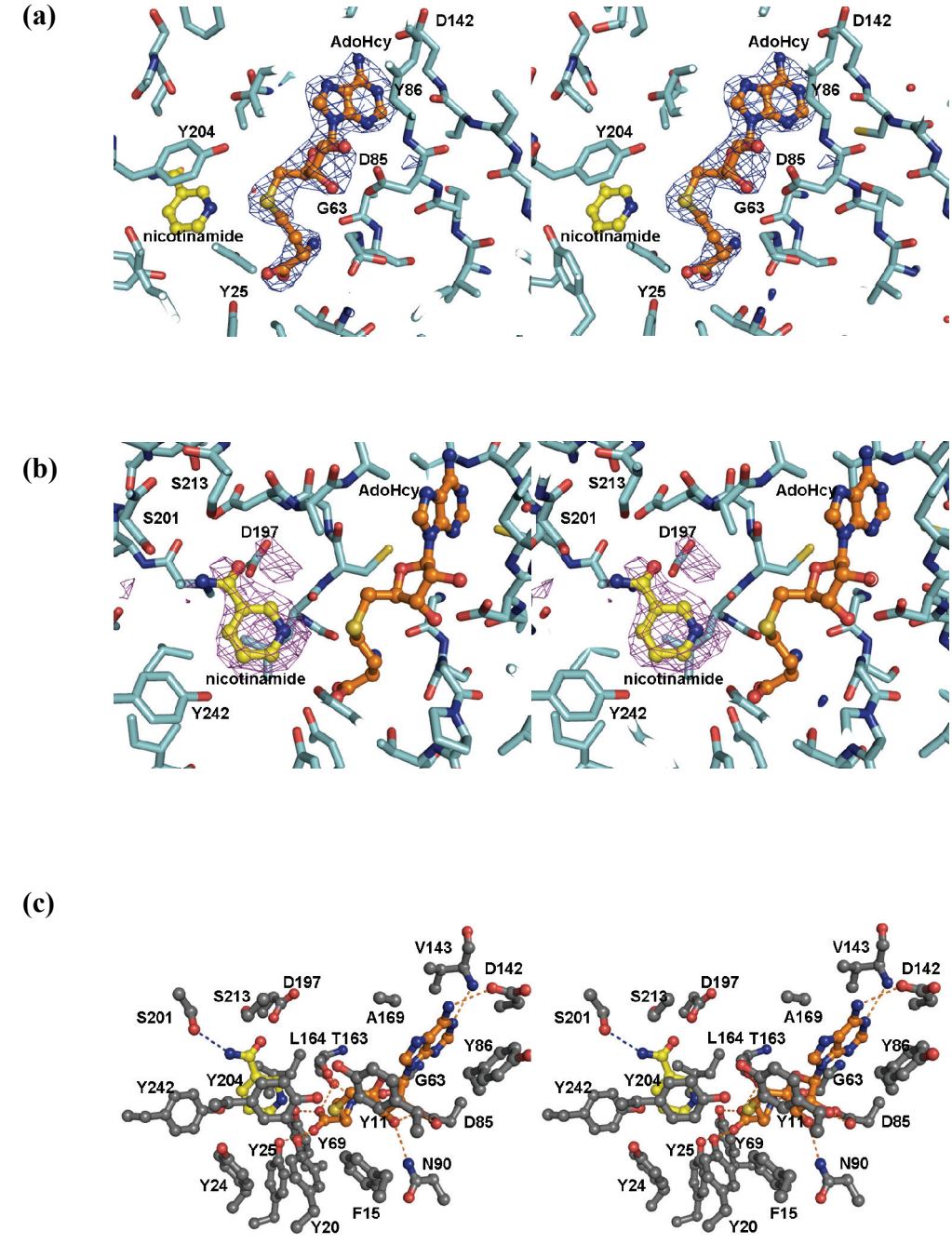

(a)
AdoHcy
D142
Y86
Y204
D85
G63
nicotinamide
Y25
(b)
S213
S201
D197
AdoHcy
nicotinamide
Y242
(c)
V143
D142
D197
S213
S201
A169
L164
T163
Y86
Y242
Y204
G63
Y11
D85
Y25
Y69
N90
Y24
F15
Y20

Figure V-2 NNMT active site.

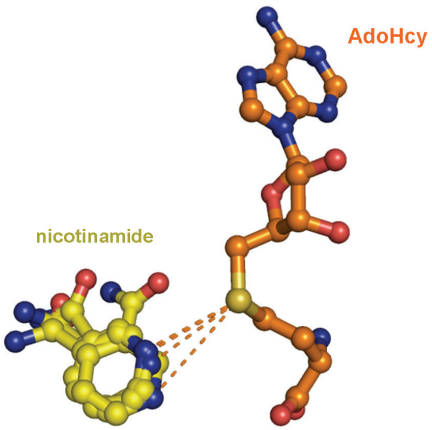
(a) AdoHcy bound to the active site of molecule A in the NNMT crystal structure. Stereoview of 2.7Å resolution omit $|F_o|$ - $|F_c|$ density contoured at 3σ is shown; similar density is present for the other three molecules in NNMT crystal structure. AdoHcy (orange), nicotinamide (yellow), and nearby NNMT residues (cyan) are shown as ball –and-stick structures.

(b) Nicotinamide bound to the active site of molecule A in NNMT crystal structure. Stereoview of 2.7Å resolution omit $|F_o|$ - $|F_c|$ density contoured at 1.5σ is shown.

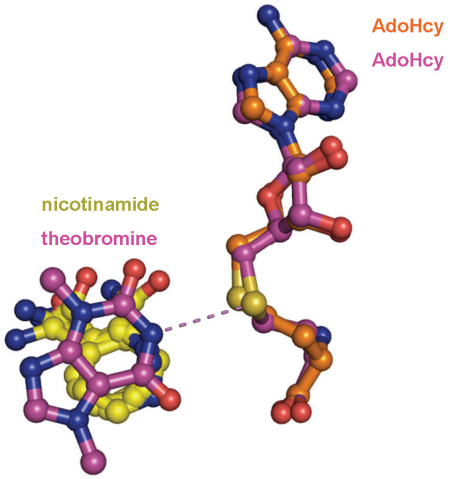
(c) Stereoview of the active site of molecule A in the NNMT crystal structure. AdoHcy (orange), nicotinamide (yellow), and nearby NNMT residues (dark gray) are shown as ball–and-stick structures. The hydrogen bonds between AdoHcy and nearby NNMT residues are shown as orange dashed lines, and the hydrogen bonds between nicotinamide and nearby NNMT residues are shown as blue dashed lines.

Figure V-3

(a)



(b)



(c)

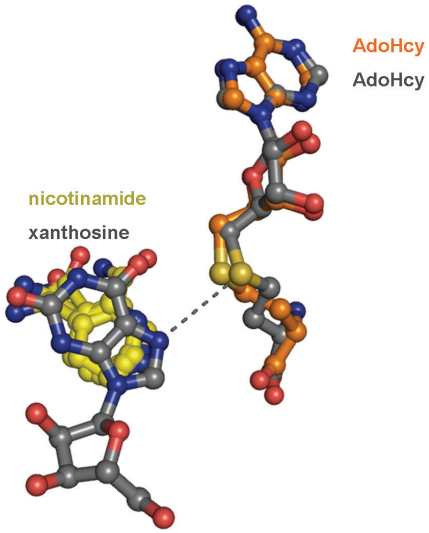


Figure V-3 AdoHcy and

acceptor substrate binding to AdoMet-dependent MTases. (a) AdoHcy and nicotinamide acceptor substrate bound to NNMT. Superposition of the AdoHcy bound to four molecules in the NNMT crystal shows that the four nicotinamides are bound in similar position and orientations. Dashed lines represent the possible path of methyl transfer between the sulfur atom of AdoHcy and the nitrogen atom of nicotinamide.

(b) AdoHcy and threobromine acceptor substrate bound to DXMT (McCarthy and McCarthy, 2007). Superposition of the DXMT and NNMT structures shows that threobromine and nicotinamide approach their respective AdoHcy from the same general direction. Dashed lines represent the possible path of methyl transfer between the sulfur atom of AdoHcy and the nitrogen atom of threobromine.

(c) AdoHcy and xanthosine acceptor substrate bound to XMT (McCarthy and McCarthy, 2007). Superposition of the XMT and NNMT structures shows that xanthosine and nicotinamide approach their respective AdoHcy in similar fashion. Dashed lines represent the possible path of methyl transfer between the sulfur atom of AdoHcy and the nitrogen atom of xanthosine.

CHAPTER VI

Expression and Purification of Arsenic Methyltransferase

6.1 Introduction

Arsenic is a metalloid widely found in nature. Its exposure can lead to neurotoxicity, carcinogenesis, hepatic injury, cardiac failure, leucopenia, and even death (Yoshida *et al.*, 2004; Simeonova and Luster, 2004). In particular, arsenic contamination of ground water throughout the world is a public health concern. In addition, As_2O_3 is an active therapeutic agent for promyelocytic leukemia (Sanz *et al.*, 2005). Methylation of inorganic arsenic to methylated and dimethylated arsenic has been regarded as a major mechanism of arsenic biotransformation and detoxification (John *et al.*, 2008). However, methylated metabolites of arsenic, especially those in the trivalent state, have also been found to even exceed arsenite and arsenate in potency as cytotoxins or DNA-damaging agents and as enzyme inhibitors (Thomas *et al.*, 2001).

