Published in final edited form as:

Proteins. 2007 April 1; 67(1): 198–208. doi:10.1002/prot.21272.

Structural Basis of Allele Variation of Human Thiopurine-S-Methyltransferase

Hong Wu¹, John R. Horton², Kevin Battaile³, Abdellah Allali-Hassani¹, Fernando Martin¹, Hong Zeng¹, Peter Loppnau¹, Masoud Vedadi¹, Alexey Bochkarev^{1,4,5,6}, Alexander N. Plotnikov^{1,7,*}, and Xiaodong Cheng^{2,*}

¹ Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L5, Canada

² Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322

³ IMCA-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439

⁴ Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada

⁵ Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada

⁶ Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma

⁷ Department of Physiology, University of Toronto, Toronto, Ontario, Canada

Abstract

Human thiopurine S-methyltransferase (TPMT) exhibits considerable person-to-person variation in activity to thiopurine drugs. We have produced an N-terminal truncation of human TPMT protein, crystallized the protein in complex with the methyl donor product S-adenosyl-L-homocysteine, and determined the atomic structure to the resolution of 1.58 and 1.89 Å, respectively, for the selenomethionine incorporated and wild type proteins. The structure of TPMT indicates that the naturally occurring amino acid polymorphisms scatter throughout the structure, and that the amino acids whose alteration have the most influence on function are those that form intra-molecular stabilizing interactions (mainly van der Waals contacts). Furthermore, we have produced four TPMT mutant proteins containing variant alleles of TPMT*2, *3A, *3B, and *3C and examined the structure-function relationship of the mutant proteins based on their expression and solubility in bacteria and their thermostability profile.

Keywords

thiopurine drugs; S-methylation; pharmacogenetics; polymorphisms; crystallography; structural genomics

INTRODUCTION

Thiopurine S-methyltransferase (TPMT) catalyzes thiopurine and thiopyrimidine S-methylation, an important metabolic pathway for thiopurine drugs, such as 6-mercaptopurine

*Correspondence to: Xiaodong Cheng, Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322. E-mail: xcheng@emory.edu or Alexander N. Plotnikov, Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L5, Canada.

(6-MP), 6-thioguanine (6-TG), and azathioprine, currently used to treat childhood acute lymphoblastic leukemia, autoimmune disorders, inflammatory bowel disease, and transplant-rejection.¹ While its natural substrate is still unknown, TPMT has long been recognized for its major role in the metabolic transformation of thiopurine drugs. The drugs were designed as purine analogues, aiming to disrupt normal DNA synthesis, metabolized by the same pathways that control normal nucleotide homeostasis. In the cell, they are converted into 6-thioguanine nucleotides (6-TGNs), being incorporated into cellular DNA and trigger programmed cell death.^{2,3} S-methylation catalyzed by TPMT reduces the bioavailability of thiopurines, therefore prevents the intracellular conversion into cytotoxic 6-TGNs.

The human enzyme TPMT displays considerable person-to-person variation in activity on thiopurine drugs, mainly due to genetic polymorphisms. This is a major factor in the large individual variation seen in thiopurine toxicity and therapeutic efficacy. To date, 21 variant alleles of TPMT have been identified and these variants are associated with different levels of TPMT activity, from virtually undetectable to nearly wild-type level, and immunoreactivity.⁴ Among all these alleles, *TPMT**3C, *3A, and *2 are the predominant variant alleles accounting for more than 95% of inherited TPMT deficiency, while the others are observed in very few individuals. The frequency of these polymorphisms is related to ethnicity.^{5–7} The most common variant allele in Caucasians (*TPMT**3A) has a frequency of approximately 5% and 1/300 Caucasians are homozygous.⁸ Patients genetically low in TPMT activity are at greatly increased risk for potentially life-threatening thiopurine toxicity; conversely, patients with high levels of enzyme activity may be undertreated with standard doses of these drugs.

We report here the crystal structure of human TPMT in a complex with the reaction by-product S-adenosyl-L-homocysteine (AdoHcy) (Table I). The structure of TPMT indicates that (1) the naturally occurring alterations in amino acid sequence scatter throughout the structure, and (2) the amino acids whose alteration have the most influence on function are those that form intramolecular stabilizing interactions via van der Waals contacts.

