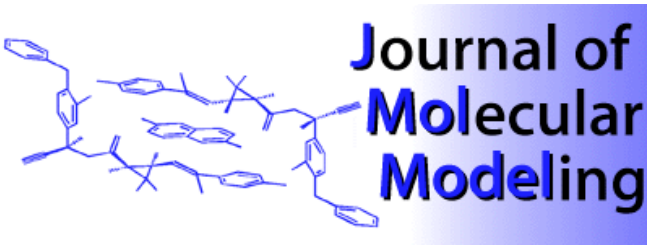
J. Mol. Model. **1998**, *4*, 211 – 220



FULL PAPER

A Three Dimensional Model of Human Thiopurine Methyltransferase; Ligand Interactions and Structural Consequences of Naturally Occurring Mutations

Roy A. Lysaa, Ingebrigt Sylte, and Jarle Aarbakke

Department of Pharmacology, Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037 Tromsø, Norway. Tel.: +4777644875, Fax: +4777645310. E-mail: sylte@fagmed.uit.no

Received: 13 May 1998 / Accepted: 25 June 1998 / Published: 10 July 1998

Abstract The three-dimensional model of human thiopurine methyltransferase (hTPMT) was constructed by molecular modeling. A multiple alignment of AdoMet dependent methyltransferases based on a structural superposition of the AdoMet binding domain of *Hhai*, *TaqI* and rCOMT was used in the modeling procedure. The reliability of the model was examined by comparing its conformation and packing properties with those of *Hhai*, *TaqI* and rCOMT and structures in the PDB-database. The examined criteria indicated a reliable model structure. The model gave insight into the structural effects of naturally occurring mutations of the hTPMT allele, and was used to characterize the ligand interactions of the protein. The residues Gln42 and Glu91 were predicted to participate in AdoMet binding through H-bond interactions whereas Phe146 participates through Van der Waal interaction. The cationic methyl-sulphonium group of AdoMet was located close to the aromatic residue Phe40. The model also indicated that substrates interact with hTPMT situated in a pocket consisting of the hydrophobic residues Phe40, Met148, Val184, Val220 and the charged residues Lys145, Glu218, Lys219. These residues were also included in a predictive explanation for the inhibitor/substrate preference of the enzyme. The most frequent of naturally occurring mutations was predicted to cause alterations on the surface of the protein with minor/none structural consequences. The mutation Ala80-Pro seemed directly to cause an inactive enzyme by disrupting the structure of the binding site of AdoMet.

Keywords Thiopurine methyltransferase, Homology model, Mutations, Ligand interactions

Introduction

Human thiopurine methyltransferase (hTPMT) is a cytosolic enzyme of unknown endogenous function that catalyzes the S-methylation of aromatic and heterocyclic sulfhydryl compounds utilizing S-adenosylmethionine (AdoMet) as methyl donor [1, 2]. The anti-cancer thiopurine agents 6-mercaptopurine (6-MP) and 6-thioguanine (6-TG), and the immunosuppressive thiopurine drug azathioprine (AZA) are all inactive prodrugs that require metabolism to thiopurine nucleotides in order to exert cytotoxity [3, 4]. The competitive metabolic pathways via either TPMT or xanthine oxidase reduce formation of the active thiopurine nucleotides.

Correspondence to: I. Sylte

The level of TPMT activity in human tissues is regulated by genetic polymorphism, about 89 % of a population are homozygous for high enzyme activity, around 11 % are heterozygous with an intermediate activity, while one of about 300 subjects is homozygous for low TPMT activity [5, 6]. The TPMT deficiency is caused by a structurally defect TPMT enzyme due to genetically inherited mutations of the hTPMT gene, and at least eight mutant alleles of the hTPMT gene have been identified [7, 8]. TPMT-deficient patient treated with conventional doses of 6-MP, 6-TG and AZA have a pronounced risk of potentially life-threatening haematopoietic toxicity [9-11].

Clinically significant drug-drug interactions have been predicted in patients treated simultaneously with thiopurines and inhibitors of TPMT activity [12]. Benzoic acid derivatives [2], diuretics [13], 2-OH-purines [14] and sulphasalazine [15] have all been reported to inhibit hTPMT activity in vitro. A detailed knowledge about the three-dimensional structure of hTPMT is necessary in order to clarify the molecular mechanisms for hTPMT inhibition and is also helpful in predicting possible drug-drug interactions in thiopurine treatment. A detailed knowledge about the three-dimensional structure of the enzyme will also give insight into the structural consequences of mutated hTPMT alleles.