Recently, arsenic methyltransferase (AS3MT) has been characterized to catalyze the multistep process that converts inorganic arsenic to its methylated metabolites with S-adenosylmethionine (AdoMet) as methyl donor (Lin *et al.*, 2002). The critical role of AS3MT in methylation of inorganic arsenic has been further shown by the observation that expression of rat AS3MT in human urothelial (UROtsa) cells which do not express AS3MT enables the cells to methylate arsenic and affects the cytotoxicity (Drobna *et al.*, 2005). Moreover, the knock-down of AS3MT expression in human hepatocellular carcinoma (HepG2) cells leads to significantly decreased methylation of arsenic (Drobna

et al., 2006). More recently, a pharmacogenetics study of human AS3MT identified 26 single nucleotide polymorphisms, three of which are nonsynonymous cSNPs (Arg173Trp, Met287Thr, and Thr306Ile) (Wood *et al.*, 2006). The study also indicated that the human AS3MT polymorphisms might contribute to individual variation in biology and pathology of chronic arsenic exposure as well as the response to arsenic-containing therapeutic agents (Vahter, 2000; Drobna *et al.*, 2004; Wood *et al.*, 2006).

Together, these findings suggest a significant role of AS3MT in arsenic methylation. To elucidate the mechanism of arsenic methylation and the relationship between AS3MT polymorphisms and individual variation in arsenic metabolism, we are pursuing the crystal structure of AS3MT. We report here the expression and purification of AS3MT, for the purpose of crystallographic studies.

6.2 Materials and Methods

6.2.1 Cloning of AS3MT

The full-length human AS3MT (hAS3MT) and mouse AS3MT (mAS3MT) coding regions were amplified by PCR from plasmids provided by Dr. Weinshilboum from the Mayo Clinic. They were cloned into the pET-28a (Novagen) vector to give expression constructs coding for AS3MTs with an N-terminally fused His_6 tag. In addition, the coding region which produces the Trp^{173} variant of hAS3MT was also amplified and cloned into pET-28a (Novagen) vector to express Trp^{173} allozyme with an N-terminally fused His_6 tag. The expression plasmids of hAS3MT Thr^{287} and Ile^{306} variants were prepared using the QuikChange site directed mutagenesis kit (Stratagene).

To design a hAS3MT mutant with reduced surface entropy which might facilitate crystallization, lysines which were predicted to be located on the protein surface were targeted for mutagenesis. In this process, a sequence alignment of human, mouse and rat AS3MT was combined with a structural homology model of hAS3MT computed with ESyPred3D Web Server 1.0 (http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/). This procedure led to the identification of 1) Lys105 located on the surface of loop as a residue which might prevent crystallization, and 2) Lys113 as a nonconserved charged residue on the region in the protein surface in which exposed charged residues had to be mutated in another AdoMet-dependent MTase for successful crystallization (http://sgc.utoronto.ca/SGC-WebPages/StructureDescription/MM.php?pdb=2IIP). Thus to improve AS3MT crystallizability, the Lys113Ala single mutation and Lys105Ala/Lys113Ala double mutations were introduced into the expression construct of His_6-tagged Thr^{287} allozyme, which is more stable and active than wild-type protein, using the QuikChange site directed mutagenesis kit (Stratagene).

6.2.2 Expression and Purification of AS3MT

The expression constructs of AS3MT were transformed into *E.coli* BL21(DE3)plysS cells. They were grown at 37 ^{o}C in Luria Broth (LB) to an OD_{600} of ~0.6 before induction with 1 mM IPTG. After incubation for an additional 3-4 h at 37 ^{o}C, cells were harvested by centrifugation and resuspended in lysis buffer (50mM Tris-HCl, pH7.9, 500mM NaCl, and 5mM imidazole) with the protease inhibitor cocktail for sonication. After centrifugation, the supernatant was loaded onto Ni^{2+}-NTA resin (Qiagen) and washed with lysis buffer. The expressed AS3MT protein was eluted with 50mM Tris-HCl,

pH7.9, 500mM NaCl, and 150mM imidazole. The eluted fractions were pooled and dialyzed into 50mM Tris-HCl, pH8.2, 2mM DTT. Subsequently, the fractions from dialysis were flash-frozen by liquid nitrogen and then stored at -80 ^{o}C. Final purification was by gel filtration on a Superdex 200 10/300GL column (Amersham Biosciences) equilibrated with 20mM Tris-HCl, pH8.0, 100mM NaCl, 2mM DTT, 1mM AdoHcy, and 5% glycerol for hAS3MT purification, or 20mM Tris-HCl, pH8.0, 100mM NaCl, and 2 or 5mM DTT for mAS3MT purification (**Figure VI-1a, b**). The estimated protein yields of AS3MT range from 30 to 50 mg/L of *E. coli*.

6.2.3 Crystallization Trials of AS3MT

Taking into account that the hAS3MT Thr^{287} allozyme displayed enhanced activity and protein level with respect to wild-type enzyme (Wood *et al.*, 2006), the His_6-tagged Thr^{287} allozyme was first used for initial crystallization screens with commercially available screening kits (Hampton Research). Co-crystallization efforts with AdoHcy (2mM), AdoMet (5mM), and sodium arsenite (2mM) were carried out. However, the experiments did not yield promising initial crystallization conditions. Subsequently, the His_6-tagged Thr^{287} allozyme was cleaved by thrombin to remove the N-terminally fused His_6 tag. The non-tagged Thr^{287} allozyme, and tagged Lys113Ala single mutant and Lys105Ala/Lys113Ala double mutant (described above) were also used to search for the initial crystallization conditions. Although promising crystallization conditions have not yet been identified, efforts are ongoing.

As an alternative approach, since mAS3MT share 76% sequence identity with hAS3MT, the crystal structure of mAS3MT would also be useful to elucidate the mechanism for arsenic metabolism and help understand AS3MT polymorphisms. Therefore, mAS3MT, which has been shown to be functional through enzyme activity assays (specific activity: 7.0×10^5 counts/minute/mg protein for bacterially expressed mAS3MT; 1.1×10^4 counts/minute/mg protein for Cos-1 expressed hAS3MT) in Dr. Weinshilboum's lab from Mayo Clinic, was used for crystallization screening as well. Co-crystallization in the presence of AdoHcy (2mM), AdoMet (5mM), and sodium arsenite (2mM) have been carried out. Currently, no crystals have been obtained. Efforts to crystallize either hAS3MT or mAS3MT continue.