MATERIALS AND METHODS

GST-TPMT

Complete human TPMT cDNA sequence (Clone ID No. 2689713, Research Genetics) was cloned into pGEX-4T-3 (Pharmacia Biotech) that produced GST-TPMT fusion. While the full-length TPMT protein (after thrombin cleavage of the GST-fusion) failed to crystallize, we generated a series of N-terminal truncations, based on the sequence alignment between human TPMT and its *Pseudomonas syringae* orthologue - whose solution NMR structure indicates an unstructured N terminus.⁹ Finally, an N-terminal truncation lacking the first sixteen amino acids, GST-TPMTΔ16, was expressed in BL21(DE3) or B834(DE3) cells by overnight auto-induction at 25°C.¹⁰ A 0.5 L auto-induced culture yielded >8 mg GST-TPMTΔ16, which was purified utilizing glutathione-affinity chromatography, equilibrated with 150 mM NaCl buffer (20 mM Tris-HCl, pH 8.0, 5% glycerol, 1 mM EDTA) and subsequently digested by thrombin at the presence of 2.5 mM CaCl₂, and further purified by HiTrapQ (equilibrated with 20 mM NaCl buffer), and S75 Sepharose sizing chromatography (150 mM NaCl buffer). In some preparations, buffers contained 1 mM dithiothreitol or 7 mM β-mercaptoethanol. In the last purification step and during concentration, AdoHcy was added to the protein at approximately 2:1 molar ratio.

Interestingly, gel filtration suggests that TPMT exists as a mixture of monomer and dimer in an equilibrium fashion. On a Superdex75 column (Pharmacia), TPMT was loaded, in the absence of any reducing reagent, onto the column at an initial concentration of approximately 2 mg/mL and eluted in two peaks, at elution volumes of 69.8 mL (equivalent to a 31.3 kDa globular protein) and 58.4 mL (62.0 kDa) (Supplemental Fig. S1A). With the addition of

reducing reagent, part of the dimeric peak shifted to the monomer peak but not completely. The concentration and age of the protein also influenced the relative portions of the two peaks: with more aged and/or concentrated the protein there was a greater proportion of dimer to monomer peak.

TPMT Δ 16 crystals (with mixture of monomer and dimer) were grown in 28–34% PEG3350, 200 mM potassium thiocyanate, and 100 mM Bis-Tris propane pH 6.8–7.4. Crystals were flash-frozen in mother liquor containing 20% ethylene glycol for data collection. The diffraction data were indexed in the space group $P4_2$ that contains two protein molecules in the asymmetric unit. Although the diffraction data could also be indexed in the higher symmetry space group $P4_22_12$ with the same cell dimensions but one molecule per asymmetric unit, we were not able to correctly place the side chain of C234, a surface residue sits in the interface with its symmetry-related mate. We reasoned that the interaction via the surface cysteine residue between two neighboring molecules might be responsible for the monomer-dimer equilibrations in solution and the orthorhombic 2-fold in crystal averages out the difference. To distinguish the two sites, we indexed the diffraction data in a lower symmetry space group, $P4_2$, with the asymmetric unit cell containing two TPMT molecules. This enabled us to assign one molecule link covalently through C234, while the other does not (Supplemental Fig. S1B). It is possible that the reducing reagent, forming a covalent disulfide bond with C234, might prevent crystallization. In addition, we modeled a small molecule Bis-Tris propane bound in the interface of two molecules that would be on the crystallographic 2-fold symmetry in space group $P4_22_12$.

The structure was determined by molecular replacement REPLACE¹¹ using the protein coordinates from the PDB 1PJZ of *P. syringae* TPMT solution NMR structure.⁹ The initial search model was modified by replacing nonidentical residues in *P. syringae* TPMT with those of human enzyme and visually assigned the best rotamer, and some residues were deleted in the loop regions. Two AdoHcy, three Bis-Tris propane, and four thiocyanate ion molecules were built manually into densities of different maps. Refinement proceeded with the program CNS.¹² Several rounds of fitting and refinement ensued and alternate conformations of several amino acid side chains included. Because of high resolution and an excess number of unique reflections (40,150) over the number of atomic parameters (3831 atoms $\times$ 4 = 15,324), the non-crystallographic symmetry was not imposed during the refinement. It is evident that the two protein molecules are highly similar, and the ~ 0.1 -Å root-mean-square deviation (RMSD) between them (comparing 229 pairs of C α atoms) is probably not due to real differences, just a reflection of errors in the data (a fact that the data could be indexed in a high symmetry space group) and the finite precision of the refinement process at the 1.89 Å resolution. The refined crystal structure of human TPMT can be superimposed with the NMR structure of *P. syringae* TPMT within 1.1 Å RMSD in the core β -sheet (comparing 40 pairs of C α atoms of strands 1–2 and 6–9). Loop regions outside of the core show larger deviations.

