Thiopurine S-methyltransferase

characterization of variants and ligand binding

Annica Blissing

Division of Chemistry
Department of Physics, Chemistry and Biology
Linköping University, SE-581 83 Linköping, Sweden

Linköping 2017

Copyright © Annica Blissing, 2017

Thiopurine S-methyltransferase
characterization of variants and ligand binding

ISBN 978-91-7685-530-0

ISSN 0280-7971

Distributed by:

Division of Chemistry

Department of Physics, Chemistry and Biology

Linköping University

SE-581 83 Linköping, Sweden

Printed in Sweden by LiU-Tryck, Linköping 2017.

*Thanks to supervisors, colleagues, friends, and family,
who have helped in the making of this thesis*

”

Nothing is impossible! Not if you can imagine it.
That's what being a scientist is all about!

Professor Hubert J. Farnsworth

Populärvetenskaplig sammanfattning

Proteinet TPMT och tiopurinläkemedel

I kroppens celler finns proteinet tiopurinmetyltransferas, vilket vanligtvis förkortas TPMT. Trots att det är vanligt förekommande är proteinets naturliga funktion fortfarande okänd. Vi vet dock att TPMT inaktiverar en viss typ av läkemedel som kallas tiopuriner, därav proteinets namn. Tiopuriner är en typ av cytostatika (cellgifter) som använd vid behandling av sjukdomstillstånd relaterade till ett överaktivt immunförsvar och vissa former av leukemi. Tiopurinläkemedel är förstklassiga imitatörer; de efterliknar nämligen byggstenarna till vårt DNA. Cellen kan inte skilja imitator från original utan bygger in tiopurinerna i nytt DNA. DNA som innehåller tiopuriner är odugligt, vilket gör att cellen till slut dör. Eftersom cellen kopierar sitt DNA före delning så kommer tiopurinläkemedel i större utsträckning påverka celler som delar sig ofta, och man får därmed en riktad effekt mot överaktiva immunceller och cancerceller.

TPMT komplicerar läkemedelsbehandling

Proteinet TPMT inaktiverar tiopurinläkemedel, och det får tyvärr konsekvenser för patienter som behandlas med dessa substanser. Inaktiveringen sker genom att proteinet modifierar tiopurinerna genom så kallad metylering. De modifierade läkemedelsmolekylerna liknar inte längre de naturliga DNA-baserna, och läkemedlets celldödande effekt minskar därmed.

För att kompensera för TPMTs härjningar måste läkemedelsdosen höjas så att tillräcklig effekt uppnås. Dock finns det flertalet naturligt förekommande varianter av proteinet, vilket ytterligare komplicerar behandlingen. Att det finns olika varianter av ett protein är en naturlig

del av den genetiska variationen i en population, men alla TPMT-varianterna inte är lika benägna att inaktivera tiopurinläkemedel. Vissa varianter är ambitiösa och inaktiverar en större mängd tiopuriner – de är alltså högaktiva. Andra är betydligt sämre på att inaktivera läkemedelsmolekyler – de är med andra ord lågaktiva. En konsekvens av varianternas olika egenskaper är att det förekommer individuella skillnader i hur man reagerar på tiopurinläkemedel, beroende på vilken typ av TPMT-protein man har anlag för. Det gör doseringen av tiopuriner komplicerad. En patient som har gener för en lågaktiv TPMT-variant riskerar att överdoseras, vilket kan leda till att immunförsvaret slås ut helt. En patient som har högaktivt TPMT-protein bildar istället stora mängder inaktiverade läkemedelsmolekyler, vilka kan skada levern vid höga halter. För att göra behandling med tiopurinläkemedel säkrare görs rutinmässigt en kontroll av vilken typ av TPMT-protein patienten har, och därefter skraddarsys doseringen för varje patient. Man övervakar också patienten med regelbundna provtagningar för att kontrollera att mängden läkemedel är på rätt nivå.

Biokemin som hjälpmedel

I vår forskning studerar vi utvalda TPMT-varianter för att förklara vad som orsakar deras olika biofysikaliska egenskaper. Vi har bland annat kunnat visa att den molekyllära strukturen hos en variant, TPMT*6 (Y180F), är lite mindre stabil än den vanligast förekommande proteinvarianten (den så kallade vildtypen). Detta gör att TPMT*6 förmodligen bryts ner snabbt i cellen, trots att denna variant faktiskt har normal funktion. Vi har också utvecklat metoder för att förbättra framtida studier av TPMTs egenskaper, varav den ena är ett nytt sätt att mäta reaktionshastigheten för metyleringen av tiopuriner. Den andra metoden baseras på en fluorescerande substans som interagerar med TPMT och ger olika stark signal beroende på proteinets egenskaper, vilket gör att det finns många tänkbara användningsområden. Som exempel lämpar sig metoden för studier av TPMTs interaktion med andra läkemedelsmolekyler än tiopuriner. Interaktion mellan olika läkemedel kan ge svåra biverkningar, och vår förhoppning är att metoden kan bidra till att förbättra både säkerheten och effektiviteten vid behandling med tiopurinläkemedel. Metoderna vi utvecklat kan också användas till grundforskning och studier av TPMTs struktur och funktion, vilket kan bidra till att förstå och kartlägga dess molekyllära egenskaper och hur dessa påverkar behandling med tiopurinläkemedel.

Abstract

Thiopurine S-methyltransferase (TPMT) belongs to the Class I S-adenosylmethionine-dependent methyltransferase (SAM-MT) super family of structurally related proteins. Common to the members of this large protein family is the catalysis of methylation reactions using S-adenosylmethionine (SAM) as a methyl group donor, although SAM-MTs act on a wide range of different substrates and carry out numerous biologically important functions. While the natural function of TPMT is unknown, this enzyme is involved in the metabolism of thiopurines, a class of pharmaceutical substances administered in treatment of immune-related disorders. Specifically, methylation by TPMT inactivates thiopurines and their metabolic intermediates, which reduces the efficacy of clinical treatment and increases the risk of adverse side effects. To further complicate matters, TPMT is a polymorphic enzyme with over 40 naturally occurring variants known to date, most of which exhibit lowered methylation activity towards thiopurines. Consequently, there are individual variations in TPMT-mediated thiopurine inactivation, and the administered dose has to be adjusted prior to clinical treatment to avoid harmful side effects.

Although the clinical relevance of TPMT is well established, few studies have investigated the molecular causes of the reduced methylation activity of variant proteins. In this thesis, the results of biophysical characterization of two variant proteins, TPMT*6 (Y180F) and TPMT*8 (R215H), are presented. While the properties of TPMT*8 were indistinguishable from those of the wild-type protein, TPMT*6 was found to be somewhat destabilized. Interestingly, the TPMT*6 amino acid substitution did not affect the functionality or folding pattern of the variant protein. Therefore, the decreased *in vivo* functionality reported for TPMT*6 is probably caused by increased proteolytic degradation in response to the reduced stability of this protein variant, rather than loss of function.

Also presented herein are novel methodological approaches for studies of TPMT and its variants. Firstly, the advantages of using 8-anilinonaphthalene-1-sulfonic acid (ANS) to probe TPMT tertiary structure and active site integrity are presented. ANS binds exclusively to the native state of TPMT with high affinity ($K_D \sim 0.2 \mu\text{M}$) and a 1:1 ratio. The stability of TPMT was dramatically increased by binding of ANS, which was shown to co-localize with the structurally similar adenine moiety of the cofactor SAM. Secondly, an enzyme activity assay based on isothermal titration calorimetry (ITC) is presented. Using this approach, the kinetics of 6-MP and 6-TG methylation by TPMT has been characterized.

Keywords: Thiopurine S-methyltransferase (TPMT); protein stability; enzyme kinetics; ligand binding; isothermal titration calorimetry (ITC); 8-anilinonaphthalene-1-sulfonic acid (ANS)

Abbreviations

SAM	S-adenosylmethionine
SAH	S-adenosylhomocysteine
SAM-MT	SAM-dependent methyltransferase
TPMT	thiopurine S-methyltransferase
6-MP	6-mercaptopurine
6-TG	6-thioguanine
TGN	thioguanine nucleotide
CD	circular dichroism
ANS	8-anilinonaphthalene-1-sulfonic acid
ITC	isothermal titration calorimetry

Papers

- [I] Patricia Wennerstrand, Annica Blissing, and Lars-Göran Mårtensson. “In vitro Protein Stability of Two Naturally Occurring Thiopurine S-methyltransferase Variants: Biophysical Characterization of TPMT*6 and TPMT*8”. 2017. Manuscript.
- [II] Annica Blissing and Lars-Göran Mårtensson. “Characterization of ligand binding and enzyme activity of Thiopurine S-methyltransferase by isothermal titration calorimetry”. 2017. Manuscript.
- [III] Patricia Wennerstrand, Annica Theresia Blissing, Lars-Göran Mårtensson, and Patrik Lundström. “Partially Assigned Chemical Shifts of Human Thiopurine S-methyltransferase Reveal Flexibility in Native Structure”. 2012. Progress report.

Papers [I] and [III]. Both first authors contributed equally in performing experiments and writing of the manuscript.

Paper [II]. Both authors contributed equally in performing experiments and writing of the manuscript.

The following papers are not included in the thesis:

- [IV] Patrik Lundström, Alexandra Ahlner, and Annica Theresia Blissing. “Isotope labeling methods for large systems”. In: *Isotope labeling in Biomolecular NMR*. Springer, 2012, pp. 3–15.
- [V] Patrik Lundström, Alexandra Ahlner, and Annica Theresia Blissing. “Isotope labeling methods for relaxation measurements”. In: *Isotope labeling in Biomolecular NMR*. Springer, 2012, pp. 63–82.
- [VI] Ulrich Weininger, Annica T Blissing, Janosch Hennig, Alexandra Ahlner, Zhihong Liu, Hans J Vogel, Mikael Akke, and Patrik Lundström. “Protein conformational exchange measured by ^1H $R_{1\rho}$ relaxation dispersion of methyl groups”. In: *Journal of biomolecular NMR* 57.1 (2013), pp. 47–55.

Contents

1	Introduction	1
1.1	Proteins	1
1.2	Secondary structure	2
1.3	Tertiary structure	3
1.4	Higher order structures	3
1.5	Catalysis	4
1.6	Examples of catalytic strategies	5
1.7	SAM-MTs – meeting the family	6
1.8	Thiopurine S-methyltransferase (TPMT)	7
1.9	The pursuit of purines	12
1.10	Metabolic activation	13
1.11	Diverting efficacy	15
1.12	TPMT catalysis	16
1.13	Influencing activity — molecular causes	16
1.14	Influencing activity — drug interactions	17
2	Methods	23
2.1	Interactions of light and matter	23
2.2	Absorption spectroscopy	24
2.3	Circular dichroism	25
2.4	Illuminating protein chemistry	26
2.5	Nuclear magnetic resonance spectroscopy	28
2.6	Isothermal titration calorimetry	30
2.7	Enzyme kinetics	31
2.8	Enzyme kinetics meets ITC	33
3	Summary and Discussion	35
3.1	Biophysical characterization of variants	35
3.2	What’s the matter with SAM?	38
3.3	Improving methodology	39
	References	43

1

Introduction

1.1 Proteins – more than the sum of their parts

Generally, 20 amino acids are all it takes to create the miraculous entities that are proteins. The specific order of amino acids and the length of the polypeptide chain constitute the primary structure, and these properties are the determinants of a specific protein. The amino acids are linked together through the formation of a peptide bond (Figure 1.1), which is perhaps the key to the greatness of protein chemistry: the resonance structures of the peptide bond contribute to the double-bond characteristics that restrict the conformational flexibility of a polypeptide chain. This is the foundation of the higher order structural organization that determines the folding pattern and, ultimately, the function of a protein.

Figure 1.1 A (di)peptide segment with the peptide bond highlighted by the aquamarine rectangle. The side chain moieties, R_1 and R_2 , vary depending on the amino acid identity.