Figure VI-1

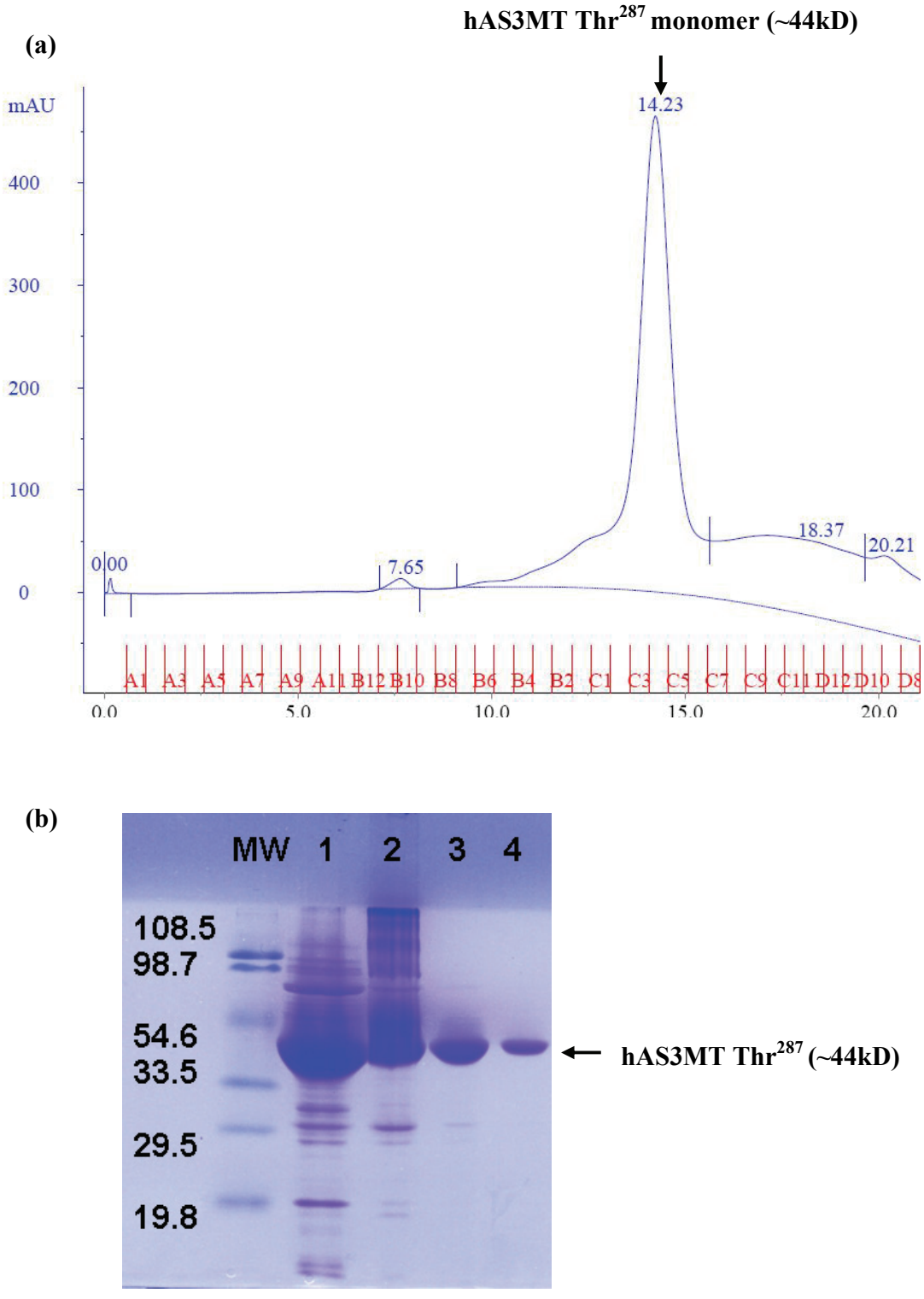

hAS3MT Thr287 monomer (~44kD)
(a)
mAU
400
300
200
100
0
14.23
18.37
20.21
7.65
0.00
A1 A3 A5 A7 A9 A11 B12 B10 B8 B6 B4 B2 C1 C3 C5 C7 C9 C11 D12 D10 D8
0.0
5.0
10.0
15.0
20.0
(b)
MW 1 2 3 4
108.5
98.7
54.6
33.5
29.5
19.8
hAS3MT Thr287 (~44kD)

Figure VI-1 Purification of AS3MT.

(a) Size-exclusion chromatography elution profile of hAS3MT Thr^{287}. A Superdex 200 column 10/300GL (Amersham Biosciences) was equilibrated with 20mM Tris-HCl, pH8.0, 100mM NaCl, 2mM DTT, 1mM AdoHcy, and 5% glycerol. The theoretic molecular weight for the hAS3MT Thr^{287} monomer, including the His_6 tag, is ~44kD. The hAS3MT Thr^{287} elutes at 14.23ml, giving a peak which corresponds to a monomer.

(b) Coomassie-stained SDS-PAGE of purified hAS3MT Thr^{287}. Lane 1 contains the concentrated protein after elution from the Ni^{2+}-NTA resin. Lanes 2, 3, and 4 show hAS3MT Thr^{287} after purification from the Superdex 200 10/300GL column (Amersham Biosciences), at different loading amounts. MW represents molecular weight markers (sizes indicated in kD).

CHAPTER VII

Summary and Future Directions

As introduced in Chapter I, the widely distributed MTases in living organisms play critical roles in the important and diverse methylation reactions. Among the tremendous number of MTases, the main subfamily of AdoMet-dependent MTases catalyze the methylation reactions utilizing the methyl donor AdoMet (Martin and McMillan). To elucidate molecular mechanisms of catalysis and function for the three AdoMet-dependent MTases, TPMT, NNMT, and AS3MT, we have determined a set of TPMT complex structures (Chapter II-IV) and the structure of NNMT bound to AdoHcy and nicotinamide (Chapter V) by X-ray crystallography, and attempted to crystallize AS3MT (Chapter VI). In particular, we have answered several questions about the enzyme catalysis and function of TPMT and NNMT:

1) What is the structural basis of substrate recognition in TPMT and what residues are important for TPMT catalysis?

2) What is the molecular mechanism of TPMT polymorphisms, especially the unique TPMT*5 that exhibits low enzyme activity with a near wt protein level (Salavaggione *et al.*, 2005)?

3) How do different substrates bind to TPMT and what is the structural basis of the varied potency of different TPMT inhibitors?

4) What are the structural features which are important for substrate recognition and the catalytic mechanism of NNMT?

The objective of this chapter is to summarize the findings from the work reported here and to discuss the future directions of continuing studies.

7.1 Summary

1) What is the structural basis of substrate recognition in TPMT and what residues are important for TPMT catalysis?

As previously mentioned, TPMT is a cytosolic AdoMet-dependent MTase that inactivates thiopurine drugs such as 6MP, azathioprine and 6TG, which are used to treat cancer and other disorders (Wang and Weinshilboum, 2006; Cheok and Evans, 2006). Although TPMT has been recognized as an important enzyme in thiopurine metabolism for a long time, its natural substrate remains unknown. Prior to the study reported here, the structural features of the TPMT fold and AdoMet binding was characterized by the crystal structure of human TPMT bound to AdoHcy and the NMR structure of a bacterial TPMT orthologue (Scheuermann *et al.*, 2003; Wu *et al.*, 2007), but these structures didn't reveal the details of methyl acceptor substrate binding. The crystal structures of murine TPMT (mTPMT) bound to AdoHcy, with and without 6MP reported here provide the first detailed look at the structural features of 6MP binding to TPMT and identify residues involved in substrate binding and TPMT activity.

The crystal structures of both mTPMT complexes (mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP) confirmed the class I AdoMet-dependent MTase core fold (**Figure II-1a**) and similar binding mode of AdoMet with other class I MTase structures (**Figure II-1b; Figure II-2a, b**). More importantly, the mTPMTwt-AdoHcy-6MP structures revealed

that the relative positions and orientations of the bound 6MP and AdoHcy are consistent with a conserved S_N2-like mechanism, as seen in several other AdoMet-dependent MTases. Specifically, the 6MP-AdoHcy sulfur-sulfur distance (3.4-4.1 Å) is consistent with in-line positioning of a methyl group such that it can interact with the two sulfur atoms simultaneously (**Figure II-4**). In the mTPMTwt-AdoHcy-6MP structure, the weaker electron density for 6MP (**Figure II-1c, d**) relative to AdoHcy, together with the observation of a larger amount of buried accessible surface area and greater number of ligand-protein hydrogen bonds formed upon AdoHcy binding compared to 6MP binding are consistent with the observation that TPMT has substantially higher affinity for AdoMet (low micromolar) than for 6MP (low millimolar) (**Table II-2**). Moreover, modeling using the mTPMTwt-AdoHcy-6MP structure as a starting point also revealed that the active site is capable of accommodating the thiopurine metabolites which are significantly larger than 6MP, such as the nucleotides 6-thioinosine-5'-monophosphate (6TIMP) and 6-thioguanosine-5'-monophosphate (6TGMP) (Krynetski *et al.*, 1995).

In addition, the ternary complex structure identified Arg147 and Arg221 as the only hydrophilic residues within 4 Å of 6MP (**Figure II-2a, b**). Thus these two arginine residues may be involved in the deprotonation of 6MP, a critical step in methylation by AdoMet-dependent MTases. To test the possible functional roles of these two residues, subsequent structure-based mutagenesis of hTPMT was carried out. The partially compromised activity of the mutants suggested that either Arg152 (Arg147 in mTPMT) or Arg226 (Arg221 in mTPMT) may participate in the TPMT reaction with one residue compensating when the other is altered, but that neither is crucial for function.

Comparison of the mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP structures revealed an active site loop (Arg31-Gln55) that is flexible in the absence of 6MP and becomes ordered upon 6MP binding (**Figure II-2c, d**). In summary, the crystal structures of mTPMTwt-AdoHcy and mTPMTwt-AdoHcy-6MP elucidated the structural basis of substrate recognition in TPMT and identified Arg152 (Arg147 in mTPMT) and Arg226 (Arg221 in mTPMT) as residues which play a role in 6MP binding.

*2) What is the molecular mechanism of TPMT polymorphisms, especially the unique TPMT*5?*

TPMT is one of the best-studied examples of pharmacogenetics, in which genetic polymorphisms lead to individual variation in drug response. To date, at least 28 TPMT polymorphisms have been reported, most of which are associated with decreased enzyme activity (Salavaggione *et al.*, 2005; Schaeffeler *et al.*, 2006; Lindqvist *et al.*, 2007; Garat *et al.*, 2008; Tamm *et al.*, 2008). The study of 13 TPMT variants showed that nearly all had an enzyme activity which correlated with the amount of immunoreactive protein. The only exception is TPMT*5 (Leu49Ser in hTPMT and Leu44Ser in mTPMT), which exhibited very low enzyme activity but only modestly decreased protein level (Salavaggione *et al.*, 2005).

The crystal structures of mTPMT*5 bound to AdoHcy, with and without 6MP (mTPMT*5-AdoHcy and mTPMT*5-AdoHcy-6MP), have been determined. These mTPMT*5 crystal structures and the computational modeling of other TPMT variants using wt mTPMT structures gave us a structural understanding of the molecular

consequences of TPMT polymorphisms, which is that the protein level and enzyme activity of the hTPMT polymorphisms appear to correlate with whether the introduced residue is compatible with the native protein structure (**Figure III-7, Table III-3**).