His6 Tagged TPMT

In a separate study, a DNA fragment encoding residues 15–245 of human TPMT (Mammalian Gene Collection clone BC009596) was cloned into a pET28a (Novagen) vector modified to be compatible with a ligation independent cloning vector (pET28a-LIC), generating His6-TPMT Δ 14; the His6 tag is in a 19-residue N-terminal fusion peptide (MGSSHHHHHSSGLVPR*GS with * as the thrombin cleavage site). *E. coli* BL21 (DE3) codon plus RIL strain (Stratagene) carrying the His6-TPMT Δ 14 expression plasmid were grown at 37°C to OD₆₀₀ of 1.5 in Terrific Broth in the presence of 50 μ g/mL of kanamycin, and induced by the addition of 1 mM isopropyl-1-thio-D-galactopyranoside and incubated overnight at 15°C. Harvested cells were re-suspended in lysis buffer (phosphate-buffered-saline, pH 7.4, 250 mM NaCl, 5 mM imidazol, 2 mM β -mercaptoethanol, 5% glycerol). The

His-tagged protein was purified by a 5 mL HiTrap Chelating column (Amersham Biosciences), followed by a Superdex200 (26 × 60) sizing column (Amersham Biosciences). Thrombin (Sigma) was added to combined fractions containing His-TPMTΔ14 and incubated overnight at 4°C. The cleaved protein was further purified to homogeneity by ion-exchange chromatography on Source 30Q column (10 × 10) (Amersham Biosciences), equilibrated with 20 mM Tris-HCl, pH 8.0, and eluted with linear gradient of NaCl up to 500 mM concentration, yielding 12 mg protein from 1 L of culture.

The seleno-methionine (Se-Met) derivative of His6-TPMTΔ14 was expressed in the same *E. coli* strain in M9 minimal medium in the presence of 50 mg/L of Se-Met. Purified TPMT protein (Se-Met derivative) was crystallized in the presence of AdoHcy (Sigma) (protein:AdoHcy molar ratio at 1:5) at 20°C by mixing 1.5 μL of the protein solution (97 mg/mL) with 1.5 μL of the reservoir solution containing 24% PEG3350, 0.2 M KCNS, 0.1 M Bis-Tris propane, pH 7.0. Single-wavelength anomalous X-ray diffraction data (with 13.5 fold redundancy) were collected from a single Se-Met derivative crystal. Data were processed and scaled with d*TREK software.¹³ The crystal was indexed in space group P4₂2₁2, with one molecule per asymmetric unit (approximately 48% solvent content). SHELXD¹⁴ found four selenium sites. Solve¹⁵ calculated initial phases, which were modified by solvent flattening using Resolve.¹⁶ ARP/wARP¹⁷ built a preliminary model of 224 amino acids and 178 water molecules. Alternating rounds manual fitting performed using the program Coot¹⁸ and automated refinement by REFMAC¹⁹ were performed. An anomalous difference map from the final model showed discrete regions of density around the four selenium atoms in the structure (not shown). A Ramachandran plot calculated using PROCHECK²⁰ has 92% of residues in most favored regions and 8% of residues in additionally allowed regions.

In addition, we generated four TPMT mutants containing the variant alleles of TPMT*3A (A154T/Y240C), TPMT*3B (A154T), TPMT*3C (Y240C), and TPMT*2 (A80P). The mutants were constructed by PCR-mediated mutagenesis on the His6-TPMTΔ14 construct. All four mutant proteins were expressed in *E. coli* cells at comparable levels to that of wild-type protein. Circular Dichroism (CD) was used to perform thermal-melt profiling. The protein was heated to 95°C and scanned at fixed wavelength using a CD Spectrophotometer (Model 215, AVIV Instruments).

RESULTS

Overall Structure

We used two different ways to solve the structure of an N-terminal deletion form of human TPMT in complex with AdoHcy (see Materials and Methods). In the P4₂ space group (PDB 2H11), the two molecules (A and B) are highly similar with ~0.1-Å RMSD comparing 229 pairs of Cα atoms. A surface residue C234 adopts different conformations in molecules A and B. The molecule A and its symmetry-related molecule link covalently through C234 mediated disulfide-bond (2.0 Å), while the sulfur atom of C234 of molecule B is 3.4-Å away from the corresponding atom of symmetry-related molecule B (Supplemental Fig. S1B). In the P4₂2₁2 space group (PDB 2BZG) with the same cell dimensions but one TPMT monomer (molecule C) in the asymmetric unit, the sulfur atom of C234 of molecule C is in close contact (1.06 Å) with that of symmetry-related molecule, indicating manual adjustment of the side chain rotamer might be needed. Besides minor differences in the crystal packing interactions (for example, a small molecule Bis-Tris propane bound in the interface of molecules A and B disrupts the perfect 2-fold symmetry), molecule C in the P4₂2₁2 space group (PDB 2BZG) is highly similar to molecules A and B in the P4₂ space group (0.15 Å RMSD) and thus we describe structure of molecule A of PDB 2H11 below.

TPMT is a single domain protein [Fig. 1(A)] with a classic Class-I methyltransferase (MTase) fold.²¹ The domain contains a nine-stranded β -sheet [green in Fig. 1(A)] flanked on each side by three helices [α B, α C, α D and α E, α F, α G; cyan in Fig. 1(A)]. Two additional helices [yellow in Fig. 1(A)] are inserted into the MTase fold: the N-terminal helix α A and the helix α H between hairpin strands β 8 and β 9. As expected from comparison to other MTases, AdoHcy is bound at the carboxyl ends of the parallel strands β 1- β 2- β 5, the hallmark of a nucleotide-binding site. The N-terminal region helps to constrain the bound AdoHcy in an acidic pocket, through which the AdoHcy adenine ring is visible [Fig. 1(B)]. Helix α A and the flanking loops provide four residues (L26, W29, W33, and F40) forming an aromatic lid to the AdoHcy binding site [Fig. 1(C)].