1.2 Secondary structure

The sequential linkage of amino acids via peptide bonds is referred to as the main chain, or backbone, of a protein. The backbone can be arranged into ordered structures stabilized by hydrogen bond formation, so called secondary structure. The main chain conformation and hydrogen bonding pattern define the particular type of secondary structure, the most common being α -helices and β -strands (Figure 1.2). In α -helical structure, the backbone is wrapped into a helix, making one turn every 3.6 residues, which allows for stabilization of the structure by hydrogen bonding between backbone amides and carbonyls of residues aligned along the helix axis. The sense of rotation of an α -helix is most often right-handed, or clockwise, since this arrangement involves less steric interference between side chains.

The β secondary structure consists of (extended) polypeptide strands laying alongside each other, forming sheets. The amino acid side chains are oriented perpendicularly to the plane of the β -sheet, pointing either upward or downward in an alternating fashion. The structure is stabilized by hydrogen bonding between the backbone amides of one strand and the backbone carbonyls of the adjacent strand. The direction of the strands relative to one another can be either parallel or anti-parallel, and the two modes differ somewhat in their hydrogen bonding geometry. The anti-parallel β structure aligns adjacent strands so that the hydrogen bonding partners face each other dead-on, and the resulting hydrogen bonds are formed between only two residues, one on each strand. The parallel β structure on the other hand produces slightly offset hydrogen bonds, where one residue interacts with two others on the neighboring strand.

Figure 1.2 The different kinds of secondary structure are α -helix (left), parallel β -sheet (middle), and anti-parallel β -sheet (right).

1.3 Tertiary structure – hold the fold

The tertiary structure, or fold, of a protein is the spatial arrangement and packing of α and β secondary structures connected by loops. The folding of a polypeptide chain is thought to proceed through a funnel-like energy landscape, where the width of the funnel corresponds to the number of conformations that are energetically available (or the entropy), and the depth represents the change in overall free energy. While the tertiary structure is stabilized by hydrogen bonding, hydrophobic, and electrostatic interactions, the driving force behind protein folding is the enthalpic contributions from van der Waals interactions formed upon burial of hydrophobic amino acid side chains, since the van der Waals interactions formed between non-polar groups make larger enthalpic contributions than those formed between non-polar groups and solvent water [1]. This initial hydrophobic collapse limits the number of conformations available, directing the polypeptide chain towards formation of the secondary structure hydrogen bonding patterns. The formation of secondary structure effectively buries hydrogen bonds, which further decreases the free energy of the emergent protein. In fact, burial of hydrogen bonds or polar groups in the hydrophobic core of a protein contributes more to overall stability than burial of the volume equivalent of non-polar groups [2, 3]. As folding progresses, the secondary structure elements are assembled into native-like tertiary structures, where the exclusion of water due to tight packing of the structural elements further decreases the free energy of the nascent protein as it descends through the energy landscape towards its native structure.

1.4 Higher order structures

The tertiary structure of a protein determines its function. The three-dimensional structure is the framework that supports and positions functional groups, such as catalytically active side chains, or residues involved in molecular recognition. In theory, there is an infinite number of ways to fold a polypeptide, but in reality, nature sticks to what works and uses a limited number of folding patterns. Instead, function is varied by utilizing different types of amino acids that provide alternative chemistry, without needlessly changing the overall tertiary structure. However, in some cases, the desired function cannot be accomplished by one protein alone, but requires the assembly of two or more interacting proteins. These interactions are stabilized by the same forces as the

constituent proteins themselves, and together they form what is referred to as quaternary structure. It is also possible for a (long enough) polypeptide chain to fold into several distinct domains, each with their own particular structure and function.

Since the three-dimensional structure is so vitally important, changes or perturbations to the folding pattern will inevitably affect protein function, often negatively. The inability of a protein to assume or maintain the proper fold will not only interfere with its function, it could also cause problems related to the misfolded structures themselves. Misfolded or damaged proteins are usually recognized by the quality control system of the cell and promptly degraded, but sometimes they aggregate to form larger, insoluble assemblies that can interfere with cellular functions. Several medical conditions, for instance Alzheimer's or Parkinson's disease, are caused by misfolded protein aggregates. Misfolding is sometimes caused by substitution of a single amino acid brought on by genetic mutation, but it can also be related to aging and impaired clearance of defective proteins, or a combination of factors [4, 5].

The three-dimensional structures of proteins are far from static. There is significant flexibility in protein structure, and the mobility ranges from motions such as molecular vibrations and local fluctuations, to movement of secondary structure elements and entire protein domains. Essentially, structural dynamics is an inherent property of proteins that is closely related to their function.

1.5 Catalysis – conquering activation energy

Many biochemically important reactions do not occur at physiological temperature due to an insurmountable activation energy. For life as we know it to exist, these reactions have to be catalyzed to increase the rate of reaction at tolerable temperatures, and the task falls on proteins. Proteins that are able to act as biological catalysts are called enzymes. Enzymes are able to form products with highly specific structural and chemical properties, and they do so by stabilizing the transition state of the reaction. The functional groups of the enzyme provide energetically favorable interactions that complement the structural and chemical properties of the transition state, thereby utilizing binding energy that partially counteracts the activation energy of the overall reaction. By lowering the activation energy through stabilization of the transition state, more reactants per unit time are able to surpass the energy barrier, and

the rate of the enzyme catalyzed reaction is increased compared to the uncatalyzed reaction. Also, the chemical and structural properties of a particular transition state allows it to make the most energetically favorable contacts with the enzyme, which leads to subsequent formation of a specific product at an increased rate. Essentially, enzymes produce the desired product through careful selection of the particular transition state that can proceed to form that product. Consequently, catalytic efficiency and specificity are both the result of transition state stabilization.

1.6 Examples of catalytic strategies

A few general catalytic strategies are frequently used in biological catalysis, although enzymes may use more than one of the strategies described below to carry out a biochemical reaction. Some reactions can be accomplished by simply providing the necessary conditions. Here, the enzyme ensures *proximity and orientation* of the reactant species relative to one another in the active site. By confining the reactants to the active site their effective concentrations are increased, which further increases the rate of the reaction. Also, the active site of a protein can provide a particular micro-environment, for instance the exclusion of solvent water to lower the dielectric constant, which enhances electrostatic interactions and increases the strength of nucleophiles in *general acid-base catalysis*.

Proteins can also form energetically unfavorable products through *coupled reactions*, where an energetically unfavorable process is linked to a favorable one, such as the hydrolysis of adenosine triphosphate (ATP), or the demethylation of S-adenosylmethionine (SAM). A biochemical process can also be facilitated by using an alternative reaction pathway. *Covalent catalysis* includes the formation of a covalent bond between enzyme and substrate, forming an intermediate whose transition state is associated with a lower free energy than that of the uncatalyzed reaction.

Another catalytic strategy is *cofactor catalysis*. Cofactors are non-protein substances, often vitamins or vitamin derivatives. Here, the protein serves as a structural scaffold that positions the reactant species, but the catalytic power is provided by the reactive groups of the cofactor. Some enzymes employ *metal ion catalysis* to achieve similar results, with the possibility of additional benefits such as charge coordi-

nation and electrostatic stabilization by the metal ion. The use of metal ions and non-protein cofactors allows proteins to access other types of chemistry that cannot be provided by the sole use of amino acids.

1.7 SAM-MTs – meeting the family

S-adenosylmethionine (SAM) is a metabolic intermediate widely used in biological methylation reactions. It is preferred over other potential methyl donors due to the large change in free energy ($-17 \text{ kcal mol}^{-1}$) [6] upon conversion of SAM to S-adenosylhomocysteine (SAH) (Figure 1.3). Therefore, SAM is diligently used in methylation reactions catalyzed by SAM-dependent methyltransferases (SAM-MTs). This large protein family is divided into subclasses depending on type of substrate (for instance small-molecules, proteins, or DNA) and the specific methyl group acceptor atom (on the substrate). The small-molecule Class I SAM-MTs (EC 2.1.1.X) methylate small organic compounds, and the substrate acceptor atom can be either nitrogen, oxygen, carbon, or sulfur. Like all SAM-MTs, the Class I proteins share a highly conserved Rossmann-like α/β core fold topology consisting of a central 7-stranded β -sheet flanked by 3 helices on each side [6], as shown in Figure 1.4A. The core fold is frequently augmented by additional structural inserts. Proteins within the same subclass usually share a common insert pattern and small-molecule SAM-MTs are no exception. The proteins within this particular subclass tend to have structural inserts at the N-terminal, an active site cover, and additional inserts between $\beta 5$ and αE , and $\beta 6$ and $\beta 7$ of the core fold [7].

Despite the high degree of structural preservation among small-molecule SAM-MTs, the sequence identity is very low. The lowest degree of sequence conservation occurs in the structural additions to the core fold, which has enabled the evolution of diverse substrate specificity among

Figure 1.3 The chemical structures of SAM (left) and SAH (right).

SAM-MTs. However, a few conserved sequence motifs exist within this protein family [7, 8]. These motifs overlap with the SAM-MT core fold, and their purpose seem to be preservation of SAM binding properties as well as conservation of packing and folding interactions that support and position functionally important residues. Motifs I-IV are involved in SAM binding and MTase activity, while Motifs V-VI are more important for structural stability and integrity [8].

1.8 Thiopurine S-methyltransferase (TPMT)

Thiopurine S-methyltransferase (TPMT, EC 2.1.1.67, 245 amino acids, 28 180 g mol⁻¹) is a cytosolic single-domain protein that belongs to the Class I SAM-MTs described above. The crystal structure of the human protein was published in 2007 (PDB ID:2BZG) [9] and determined that TPMT conforms to the Class I SAM-MT structural norm, sporting several of the structural inserts becoming of a well-behaved family member. Figure 1.4 shows the topology of TPMT and localization of the structural inserts in the tertiary structure, and the details of the TPMT structural inserts to the core fold are summarized in Table 1.1. The monomeric structures of the related proteins used for comparison in Table 1.1 are shown in Figure 1.5. The conserved SAM-MT sequence motifs and their TPMT equivalents are discussed in Table 1.2, and Figure 1.6 shows the conserved SAM-MT sequence motif equivalents mapped onto TPMT tertiary structure.

TPMT is of great clinical importance since it methylates (and inactivates) thiopurine pharmaceuticals administered in management of immune-related disorders, and the protein is also linked to adverse side effects from treatment [10, 11]. However, despite the extensive knowledge of the structural and functional kinship of SAM-MTs, the natural substrate of TPMT is still unknown.

Table 1.1 *TPMT structural additions to the SAM-MT core fold*

Localization of insert ^a	TPMT structural insert	Comparison with other SAM-MTs ^b
N-terminal (preceding α Z)	Long, unstructured N-terminal followed by α -helix containing residues in direct contact with SAM (W29, W33, F40), forming a hydrophobic lid.	Insert forms the active site cover. Similar active site covers are common among structurally related proteins (GNMT, HNMT, IOMT, PNMT, PIMT, NNMT).
Between α B and β 3	Two anti-parallel β -strands at the edge of the central β -sheet.	Inserts at this position also occurs in other SAM-MTs, albeit with different secondary structure elements (helices in PNMT, GNMT, NNMT).
Between β 4 and α D	Mini-helix supporting SAM-binding residues.	Similar helical inserts occur in several other SAM-MTs (GNMT, IOMT, PNMT, CmaA1, NNMT).
Between β 5 and α E	Long loop that interacts with other structural additions to cover the active site.	Inserts at this position are common in structurally related proteins, albeit with different secondary structure elements (GNMT, HNMT, IOMT, PNMT, GAMT, CmaA1, NNMT).
Between β 6 and β 7	α -helix that interacts with other structural inserts to cover the active site.	Large loops (COMT, GAMT, PNMT, GAMT, NNMT), helices (CmaA1) or entire domains (HNMT) are common in structurally related proteins.