In the mTPMT*5-SAH-6MP structure, the observation that 6MP is bound further away from the AdoHcy (the sulfur-sulfur distances of 4.9 Å, **Figure III-1d**) may also help explain the low enzyme activity of TPMT*5. Significantly, mTPMT*5 structures revealed a more flexible active site loop compared to that of wt mTPMT (**Figure III-3**). Comparison of the local conformations for the Leu44 in wt mTPMT and the Ser44 in mTPMT*5 structures revealed that the replacement of the hydrophobic Leu44 side chain (**Figure III-2a**) with the hydrophilic Ser44 (**Figure III-2b**) leads to conformational differences near and at the active site, thus compromising the AdoMet and more substantially the 6MP binding to TPMT, and causing the dramatic decrease in enzyme activity. Furthermore, the molecular dynamics simulation calculations of wt mTPMT, mTPMT*5 and modeled mTPMT*5 also confirmed that the *5 substitution is not compatible with the native wt mTPMT protein conformation (**Figure III-4**).

3) How do different substrates bind to TPMT and what is the structural basis of the varied potency of different TPMT inhibitors?

In an effort to discover potential inhibitors and new substrates of TPMT for experimental and clinical purposes, benzoic acid compounds and nonheterocyclic aromatic thiols were identified as potential inhibitors and substrates of TPMT, respectively (Woodson *et al.*, 1983; Ames *et al.*, 1986). The crystal structures of mTPMT with several benzoic acid

inhibitors, 3,4,5,-triiodobenzoic acid (mTPMTwt-AdoHcy-TIB), 4-iodobenzoic acid (mTPMTwt-AdoHcy-4IB), and 4-tert-butylbenzoic acid, and with a thiophenol substrate (mTPMTwt-AdoHcy-TP), reported here provides insight into the different acceptor substrates binding to TPMT and the different degree of inhibition by these benzoic acid inhibitors.

The two molecules in the mTPMTwt-AdoHcy-TP crystal structure showed that the thiophenols are bound in essentially the identical position and orientation in the active site (**Figure IV-2**). The thiophenol electron density is well-defined (**Figure IV-1c**), which is consistent with the observation that the thiophenol affinity is substantially higher than that of 6MP (Woodson *et al.*, 1983, Peng *et al.*, 2008). Comparison of the mTPMTwt-AdoHcy-TP and mTPMTwt-AdoHcy-6MP structures showed that both thiophenol and 6MP are bound to TPMT in the same positions, with the same van der Waals interactions.

In the inhibitor complexes, the mTPMTwt-AdoHcy-TIB structure displayed stronger density for 3,4,5,-triiodobenzoic acid than that for 4-iodobenzoic acid in the mTPMTwt-AdoHcy-4IB structure (**Figure IV-1a, b**). Additionally, co-crystallization and soaking of 4-tert-butylbenzoic acid didn't give any detectable electron density for 4-tert-butylbenzoic acid. These experimental findings are consistent with the relative inhibitory ability of the three compounds: 3,4,5,-triiodobenzoic acid is the best inhibitor while 4-tert-butylbenzoic acid is the weakest(**Table IV-2**) (Ames *et al.*, 1986). Further analysis of the mTPMTwt-AdoHcy-TIB and mTPMTwt-AdoHcy-4IB structures showed that the

3,4,5,-triiodobenzoic acid binding involves more interactions between its carboxylate and TPMT than seen for 4-iodobenzoic acid, and that the additional iodine substituents aid the orientation of the inhibitor (**Figure IV-3**). These structural details suggest that the flexibility of the side chains of a few residues on the active site loop (Arg31-Gln55) and nearby Glu220-Trp225 helix (Peng *et al.*, 2008, **Figure IV-4**) results in active site plasticity that allows for the binding of different benzoic acid inhibitors.

4) What are the structural features which are important for substrate recognition and the catalytic mechanism of NNMT?

NNMT is an AdoMet-dependent MTase that catalyzes the N-methylation of nicotinamide, pyridines and other structural analogs (Williams and Ramsden, 2005b). As previously described, NNMT plays a significant role in the regulation of metabolic pathways and cancers. Before the study reported here, the crystal structures of human and murine NNMT bound to AdoHcy had been deposited into Protein Data Bank (unpublished, PDB accession codes: 2IIP and 2I62). However, the crystal structure of NNMT bound to nicotinamide is still invaluable to elucidate structural features important for the catalytic mechanism and substrate recognition, in particular to identify residues that may be involved in catalysis and substrate binding and to provide a basis for structure-based mutagenesis and functional studies. The crystal structure of NNMT bound to both AdoHcy and nicotinamide reported here reveals details of the nicotinamide binding site and inspired structure-based mutagenesis to probe catalytic mechanism of NNMT.

In our NNMT complex structure, nicotinamide binds such that the nicotinamide-AdoHcy nitrogen-sulfur distances of 3.5-4.5 Å are consistent with an S_N2-like reaction mechanism, as seen in other AdoMet-dependent MTases (**Figure V-3**). The ternary complex structure also allows several NNMT active site residues (Ser201, Ser213 and Asp197, Tyr20, and Tyr242), which either directly interaction with nicotinamide or are in close proximity (**Figure V-2**), to be proposed as possible candidates for substrate deprotonation.

7.2 Future Directions

1) Structural-based mutagenesis to probe catalytic mechanism of NNMT

The crystal structure of NNMT bound to AdoHcy and nicotinamide presented here provides a structural basis for designed structure-function experiments. The structure provides the foundation for understanding NNMT catalysis through site-directed mutagenesis of NNMT of residues proposed as candidates for the substrate deprotonation step of NNMT catalysis. A number of mutants have been designed: 1) single mutants (Asp197Ala, Ser201Ala, Ser213Ala, Tyr20Ala and Tyr20Phe), double mutants (Ser201Ala/Ser213Ala, Ser201/Asp197, and Ser213Asp197) and a triple mutant (Ser201Ala/Ser213Ala/Asp197Ala) to probe the possible roles of Tyr20, Asp197, Ser201, and Ser213 in the nicotinamide deprotonation step; 2) single mutants (Tyr25Ala, and Tyr69Ala) and a double mutant (Tyr25Ala/Tyr69Ala) to investigate whether Tyr25 or Tyr69 modulate AdoHcy activation of Tyr20, a possible catalytic base; 3) single mutants Tyr242Ala and Tyr242Phe to test whether Tyr242 may be important in nicotinamide deprotonation; and 4) single mutants Tyr24Ala and Tyr204Ala to probe the importance of these residues and their hydrophobic interactions in nicotinamide binding. After

expression and purification of the mutated proteins, enzyme activity assays will allow evaluation of any changes in enzyme activity caused by the amino acid substitutions and expand our understanding of the enzyme's catalysic mechanism. In addition, the complex structure provides a scaffold that would be useful in the design of new NNMT inhibitors as possible leads for drug discovery.

2) The crystal structure of AS3MT

AS3MT is an important AdoMet-dependent MTase involved in methylation of inorganic arsenic. Our objective is to obtain the crystal structure of AS3MT for elucidation of the mechanism of arsenic methylation and to probe the relationship between AS3MT polymorphisms and individual variation in arsenic metabolism. As described in Chapter VI, wild-type human and murine AS3MT, and designed human AS3MT mutants, have been expressed and purified. Although the AS3MT mutants are designed to reduce surface entropy and thus to possibly improve crystallizability, crystallization trials of these various AS3MT constructs have yet to produce useful crystals. Considering the high yield, purity, and enzymatic activity of the AS3MT protein preparations, crystallization of AS3MT should be feasible. Efforts to crystallize AS3MT are ongoing, and as an alternative approach additional protein modifications will be explored.

One concern may be complications posed by the cysteine content of AS3MT. The human and mouse AS3MT sequences contain 14 and 12 cysteine residues, respectively. Such a large number of cysteines may have an effect on structural homogeneity, and thus the crystallization of AS3MT. From the sequence alignment of human and mouse

AS3MT (**Figure VII-1**), 12 of the cysteine residues are conserved in the two proteins. A previous study showed that rat AS3MT also has the 12 conserved cysteine residues (Li *et al.*, 2005). Among the dozen conserved AS3MT cysteine residues, four in the N-terminal region have been tested (residues 32, 61, 85, and 156 in rat). Among them, only the substitution of Cys156 with Ser (Cys157 in mouse) led to the loss of enzyme activity (Li *et al.*, 2005). Analysis of the human AS3MT computational homology model indicates that Cys85 and Cys226 (Cys86 and Cys227 in mouse) are located in α helices; Cys85 is buried while Cys226 is solvent-exposed. Therefore, the mutation of Cys85 may interfere with proper AS3MT folding while substitution of Cys226 would be tolerated. Cys61, Cys72 and Cys206 (Cys62, Cys73 and Cys207 in mouse) are located in surface loop regions and are solvent-exposed. Thus these residues are also targeted for mutagenesis in mouse AS3MT, to generate additional constructs for future AS3MT crystallization efforts.