To date, the x-ray crystal structure of five AdoMet dependent methyltransferases have been reported, *Hhai* DNA-methyltransferase [16], *TaqI* DNA-methyltransferases [17], *HaeIII* DNA-methyltransferase [18], rat catechol O-methyltransferases (rCOMT) [19] and rat glycine N-methyltransferase (rGNMT) [20]. These x-ray crystal structures show that the AdoMet binding domain of the DNA-methyltransferases have an overall three-dimensional structure similar to that of the small molecule methyltransferases rCOMT and rGNMT. However, compared to the DNA-methyltransferases and rCOMT, rGNMT has an additional domain (S-domain) close to the active site, which may reflect its capability in binding polycyclic aromatic hydrocarbons. If the S-domain is removed, the folding pattern of rGNMT is similar to that of rCOMT and to the AdoMet binding domain of the DNA-methyltransferases [20]. The folding similarities suggest that methyltransferases in general have a common three-dimensional structure of the AdoMet binding domain, and that these structural equivalences permits the prediction of the tertiary structure of other methyltransferases.

In the present study we constructed a three-dimensional model of hTPMT from the amino acid sequence [21] by the use of homology modeling. The sequence alignment used for the homology modeling procedure was based on a structural superposition of the AdoMet binding domain of *Hhai*, *TaqI* and rCOMT [22].

```
                                                  α3                                  β1
rcomt  MGDTKEQRILRYVQQNAKP--GDAQSVLEAIDTYCTQKE-WAMNVGDAKGQIMDAVI-----  54 ------REYS-PSLVLELGA  67
hhai   ---------------------------  299 YKQFGNSVVINVLQYIAYNIGSSLNFKPY 327 MIEIKDKQLT-GLRFIDLFA  19
TaqI   MGLPP----LLSLPSNSAPR------------SLGRVETP-PEVVDFMVSLAEAPR-----  39 ----------GGRVLEPAC  48
rgamt  MSSSA-ASPLFAPGEDCGPAWRAAPAAYDTSDTHLQILGKPVMERWETPYMHSLAAAAA---  58 ---------SRGGRVLEVGF  69
htpmt  -MDGTRTSLDIEEYSDTEVQKNQVLTLEEWQDKWVN---GKTAFHQEQGHQLLKKHLDTFL-  57 ------KGKSGGLRVFFPLC  71

          α4          β2        α5                  β3         α6          β4
rcomt  -YCGYSAVRMARLLQPGARLLTMEMNPDYAAITQQMLNF--AGLQDKVTILNGASQDLIPQLKKKYDV----DTLDMVFL------- 140
hhai   GLGGFRLALESCG---AECVYSNEWD-KYAQEVYEMN----FG-----EKPEG-D---ITQVNEKTIP----DH-DILCA-------  77
TaqI   AHGPFLRAFREAHG-TAYRFVGVEID-PKALDLPP------WA-----EGILA-D---FLLWEPGEA-------FDLILGNPPYGIV 111
rgamt  GMAIAASRVQQAP---IKEHWIIECN-DGVFQRLQN-----WALKQP-HKVVPLKG--LWEEEAPTLP---DGHFDGILY------- 134
htpmt  GKAVEMKWFADRG----HSVVGVEIS-ELGIQEFFTEQNLSYS-EEPITEIPGTK---VFKSSSGNISLYCCSIFDLPRT------- 142

                                        α7          β5                  α8          β6
rcomt  ---------------DHWKDR-----------YLPDTLLLEKCGLL-RKGTVLLADN----VIVPGTPDFLAYVRG-SSS---FEC 191
hhai   -------GFPCQAFSISGKQKGFEDSRGTLF-FDIARIVREK-------KPKVVFMENVKNFASHD-NGNTLEVVKNTMNELDYSFH 148
TaqI   GEASKYPIHVFKAVKDLYKKAFSTWKGKYNLYGAFLEKAVR----LL-KPGGVLVFVVPATWLV----LEAFALLREFLAREGKTSV 189
rgamt  ---------------DTYPLSEETWHTHQ---FNFIKTHAFR---LL-KPGGILTYCN----LTSWGELMKSKYTDI-TAM---FEE 191
htpmt  ---------------IGKFDMIWDRGALVAINP-GDRKCYADTMFSLLGKKFQYLLCVLSYDPTKHPGPPFYVPHAEI-ERL---FGK 210

          β6          β7
rcomt  THYSSYLEYMKVV-DGLEKAIYQGP-SSPDKS------------------------------ 221
hhai   AKVLNALDYG--IPQKRERIYMICF-RNDLNIQNFQFPKPFELNT----------------- 190
TaqI   YYLG---EV---FPQKKVSAVVIRFQKSGKGLSLWDTQESESGFTPILWAEYPHWEGEIIRF 245
rgamt  TQVPALLEAGF---QRENICTEVMALVPPADCRYYAFPQMITPLVTKH-------------- 236
htpmt  ICNIRCLEKV----DAFEERHKSWGIDCLFEKLYLLTEK---------------------- 245
```