^a Localization of insert relative to SAM-MT core fold.^b Note that the examples presented herein are only a representative selection and do not constitute a complete account of structurally related proteins. Proteins with similar inserts are shown in parenthesis.

Abbreviations: Phenylethanolamine N-methyltransferase (PNMT, PDB ID:1HNN), Glycine N-methyltransferase (GNMT, PDB ID:1XVA), Isoflavone O-methyltransferase (IOMT, PDB ID:1FPX), Histamine methyltransferase (HNMT, PDB ID:1JQD), Catechol O-methyltransferase (COMT, PDB ID:1VID), Guanidinoacetate methyltransferase (GAMT, PDB ID:1KHH), Protein isospartyl methyltransferase (PIMT, PDB ID:1I1N), Mycolic acid cyclopropane synthase (CmaA1, PDB ID:1KPH), Nicotinamide N-methyltransferase (NNMT, PDB ID:3ROD)

Table 1.2 *TPMT equivalents of conserved SAM-MT sequence motifs*

Motif and description ^a	TPMT equivalent ^b	Description of TPMT equivalents ^c
I β1 and succeeding loop. GxGxG consensus sequence. Conserved acidic residue (D or E).	R ₆₄ -V ₆₅ -F ₆₆ -F ₆₇ -P ₆₈ - L₆₉ - C₇₀ - G₇₁	No GxGxG consensus sequence. Instead, the correct main chain conformation is achieved by P68, G71, and A73. Cysteine residue frequently occurs at C70 position equivalent. No acidic residue or equivalent.
II β2 and adjoining turn/loop. Partially conserved acidic residue and conserved hydrophobic SAM-binding residue.	S ₈₅ -V ₈₆ -V ₈₇ -G ₈₈ -V ₈₉ - E₉₀ - I₉₁ -S ₉₂	E90 and I91 correspond to conserved acidic and hydrophobic residues, respectively. GVEIS sequence corresponds to the (G-[non-polar]-[E/D]-[non-polar]-S) frequently occurring in structurally related proteins.
III β3. Partially conserved acidic residue.	I ₁₂₇ -S ₁₂₈ -L ₁₂₉ -Y ₁₃₁ -C ₁₃₂	No acidic residue equivalent.
IV β4 and succeeding loop. Conserved D/E/N residue at the N-terminal end of strand.	K ₁₄₅ -F ₁₄₆ -D ₁₄₇ -M ₁₄₈ -I ₁₄₉ -W ₁₅₀ -D ₁₅₁ - R₁₅₂ - G₁₅₃ -A ₁₅₄ -L ₁₅₅ -V ₁₅₆ -A ₁₅₇ -I ₁₅₈	D147 corresponds to well-preserved acidic/polar residue. Usually preceded by a non-polar residue (often F) in structurally related proteins. TPMT sports a mini-helix structural insert in the loop corresponding to that of Motif IV.
V αD. Contains large hydrophobic residues that support SAM-binding structures.	R ₁₆₃ -K ₁₆₄ -C ₁₆₅ -Y ₁₆₆ -A ₁₆₇ -D ₁₆₈ -T ₁₆₉ -M ₁₇₀ -F ₁₇₁ -S ₁₇₂	Y166 is situated in the middle of the corresponding helix, facing inwards toward the protein core.
VI β5 and preceding tight-turn. Nearly invariant Glycine residue.	G ₁₇₅ -K ₁₇₆ -K ₁₇₇ -F ₁₇₈ -Q ₁₇₉ -Y ₁₈₀ -L ₁₈₁ -L ₁₈₂ -C ₁₈₃ -V ₁₈₄ -L ₁₈₅	G175 corresponds to the conserved Glycine residue.

^a Note that the localization is specified relative to the SAM-MT core fold.

^b Residues in direct contact with SAM in bold

^c Comparisons have been made based on relevant publications and DALI structural similarity search [12] results.

Figure 1.4 Topology of (A) the SAM-MT core fold and (B) TPMT, compared to (C) the TPMT tertiary structure (PDB ID:2BZG) viewed from either side of the central β -sheet. TPMT structural inserts are highlighted in aquamarine. In (A) and (B), the structural elements are labeled numerically (β -strands) and alphabetically (α -helices). Helices lying above and below the plane of the central β -sheet are shown in light and dark gray, respectively. Note that TPMT numbering differs from that of the core fold due to the presence of structural inserts.

Figure 1.5 Structural inserts of selected proteins that share the SAM-MT core fold. The monomeric structures of the proteins discussed in Table 1.1 are shown in ribbon representation. To facilitate comparison, all structures are shown with their respective SAM-MT core fold oriented similar to the left image of Figure 1.4C (C-terminal end of central β -sheet to the right). Additions to the core fold are highlighted in aquamarine. Other deviations from the core fold is shown in pink. Regions shown in gray correspond to the SAM-MT core fold. SAM and SAH are shown in stick representation and colored by element.

Figure 1.6 Conserved SAM-MT sequence motifs mapped onto TPMT tertiary structure (PDB ID:2BZG). The sequence motifs are shown in green (Motif I), blue (Motif II), purple (Motif III), red (Motif IV), orange (Motif V), and pink (Motif VI). S-adenosylhomocysteine (SAH) is depicted in stick representation and colored by element. Regions shown in gray are not part of the conserved sequence motifs.

1.9 The pursuit of purines

Thiopurine pharmaceutical substances are routinely used as chemotherapeutic agents in treatment of immune-related conditions, such as childhood acute lymphoblastic leukemia (ALL), irritable bowel syndrome (IBS), autoimmune disorders, and to prevent rejection in patients receiving organ transplants [13, 14]. Thiopurines such as 6-mercaptopurine (6-MP) and 6-thioguanine (6-TG) require stepwise metabolic activation to generate thioguanine nucleotides (TGNs). TGNs are antimetabolites of the natural purine nucleobases in the sense that they share structural but not chemical properties. By masquerading as the natural substrates, TGNs are able to interfere with biochemical processes involving purine nucleotides, such as cell signaling and replication, which ultimately causes apoptosis since the cell cannot survive without viable DNA [15, 16, 17]. Essentially, thiopurines are the biochemical equivalents of dead-ends in this context. However, there is the matter of TPMT.

Methylation by TPMT interferes with the therapeutic efficacy of thiopu-

rines since the methylated products are not active cytotoxic agents. In fact, a substantial portion of thiopurines administered in clinical treatment are inactivated by TPMT, and erythrocyte TGN concentrations have been shown to be inversely proportional to TPMT activity [11, 18]. To add insult to injury, there are over forty naturally occurring polymorphic protein variants known to date [19], and the majority of the variants exhibit decreased enzyme activity towards thiopurine substrates. The main reason for this diversity is the occurrence of genetic variations caused by single-nucleotide polymorphisms (SNPs) that result in amino acid substitution in the encoded protein, although mutations involving non-coding regions have also been shown to affect *in vivo* methylation activity [20, 21].

Individual variations in TPMT activity can influence thiopurine therapy in different ways. Aside from decreasing the pool of active TGN metabolites, buildup of methylated product as a result of high TPMT activity could cause hepatotoxicity, while on the other hand, decreased TPMT activity might result in myelosuppression due to accumulation of active TGNs [14]. Interestingly, provided that the thiopurine dosage is appropriately reduced to avoid excessive toxicity, patients with low or intermediate TPMT activity have been found to experience better treatment outcome and lower risk of relapse [22, 23]. Conversely, patients with normal to high TPMT activity may instead be at risk of undertreatment and relapse [11]. To account for individual variations, the patient's TPMT genotype and enzyme activity phenotype are routinely measured prior to thiopurine treatment, and the dosage is set accordingly. This type of individualization of medical treatment based on genetically predicted drug response and tolerance is referred to as pharmacogenetics.

1.10 Metabolic activation

The stepwise conversion of 6-MP to active TGN metabolites is shown in Figure 1.7 (vertical reactions). Briefly, hypoxanthine phosphoribosyl-transferase (HPRT, EC 2.4.2.8) covalently links 6-MP to phosphoribosyl pyrophosphate (PRPP), forming thioinosine monophosphate (TIMP), which is in turn used as substrate by inosine monophosphate dehydrogenase (IMPDH, EC 1.1.1.205) to form thioxanthosine monophosphate (TXMP), which is then converted to thioguanosine monophosphate (TGMP) by guanosine monophosphate synthetase (GMPS, EC 6.3.5.2). TGMP is phosphorylated by nucleoside-phosphate kinase

Figure 1.7 (left) The stepwise conversion of 6-mercaptopurine to active cytotoxic TGNs (vertical reactions), and the competing inactivation pathways (horizontal reactions). Abbreviations: 6-mercaptopurine (6-MP), 6-methyl mercaptopurine (6-MeMP), thiopurine S-methyltransferase (TPMT), hypoxanthine phosphoribosyltransferase (HPRT), thioinosine monophosphate (TIMP), methyl thioinosine monophosphate (MeTIMP), inosine monophosphate dehydrogenase (IMPDH), thioxanthosine monophosphate (TXMP), guanosine monophosphate synthetase (GMPS), thioguanosine monophosphate (TGMP), methyl thioguanosine monophosphate (MeTGMP), nucleoside-phosphate kinase (NPK), thioguanosine diphosphate (TGDP), nucleoside-diphosphate kinase (NDPK), thioguanosine triphosphate (TGTP).

(NPK, EC 2.7.4.4) to form thioguanosine diphosphate (TGDP), which is in turn phosphorylated by nucleoside-diphosphate kinase (NDPK, EC 2.7.4.6) to form thioguanosine triphosphate (TGTP). Both TGDP and TGTP are capable of acting as substrates in various cellular processes but lack the functionality of the natural purine nucleoside substrates.

The biochemical conversion of 6-TG into active TGN metabolites is less elaborate, as 6-TG is directly converted to TGMP by HPRT. TGMP is then further processed as described above.

1.11 Diverting efficacy

Several TPMT catalyzed reactions compete for thiopurine metabolic intermediates as substrates (horizontal reactions in Figure 1.7). Both 6-MP and 6-TG can be methylated by TPMT to form the inactive metabolites 6-methyl mercaptopurine (6-MeMP) and 6-methyl thioguanine (6-MeTG), respectively. TIMP and TGMP have also been shown to be methylated by TPMT [24] (forming MeTIMP and MeTGMP, respectively), resulting in further depletion of the metabolic intermediates available for generation of active TGNs. However, MeTIMP has been shown to inhibit the formation of PRPP [25, 26]. This effectively inhibits *de novo* purine biosynthesis, making the purine salvage pathway, capable of generating active TGNs, the only source of nucleotides available to the cell. Consequently, methylation of TIMP by TPMT actually adds to the cytotoxicity of thiopurines.

1.12 TPMT catalysis

The methylation of 6-MP by TPMT has been studied extensively. The essence of TPMT catalysis is the positioning of the substrate relative to the reactive species, SAM. The reaction proceeds through an S_N2 type mechanism [9, 27], where the reactive methyl group of SAM is transferred to a nucleophilic center (sulfur atom) on the 6-MP substrate, forming the products SAH and 6-MeMP. A solvent channel provides access to the active site of the enzyme, where the imidazole moiety of bound 6-MP can be oriented to either side [9, 28], but formation of product should be possible regardless of the relative orientation. Studies of murine TPMT have shown that after binding of 6-MP, deprotonization of the active thiol tautomeric form is necessary to achieve an activation energy (20 kcal mol^{-1}) [27] consistent with the observed reaction rate [28]. Although the precise circumstances have not been determined, deprotonization through water mediated hydrogen bonding has been proposed [9, 27].