In summary, structural studies of the AdoMet-dependent MTases reported here have not only revealed the structural basis of substrate recognition in TPMT and NNMT, but also identified residues important for their catalysis function. Specifically, the determination of TPMT*5 crystal structures has aided the understanding of TPMT polymorphisms. Furthermore, structural characterization of TPMT complexes with benzoic acid inhibitors provided structural insight into the varied potency of different TPMT inhibitors, and may provide useful information in identifying new small molecule modulators for optimization of thiopurine-based therapy. In addition, the structure-based mutagenesis of NNMT that is being pursued in ongoing studies will provide new insights into substrate recognition and the catalytic mechanism of NNMT. Significantly, the efforts to

crystallize AS3MT will have a high probability of yielding the crystal of AS3MT, which will be invaluable for the understanding of AS3MT catalysis and polymorphisms.

Figure VII-1 Sequence alignment of human and mouse AS3MT. The sequences were obtained from the NCBI GenBank: human AS3MT (Q9HBK9) and mouse AS3MT (AAH13468). This figure is prepared by ESPript (Gouet *et al.*, 1999).

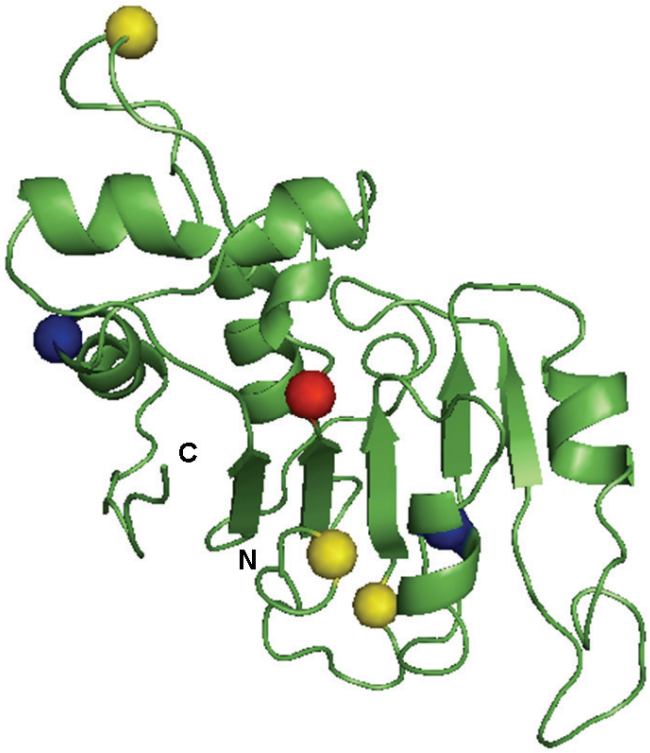


Figure VII-2 The computational homology model of human AS3MT. This model is only composed of residues 59-245, since it is restricted by the extending of sequence homology to the YcgJ protein from *Bacillus subitilis* template (PDB accession code 1XXL, unpublished). Cys156 (red) is a critical residue for enzyme activity. Cys85 and Cys226 (blue) are located in α helices. Cys61, Cys72 and Cys206 (yellow) are located in the loop regions and solvent-exposed. This model is prepared through ESyPred3D Web Server 1.0 mentioned in Chapter VI.

BIBLIOGRAPHY

Aksoy, S., Szumlanski, C. L., and Weinshilboum, R. M. (1994) Human liver nicotinamide N-methyltransferase. cDNA cloning, expression, and biochemical characterization. *J Biol Chem.* **269,** 14835-14840.

Ames, M. M., Selassie, C. D., Woodson, L. C., Van Loon, J. A., Hansch, C., and Weinshilboum, R. M. (1986) Thiopurine methyltransferase: structure-activity relationships for benzoic acid inhibitors and thiophenol substrates, *J. Med. Chem.* **29,** 354-358.

Anantharaman, V., Koonin, E. V., and Aravind, L. (2002) SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol.* **4,** 71-75.

Attieh, J. M., Hanson, A. D., and Saini, H. S. (1995) Purification and characterization of a novel methyltransferase responsible for biosynthesis of halomethanes and methanethiol in Brassica oleracea. *J Biol Chem*. **270,** 9250-9257.

Bekker, H., Berendsen, H. J. C., Dijkstra, E. J., Achterop, S., van Drunen, R., van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., and Renardus, M. K. R. (1993) Gromacs: A parallel computer for molecular dynamics simulations. In Physics Computing 92 Singapore,.de Groot, R. A., Nadrchal, J., eds. World Scientific.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren,W. F., and Hermans, J. (1981) Interaction models for water in relation to protein hydration. In: Intermolecular Forces. Pullman, B. ed. . D. Reidel Publishing Company Dordrecht, 331–342.

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **91,** 43–56.

Bergfors, T. M. (1999) Protein crystallization: techniques, strategies, and tips: a laboratory manual. International University Line, La Jolla, CA, 4-5.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4,** 187-217.

Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) Crystallographic R Factor Refinement by Molecular Dynamics. *Science.* **235,** 458-460.

Brunger, A. T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472-474.

Brunger, A. T. (1993) Assessment of phase accuracy by cross validation: The free R value. *Acta Cryst. D* **49**, 24-36.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D* **54,** 905-921.

Cantoni, G. L. (1975) Biological methylation: selected aspects. *Annu. Rev. Biochem.* **44,** 435-451.

Case, D. A., and Karplus, M. (1979) Dynamics of ligand binding to heme proteins. *J Mol Biol.* **132,** 343-368.

Chayen, N. E. (2005). Methods for separating nucleation and growth in protein crystallization. *Prog. Biophys. Mol. Biol.* **88**, 329–337.

Cheng, Q., Yang, W., Raimondi, S. C., Pui, C. H., Relling, M. V., and Evans, W. E. (2005) Karyotypic abnormalities create discordance of germline genotype and cancer cell phenotypes. *Nat Genet.* **37**,878-882.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. (1993) Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell.* **74,** 299-307.

Cheng, X., and Blumenthal, R. M. (1999) S-Adenosylmethionine-Dependent Methyltransferase: Structures and Functions. World Scientific Publishing, Singapore.

Cheok, M. H., and Evans, W. E. (2006) Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy. *Nature Rev.Cancer.* ***6,*** 117-129.

Chernov, A. A. (2003). Protein crystals and their growth. *J. Struct. Biol.* **142**, 3–21.

Kendrew, J. C., Bodo,G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by X-Ray analysis. *Nature* **181**, 662–666.

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., and McCann, P. P. (1996) S-Adenosylmethionine and methylation. *FASEB J.* **10,** 471-480.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz JR., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *Journal of the American Chemical Society.* **117,** 5179–5197.

Coulthard, S., and Hogarth, L. (2005) The thiopurines: an update. *Invest New Drugs.* **23,** 523-532.

Cournoyer, B., Watanabe, S., and Vivian, A. (1998) A tellurite-resistance genetic determinant from phytopathogenic pseudomonads encodes a thiopurine methyltransferase: evidence of a widely-conserved family of methyltransferases. *Biochim Biophys Acta.* **1397**,161-168.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., and McPherson, A. (1994) Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr.* **50**, 414-423.

Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615-W619.

Deininger, M., Szumlanski, C. L., Otterness, D. M., Van Loon, J., Ferber, W., and Weinshilboum, R. M. (1994) Purine substrates for human thiopurine methyltransferase. *Biochem Pharmacol.* **48**, 2135-2138.

DeLano, W. L. (2004) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA. (www.pymol.org)

Dixon, M. M., Huang, S., Matthews, R. G., and Ludwig, M. (1996) The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B_{12}. *Structure.* **4,** 1263-1275.

Dodson, G. G., Lane, D. P., and Verma, C. S. (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Rep.* **9,** 144-150.

Drenth, J. (1994) Principles of protein X-ray crystallography, Springer-Verlag, New York.

Drobna, Z., Waters, S. B., Walton, F. S., LeCluyse, E. L., Thomas, D. J. and Styblo, M. (2004) Interindividual variation in the metabolism of arsenic in cultured primary human hepatocytes. *Toxicol Appl Pharmacol.* **201**, 166-177.

Drobna, Z., Waters, S. B., Devesa, V., Harmon, A. W., Thomas, D. J., and Styblo, M. (2005) Metabolism and toxicity of arsenic in human urothelial cells expressing rat arsenic (+3 oxidation state)-methyltransferase. *Toxicol Appl Pharmacol.* **207,** 147-159.

Drobna, Z., Xing, W., Thomas, D. J., and Styblo, M.(2006) shRNA silencing of AS3MT expression minimizes arsenic methylation capacity of HepG2 cells. *Chem Res Toxicol.* **19**, 894-898.

Ealick, S. E. (2000) Advances in multiple wavelength anomalous diffraction crystallography. *Current opinion in chemical biology* **4,** 495-499.

Eichelbaum, M., Ingelman-Sundberg, M., and Evans, W. E. (2006) Pharmacogenomics and individualized drug therapy. *Annu Rev Med.* **57,** 119-137.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D* **60**, 2126-2132.

Fontecave, M., Atta, M., and Mulliez, E. (2004) S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci.* **29,** 243-249.

Garat, A., Cauffiez, C., Renault, N., Lo-Guidice, J. M., Allorge, D., Chevalier, D., Houdret, N., Chavatte, P., Loriot, M. A., Gala, J. L., and Broly, F. (2008) Characterisation of novel defective thiopurine S-methyltransferase allelic variants. *Biochem Pharmacol.* **76,** 404-415.