The β -sheet contains five parallel strands with a pair of antiparallel β -hairpins on either end (8↓9↑7↓6↓1↓2↓5↓4↑3↓), symmetrically related to the central strand β 1 [Fig. 2(A)]. This symmetry may be a reflection of the fact that the adenine moiety of AdoMet or AdoHcy and the substrate 6-MP have nearly identical chemical structure (except an exocyclic amino group NH_2 vs. a SH group), and both would be bound at the carboxyl ends of the parallel strands [Fig. 2(A)]. The symmetry was previously noticed for DNA MTases where the DNA-adenosyl and AdoMet-adenosyl binding sites are structurally comparable.²²

Active Site Conformation and a Proposed Catalytic Mechanism for S-Methylation

Thus far, we are unable to produce co-crystals of TPMT in complex with 6-MP, 6-TG, or 2-thiopurine (data not shown), largely due to their extremely low solubility in water (while the drugs can be dissolved in an organic solvent, mixing it with the protein solution caused these compounds to come out of solution). However, the current structure provides some insights into the mechanism of substrate recognition and catalysis. An internal solvent channel, with two open ends on opposite sides of the enzyme, is interconnected to the AdoHcy binding pocket [Fig. 1(B)]. Small molecule substrates can likely diffuse into and out of the active site through this channel. Interestingly, in the absence of substrate, a five-member water ring is observed in the active site [Fig. 3(A)]. The water ring is nearly planar and is flanked by P196 and F40. The water molecule w1 is only ~ 3.1 Å away from the sulfur atom of AdoHcy, where the transferable methyl group would be attached in AdoMet. We modeled 6-MP into the active site by aligning the plane of 6-MP with the plane of the water ring and positioned the target exocyclic sulfur atom at water site w1 [Fig. 3(B)]. As a result, the four ring nitrogen atoms of 6-MP are superimposed nearly perfectly onto the four water molecules [Fig. 3(B)]. In this model, the target sulfur atom is in a position to become the nucleophile attacking the methyl group of AdoMet. It is in line with the methyl group and the sulfur atom of AdoMet; such linear arrangement of the nucleophile, the methyl group, and the leaving thioester group in the transition state is required by the classic $\text{S}_{\text{N}}2$ reaction mechanism used by most (if not all) other MTases. Like other target atoms (C, N, O) that usually require activation, the immediate hydrophilic microenvironment (the target sulfur position is assumed a water binding site) would ensure that the methylable target sulfur becomes deprotonated and the proper positioning of the target and the donor sulfur atoms is necessary for methyl transfer between the two sulfur atoms. There is no charged or polar residue in the immediate vicinity of near the target sulfur atom; however, a base residue (K32 or R226 or both; Fig. 3(C, D)) removing the proton attached to either N7 or N9 ring nitrogen would produce deprotonated thiolic anion (S^-).

The oblate shaped active site is large enough to accommodate an alternative conformation of 6-MP by rotating the purine ring 180° along its S — C bond [comparing Fig. 3(C) and 3(D)]. Dynamic trajectory calculations suggest favorable free energies of binding for both conformations (supplemental Table S1). Both conformations suggest the binding of 6-MP with K32 of α A (with the ring nitrogen atom N1 or N7) and R226 of helix α H (with the ring nitrogen

atoms N3 and N9), either directly or through the recruitment of bridging water molecules observed in the crystal structure.

Structural Analysis of Amino Acid Sequence Alterations

There are 21 variant alleles of TPMT that are associated with 17 alterations in amino acid sequence⁴ (Table II). The positions of the point mutants scatter throughout the entire region/structure, from the first residue (M1V) to near the carboxyl terminus (Y240C), of total 245 residues [Fig. 2(B)]. Most human TPMT variants alter residues that are highly conserved throughout vertebrates (supplemental Fig. S2).⁴ The structure of TPMT offers a rationale for their functions (except the first residue which is not present in the current structure). We group them into three categories: (A) those involved in structural packing via van der Waals contacts (L49S, A80P, C132Y, A154T, Y180F, and Y240C) [Fig. 4(A–E)]; (B) those involved in intramolecular polar interactions (E28V, K119T, K122T, S125L, G144R, R163H, R215H, and H227Q) (Supplemental Fig. S3); and (C) those directly or indirectly involved in cofactor binding (G71R, Q42E) (Supplemental Fig. S3). Some mutants can be grouped into more than one category. For example, the C β atom of A154 is in van der Waals contacts with P68, L69 (interacting with AdoHcy), and D151 (hydrogen bonding interactions with Y166 and Y180) [Fig. 4(A)]. Because of its proximity to the cofactor binding site and active site, substitution of A154 with threonine (A154T; TPMT*3B) might alter cofactor binding affinity and enzyme activity. Figure 4 illustrates these interactions in detail.