Despite the agreement between the calculated activation energy and observed reaction rates of the murine enzyme, the published kinetic parameters of human TPMT vary significantly, as does the source of protein, method of purification, choice of activity assay, reaction conditions (including choice of thiopurine substrate and reactant concentrations), use of units, and general presentation of results [24, 28, 29, 30, 31, 32]. Both SAH and 6-MeMP coproducts inhibit the enzyme, albeit through different mechanisms; SAH is a potent competitive inhibitor with a K_i of $0.75 \mu\text{M}$ with regard to SAM, and 6-MeMP is a non-competitive inhibitor with a K_i of $560 \mu\text{M}$ with regard to 6-MP [10].

1.13 Influencing activity — molecular causes

The inherent enzyme activity and cellular levels of variant proteins are important factors that directly influence the extent of TPMT methylation. Most of the known protein variants exhibit decreased *in vivo* enzyme activity towards thiopurine substrates, but the molecular causes of reduced functionality have only been investigated for a handful of protein variants (to date), namely TPMT*2 (A80P) [33], TPMT*3A (A154T/Y240C) [34, 35], TPMT*3B (A154T) [34, 35], TPMT*3C (Y240C) [34, 35], TPMT*5 (L49S) [33], TPMT*16 (R163H) [30], TPMT*21 (L69V) [29], TPMT*24 (Q179S) [29], TPMT*25 (C212R) [29],

and TPMT*31 (I204T) [31]. The variant proteins are often found to be destabilized (due to amino acid substitution), making them susceptible to proteolytic degradation by the quality control system of the cell [34, 35, 36], although one exception is TPMT*5, which exhibits nearly complete loss of function rather than reduced structural stability [33].

1.14 Influencing activity — drug interactions and rogue methylation

Methylation activity can also be affected by interference from substances competing for the enzyme. TPMT has been shown to bind a large number of heterocyclic compounds with polar substituents (albeit with varying affinity), and the corresponding thiols often act as substrates. Several clinically relevant substances inhibit the enzyme, and have been shown to affect treatment when administered concomitantly with thiopurines [37, 38, 39, 40]. Substances capable of acting as substrates or inhibitors of the enzyme are summarized in Tables 1.3 and 1.4, respectively. The large number of substances interacting with TPMT coupled with its central role as a metabolizing enzyme makes thiopurine chemotherapy vulnerable to drug interactions that could potentially affect patient safety.

Table 1.3 *Known in vitro TPMT substrates.*

Substance	K_m	k_{cat}^a	k_{cat}/K_m	Ref
	(μM)	($nmol\ min^{-1}\ mg^{-1}$)	($ml\ min^{-1}\ mg^{-1}$)	
S-adenosylmethionine (SAM)	2.7	6.2	2.3	[32]
6-mercaptopurine (6-MP)	300	7.9	0.026	[32]
	10.6	48	4.5	[24]
	680	73.8	0.11	[28]
	383	15.0	0.039	[41]
6-thioguanine (6-TG)	550	9.1	0.017	[32]
	18.1	55	3.0	[24]
	63	14	0.22	[29]
	15.2	19.2	1.3	[30]
	557	18	0.032	[41]
6-thioinosine	55.1	89	1.6	[24]
	1170	10.7	0.0091	[41]
6-thiodeoxyinosine	12.7	22	1.7	[24]
thioinosine monophosphate (TIMP)	25.7	31	1.2	[24]
	1270	27.2	0.021	[41]
thioinosine triphosphate	890	1.7	0.0019	[41]
6-thioguanosine	26	32	1.2	[24]
	761	6.5	0.0085	[41]
6-thiodeoxyguanosine	131.4	120	0.91	[24]
thioguanosine monophosphate (TGMP)	27.1	59	2.2	[24]
	1040	1.3	0.0013	[41]
6-thiouracil (TU)	2000	3.6	0.0018	[32]
2,6-dithiopurine (2,6-DTP)	34	4.6	0.14	[42]
6-selenopurine	29.1	32.5	1.1	[41]
6-selenoinosine	58.1	33.2	0.57	[41]
6-selenoguanosine	139	20.8	0.15	[41]
7-methyl-6-MP	231	39	0.17	[41]
9-(n-butyl)-6-MP	292	29.5	0.10	[41]
9-ethyl-6-MP	372	24.3	0.065	[41]
9-(n-propyl)-6-TG	159	2.8	0.018	[41]
β -mercaptoethanol (β -ME)	168 000	2.1	—	[32]
1-methyltetrazole-5-thiol (MTT)	260	3.1	0.012	[43]

Substance	K_m	k_{cat} ^a	k_{cat}/K_m	Ref
	(μM)	($nmol\ min^{-1}\ mg^{-1}$)	($ml\ min^{-1}\ mg^{-1}$)	
2-methyl-1,3,4-thiadiazole-5-thiol	68	11.9	0.18	[43]
(MTD)	63	N.A.	N.A.	[44]
diethyldicarbamate (DDC)	95	0.05	—	[45]
2-nitro-5-thiobenzoic acid (TNB)	N.A.	N.A.		[46]
thiophenol	3.9	N.A.		[47]
	0.36	N.A.		[40]
– derivatives				
4-NHCOCH ₃	2.1			[47]
2-OCH ₃	2.6			
3-OCH ₃	2.1			
4-OCH ₃	3.2			
2-NH ₂	2.5			
4-CH ₃	5.9			
4-NO ₂	6.5			
2-COOH	7.8			

^a the unit of k_{cat} is the one most frequently used in TPMT research

N.A. Not determined or not specified

— Value too small

To enable comparison with our own work, the results of the literature study presented herein has been limited to publications where (1) the relevant components of the biochemical system under study has been measured in (relative) isolation, and (2) the TPMT concentration has been estimated or determined using an established method. Typically, this includes publications using recombinantly expressed TPMT or protein purified from tissue sources. The kinetic parameters have been converted to similar units to facilitate comparison, although it is worth noting that the experimental conditions may vary. Also note that the literature review presented herein does not constitute a qualitative evaluation of the published work included, nor is it to be considered a complete account of all known interacting substances as some may have been overlooked.

Table 1.4 *Known inhibitors of TPMT.*

Substance	K_i^a (μM)	IC_{50} (μM)	Type of inhibition ^b	Ref
2-OH-6-MP	K_{is} 27 K_{ii} 183		Non-competitive or mixed	[41]
2,8-OH-6-MP	K_{is} 95 K_{ii} 340		Non-competitive or mixed	[41]
S-adenosylhomocysteine (SAH)	0.75		Competitive with regard to SAM	[32]
6-thioxanthine (6-TX)	329		N.A.	[48]
6-methyl mercaptopurine (6-MeMP)	560		Non-competitive	[40]
tropolone	K_{is} 850 K_{ii} 1630		Non-competitive	[32]
NSAIDs			Non-competitive	[39]
mefenamic acid	39	39		
tolfenamic acid	50	63		
naproxen	52	79		
ketoprofen	172	1013		
olsalazine	208	1474		
diclofenac	722	1582		
ibuprofen	1043	1968		
lornoxicam	1410	2135		
flurbiprofen	1524	1649		
celecoxib	2413	2416		
piroxicam	2589	2589		
meloxicam	4238	4292		
nabumetone	4300	4341		
paracetamol	5162	5168		
benzoic acid	K_{is} N.A. K_{ii} 870		Non-competitive or mixed	[40]
salicylic acid	K_{is} 220 K_{ii} 530		Non-competitive or mixed	[40]
acetylsalicylic acid	K_{is} 1400 K_{ii} 5400		Non-competitive or mixed	[40]
3,4-dimethoxy-5-hydroxybenzoic acid (DMHBA)	K_{is} 10 K_{ii} 19 K_{is} 6 K_{ii} 30	17.6	Non-competitive or mixed Non-competitive or mixed	[40] [38]
sulphasalazine	K_{is} 43 K_{ii} 141	81	Non-competitive	[49]

Substance	K_i^a (μM)	IC_{50} (μM)	Type of inhibition ^b	Ref
5-ASA [†]	K_{is} 1260 K_{ii} 979	986	Non-competitive	[49]
4-ASA	K_{is} 4640 K_{ii} 1510	2900	Non-competitive	[49]
3-ASA		99		[49]
furosemide	K_{is} 31 K_{ii} 140	170	Non-competitive or mixed	[38]
trichlormethiazide	K_{is} 530 K_{ii} 270	1000	Non-competitive or mixed	[38]
bendroflumethiazide	K_{is} 160 K_{ii} 180	360	Non-competitive or mixed	[38]
methotrexate (MTX) [‡]	N.A.			[37]
3-[N-morpholino]-propanesulfonic acid (MOPS)	N.A.			[38]
4-[2-hydroxyethyl]-1- piperazineethanesulfonic acid (HEPES)	N.A.			[38]

^a The inhibition constants are presented differently depending on the mode of inhibition, where K_i is the equilibrium inhibition constant in case of competitive inhibition. In case of non-competitive or mixed inhibition, K_{is} and K_{ii} correspond to inhibition affecting the slope and intercept, respectively, of a Lineweaver-Burk plot.

^b Inhibition with regard to 6-MP if not otherwise specified.

[†] Balsalazide and olsalazine release 5-ASA as active substance

[‡] No inhibition parameters available but authors report a K_D of 40 μM

N.A. Not determined or not specified

To enable comparison with our own work, the results of the literature study presented herein has been limited to publications where (1) the relevant components of the biochemical system under study has been measured in (relative) isolation, and (2) the TPMT concentration has been estimated or determined using an established method. Typically, this includes publications using recombinantly expressed TPMT or protein purified from tissue sources. The inhibition parameters have been converted to similar units to facilitate comparison, although it is worth noting that the experimental conditions may vary. Also note that the literature review presented herein does not constitute a qualitative evaluation of the published work included, nor is it to be considered a complete account of all known interacting substances as some may have been overlooked.

2

Methods

2.1 Interactions of light and matter

Spectroscopy is the study of interactions between matter and electromagnetic radiation. An atom or molecule may absorb electromagnetic radiation if the energy, E , of the incident light

$$E = h\nu$$

(where h is Planck's constant and ν is the frequency in Hz) is equal to the energy difference between two states of that atom or molecule, according to

$$\Delta E = E_e - E_g$$

where ΔE is the energy difference, E_e is the energy of the excited state and E_g is the energy of the ground state. Following *absorption* and excitation, a molecule can return to its ground state through different energy conversions. The excess energy can be released through radiationless vibrational transitions and given off as heat, or the energy can be emitted as photons of the same or lower energy as those absorbed, a process known as *fluorescence emission*. The energy transitions associated with absorption and fluorescence, respectively, are illustrated in Figure 2.1.

Figure 2.1 Schematic illustration of the transitions involved in absorption and fluorescence emission. Absorption of electromagnetic energy induces a transition (solid line arrows shown in green) from the electronic ground state to one of the vibrational sublevels (horizontal lines) of an excited electronic state. The absorbing molecule can then return to its electronic ground state through non-radiative processes (dashed line arrows shown in blue), or through radiative processes such as fluorescence emission, where the excess energy is given off as electromagnetic radiation (dashed line arrows shown in red) with a wavelength corresponding to the difference in energy between the excited electronic state and one of the vibrational sublevels of the electronic ground state of the molecule.

2.2 Absorption spectroscopy

The extent of absorption by an irradiated biomolecule is directly proportional to its concentration. The relationship is described by Beer-Lambert's law

$$A = \varepsilon \cdot c \cdot l$$

where A is the absorbance, ε is the molar absorption coefficient ($\text{l mol}^{-1} \text{cm}^{-1}$), c is the concentration of the absorbing species (mol l^{-1}), and l is the length of the sample traversed by the incident light (cm). The molar absorption coefficient, ε , is a measure of the probability of an electronic transition occurring at a specific wavelength. It follows

that the extent of absorption by a particular molecule varies with the wavelength of the irradiating light, and these properties are what constitute the absorption spectrum of a certain molecular species. The many vibrational and electronic states available to biomolecules make several different energy transitions possible, and the resulting spectra are quite complex. On the other hand, the available energy transitions vary with the structural properties of a molecule. The peptide bond absorbs strongly in the far-UV region of the spectrum at 190–240 nm, and the side chains of aromatic residues absorb in the near-UV region at 250–320 nm. The absorption of non-protein substances associated with protein function (such as cofactors and prosthetic groups) can also be utilized when monitoring molecular changes involved in protein function or catalysis.