Gee, C. L., Drinkwater, N., Tyndall, J. D., Grunewald, G. L., Wu, Q., McLeish, M. J., and Martin, J. L. (2007) Enzyme adaptation to inhibitor binding: a cryptic binding site in phenylethanolamine N-methyltransferase. *J Med Chem.* **50,** 4845-4853.

Gilliland, G. L., and Ladner, J. E. (1996) Crystallization of biological macromolecules for X-ray diffraction studies. *Curr Opin Struct Biol.* **6,** 595-603.

Glusker, J. P., Lewis, M., and Rossi, M. (1994) Crystal structure analysis for chemists and biologists. VCH Publishers,Inc, New York.

Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics*. **15,** 305-308.

Grillo, M. A., and Colombatto, S. (2008) S-adenosylmethionine and its products. *Amino Acids*. **34,** 187-193.

Hall, P. R. (2004) The transcarboxylase multienzyme complex: structural studies of the components of a 1.2 million-dalton carboxylation machine. Degree: Doctor of Philosophy, Biochemistry, Case Western Reserve University.

Hamacher, K., Trylska, J., and McCammon, J. A. (2006) Dependency map of proteins in the small ribosomal subunit. *PLoS Comput Biol.* **2**, e10.

Hartford, C., Vasquez, E., Schwab, M., Edick, M. J., Rehg, J. E., Grosveld, G., Pui, C. H., Evans. W. E., and Relling, M. V. (2007) Differential effects of targeted disruption of thiopurine methyltransferase on mercaptopurine and thioguanine pharmacodynamics. *Cancer Res.* **67**, 4965-4972.

Harvey, S. C., Prabhakaran, M., Mao, B., and McCammon, J. A. (1984) Phenylalanine transfer RNA: molecular dynamics simulation. *Science.* **223,** 1189-1191.

Hauptman, H. (1997) Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol.* **7**, 672-680.

Helliwell, J. R. (2005) Protein crystal perfection and its application. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 793-798.

Hendrickson, W., and Ogata, C. (1997) Phase determination from multiwavelength anomalous diffraction measurements. *Meth Enzymol,* **276,** 494-523.

Hernandez, J. S., Van Loon, J. A., Otterness, D. M., Guerciolini, R., and Weinshilboum, R. M. (1991) Thiopurine methyltransferase regulation in rat kidney: immunoprecipitation studies. *Xenobiotica*. **21**, 451-459.

Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol.* **233,** 123-138.

Honchel, R., Aksoy, I. A., Szumlanski, C., Wood, T. C., Otterness, D. M., Wieben, E. D., and Weinshilboum R. M. (1993) Human thiopurine methyltransferase: molecular cloning and expression of T84 colon carcinoma cell. *Mol Pharmacol.* **43,** 878-887.

Horton, J. R., Sawada, K., Nishibori, M., and Cheng, X. (2005) Structural basis for inhibition of histamine N-methyltransferase by diverse drugs, *J. Mol. Biol.* **353,** 334-344.

Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M., and Takusagawa, F. (2000) Mechanisms for auto-inhibition and forced product release in glycine N-methyltransferase: crystal structures of wild-type, mutant R175K and S-adenosylhomocysteine-bound R175K enzymes. *J Mol Biol.* **298,** 149-162.

Hughes, M. F. (2002) Arsenic toxicity and potential mechanisms of action. *Toxicol Lett.* **133**, 1-16.

Jacobs, S. A., Harp, J. M., Devarakonda, S., Kim, Y., Rastinejad, F., and Khorasanizadeh, S. (2002) The active site of the SET domain is constructed on a knot. *Nat Struct Biol.* **9,** 833-838.

Jancarik, J., and Kim, S-H. J. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Appl. Cryst.* **24**, 409-411.

Jeruzalmi, D. (2006) First analysis of macromolecular crystals: biochemistry and x-ray diffraction. *Methods Mol. Biol.* **364,** 43-62**.**

John, J. P., Oh, J. E., Pollak, A., and Lubec, G. (2008) Identification and characterisation of arsenite (+3 Oxidation State) methyltransferase (AS3MT) in mouse neuroblastoma cell line N1E-115. *Amino Acids*. **35,** 355-338.

Jorgensen,W. L., Chandrasekhar, J., Madura, J. D., Impey, R.W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.

Joseph, D., Petsko, G. A., and Karplus, M. (1990) Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science.* **249,** 1425-1428.

Karplus, M., and Petsko, G. A. (1990) Molecular dynamics simulations in biology. *Nature.* **347,** 631-639.

Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nat Struct Biol.* **9,** 646-652.

Karplus, M., and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci U S A.* **102,** 6679-6685.

Kazmirski, S. L., Zhao, Y., Bowman, G. D., O'donnell, M., and Kuriyan, J. (2005) Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic

and archaeal proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A.* **102,** 13801-13806.

Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 484-491.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell.* **76,** 357-369.

Korolev, S., Ikeguchi, Y., Skarina, T., Beasley, S., Arrowsmith, C., Edwards, A., Joachimiak, A., Pegg, A. E., and Savchenko, A. (2002) The crystal structure of spermidine synthase with a multisubstrate adduct inhibitor, *Nat. Struct. Biol.* **9,** 27-31.

Kozbial, P. Z., and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins. *BMC Struct Biol*. **5,**19.

Krynetski, E. Y., Schuetz, J. D., Galpin, A. J., Pui, C. H., Relling, M. V., and Evans, W. E. (1995a) A single point mutation leading to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc Natl Acad Sci U S A.* **92,** 949-953.

Krynetski, E. Y., Krynetskaia, N. F., Yanishevski, Y., and Evans, W. E. (1995b) Methylation of mercaptopurine, thioguanine, and their nucleotide metabolites by heterologously expressed human thiopurine S-methyltransferase, *Molec. Pharmacol.* **47,** 1141-1147.

Krynetski, E. Y., and Evans, W. E. (1999) Parmacogenetics as a molecular basis for individualized drug therapy: the thiopurine S-methyltransferase paradigm. *Pharm Res.* **16**, 342-349.

Krynetski, E., and Evans, W. E. (2003) Drug methylation in cancer therapy: lessons from the TPMT polymorphism. *Oncogene.* **22**, 7403-7413.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* **26,** 283-291.

Lee, H. J., Srinivasan, D., Coomber, D., Lane, D. P., and Verma, C. S. (2007) Modulation of the p53-MDM2 interaction by phosphorylation of Thr18: a computational study. *Cell Cycle.* **6,** 2604-2611.

Lennard, L., Van Loon, J. A., Lilleyman, J. S., and Weinshilboum, R. M. (1987) Thiopurine pharmacogenetics in leukemia: correlation of erythrocyte thiopurine methyltransferase activity and 6-thioguanine nucleotide concentrations. *Clin Pharmacol Ther.* **41,** 18-25.

Lennard, L., Lilleyman, J. S., Van Loon, J., and Weinshilboum, R. M. (1990) Genetic variation in response to 6-mercaptopurine for childhood acute lymphoblastic leukaemia. *Lancet.* **336,** 225-229.

Lennard, L. (1992) The clinical pharmacology of 6-mercaptopurine. *Eur. J. Clin. Pharmacol.* ***43*****,** 329-339.

Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995) Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Computer Physics Communications* **91,** 215–231.

Levitt, M. (2001) The birth of computational structural biology. *Nat Struct Biol.* **8,** 392-393.

Li J, Waters SB, Drobna Z, Devesa V, Styblo M, Thomas DJ. (2005) Arsenic (+3 oxidation state) methyltransferase and the inorganic arsenic methylation phenotype. *Toxicol Appl Pharmacol.* **204,** 164-169.

Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J., Howard, A., Eisenstein, E., and Herzberg, O. (2003) Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot. *Proteins.* **51,** 56-67.

Lin, S., Shi, Q., Nix, F. B., Styblo, M., Beck, M. A. (2002) Herbin-Davis, K. M., Hall, L. L., Simeonsson, J. B. and Thomas, D. J., A novel S-adenosyl-L-methionine: arsenic(III) methyltransferase from rat liver cytosol. *J Biol Chem.* **277**: 10795-10803.

Lindahl, E., Hess, B., and van der Spoel, D. (2001) Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7,** 306–317.

Lindqvist, M., Skoglund, K., Karlgren, A., Söderkvist, P., Peterson, C., Kidhall, I., and Almer, S. (2007) Explaining TPMT genotype/phenotype discrepancy by haplotyping of TPMT*3A and identification of a novel sequence variant, TPMT*23. *Pharmacogenet Genomics.* **17**, 891-895.

Liu, H., M¨uller-Plathe, F., and van Gunsteren, W. F. (1995) A force field for liquid dimethyl sulfoxide and liquid proporties of liquid dimethyl sulfoxide calculated using molecular dynamicssimulation. *J. Am. Chem. Soc.* **117,** 4363–4366.

Lu, S. C. (2000) S-adenosylmethionine. *Int J Biochem Cell Biol.* **32,** 391-395.

MacKerell, JR. A. D., Brooks, B., Brooks, C. L. III, Nilsson, L., Roux, B., Won, Y., and Karplus M. (1998) CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, in The Encyclopedia of Computational Chemistry, 1, 271-277. P. v. R. Schleyer et al., editors, John Wiley & Sons: Chichester.