The category A mutants are located in the central β -strands (C132Y, β 5; Y180F, β 7; Y240C β 9), immediately outside the carboxyl end of strand β 6 (A154T), or in helices (A80P, α C; L49S, α B) whose side chains intercalate with each other in the interfaces of strands to form the hydrophobic cores of the molecule [Fig. 2(A)]. These intra-molecular hydrophobic and internal (or buried) polar interactions (such as the hydroxyl groups of Y180 and Y240) form a contiguous network that is essential for the stability of the molecule and for maintaining enzymatic function. Y180 is involved in a hydrogen-bonding network of Y180-D151-Y166 [Fig. 4(A)]. D151 is the second to last residue of strand β 6 and is buried in the center of molecule. In other MTases, the corresponding residue at the carboxyl end of strand β 6 often performs some catalytic functions.²³

The category B mutants, such as K119T, S125L, R163H, and H227Q, are located on the protein surface away from the active site. Their side chains are often involved in intramolecular polar interactions with other surface residues that likely confer additional stability to the molecule.

Mutational Analysis

In vitro, the solubility of A154T mutant (TPMT*3B), or its related double mutation (A154T/Y240C; TPMT*3A), was significantly lower even though the expression level in *E. coli* is at comparable level with that of wild-type protein (data not shown). A154 is involved in part of the core interactions adjacent to the cofactor binding site and active site [Fig. 4(A)], with a distance of 6.4 Å between the C α of A154 and the sulfur of AdoHcy. The central location of this residue is consistent with the observation that A154T causes the most severe change in immunoreactive protein level and activity (only 1–2% of WT).⁴ More recently, Wang et al.²⁴ demonstrated that recombinant TPMT allelic variant proteins (TPMT*3A, *3B, and *3C) expressed in *E. coli* had a tendency to aggregate, correlated with the extent of aggresome formation in cultured cells, and suggested the naturally occurring TPMT amino acid polymorphisms can promote protein aggregation.

Mutant A80P (TPMT*2) eluted from the sizing column close to the void volume (data not shown), indicating the amino acid change alter the structure of the protein and lead to aggregation. A80P retained approximately one fourth of protein level and activity.⁴ The residue

is approximately 20 Å away from the AdoHcy sulfur. Substitution of A80 with proline (A80P) would introduce a deformation into the last turn of helix α C and destabilize the interactions with the two central strands β 2 (V86) and β 5 (I128) [Fig. 4(B) and 4(C)].

Y240, located in the middle of the strand β 9, is involved in both hydrophobic and polar interactions with helix α G and strand β 8. Substitution of Y240 with the smaller cysteine (Y240C; TPMT*3C) would be expected to lose hydrogen bonding with H201 and K238 and introduce a cavity [Fig. 4(D)]. In addition, a cysteine at residue 240 could form a disulfide linkage with C216 of strand β 8, altering local protein structure. CD spectra showed that only at high temperature of approximately 50°C or above, were Y240C and A154T more thermolabile (approximately 10°C) than the wild-type enzyme [Fig. 4(E)]. Our data are consistent with the trend observed for two polymorphic forms of human histamine MTase (HNMT),²⁵ in which both variants of HNMT are stable over a wide range of temperature (24–45°C). In addition to the differences in thermal properties of the variants, differences in protein quantity (due to expression or degradation) and other factors such as protein-protein interactions might contribute to the larger differences in activity observed *in vivo*.

DISCUSSION

It is expected that the amino acid sequence alterations change the catalytic activity and the protein stability of the individual alleles. Although some of the alterations might not alter the structure of proteins, others may. It is thus important to note that in the absence of mutant protein structure, our analysis of structure-function of TPMT was based on the assumption that the structures of alleles are similar to that of the wild type enzyme. It would be necessary to generate several crystal structures of important alleles to have a complete understanding. Unfortunately, low solubility (A154T and its related double mutant A154T/Y240C) or aggregation (A80P) prevented further structural study of alleles.

Salavaggione et al.⁴ investigated the functional effects of 14 TPMT mutations (including the wild-type) on activity, protein level, and stability (i.e., rates of degradation) (Table II). For most mutant allozymes, accelerated rates of degradation are responsible, at least in part, for reduced levels of enzyme that correlate almost linearly with reduced activity. This observation is consistent with the structural locations of the individual mutations. The category A mutants that affect the hydrophobic core of the molecule would destabilize the molecule, making it more susceptible to degradation.