2.3 Circular dichroism

In linear (or plane) polarized light, the electric field vector oscillates in a plane perpendicular to the propagation axis. In circular polarized light on the other hand, the electric field vector rotates about the propagation axis and maintains constant magnitude. Circular dichroism (CD) is the differential absorption of the left- and right-handed components of circularly polarized light by an optically active chiral molecule. The differential absorption, ΔA , is expressed as

$$\Delta A = A_L - A_R$$

where A_L and A_R are the absorption of left- and right circularly polarized light, respectively. The differential absorption can also be expressed in the form of Beer-Lambert's law

$$\Delta A = (\varepsilon_L - \varepsilon_R) \cdot c \cdot l$$

where ε_L and ε_R are the molar absorption coefficients ($\text{l mol}^{-1} \text{cm}^{-1}$) of left- and right circularly polarized light, respectively, c is the concentration of the absorbing species (mol l^{-1}), and l is the length of the sample traversed by the incident light (cm).

Proteins are inherently chiral (due to the asymmetric carbon centers of their constituent amino acids) and will therefore interact with circularly polarized light. When subjected to circularly polarized light, differential

absorption of the left- and right-handed components affects the magnitude of the respective electric field vector differently, causing the light traversing the sample to exhibit elliptical characteristics. Therefore, the circular dichroism of biomolecules is usually expressed in terms of ellipticity. The molar ellipticity, θ , ($\text{deg cm}^2 \text{dmol}^{-1}$) is a measure of the circular dichroism related to the concentration of the absorbing species, and the mean residue ellipticity (MRE) ($\text{deg cm}^2 \text{dmol}^{-1} \text{res}^{-1}$) is the molar ellipticity corrected for the length of the polypeptide.

CD spectroscopy is particularly useful when studying properties involving the secondary structures of proteins. The various secondary structures involve different conformations of the polypeptide backbone, which affect the spectroscopic properties and the appearance of the far-UV CD spectrum (190–240 nm) [50]. There are distinct spectral features associated with each type of secondary structure: a positive peak around 190 nm together with negative peaks at 208 nm and 222 nm is typical of α -helical structure; a positive peak around 195 nm combined with a negative peak around 216 nm is indicative of β secondary structure; and random coil shows a negative peak around 195 nm but very little ellipticity above 210 nm. The correlation between main chain conformation and spectral appearance is so strong that it is possible to estimate the secondary structures content of a protein based on its far-UV CD spectrum [51]. Also, since the secondary structure is sensitive to perturbations, it is also possible to use CD spectroscopy to study changes to protein structure, such as the effects of mutation, binding events, or structural decay during thermal unfolding. Although much less detailed, the near-UV CD spectrum (250–320 nm) can provide information on the tertiary structure of proteins, since the spectral properties of the aromatic side chains are affected by their physical environment.

2.4 Illuminating protein chemistry

Fluorescence is the radiative process by which a molecule in an excited electronic state returns to its ground state through spontaneous photon emission. Following absorption, the molecule returns to the lowest vibrational energy level of its excited electronic state. The remaining excess energy is then emitted as electromagnetic radiation corresponding to the difference between the lowest vibrational level of the excited electronic state and one of the vibrational sublevels of the electronic ground state. Since there are several vibrational states available to the mole-

cule, photons of different wavelengths can be emitted, and because the excited electronic state has undergone relaxation through non-radiative processes, the emitted photons are usually of lower energy than those initially absorbed by the molecule. This loss of energy through radiationless transitions prior to photon emission shifts the emission spectrum to longer wavelengths compared to the absorption spectrum of the same molecular species. This spectral shift is known as *Stokes shift*.

Not every photon absorbed leads to a photon emitted. The quotient between the number of photons emitted and the number of photons absorbed gives the *quantum yield*, and this ratio varies with the particular compound and the nature of its chemical environment. An excited molecule can also return to its electronic ground state through radiationless energy transfer between colliding molecules. This kind of attenuation of fluorescence emission through non-radiative processes is called *fluorescence quenching*, and will inevitably affect both the quantum yield and the fluorescence lifetime. Fluorescence energy can also be transferred if the photons emitted are absorbed by a nearby molecule, also known as *fluorescence resonance energy transfer*. Naturally, this process also decreases quantum yield and fluorescence lifetime.

The major contributions to the intrinsic fluorescence of proteins come from Tryptophan residues. Tryptophan has its maximum absorption at 280 nm and typically emits at 320–350 nm, depending on its physical environment. This makes intrinsic Tryptophan fluorescence very useful to probe changes to protein structure such as unfolding events, or perturbation of local structure due to mutation. However, measuring intrinsic Tryptophan fluorescence is not always feasible or appropriate. In these cases, another compound with suitable fluorescence properties and biocompatibility can be utilized, for instance 8-anilino-1-naphthalene-sulfonic acid (ANS). When bound, the conformational flexibility of ANS is restricted, which allows for energy transitions that lead to enhanced fluorescence emission and a dramatically increased quantum yield compared to the unbound species. The bound fluorophore is also sheltered from quenching through molecular collisions, which severely decreases the fluorescence quantum yield of free ANS in aqueous solution. Because of its beneficial spectroscopic properties, ANS has been extensively used in protein chemistry to monitor the exposure of hydrophobic surfaces and molten globule-like states during unfolding events. While some proteins interact with ANS only in their partially unfolded states [52, 53], others bind ANS exclusively to their native states [33, 54, 55]. In some

of the latter cases, ANS binding competes with the natural ligand for the active site [56], and ANS has been shown to bind preferentially to the adenosine-binding site of NAD⁺-dependent dehydrogenases [57].

2.5 Nuclear magnetic resonance spectroscopy

The angular momentum of magnetic nuclei is called spin. The spin states $+\frac{1}{2}$ and $-\frac{1}{2}$ of atomic nuclei such as the ^1H , ^{13}C , and ^{15}N isotopes are energetically degenerate, but when exposed to a strong magnetic field the spins align either with or against the external field, causing the energies of the two spin states to diverge. The difference in energy between the two states depends on the particular type of nucleus and the strength of the external magnetic field, as described by

$$\Delta E = \hbar\gamma B_0$$

where ΔE is the energy difference between the spin states, $\hbar$ is the reduced Planck constant, γ is the gyromagnetic ratio of the nucleus, and B_0 is the strength of the external magnetic field. A nucleus populating a lower energy state can be excited by absorption of electromagnetic radiation corresponding to the energy difference, ΔE , of the transition. The energy absorbed by a given nucleus corresponds to its *Larmor frequency*, and the magnetic moment of the nucleus will precess around the direction of the external magnetic field at that frequency. However, movement of electrons in molecular orbitals generates local magnetic fields that may be oriented differently relative to the external magnetic field. These local magnetic fields will therefore decrease or augment the total magnetic field felt by a particular nucleus, affecting its resonance frequency. The difference in resonance frequency of a nucleus relative to a standard is referred to as *chemical shift*.

The *scalar coupling* (or *spin-spin coupling*) is the electron-mediated interaction between magnetic nuclei. Coupling splits the energy levels of the bonded nuclei due to the energetics of the different spin configurations involved, parallel or anti-parallel, where the former is associated with slightly higher energy due to repulsion. Since the relative orientations of the coupled nuclei affect the energy separation of the different spin states, the resulting spectral peaks are split into multiplets. The scalar coupling depends not only the type of neighboring nucleus, but also on the dihedral bond angles, and bond distance. It is also possible

for the magnetic dipoles of two neighboring nuclei to affect each other's spin populations through space, although this *dipole-dipole interaction* is strongly dependent on the physical separation of the nuclei.

The possibility to affect atomic nuclei through chemical bonds and spatial interactions enables magnetization polarization to be transferred. This is the basis of multi-dimensional nuclear magnetic resonance (NMR) spectroscopy, which is incredibly useful for investigation of the structural and dynamic properties of proteins. Heteronuclear single quantum correlation (HSQC) is an example of a two-dimensional NMR experiment that determines the through-bond correlations between ^1H and ^{15}N (such as the protein backbone and side chain amines and amides), producing cross peaks in the resulting 2D spectrum. Since the NH-groups of a folded protein are exposed to vastly different chemical environments throughout the protein structure (depending on the specific amino acid sequence and folding pattern), they will produce an HSQC spectrum with well dispersed peaks. On the other hand, the nuclei of an unfolded protein experience similar local chemistry since they are all exposed to the uniform environment of the surrounding solvent, causing the cross peaks of the HSQC spectrum to cluster together. The sensitivity of the chemical shifts to the amino acid sequence, folding pattern, and structural properties results in unique HSQC spectral appearance for different proteins. The chemical environment experienced by a nucleus can also change due to interactions with a ligand or substrate, or structural rearrangements associated with binding, making it possible to study the effects of these events using HSQC experiments.

Multi-dimensional NMR spectroscopy can also be used to determine the solution structure of proteins. Simply put, the spatial arrangement of different nuclei can be established by through-space magnetization transfer, and the protein structure can be calculated based on knowledge of which of nuclei are in close proximity of each other. The dependence of the scalar coupling on the dihedral angles of the protein backbone can also be utilized in structural calculations. However, structural determination requires the assignment of spectral peaks to specific nuclei within the protein. This is instead determined using multi-dimensional through-bond magnetization transfer experiments, where the purpose is to identify the specific nucleus corresponding to a particular spectral peak and determine its through-bond sequential correlations.

2.6 Isothermal titration calorimetry

Calorimetry can be used to study biomolecular interactions by monitoring changes in sample temperature, since the uptake or release of heat upon addition of ligand to a protein sample is proportional to the extent of binding. In *isothermal titration calorimetry* (ITC), the measured parameter is the difference in energy required to return the reference and sample cells, respectively, to the initial temperature following addition of ligand.

The non-covalent binding of a ligand, L , to a protein, P , to form the protein-ligand complex, PL , is described by

The equilibrium dissociation constant, K_D , of the interaction is given by

$$K_D = \frac{[P][L]}{[PL]}$$

and describes the extent of protein-ligand complex formation relative to the concentrations of the reactant species at equilibrium. The extent of binding depends on the strength of the interactions within the protein-ligand complex relative to the strength of their respective interactions with water when free in solution. While the formation of weak interactions between protein and ligand make enthalpic contributions that decrease the free energy of the system, binding constricts the conformational entropy of the ligand which increases the free energy. However, solvent water associated with the binding site has to be released to allow the interactions between protein and ligand to form, and this release of water increases net entropy of the system. Consequently, the extent of binding can be related to the thermodynamics of protein-ligand complex formation, as described by

$$\Delta G = -RT \ln(K_D) = \Delta H - T\Delta S$$

where ΔG is the difference in Gibbs free energy, R is the universal gas constant, T is the temperature in Kelvin, K_D is the equilibrium dissociation constant, ΔH is the difference in enthalpy, and ΔS is the difference in entropy.

By monitoring the titration of a ligand to a protein sample using ITC, it is possible to obtain information on the affinity, thermodynamic parameters (enthalpy, entropy, free energy), and stoichiometry of binding from the resulting isotherm, and the technique requires no labeling or extrinsic probes.