Mark, A. E., van Helden, S. P., Smith, P. E., Janssen, L. H. M., and van Gunsteren, W. F. (1994) Convergence properties of free energy calculations: α-cyclodextrin complexes as a case study. *J. Am. Chem. Soc.* **116,** 6293–6302.

Martin, J. L., Begun, J., McLeish, M. J., Caine J. M., and Grunewald, G. L. (2001) Getting the adrenaline going: crystal structure of the adrenaline-synthesizing enzyme PNMT, *Structure* **9,** 977-985.

Martin, J. L., and McMillan F. M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol.* **12**, 783-793.

Mass, M. J., Tennant, A., Roop, B. C., Cullen, W. R., Styblo, M., Thomas, D. J., and Kligerman, A. D. (2001) Methylated trivalent arsenic species are genotoxic. *Chem Res Toxicol.* **14,** 355-356.

Mato, J. M., Alvarez, L., Ortiz, P., and Pajares, M. A. (1997) S-adenosylmethionine synthesis: Molecular mechanisms and clinical implications. *Pharmacol. Ther.* **73,** 265-280.

McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins. *Nature.* **267,** 585-590.

McCarthy, A. A., and McCarthy, J. G. (2007) The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.* **144,** 879-889.

McLeod, H. L., Krynetski, E. Y., Relling, M. V., and Evans, W. E. (2000) Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia.* **14,** 567-572.

Meza, M. M., Yu, L., Rodriguez, Y. Y., Guild, M., Thompson, D., Gandolfi, A. J., and Klimecki, W. T. (2005) Developmentally restricted genetic determinants of human arsenic metabolism: association between urinary methylated arsenic and CYT19 polymorphisms in children. *Environ Health Perspect.* **113**,775-781.

Michel, G., Sauvé, V., Larocque, R., Li, Y., Matte, A., and Cygler M. (2002) The structure of the RlmB 23S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure.* **10,**1303-1315.

Min, J., Zhang, X., Cheng, X., Grewal, S. I., and Xu, R. M. (2002) Structure of the SET domain histone lysine methyltransferase. *Nat Struct Biol.* **9,** 828-832.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J.(1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D* **53,** 240-255.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., and Karplus, M. (1986) Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5' d(C-G-T-A-C-G)$_2$. *J Mol Biol.* **188,** 455-475.

Norberg, J., and Nilsson, L. (2003) Advances in biomolecular simulations: methodology and recent applications. *Q Rev Biophys.* **36,** 257-306.

Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D. G., Shibata, T., Inoue, Y., Kuramitsu, S., and Yokoyama, S. (2002) An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallogr D Biol Crystallogr.* **58,**1129-1137.

Ohsawa, N., Tsujita, M., Morikawa, S., and Itoh, N. (2001) Purification and characterization of a monohalomethane-producing enzyme S-adenosyl-L-methionine:

halide ion methyltransferase from a marine microalga, Pavlova pinguis. *Biosci Biotechnol Biochem.* **65,** 2397-2404.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276**, 307-326.

Paterson, A. R. P., and Tidd, D. M. (1975) 6-thiopurines (Springer Verlag, New York).

Peng, Y., Feng, Q., Wilk, D., Adjei, A. A., Salavaggione, O. E., Weinshilboum, R. M., and Yee, V. C. (2008) Structural basis of substrate recognition in thiopurine s-methyltransferase. *Biochemistry*. **47,** 6216-6225.

Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 1718-1725.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L, and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry.* **26,** 1781-1802.

Ranjard, L., Prigent-Combaret, C., Nazaret, S., and Cournoyer, B. (2002) Methylation of inorganic and organic selenium by the bacterial thiopurine methyltransferase. *J Bacteriol.* **184**, 3146-3149.

Relling, M.V, and Dervieux, T. (2001) Pharmacogenetics and cancer therapy. *Nat Rev Cancer.* **1,** 99-108.

Remy, C. N. (1963) Metabolism of thiopyrimidines and thiopurines. S-Methylation with S-adenosylmethionine transmethylase and catabolism in mammalian tissues. *J Biol Chem.* **238,** 1078-1084.

Rod, T. H., Radkiewicz, J. L., and Brooks, C. L. 3rd. (2003) Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc Natl Acad Sci U S A.* **100,** 6980-6985.

Rupp, B., and Wang, J. (2004). Predictive models for protein crystallization. *Methods* **34**, 390–407.

Rutherford, K., and Daggett, V. (2008) Four human thiopurine s-methyltransferase alleles severely affect protein structure and dynamics. *J Mol Biol.* **379,** 803-814.

Sahasranaman, S., Howard, D., and Roy, S. (2008) Clinical pharmacology and pharmacogenetics of thiopurines *Eur J Clin Pharmacol.* **64,** 753-767.

Salavaggione, O. E., Wang, L., Wiepert, M., Yee, V. C., and Weinshilboum, R. M. (2005) Thiopurine S-methyltransferase pharmacogenetics: variant allele functional and comparative genomics. *Pharmacogenet Genomics.* **15**, 801-815.

Sanz, M. A., Fenaux, P., Lo Coco, F., and European APL Group of Experts (2005) Arsenic trioxide in the treatment of acute promyelocytic leukemia. A review of current evidence. *Haematologica* **90,** 1231-1235.

Sartini, D., Muzzonigro, G., Milanese, G., Pierella, F., Rossi, V., and Emanuelli, M. (2006) Identification of nicotinamide N-methyltransferase as a novel tumor marker for renal clear cell carcinoma. *J Urol.* **176**, 2248-2254.

Sartini, D., Santarelli, A., Rossi, V., Goteri, G., Rubini, C., Ciavarella, D., Lo Muzio, L., and Emanuelli, M. (2007) Nicotinamide N-methyltransferase upregulation inversely correlates with lymph node metastasis in oral squamous cell carcinoma. *Mol Med.* **13,** 415-421.

Sasaki, T., Goto, E., Konno, Y., Hiratsuka, M., and Mizugaki, M. (2006) Three novel single nucleotide polymorphisms of the human thiopurine s-methyltransferase gene in Japanese individuals. *Drug Metab Pharmacokinet.* **21**:332-336.

Schaeffeler, E., Eichelbaum, M., Reinisch, W., Zanger, U. M., and Schwab, M. (2006) Three novel thiopurine S-methyltransferase allelic variants (TPMT*20, *21, *22) - association with decreased enzyme function. *Hum Mutat.* **27**, 976.

Scheuermann, T. H., Lolis, E., and Hodsdon, M. E. (2003) Tertiary structure of thiopurine methyltransferase from Pseudomonas syringae, a bacterial orthologue of a polymorphic, drug-metabolizing enzyme, *J. Mol. Biol.* **333,** 573-585.

Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C., and Warren, M. J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nat Struct Biol.* **5,** 585-592.

Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003a) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci.* **28,** 329-335.

Schubert, H. L., Phillips, J. D., and Hill, C. P. (2003b) Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase, *Biochemistry* **42,** 5592-5599.

Schuettelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG – a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* **D60,** 1355-1363.

Scott, W. R. P., Hunenberg, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Kruger, P., and van Gunsteren, W. F. (1999) The GROMOS

biomolecular simulation program package. *Journal of Physical Chemistry (A)* **103,** 3596–3607.

Simeonova, P. P., and Luster, M. I. (2004) Arsenic and atherosclerosis. *Toxicol Appl Pharmacol.* **198,** 444-449.

Somani, S., Chng, C. P., and Verma, C. S. (2007) Hydration of a hydrophobic cavity and its functional role: a simulation study of human interleukin-1β. *Proteins.* **67,** 868-885.

Styblo, M., Del Razo, L. M., Vega, L., Germolec, D. R., LeCluyse, E. L., Hamilton, G. A., Reed, W., Wang, C., Cullen, W. R., and Thomas, D. J. (2000) Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells. *Arch Toxicol.* **74,** 289-299.

Styblo, M., Drobna, Z., Jaspers, I., Lin, S. and Thomas, D. J. (2002) The role of biomethylation in toxicity and carcinogenicity of arsenic: a research update. *Environ Health Perspect*. **110**: 767-771.

Szumlanski, C., Otterness, D., Her, C., Lee, D., Brandriff, B., Kelsell, D., Spurr, N., Lennard, L., Wieben, E., and Weinshilboum, R. (1996) Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. *DNA Cell Biol.* **15,** 17-30.

Tai, H.L., Krynetski, E.Y., Yates, C.R., Loennechen, T., Fessing, M.Y., Krynetskaia, N.F., and Evans, W.E. (1996) Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *Am. J. Hum. Genet.* 58, 694-702.

Tai, H. L., Krynetski, E. Y., Schuetz, E. G., Yanishevski, Y., and Evans, W. E. (1997) Enhanced proteolysis of thiopurine S-methyltransferase (TPMT) encoded by mutant

alleles in humans (TPMT*3A, TPMT*2): mechanisms for the genetic polymorphism of TPMT activity. *Proc Natl Acad Sci U S A* **94**, 6444-6449.