An exception to the near-linear correlation of TPMT level and enzyme activity is L49S, which displays a 30% decrease in protein level but retains only ~2% activity. L49 is deeply buried in the center of the molecule [Fig. 4(B)] and packs against aromatic and aliphatic residues from helices α B (L53 and F56), α C (W78 and F79), and three central strands- β 1 (F67), β 6 (W150), and β 7 (L181) important for cofactor binding and active site formation. The smaller serine substitution would introduce an unbalanced polar atom into a less packed internal cavity, causing a disproportionate decrease in enzyme activity.

CONCLUSIONS

Our structural studies of human TPMT helped to understand and predict the effects of alteration in amino acid sequence on TPMT function. The example provided by TPMT, a pharmacogenetically important drug-metabolizing enzyme, demonstrates that naturally occurring genetic polymorphisms/mutations which alter encoded amino acid sequence may influence function predominantly by decreasing the quantity of enzyme protein.⁴ Our study also supports the notion that the remote stabilizing interactions (such as L49S), although distal from the active site, form a clear focal point centered near the active site, likely to be involved

in substrate binding and/or catalysis, as demonstrated by molecular dynamics simulations of HhaI DNA MTase.²⁶

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Geoffrey G. Wilson of New England Biolabs for comments on the manuscript; Aiping Dong for help with data collection; and Dr. Yong Zhao for help with bio-informatics. The study was partly supported by US Public Health Service grants (to XC) and by the Structural Genomics Consortium, a public private charitable partnership that receives funds from Canada Foundation for Innovation, Canadian Institutes for Health Research, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, the Ontario Innovation Trust, the Ontario Research and Development Challenge Fund and the Wellcome Trust. Financial support for the beamline operation of Emory's shares at the beamline SERCAT-22 of the Advanced Photon Source (APS) at Argonne National Laboratory comes from the Dean's Office of Emory University School of Medicine. The companies of the Industrial Macromolecular Crystallography Association through a contract with the Center for Advanced Radiation Sources at the University of Chicago supported use of the IMCA-CAT beamline 17-ID at the APS. Use of the APS was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. Figures were drawn using the program Pymol, a user-sponsored molecular modeling system with an OPEN-SOURCE foundation (<http://pymol.sourceforge.net>).

References

1. Coulthard S, Hogarth L. The thiopurines: an update. *Invest New Drugs* 2005;23:523–532. [PubMed: 16267626]
2. Tidd DM, Paterson AR. A biochemical mechanism for the delayed cytotoxic reaction of 6-mercaptopurine. *Cancer Res* 1974;34:738–746. [PubMed: 4856046]
3. Tidd DM, Paterson AR. Distinction between inhibition of purine nucleotide synthesis and the delayed cytotoxic reaction of 6-mercaptopurine. *Cancer Res* 1974;34:733–737. [PubMed: 4856045]
4. Salavaggione OE, Wang L, Wiepert M, Yee VC, Weinshilboum RM. Thiopurine S-methyltransferase pharmacogenetics: variant allele functional and comparative genomics. *Pharmacogenet Genomics* 2005;15:801–815. [PubMed: 16220112]
5. Krynetski EY, Schuetz JD, Galpin AJ, Pui CH, Relling MV, Evans WE. A single point mutation leading to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc Natl Acad Sci USA* 1995;92:949–953. [PubMed: 7862671]
6. Tai HL, Krynetski EY, Yates CR, Loennechen T, Fessing MY, Krynetskaia NF, Evans WE. Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *Am J Hum Genet* 1996;58:694–702. [PubMed: 8644731]
7. Szumlanski C, Otterness D, Her C, Lee D, Brandriff B, Kelsell D, Spurr N, Lennard L, Wieben E, Weinshilboum R. Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. *DNA Cell Biol* 1996;15:17–30. [PubMed: 8561894]
8. Yan L, Zhang S, Eiff B, Szumlanski CL, Powers M, O'Brien JF, Weinshilboum RM. Thiopurine methyltransferase polymorphic tandem repeat: genotype–phenotype correlation analysis. *Clin Pharmacol Ther* 2000;68:210–219. [PubMed: 10976552]
9. Scheuermann TH, Lolis E, Hodsdon ME. Tertiary structure of thiopurine methyltransferase from *Pseudomonas syringae*, a bacterial orthologue of a polymorphic, drug-metabolizing enzyme. *J Mol Biol* 2003;333:573–585. [PubMed: 14556746]
10. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 2005;41:207–234. [PubMed: 15915565]
11. Tong L, Rossmann MG. Rotation function calculations with GLRF program. *Methods Enzymol* 1997;276:594–611. [PubMed: 9048382]
12. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR

- system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–921. [PubMed: 9757107]
13. Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* 1999;55:1718–1725. [PubMed: 10531521]
 14. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 2002;58:1772–1779. [PubMed: 12351820]
 15. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 1999;55:849–861. [PubMed: 10089316]
 16. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 2000;56:965–972. [PubMed: 10944333]
 17. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999;6:458–463. [PubMed: 10331874]
 18. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132. [PubMed: 15572765]
 19. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255. [PubMed: 15299926]
 20. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 1993;231:1049–1067. [PubMed: 8515464]
 21. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyl-transfer: a chronicle of convergence. *Trends Biochem Sci* 2003;28:329–335. [PubMed: 12826405]
 22. Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* 1995;253:618–632. [PubMed: 7473738]
 23. Fauman, EB.; Blumenthal, RM.; Cheng, X. Structure and evolution of AdoMet-dependent methyltransferases. In: Cheng, X.; Blumenthal, RM., editors. *S-adenosylmethionine-dependent methyltransferases: structures and functions*. River Edge, NJ: World Scientific; 1999. p. 1-38.
 24. Wang L, Nguyen TV, McLaughlin RW, Sikkink LA, Ramirez-Alvarado M, Weinshilboum RM. Human thiopurine *S*-methyl-transferase pharmacogenetics: variant allozyme misfolding and aggresome formation. *Proc Natl Acad Sci USA* 2005;102:9394–9399. [PubMed: 15967990]
 25. Horton JR, Sawada K, Nishibori M, Zhang X, Cheng X. Two polymorphic forms of human histamine methyltransferase: structural, thermal, and kinetic comparisons. *Structure* 2001;9:837–849. [PubMed: 11566133]
 26. Estabrook RA, Luo J, Purdy MM, Sharma V, Weakliem P, Bruice TC, Reich NO. Statistical coevolution analysis and molecular dynamics: identification of amino acid pairs essential for catalysis. *Proc Natl Acad Sci USA* 2005;102:994–999. [PubMed: 15657135]

Fig. 1.

Structure of human TPMT (**A**) Two views of TPMT in ribbon model. Strands are colored green and helices cyan. The additional N-terminal helix (α A) and the insertion helix α H between strands β 8 and β 9 are in yellow. Naturally occurring point mutations are shown in red. The bound AdoHcy is in ball-and-stick model with conventional coloration (carbon atoms grey; nitrogen blue; oxygen red; sulfur yellow). In the box, the chemical structure of 6-MP and its methylation reaction carried out by TPMT. (**B**) Four views of TPMT in surface model, showing the cofactor binding pocket and the potential substrate diffusion channel. The surface is colored blue for positive, red for negative, and white for neutral. The bound AdoHcy is in stick model.

(C) Close-up stereo view of TPMT-AdoHcy interactions. The omit electron density map for AdoHcy was contoured at 1.5σ .

Fig. 2.

Topology, secondary structures, and sequence of human TPMT. (A) Topology of human TPMT. Helices are shown in cylinders (lettered) and strands as broad arrows (numbered). Residue numbers indicate the extent of the secondary structure element. The naturally occurring alterations in amino acid sequence are circled with white letter against the black background. Helices α B, α C, and α D are on one side of the β -sheet, while the helices α E, α F, and α G (hatched) on the other side. (B) Sequence of human TPMT. Letters (A–G) for helices and numbers 1–9 for strands indicate the secondary structure elements. Residue numbering is shown above the sequence. The naturally occurring alterations in amino acid sequence and the names of TPMT variant alleles are shown below the sequence. The N-terminal residues 1–16 were not included in the crystallization.

Fig. 3.

Molecular modeling of 6-MP at the TPMT active site (A) A five-member water ring (labeled with numbers 1–5) is observed in the active site. The water ring is nearly planar and is flanked by P196 and F40. The water molecule w1 is only ~3.1 Å away from the sulfur atom of AdoHcy. (B) 6-MP was modeled in the active site by aligning the plane of 6-MP with the plane of the water ring, positioning the target exocyclic sulfur atom at water site w1 and the four ring nitrogen atoms of 6-MP onto the other four water molecules. The arrow indicates the movement of the methyl group transferred from the AdoMet methylsulfonium group to the target sulfur atom. [(C) and (D)] Two possible conformations of 6-MP as minimized in the active site (see supplemental Table S1).

Fig. 4.

Consequences of particular mutations (A) A154 is in van der Waals contacts with P68, L69 (interacting with AdoHcy), and D151 (hydrogen bonding with Y166 and Y180). The substitution of A154 with threonine (A154T) causes the most severe change in protein level and activity (only 1–2% of WT), the larger size of threonine side chain and its polarity destroy the core interactions adjacent to the cofactor binding site and active site. (B) L49, L53, and F56 of helix α B pack with the aromatic side chains of W78 and F79 (α C), W150 (β 6), and the aliphatic side chain of Leu181 (β 7) to form the hydrophobic core. Mutation of L49 to the smaller serine (L49S) would introduce a polar residue into a less packed internal cavity, causing a disproportionate decrease in enzyme activity. (C) A80 is involved in hydrophobic interactions