2.7 Enzyme kinetics

Enzyme kinetics is the measurement of the rate of enzyme catalyzed reactions under a particular set of conditions, where the measured parameter is often the formation of product or the disappearance of substrate per unit time. The Michaelis-Menten model of single-substrate enzyme kinetics assumes the formation of an enzyme-substrate (ES) complex as an intermediate step in catalysis [58, 59], as illustrated by the reaction scheme

For an enzyme catalyzed reaction, free enzyme (E) binds substrate (S) to form the ES complex with rate constant k_1 . Once formed, the ES complex can either dissociate to free enzyme and substrate with rate constant k_{-1} , or proceed to form product (P) with rate constant k_2 , or k_{cat} . Due to the reversibility of many biochemical reactions, the ES complex can also be formed by the decay of enzyme-bound product with rate k_{-2} . However, at the beginning of the reaction ($t \sim 0$), no significant amounts of product have yet been formed, which permits simplification of the model by omission of k_{-2} . The reaction rate, v , is then described by the Michaelis-Menten equation

$$v = \frac{V_{max}[S]}{[S] + K_m}$$

where V_{max} is the maximum reaction rate, $[S]$ is the substrate concentration, and K_m is the Michaelis constant defined as

$$K_m = \frac{k_{-1} + k_{cat}}{k_1}$$

The maximum reaction rate, V_{max} , is given by

$$V_{max} = k_{cat}[E]_{tot}$$

where $[E]_{tot}$ is the total enzyme concentration, or more specifically, the number of active sites. V_{max} is achieved at saturating substrate concentrations where all enzyme occurs as ES complex, and the overall reaction rate equals the turnover number, k_{cat} .

The kinetic parameters described above provide mechanistic and thermodynamic information on the catalytic process. K_m is the substrate concentration at half the maximum reaction rate, and is indicative of the extent of ES complex formation. A low K_m value suggests a high affinity for the substrate since the ES complex is readily formed at low substrate concentrations, and *vice versa*. K_m is therefore a measure of the substrate concentration needed to achieve a significant rate of catalysis. It follows that at substrate concentrations lower than K_m , the rate of catalysis is much less than k_{cat} since the enzyme is far from saturated.

According to transition-state theory, the observed rate can be related to the activation energy of the reaction as described by the equation

$$k_{obs} = \frac{k_B T}{h} e^{\frac{-\Delta G^\ddagger}{RT}}$$

where k_{obs} is the observed rate constant, k_B is Boltzmann's constant, h is Planck's constant, $\Delta G^\ddagger$ is the free energy of activation, R is the universal gas constant, and T is the temperature in Kelvin. The obtained rate constants can therefore be used to calculate the activation energies associated with the various steps involved in the enzyme catalyzed conversion of substrate to product. The turnover number k_{cat} is the maximum rate of product formation per active site, and is proportional to the difference in free energy between the ES complex and the transition state of the reaction. The specificity constant, k_{cat}/K_m , takes into account both the active site occupancy and the extent of product formation from the ES complex, and the ratio is proportional to the difference in free energy between free enzyme and substrate, and the transition state of the reaction. The value of k_{cat}/K_m is limited by the encounter-based formation of ES complex, with the rate of diffusion setting the upper limit. k_{cat}/K_M is a substrate-specific measure of the catalytic efficiency of an enzyme, where a high value indicates that

nearly every encounter results in the formation of ES complex and subsequent catalytic conversion of substrate to product. If the enzyme acts on more than one substrate, k_{cat}/K_M is indicative of which substrate forms the most favorable interactions with the enzyme and will proceed to form product to a greater extent.

2.8 Enzyme kinetics meets ITC

In cases where an enzyme catalyzed reaction proceeds at a suitable rate and involves an observable release or uptake of energy, it is possible to measure the kinetics of substrate turnover using ITC since the reaction rate, v , is directly proportional to the time-dependent change in sample heat (dQ/dt) as described by

$$v = \frac{1}{\Delta H_{app} V_{cell}} \frac{dQ}{dt}$$

where ΔH_{app} is the total molar enthalpy and V_{cell} is the volume of the sample cell [60]. ΔH_{app} is the total molar enthalpy of the overall reaction, and this parameter is determined experimentally by monitoring the complete turnover of small amounts of substrate by the enzyme, where the area of the resulting peak corresponds to the total molar enthalpy, ΔH_{app} , of the overall reaction. By measuring enzyme kinetics using ITC, the information obtained from the rate constants and the overall thermodynamic parameters can be combined to derive a complete mechanistic profile of the enzyme catalyzed reaction.

3

Summary and Discussion

3.1 Biophysical characterization of variants

Importance of the D151 hydrogen bonding network

TPMT*6 (Y180F) has been classified as a low activity variant [61], but previous publications have reported conflicting enzyme activities and protein levels in mammalian cells [62, 63]. No extensive biophysical characterization of this protein variant has been published to date. We have found that TPMT*6 is somewhat destabilized compared to the wild-type protein, although the variant is still able to adopt the correct folding pattern and exhibits normal functionality.

The destabilization of TPMT*6 is not surprising considering the very nature of the amino acid substitution and its localization in the protein structure. Y180 is situated on strand $\beta 7$ in the central β -sheet of the protein core, where the side chain participates in both hydrogen bonding and packing interactions. While the latter are retained upon substitution to Phenylalanine, deletion of the Tyrosine hydroxyl group due to mutation precludes hydrogen bonding. Buried hydrogen-bonded Tyrosine hydroxyls have been shown to contribute greatly to overall protein stability [3], which will inevitably decrease upon substitution to Phenylalanine. The Y180 hydroxyl group is hydrogen bonded to the side chain carboxyl of D151, which also coordinates interactions with the L155 backbone amide and the side chain hydroxyl of Y166. Together, these interacting residues form the D151 hydrogen bonding network (shown in Figure 3.1). Interestingly, the participating residues

also co-localize with the SAM-MT conserved sequence motifs IV (D151 and L155), V (Y166), and VI (Y180). Furthermore, the D151 side chain carboxyl constitutes a buried charge, which would be unfavorable should it not be partially neutralized by hydrogen bonding to Y166 and Y180 (N-capping of the α F mini-helix through hydrogen bonding to the L155 backbone amide neutralizes the remaining charge). The importance of the D151 hydrogen bonding network is not only illustrated by the reduced stability of TPMT*6, but also by the biophysical properties of other variant proteins where mutations have affected the hydrogen bonding interactions. Substitution of A154 to Threonine (TPMT*3B) has been shown to cause distortion of helix α G, harboring Y166, which disrupts the interactions of the D151 hydrogen bonding network and renders the TPMT*3B protein variant severely destabilized [64]. Similarly, disruption of the D151 hydrogen bonding network by substitution of A167 to Glycine (TPMT*23) also caused distortion of helix α G with deleterious effects to the structure and stability of this protein variant [65]. Also, a recently characterized TPMT variant (Y166H) has been found to be destabilized compared to the wild-type protein (L-G. Mårtensson, personal communication).

Considering the effects of perturbation to the D151 hydrogen bonding network, the reason for the low *in vivo* enzyme activity of TPMT*6 is likely related to the cellular longevity of this variant. Destabilization due to disruption of the D151 hydrogen bonding network probably causes the quality control functions of the cell to target the variant protein for degradation. Other destabilized TPMT variants have been shown to be susceptible to proteosomal degradation which reduces the cellular levels of these proteins [34, 35, 36]. Therefore, the conflicting results regarding enzyme activity and protein levels reported by Salavaggione *et al* [62] and Ujjie *et al* [63] can likely be explained by the reduced intrinsic stability of this variant, since both measured parameters depend on the extent of cellular clearance in response to decreased protein stability.

TPMT*8 (R215H) similar to wild-type

We have found that TPMT*8 (R215H) exhibits similar biophysical characteristics as the wild-type protein, with no evidence of perturbations to tertiary structure, stability, or functionality. Previous publications have reported decreased enzyme activity and lowered protein levels in mammalian cells [62, 63, 66], but these findings cannot be explained by

Figure 3.1 The D151 hydrogen bonding network. An excerpt from TPMT tertiary structure is depicted in ribbon representation in gray. SAH and the side chains of selected residues are shown in stick representation and colored by element.

the molecular properties of the variant protein, although there may be other factors at play beyond the scope of our research.

R215 is situated on strand $\beta 8$ at the edge of the central β -sheet, facing solvent. In the crystal structure (PDB ID:2BZG), R215 packs against the side chains of L217 ($\beta 8$) and L241 ($\beta 9$), although these interactions could potentially be maintained in TPMT*8 since the Histidine side chain is of adequate length. Molecular modeling simulations have shown that TPMT has an inherent surface dipole moment [29], but substitution of Arginine to Histidine does not alter side chain polarity, and should therefore not affect the electrostatic properties of the protein surface. Although it cannot be completely ruled out that there might be perturbations to local structure, the stability and functionality of TPMT*8 are not affected by mutation, and the decreased *in vivo* activity of this variant cannot be explained by our findings.

3.2 What's the matter with SAM?

Although S-adenosylmethionine (SAM) is per definition a co-substrate of TPMT, it is common (and historical) practice within the field of thiopurine research to refer to SAM as a cofactor of the protein. Therefore, biochemical formalities are set aside (although duly noted) and SAM will henceforth be referred to as a cofactor. Regardless of its label, the contributions of SAM to the overall stability of TPMT have lately been the focus of a lot of discussion within the field. However, no observations of thermodynamic stabilization of TPMT by SAM binding has (to our knowledge) been published to date, although studies have demonstrated a decrease in the cellular longevity of the protein in cofactor deficient cells [67]. We have therefore investigated properties of SAM binding to TPMT and found that despite exhibiting strong binding to the protein ($K_D \sim 2 \mu\text{M}$), SAM does not contribute significantly to the stability of TPMT under our experimental conditions. Similar (lack of) effects on T_m despite strong binding of SAM have also been observed for the structurally related histamine N-methyltransferase (HNMT) protein [68]. While one might expect a high-affinity ligand like SAM to make a noticeable contribution to the thermal stability of TPMT, it is important to note that the absence of a measurable increase in T_m does not necessarily exclude the possibility of an actual thermodynamic stabilization of the protein by ligand binding. For instance, a ligand that also interacts with the unfolded state of a protein would result in a lower observed T_m since stabilization of the unfolded state through ligand interaction shifts the equilibrium. It has been demonstrated by thermal unfolding simulations that strong enthalpically driven ligand binding to the native state could be obscured by entropically driven weak interactions with the unfolded state, which could (in some cases) even decrease the observed T_m compared to the apoprotein [69]. The same study also simulated the effects of different binding energetics on the observed T_m , and showed that an increase in T_m does not necessarily equal high binding affinity at physiological temperatures, and *vice versa*. They also showed that while entropically driven binding affected T_m dramatically, enthalpically driven binding had only marginal effects on T_m . These findings could offer an explanation as to why no apparent increased thermal stability in the presence of SAM can be observed.

It is also possible that the observed increase in cellular longevity is due to a mechanism other than thermodynamic stabilization of the enzyme-

cofactor complex. Interestingly, the exposure of reactive Cysteine thiols have been shown to increase over time in the absence of SAM [70], and molecular dynamics simulations have revealed that removal of SAH leads to loss of structure of helix αE and increases the solvent exposure of the SAM binding site [64]. While this could be a potential mechanism to allow dissociation of coproduct and re-binding of cofactor prior to the next catalytic cycle, it could also be the cause of the increased cellular longevity of TPMT in the presence of SAM. Binding of cofactor would result in decreased structural fluctuations, shifting the equilibrium towards more ordered conformations. Conversely, depletion of SAM in cells starved of cofactor precursors would cause the apoprotein to appear misfolded and targeted for degradation.

3.3 Improving methodology

Structural exploration

To gain insight into the structural properties of TPMT, we have attempted to use NMR spectroscopy. However, more elaborate studies using this measurement technique require (more or less) complete resonance assignments, which we have been unable to accomplish. Although the protein produces well-dispersed peaks in HSQC spectra, we have only been able to achieve partial resonance assignments and, unfortunately, have not been able to identify the spectral peaks of the residues surrounding the active site. Some unassigned regions overlap with regions with high B-factors in the crystal structure (PDB ID:2BZG), and other unassigned residues exhibit millisecond dynamics, which could suggest low signal intensity due to structural flexibility.