Tai, H. L., Fessing, M. Y., Bonten, E. J., Yanishevsky, Y., d'Azzo, A., Krynetski, E. Y., and Evans, W. E. (1999) Enhanced proteasomal degradation of mutant human thiopurine S-methyltransferase (TPMT) in mammalian cells: mechanism for TPMT protein deficiency inherited by TPMT*2, TPMT*3A, TPMT*3B or TPMT*3C. *Pharmacogenetics*. **9**, 641-650.

Takamori, S., Holt, M., Stenius, K., Lemke, E. A., Gronborg, M., Riedel, D., Urlaub, H., Schenck, S., Brugger, B., Ringler, P., Muller, S. A., Rammner, B., Grater, F., Hub, J. S., De Groot, B. L., Mieskes, G., Moriyama, Y., Klingauf. J., Grubmuller, H., Heuser, J., Wieland, F., and Jahn, R. (2006) Molecular anatomy of a trafficking organelle. *Cell.* **127,** 831-846.

Tama, F., and Brooks, C. L. (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct.* **35,** 115-133.

Tamm, R., Oselin, K., Kallassalu, K., Magi, R., Anier, K., Remm, M., and Metspalu, A. (2008) Thiopurine S-methyltransferase (TPMT) pharmacogenetics: three new mutations and haplotype analysis in the Estonian population. *Clin Chem Lab Med.* **46**, 974-979.

Taylor, G. (2003) The phase problem. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1881-1890.

Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr.* **D 55***,* 849-861.

Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Crystallogr.* **D 56,** 965-972.

Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Crystallogr.* **D 59**, 38-44.

Thomas, D. J., Styblo, M., and Lin, S. (2001) The cellular metabolism and systemic toxicity of arsenic. *Toxicol Appl Pharmacol.* **176,** 127-144.

Thomas, D. J., Waters, S. B. and Styblo, M. (2004) Elucidating the pathway for arsenic methylation. *Toxicol Appl Pharmacol*. **198**, 319-326.

Thomas, D. J., Li, J., Waters, S. B., Xing, W., Adair, B. M., Drobna, Z., Devesa, V., and Styblo. M. (2007) Arsenic (+3 oxidation state) methyltransferase and the methylation of arsenicals. *Exp Biol Med (Maywood).* **232**, 3-13

Trievel, R. C., Beach, B. M., Dirk, L. M, Houtz, R. L., and Hurley, J. H. (2002) Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell.***111**, 91-103.

Usón, I., and Sheldrick, G. M. (1999) Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643-648.

Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30**, 1022-1025.

Vaguine, A. A., Richelle, J., and Wodak,S.J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D. Biol. Crystallogr.* **55**, 191-205.

Vahter, M. (2000) Genetic polymorphism in the biotransformation of inorganic arsenic and its role in toxicity. *Toxicol Lett.* **112-113**, 209-217.

van Buuren, A. R., and Berendsen, H. J. C. (1993) Molecular dynamics simulation of the stability of a 22 residue alpha-helix in water and 30 % trifluoroethanol. *Biopolymers* **33,** 1159–1166.

van Buuren, A. R., Marrink, S. J., and Berendsen, H. J. C. (1993) A molecular dynamics study of the decane/water interface. *J. Phys. Chem.* **97,** 9206–9212.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, Flexible and Free. *J. Comp. Chem.* **26,** 1701–1718.

van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., and Berendsen, H. J. C. (2005) Gromacs User Manual version 3.3, www.gromacs.org.

van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamy. *J. Mol. Biol.* **229,** 105-124.

van Gunsteren, W. F., and Berendsen, H. J. C. (1987) Gromos-87 manual. Biomos BV Nijenborgh 4, 9747 AG Groningen, The Netherlands.

Vasken Aposhian, H. (1997) Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annu Rev Pharmacol Toxicol.* **37**, 397-419.

Vasken Aposhian, H., Zakharyan, R. A., Avram, M. D., Sampayo-Reyes, A. and Wollenberg, M. L., (2004) A review of the enzymology of arsenic metabolism and a new potential role of hydrogen peroxide in the detoxication of the trivalent arsenic species. *Toxicol Appl Pharmacol.* **198**, 327-335.

Vidgren, J., Svensson, L. A., and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase. *Nature.* **368,** 354-358.

Walsh, M. A., Evans, G., Sanishvili, R., Dementieva, I., and Joachimiak, A. (1999) MAD data collection - current trends. *Acta Crys.* **D55,** 1726-1732.

Wang, L., Sullivan, W., Toft, D., and Weinshilboum, R. (2003) Thiopurine S-methyltransferase pharmacogenetics: chaperone protein association and allozyme degradation. *Pharmacogenetics.*. **13**, 555-564.

Wang, L., Nguyen, T. V., McLaughlin, R. W., Sikkink, L. A., Ramirez-Alvarado, M., and Weinshilboum, R. M. (2005) Human thiopurine S-methyltransferase pharmacogenetics: variant allozyme misfolding and aggresome formation. *Proc Natl Acad Sci U S A*. **102**, 9394-9399.

Wang, L., and Weinshilboum, R. (2006) Thiopurine S-methyltransferase pharmacogenetics: insights, challenges and future directions. *Oncogene*. **25,** 1629-1638.

Watney, J. B., Agarwal, P. K., and Hammes-Schiffer, S. (2003) Effect of mutation on enzyme motion in dihydrofolate reductase. *J Am Chem Soc.* **125,** 3745-3750.

Watters, J. W., and McLeod, H. L. (2003) Cancer pharmacogenomics: current and future applications. *Biochim Biophys Acta.* **1603**, 99-111.

Weinshilboum, R. M., Raymond, F. A., and Pazmiño, P. A. (1978) Human erythrocyte thiopurine methyltransferase: radiochemical microassay and biochemical properties. Clin Chim Acta. 85, 323-333.

Weinshilboum, R. M., and Sladek, S. (1980) Mecaptopurine pharmacogenetics: monogenetic inheritance of erythrocyte thiopurine methyltransferase activity. *Am J Hum Genet.* **32**, 651-662.

Weinshilboum, R. (2001) Thiopurine pharmacogenetics: clinical and molecular studies of thiopurine methyltransferase. *Drug Metab Dispos.* **29**, 601-605.

Weinshilboum, R., and Wang, L. (2004a) Pharmacogenetics: inherited variation in amino acid sequence and altered protein quantity. *Clin Pharmacol Ther*. **75,** 253-258.

Weinshilboum, R., and Wang, L. (2004b) Pharmacogenomics: bench to bedside. *Nat Rev Drug Discov.* **3,** 739-748.

Weinshilboum, R. M., and Wang, L. (2006) Pharmacogenetics and pharmacogenomics: development, science, and translation. *Annu Rev Genomics Hum Genet.* **7,** 223-245.

Williams, A. C., Cartwright, L. S., and Ramsden, D. B. (2005) Parkinson's disease: the first common neurological disease due to auto-intoxication? *QJM.* **98,** 215-226.

Williams, A. C., and Ramsden, D. B. (2005a) Nicotinamide homeostasis: a xenobiotic pathway that is key to development and degenerative diseases. *Med Hypotheses.* **65,** 353-362.

Williams, A. C., and Ramsden, D. B. (2005b) Autotoxicity, methylation and a road to the prevention of Parkinson's disease. *J Clin Neurosci.* **12,** 6-11.

Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J., and Xiao, B. (2002) Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell.* **111,** 105-115

Wood, T. C., Salavagionne, O. E., Mukherjee, B., Wang, L., Klumpp, A. F., Thomae, B. A., Eckloff, B. W., Schaid, D. J., Wieben, E. D., and Weinshilboum, R. M. (2006) Human arsenic methyltransferase (AS3MT) pharmacogenetics: gene resequencing and functional genomics studies. *J Biol Chem.* **281**, 7364-7373.

Woodard, R. W., Tsai, M. D., Floss, H. G., Crooks, P. A., and Coward, J. K. (1980) Stereochemical course of the transmethylation catalyzed by catechol O-methyltransferase, *J. Biol. Chem.* **255,** 9124-9127.

Woodson, L. C., and Weinshilboum, R. M. (1983) Human kidney thiopurine methyltransferase. Purification and biochemical properties. *Biochem Pharmacol.* **32,** 819-826.

Woodson, L. C., Ames, M. M., Selassie, C. D., Hansch, C., and Weinshilboum, R. M. (1983) Thiopurine methyltransferase. Aromatic thiol substrates and inhibition by benzoic acid derivatives, *Mol. Pharmacol.* **24,** 471-478.

Wu, H., Horton, J. R., Battaile, K., Allali-Hassani, A., Martin, F., Zeng, H., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A. N., and Cheng, X. (2007) Structural basis of allele variation of human thiopurine-S-methyltransferase, *Proteins : Struct., Funct., and Bioinformatics.* **67,** 198-208.

Yoshida, T., Yamauchi, H., and Fan Sun, G. (2004) Chronic health effects in people exposed to arsenic via the drinking water: dose-response relationships in review. *Toxicol Appl Pharmacol*. **198,** 243-252.

Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., and Cheng, X. (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell.***111,** 117-127.

Zubieta, C., He, X. Z., Dixon, R. A., and Noel, J. P. (2001) Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nat Struct Biol.* **8,** 271-279.

Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J. P. (2003) Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family, *Plant Cell* **15,** 1704-1716.