with residues from strands $\beta 2$ (Val86) and $\beta 5$ (Ile128). Substitution of A80 with proline (A80P) would introduce a deformation into the last turn of helix αC and destabilize the contiguous interactions with the hydrophobic core involving L49 (see panel B). S125, located in the loop between strands $\beta 4$ and $\beta 5$, makes two hydrogen bonds with the main chain amide of N127 and the side chain of S123; S123 also interacts with Ile128. Replacement of S125, a smaller polar residue, by a larger aliphatic residue (S125L), though located on the surface away from the active site, would destabilize the local structure causing ~60% decrease in activity and protein level. **(D)** Y240 of strand $\beta 9$ is involved in hydrogen bonding with H201 (the first residue of αG), a water-mediated interaction with K238 ($\beta 9$), and in van der Waals contacts with I204 (αG), I214 and C216 ($\beta 8$). Substitution of Y240 with the smaller cysteine (Y240C) would be expected to lose hydrogen bonding with H201 and K238 and introduce a cavity. In addition, a cysteine at residue 240 could form a disulfide linkage with C216 of strand $\beta 8$, altering local protein structure. R215 of strand $\beta 8$ has a two-water-mediated interaction with Q213 ($\beta 8$) and T243 ($\beta 9$). The surface mutation R125H has much less effect on the protein level or activity. **(E)** CD spectra of thermo-melt spectra of TPMT wild type ($\bullet$), A154T ($\circ$) and Y240C ($\blacktriangledown$).

TABLE I
Statistics of X-ray Data Reduction and Structural Refinement

	TPMT(native)-AdoHcy	TPMT(SeMet)-AdoHcy
PDB code	2H11	2BZG
Data collection		
Beamline (wavelength)	APS 22-ID (1.0 Å)	APS 17-ID (0.97974 Å)
Space group	P4 ₂	P4 ₂ 2 ₁ 2
Unit Cell (Å)	$a=b=85.64$, $c=69.82$	$a=b=85.49$, $c=69.36$
Resolution range (Å) (highest resolution shell)	33.58–1.89 (1.94–1.89)	60.41–1.58 (1.64–1.58)
Measured reflections	147,990	480,303
Unique reflections	40,150	35,537
$\langle I/\sigma \rangle$	10.5	11.8 (4.1)
Completeness (%)	98.7 (96.2)	98.51
$R_{\text{linear}} = \Sigma I - \langle I \rangle / \Sigma \langle I \rangle$	0.091 (0.441)	0.072 (0.270)
Refinement statistics		
Resolution range (Å)	33.58–1.89 (1.94–1.89)	36.39–1.58 (1.64–1.58)
Unique reflections	40,150	33,511
R-factor= $\Sigma F_o - F_c / \Sigma F_c$	0.203 (0.270)	0.213 (0.239)
R-free	0.236 (0.292) (10% data)	0.249 (0.309) (5% data)
Non-hydrogen atoms		
Protein	3710	1859
Hetrogen	121 (2 AdoHcy, 3 BisTris propane, 4 thiocyanate ions)	26 (AdoHcy) 19 (BisTris propane)
Water	322	231
Thermal B values (Å ²)		
Overall mean B value	29.7	18.4
Root-mean-square deviation from ideality		
Bond Lengths (Å)	0.01	0.02
Bond Angles (°)	1.6	1.4

TABLE II

Thiopurine S-Methyltransferase (TPMT) Alleles

TPMT allele ^a	Amino acid change ^a	Activity ^a (% of WT)	Immunoreactive protein ^c (% of WT)	Structural location	Possible functions and comments
*1	(wild-type)	100	100	—	—
*7	H227Q	98	87	αH	(b)
*9	K119T	92	92	β4	(b)
*19	K122T			β4	(b) (similar as K119T)
*8	R215H	82	91	β8	(b)
*16	R163H			αF	(b) (involved in multiple polar interactions)
*10	G144R	69	80	Loop between αE and β6	(b)
*13	E28V	61	69	αA	(b) (might be indirectly involved in cofactor binding)
*12	S125L	41	44	Loop between β4 and β5	(b)
*17	Q42E			Loop between αA and αB	(water-mediated interaction with cofactor)
*11	C132Y	36	61	β5	(a) (c)
*6	Y180F	36	59	β7	(a) (b)
*2	A80P	26	27	αC	(a)
*3C	Y240C	17	25	β9	(a) (b)
*18	G71R			Loop after β1	(b) (directly involved in cofactor binding, predicted to be low activity)
*5	L49S	2.4	71	αB	(a)
*3B	A154T	1.7	1.0	Loop after β6	(a) (c)
*3A	A154T + Y240C	1.6	1.0		
*3D	E98STOP + A154T + Y240C				
*14	M1V			Not present in the current structure	
*15		Deletion of 25 residues 140–165			

^aThe first four columns are taken from Ref. 4. (a) hydrophobic core. (b) intramolecular polar interactions. (c) cofactor binding and/or potentially substrate binding.