It is possible that ordering of TPMT structure by cofactor binding could have been helpful when acquiring NMR spectra. The difference in thiol exposure between the apoprotein and the protein-cofactor complex observed by Lysaa *et al* was readily apparent in under 12 hours [70], which is significantly less than the time required to run triple-resonance NMR experiments on a (double labeled) protein of this size. It is therefore reasonable to suspect that a decrease in structural fluctuations by cofactor binding could have alleviated some of the issues regarding lack of connectivities in triple-resonance spectra.

The delights of ANS

In our previous studies, we have relied on far-UV CD spectroscopy to determine the effects of mutation. However, this method is limited to measurements of the overall secondary structure content. While the information obtained by CD spectroscopy is still valuable, it does not provide detailed information on the active site of TPMT. Unfortunately, it is not feasible to use intrinsic fluorescence to monitor structural events since there is essentially no red-shifting of fluorescence emission during unfolding of this protein. This is most likely due to the solvent accessible localization of four out of five Tryptophan residues in the protein structure (Figure 3.2). W29 and W33 are situated in helix αA (covering the active site), participating in packing and hydrogen bonding interactions with the cofactor SAM. W230 (helix αI) sits at the mouth of the channel leading into the active site, with the pyrrole nitrogen facing solvent and the remainder of the side chain partially buried. Similarly, W78 (helix αC) is located on the surface of the protein, with its pyrrole nitrogen facing solvent and the rest of the side chain partially buried. Only W150 ($\beta 6$) is completely buried within the protein core, engaged in packing interactions, but contributions to the emission spectrum from this residue (or other aromatic side chains) are not readily observable due to the overwhelming fluorescence intensity of the solvent accessible residues.

To be able to study the active site of TPMT, we have developed a method in which 8-anilidonaphthalene-1-sulfonic acid (ANS) is used to probe the tertiary structure of this protein. ANS binds exclusively to the native state of TPMT (with a K_D of $0.2 \mu M$) and specifically to the active site, co-localizing with the adenine moiety of SAM. Binding of ANS to TPMT dramatically increases the stability of the protein, and it is possible that co-crystallization of the full-length protein with ANS could be successful. Similarly, ANS could increase the stability of TPMT samples in NMR experiments, which might improve the quality of the resulting spectra enough to allow complete resonance assignment.

Another potential use of ANS is to serve as a probe to gauge the condition of protein samples and determine the number of active sites available prior to measurements of enzyme activity or inhibition, which is an important aspect of TPMT research given the clinical significance of this protein. Furthermore, considering the structural relationship and conservation of cofactor binding characteristics among SAM-MTs, it might

Figure 3.2 Localization of Tryptophan residues in TPMT tertiary structure. Tryptophan side chains are shown as sticks and colored in aquamarine, the protein crystal structure (PDB ID:2BZG) is shown in ribbon representation in gray, and the cofactor SAM is shown as sticks and colored by element.

be possible to study other proteins belonging to the SAM-MT superfamily using ANS. In summary, ANS is an excellent tool to probe the active site integrity of TPMT, and this method in combination with CD spectroscopy provides a comprehensive view of the overall biophysical properties of TPMT variant proteins.

The versatility of ITC

We have used isothermal titration calorimetry to characterize the interactions between TPMT and ligands frequently encountered in the field. We also present a novel TPMT enzyme activity assay based on this measurement technique. Using this approach, we have characterized

the kinetics of 6-MP and 6-TG methylation by TPMT. The results are in excellent agreement with the rates obtained using the radiochemical assay, and the calculated activation energy corresponds to previously reported values [28].

To further improve the precision and accuracy of functional characterization of TPMT and its variants, determination of the kinetic parameters could also be combined with batch-wise determination of the number of active sites available by characterization of ANS binding. This would allow the enzyme activity to be correlated with the number of active protein molecules (rather than total protein concentration), since the condition of protein samples vary between purification batches.

We have also investigated the properties of the most commonly used protein constructs. It was found that truncation had more pronounced effects on stability (as determined by thermal denaturation monitored by CD spectroscopy) and enzymatic function than the presence of a fusion His-tag, which did not alter the thermal stability and had only minor effects on k_{cat} . Truncation on the other hand significantly increased the thermal stability of the protein construct, and while K_{m} was not affected, k_{cat} was reduced, suggesting that the N-terminal might be involved in transition-state stabilization.

References

- [1] Peter L Privalov and Stanley J Gill. “Stability of protein structure and hydrophobic interaction”. In: *Advances in protein chemistry* 39 (1988), pp. 191–234.
- [2] C Nick Pace. “Polar group burial contributes more to protein stability than nonpolar group burial”. In: *Biochemistry* 40.2 (2001), pp. 310–313.
- [3] C Nick Pace, Geoffrey Horn, Eric J Hebert, John Bechert, Kevin Shaw, Lubica Urbanikova, J Martin Scholtz, and Jozef Sevcik. “Tyrosine hydrogen bonds make a large contribution to protein stability”. In: *Journal of Molecular Biology* 312.2 (2001), pp. 393–404.
- [4] Tapan K Chaudhuri and Subhankar Paul. “Protein-misfolding diseases and chaperone-based therapeutic approaches”. In: *FEBS Journal* 273.7 (2006), pp. 1331–1349.
- [5] Christopher M Dobson. “Protein misfolding, evolution and disease”. In: *Trends in biochemical sciences* 24.9 (1999), pp. 329–332.
- [6] Heidi L. Schubert, Robert M. Blumenthal, and Xiaodong Cheng. “Many paths to methyltransfer: a chronicle of convergence”. In: *Trends in Biochemical Sciences* 28.6 (Mar. 2017), pp. 329–335.
- [7] Jennifer L Martin and Fiona M McMillan. “SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold”. In: *Current Opinion in Structural Biology* 12.6 (2002), pp. 783–793.
- [8] Piotr Z. Kozbial and Arcady R. Mushegian. “Natural history of S-adenosylmethionine-binding proteins”. In: *BMC Structural Biology* 5.1 (2005), p. 19.
- [9] Hong Wu et al. “Structural basis of allele variation of human thiopurine-S-methyltransferase”. In: *Proteins: Structure, Function, and Bioinformatics* 67.1 (2007), pp. 198–208.

- [10] Richard M Weinshilboum and Susan L Sladek. “Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity”. In: *American journal of human genetics* 32.5 (Sept. 1980), pp. 651–662.
- [11] L. Lennard, J. S. Lilleyman, J. Van Loon, and R. M. Weinshilboum. “Genetic variation in response to 6-mercaptopurine for childhood acute lymphoblastic leukaemia”. In: *Lancet* 336.8709 (July 1990), pp. 225–229.
- [12] Liisa Holm and Päivi Rosenström. “Dali server: conservation mapping in 3D”. In: *Nucleic Acids Research* 38.suppl_2 (2010), W545.
- [13] L. Wang and R. Weinshilboum. “Thiopurine S-methyltransferase pharmacogenetics: insights, challenges and future directions”. In: *Oncogene* 25.11 (2006), pp. 1629–1638.
- [14] Alan Kambiz Fotoohi, Sally A. Coulthard, and Freidoune Albertioni. “Thiopurines: Factors influencing toxicity and response”. In: *Biochemical Pharmacology* 79.9 (2010), pp. 1211–1220.
- [15] Lynne Lennard. “Implementation of TPMT testing”. In: *British journal of clinical pharmacology* 77.4 (2014), pp. 704–714.
- [16] Natalia F. Krynetskaia, Eugene Y. Krynetski, and William E. Evans. “Human RNase H-Mediated RNA Cleavage from DNA-RNA Duplexes Is Inhibited by 6-Deoxythioguanosine Incorporation into DNA”. In: *Molecular Pharmacology* 56.4 (1999), pp. 841–848.
- [17] Natalia F. Krynetskaia, Xiangjun Cai, John L. Nitiss, Eugene Y. Krynetski, and Mary V. Relling. “Thioguanine substitution alters DNA cleavage mediated by topoisomerase II”. In: *The FASEB Journal* 14.14 (2000), pp. 2339–2344.
- [18] Lynne Lennard, Jon A Van Loon, John S Lilleyman, and Richard M Weinshilboum. “Thiopurine pharmacogenetics in leukemia: Correlation of erythrocyte thiopurine methyltransferase activity and 6-thioguanine nucleotide concentrations”. In: *Clinical Pharmacology & Therapeutics* 41.1 (1987), pp. 18–25.
- [19] Malin Lindqvist Appell et al. “Nomenclature for alleles of the thiopurine methyltransferase gene”. In: *Pharmacogenetics and genomics* 23.4 (2013), p. 242.
- [20] Diane M Otterness, Carol L Szumlanski, Thomas C Wood, and Richard M Weinshilboum. “Human thiopurine methyltransferase pharmacogenetics. Kindred with a terminal exon splice junction mutation that results in loss of activity”. In: *The Journal of Clinical Investigation* 101.5 (Mar. 1998), pp. 1036–1044.

- [21] Malin Lindqvist, Sofie Haglund, Sven Almer, Curt Peterson, Jan Taipalensu, Erik Hertervig, Ebbe Lyrenäs, and Peter Söderkvist. "Identification of two novel sequence variants affecting thiopurine methyltransferase enzyme activity". In: *Pharmacogenetics and Genomics* 14.4 (2004), pp. 261–265.
- [22] Mary V Relling, Michael L Hancock, James M Boyett, Ching-Hon Pui, and William E Evans. "Prognostic importance of 6-mercaptopurine dose intensity in acute lymphoblastic leukemia". In: *Blood* 93.9 (1999), pp. 2817–2823.
- [23] Martin Stanulla et al. "Thiopurine methyltransferase (TPMT) genotype and early treatment response to mercaptopurine in childhood acute lymphoblastic leukemia". In: *Jama* 293.12 (2005), pp. 1485–1489.
- [24] Eugene Y Krynetski, Natalia F Krynetskaia, Yuri Yanishevski, and William E Evans. "Methylation of mercaptopurine, thioguanine, and their nucleotide metabolites by heterologously expressed human thiopurine S-methyltransferase". In: *Molecular Pharmacology* 47.6 (1995), pp. 1141–1147.
- [25] Mario HJ Vogt, Elisabet H Stet, Ronney A De Abreu, Jos PM Bökerink, Lambert HJ Lambooy, and Frans JM Trijbels. "The importance of methylthio-IMP for methylmercaptopurine ribonucleoside (Me-MPR) cytotoxicity in Molt F4 human malignant T-lymphoblasts". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1181.2 (1993), pp. 189–194.
- [26] BS Tay, R McC Lilley, AW Murray, and MR Atkinson. "Inhibition of phosphoribosyl pyrophosphate amidotransferase from Ehrlich ascites-tumour cells by thiopurine nucleotides". In: *Biochemical Pharmacology* 18.4 (1969), pp. 936–938.
- [27] Xiao-Liang Pan, Feng-Chao Cui, and Jing-Yao Liu. "Quantum Mechanical/Molecular Mechanical Molecular Dynamics and Free Energy Simulations of the Thiopurine S-Methyltransferase Reaction with 6-Mercaptopurine". In: *The Journal of Physical Chemistry B* 115.24 (2011), pp. 8033–8037.
- [28] Yi Peng, Qiping Feng, Dennis Wilk, Araba A Adjei, Oreste E Salavaggi-one, Richard M Weinshilboum, and Vivien C Yee. "Structural Basis of Substrate Recognition in Thiopurine S-Methyltransferase". In: *Biochemistry* 47.23 (2008), pp. 6216–6225.
- [29] A Garat et al. "Characterisation of novel defective thiopurine S-methyltransferase allelic variants". In: *Biochemical pharmacology* 76.3 (2008), pp. 404–415.

- [30] Rima Hamdan-Khalil, Jean-Luc Gala, Delphine Allorge, Jean-Marc Lo-Guidice, Yves Horsmans, Nicole Houdret, and Franck Broly. "Identification and functional analysis of two rare allelic variants of the thiopurine S-methyltransferase gene, TPMT*16 and TPMT*19". In: *Biochemical pharmacology* 69.3 (2005), pp. 525–529.
- [31] Malin Lindqvist Appell, Patricia Wennerstrand, Curt Peterson, Erik Hertervig, and Lars-Göran Mårtensson. "Characterization of a novel sequence variant, TPMT*28, in the human thiopurine methyltransferase gene". In: *Pharmacogenetics and genomics* 20.11 (2010), pp. 700–707.
- [32] Lee C Woodson and Richard M Weinshilboum. "Human kidney thiopurine methyltransferase purification and biochemical properties". In: *Biochemical pharmacology* 32.5 (1983), pp. 819–826.
- [33] Patricia Wennerstrand, Paolo Dametto, Janosch Hennig, Thérèse Klingstedt, Karin Skoglund, Malin Lindqvist Appell, and Lars-Göran Mårtensson. "Structural characteristics determine the cause of the low enzyme activity of two thiopurine S-methyltransferase allelic variants: a biophysical characterization of TPMT*2 and TPMT*5". In: *Biochemistry* 51.30 (2012), pp. 5912–5920.
- [34] Liewei Wang, Tien V Nguyen, Richard W McLaughlin, Laura A Sikkink, Marina Ramirez-Alvarado, and Richard M Weinshilboum. "Human thiopurine S-methyltransferase pharmacogenetics: variant allozyme misfolding and aggresome formation". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.26 (2005), pp. 9394–9399.
- [35] Hung-Liang Tai, Eugene Y Krynetski, Erin G Schuetz, Yuri Yanishevski, and William E Evans. "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". In: *Proceedings of the National Academy of Sciences* 94.12 (1997), pp. 6444–6449.
- [36] Qiping Feng, Suda Vannaprasaht, Yi Peng, Susothorn Angsuthum, Yingyos Avihingsanon, Vivien C Yee, Wichitra Tassaneeyakul, and Richard M Weinshilboum. "Thiopurine S-methyltransferase pharmacogenetics: functional characterization of a novel rapidly degraded variant allozyme". In: *Biochemical pharmacology* 79.7 (2010), pp. 1053–1061.
- [37] Patricia Wennerstrand, Lars-Göran Mårtensson, Stefan Söderhäll, Anna Zimdahl, and Malin Lindqvist Appell. "Methotrexate binds to recombinant thiopurine S-methyltransferase and inhibits enzyme activity after high-dose infusions in childhood leukaemia". In: *European journal of clinical pharmacology* 69.9 (2013), pp. 1641–1649.

- [38] RA Lysaa, T Giverhaug, H Libaek Wold, and J Aarbakke. "Inhibition of human thiopurine methyltransferase by furosemide, bendroflumethiazide and trichlormethiazide". In: *European journal of clinical pharmacology* 49.5 (1996), pp. 393–396.
- [39] Kersti Oselin and Kaili Anier. "Inhibition of human thiopurine S-methyltransferase by various nonsteroidal anti-inflammatory drugs in vitro: a mechanism for possible drug interactions". In: *Drug Metabolism and Disposition* 35.9 (2007), pp. 1452–1454.
- [40] LC Woodson, MM Ames, CD Selassie, C Hansch, and Richard M Weinshilboum. "Thiopurine methyltransferase. Aromatic thiol substrates and inhibition by benzoic acid derivatives." In: *Molecular pharmacology* 24.3 (1983), pp. 471–478.
- [41] Monika Deininger, Carol L Szumlanski, Diane M Otterness, Jon Van Loon, Wolfgang Ferber, and Richard M Weinshilboum. "Purine substrates for human thiopurine methyltransferase". In: *Biochemical pharmacology* 48.11 (1994), pp. 2135–2138.
- [42] Wei-Guo Qing, K Leslie Powell, George Stoica, Carol L Szumlanski, Richard M Weinshilboum, and Michael C Macleod. "Toxicity and metabolism in mice of 2, 6-dithiopurine, a potential chemopreventive agent." In: *Drug metabolism and disposition* 23.8 (1995), pp. 854–860.
- [43] AL Kerremans, JJ Lipsky, J Van Loon, MO Gallego, and Richard M Weinshilboum. "Cephalosporin-induced hypoprothrombinemia: possible role for thiol methylation of 1-methyltetrazole-5-thiol and 2-methyl-1, 3, 4-thiadiazole-5-thiol." In: *Journal of Pharmacology and Experimental Therapeutics* 235.2 (1985), pp. 382–388.
- [44] Thomas C Wood, Kenneth L Johnson, Stephen Naylor, and Richard M Weinshilboum. "Cefazolin administration and 2-methyl-1, 3, 4-thiadiazole-5-thiol in human tissue: possible relationship to hypoprothrombinemia". In: *Drug metabolism and disposition* 30.10 (2002), pp. 1123–1128.
- [45] Todd A Glauser, AN Nelson, DE Zembower, JJ Lipsky, and Richard M Weinshilboum. "Diethyldithiocarbamate S-methylation: evidence for catalysis by human liver thiol methyltransferase and thiopurine methyltransferase." In: *Journal of Pharmacology and Experimental Therapeutics* 266.1 (1993), pp. 23–32.
- [46] Lisa M Cannon, Felice N Butler, Wei Wan, and Zhaohui Sunny Zhou. "A stereospecific colorimetric assay for (S, S)-adenosylmethionine quantification based on thiopurine methyltransferase-catalyzed thiol methylation". In: *Analytical biochemistry* 308.2 (2002), pp. 358–363.

- [47] Matthew M Ames, Cynthia Dias Selassie, Lee C Woodson, Jon A Van Loon, Corwin Hansch, and Richard M Weinshilboum. "Thiopurine methyltransferase: structure-activity relationships for benzoic acid inhibitors and thiophenol substrates." In: *Journal of medicinal chemistry* 29.3 (1986), pp. 354–358.
- [48] PA Blaker, M Arenas-Hernandez, MA Smith, EA Shobowale-Bakre, L Fairbanks, PM Irving, JD Sanderson, and AM Marinaki. "Mechanism of allopurinol induced TPMT inhibition". In: *Biochemical pharmacology* 86.4 (2013), pp. 539–547.
- [49] Carol L Szumlanski and Richard M Weinshilboum. "Sulphasalazine inhibition of thiopurine methyltransferase: possible mechanism for interaction with 6-mercaptopurine and azathioprine." In: *British journal of clinical pharmacology* 39.4 (1995), pp. 456–459.
- [50] Norma J Greenfield. "Using circular dichroism spectra to estimate protein secondary structure". In: *Nature protocols* 1.6 (2006), pp. 2876–2890.
- [51] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. "How to study proteins by circular dichroism". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1751.2 (2005), pp. 119–139.
- [52] Lars-Göran Mårtensson, Martin Karlsson, and Uno Carlsson. "Dramatic stabilization of the native state of human carbonic anhydrase II by an engineered disulfide bond". In: *Biochemistry* 41.52 (2002), pp. 15867–15875.
- [53] Roberto Favilla, Matteo Goldoni, Alberto Mazzini, Paolo Di Muro, Benedetto Salvato, and Mariano Beltramini. "Guanidinium chloride induced unfolding of a hemocyanin subunit from *Carcinus aestuarii*: I. Apo form". In: *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1597.1 (2002), pp. 42–50.
- [54] Ramil F Latypov, Dingjiang Liu, Kannan Gunasekaran, Timothy S Harvey, Vladimir I Razinkov, and Andrei A Raibekas. "Structural and thermodynamic effects of ANS binding to human interleukin-1 receptor antagonist". In: *Protein Science* 17.4 (2008), pp. 652–663.
- [55] M Soledad Celej, Sergio A Dassie, Eleonora Freire, M Lucia Bianconi, and Gerardo D Fidelio. "Ligand-induced thermostability in proteins: Thermodynamic analysis of ANS–albumin interaction". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1750.2 (2005), pp. 122–133.

-
- [56] Jeramia J Ory and Leonard J Banaszak. “Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate”. In: *Biophysical journal* 77.2 (1999), pp. 1107–1116.
- [57] Roland Einarsson, Hans Eklund, Eila Zeppezauer, Torne Boiwe, and Carl-Ivar Brändén. “Binding of Salicylate in the Adenosine-Binding Pocket of Dehydrogenases”. In: *European journal of biochemistry* 49.1 (1974), pp. 41–47.
- [58] Leonor Michaelis and Maud L Menten. “Die kinetik der invertinwirkung”. In: *Biochemische Zeitschrift* 49.333-369 (1913), p. 352.
- [59] Kenneth A Johnson and Roger S Goody. “The original Michaelis constant: translation of the 1913 Michaelis–Menten paper”. In: *Biochemistry* 50.39 (2011), pp. 8264–8269.
- [60] Matthew J Todd and Javier Gomez. “Enzyme kinetics determined using calorimetry: a general assay for enzyme activity?” In: *Analytical biochemistry* 296.2 (2001), pp. 179–187.
- [61] Diane Otterness et al. “Human thiopurine methyltransferase pharmacogenetics: gene sequence polymorphisms”. In: *Clinical Pharmacology & Therapeutics* 62.1 (1997), pp. 60–73.
- [62] Oreste E Salavaggione, Liewei Wang, Mathieu Wiepert, Vivien C Yee, and Richard M Weinshilboum. “Thiopurine S-methyltransferase pharmacogenetics: variant allele functional and comparative genomics”. In: *Pharmacogenetics and genomics* 15.11 (2005), pp. 801–815.
- [63] Shuta Ujiie, Takamitsu Sasaki, Michinao Mizugaki, Masaaki Ishikawa, and Masahiro Hiratsuka. “Functional characterization of 23 allelic variants of thiopurine S-methyltransferase gene (TPMT*2–*24)”. In: *Pharmacogenetics and genomics* 18.10 (2008), pp. 887–893.
- [64] Karen Rutherford and Valerie Daggett. “Four human thiopurine s-methyltransferase alleles severely affect protein structure and dynamics”. In: *Journal of molecular biology* 379.4 (2008), pp. 803–814.
- [65] Sofiene Larif, Chaker Ben Salem, Zohra Soua, Housseem Hmouda, and Kamel Bouraoui. “Insight into TPMT*23 mutation mis-folding using molecular dynamics simulation and protein structure analysis”. In: *Journal of Biomolecular Structure and Dynamics* 31.10 (2013), pp. 1066–1076.

- [66] Yuen Yi Hon, Michael Y Fessing, Ching-Hon Pui, Mary V Relling, Eugene Y Krynetski, and William E Evans. “Polymorphism of the thiopurine S-methyltransferase gene in African-Americans”. In: *Human molecular genetics* 8.2 (1999), pp. 371–376.
- [67] Miha Milek, Alenka Smid, Riin Tamm, Natasa Karas Kuzelicki, Andres Metspalu, and Irena Mlinaric-Rascan. “Post-translational stabilization of thiopurine S-methyltransferase by S-adenosyl-L-methionine reveals regulation of TPMT*1 and *3C allozymes”. In: *Biochemical pharmacology* 83.7 (2012), pp. 969–976.
- [68] Chanakan Tongsook, Johannes Niederhauser, Elena Kronegger, Grit Straganz, and Peter Macheroux. “Leucine 208 in human histamine N-methyltransferase emerges as a hotspot for protein stability rationalizing the role of the L208P variant in intellectual disability”. In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1863.1 (2017), pp. 188–199.
- [69] Travis T Waldron and Kenneth P Murphy. “Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics”. In: *Biochemistry* 42.17 (2003), pp. 5058–5064.
- [70] Roy A Lysaaa, David J Warren, Ingebrigt Sylte, and Jarle Aarbakke. “Effect of the glutathione/glutathione disulfide redox couple on thiopurine methyltransferase”. In: *Biochemical pharmacology* 61.6 (2001), pp. 707–714.

Papers

The articles associated with this thesis have been removed for copyright reasons.

For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-136558>