Figure 1 *Multiple sequence alignment of AdoMet dependent methyltransferases. The sequences shown in the figures are; rcomt: rat catechol O-methyltransferase, hhai:* Hhai *DNA-methyltransferase, TaqI:* TaqI *DNA-methyltransferase, rgamt: rat guanidinoacetate-methyltransferase and htpmt: human thiopurine methyltransferase. The nomenclature for α-helices and β-strands is as in the crystal structure of rCOMT.*

Methods

Multiple sequence alignments

The multiple sequence alignments were carried out using the progressive alignment program ClustalW [23]. The sequence alignments were performed with an initial gap open penalty of 12 and an initial gap extension penalty of 4 both in the pairwise and multiple alignments. Several efforts were made to perform an automatic sequence alignment of all sequences. However, in spite of various combinations of input parameters, this strategy was not successful and the conserved sequence motifs observed in the small molecule methyltransferases [24] were not lined up. Finally the profile method was used to build up the initial multiple sequence alignment from alignments of closely related sequences. The misalignment of rCOMT, *Hhai* and *TaqI* in the initial multiple alignment was repaired according to an alignment based on the structural superposition of these three enzymes [22]. The remaining sequences were then realigned using amino acid sequence analysis [24] and experimental studies of small molecule methyltransferases as guideposts for the alignment without violating conserved motifs observed in the crystal structures. The structural superposition of rCOMT, *Hhai* and *TaqI* did not give any information about the structure of the methyltransferases between the N-terminus and α3. For this part we have simply used the best multiple sequence alignment. The multiple alignment included the sequences presented in Figure 1 and the following sequences: human catechol O-methyltransferase, pig catechol O-methyl-transferase, human phenylethanolamine N-methyltransferase, rat phenylethanolamine N-methyltransferase, bovine phenylethanolamine N-methyltransferase, mouse phenylethanolamine N-methyltransferase, human nicotine N-methyltransferase, human guanidinoacetate methyltransferase, rat guanidinoacetate methyltransferase, mouse thioether S-methyltransferase, human carboxyl methyltransferase, parsley caffeoyl CoA O-methyltransferase, rat carboxyl methyltransferase, human isoaspartyl O-methyltransferase, rat isoaspartyl O-methyltransferase, bovine isoaspartyl O-methyltransferase, mouse isoaspartyl O-methyltransferase, *E.Coli* isospartyl O-methyltransferase.

Guideposts for the alignments

Amino acid sequence analysis have indicated a widespread occurrence of three sequence motifs among AdoMet dependent methyltransferases [24], which were used as guidelines for the multiple amino acid sequence alignments. These three sequence motifs (Figure 1) correspond to the segments VLELGAYCG (residues 62-70), DTLDMVFL (residues 133-140) and LLRKGTVLLA (residues 159-168) in rCOMT, and the segments VLEVGFGMA (residues 64-72), GHEDGILY (residues 127-134) and LLKPGGILTY (residues 159-168) in rat guanidinoacetate methyltransferase (rGAMT). The crystal structure of rCOMT shows that residues in motif I participate in binding of AdoMet, and that motif II also forms a part of the Adomet binding site. The importance of motif I for binding of AdoMet has also been confirmed by site directed mutagenesis studies of Gly67 and Gly69 of rGAMT [25]. Site directed mutagenesis studies of residues in motif III of rGAMT did not affect the binding of AdoMet and guanidinoacetate, but altered trypsin susceptibilities at Arg20.

When constructing the alignment attention was also given to other residues that evidently are known to participate in AdoMet binding. Mutation analysis described by [25, 26] gave evidence that Asp135 and Tyr137 in the segment DTYP (residues 134-138) in rGAMT function as part of the AdoMet binding site. The aspartate was found to be crucial for AdoMet binding whereas the aromatic Tyr137 was suggested to stabilize the binding of AdoMet through a cation-π interaction with the positively charged sulphonium of AdoMet. The crystal structure of rCOMT indicates that the segment DHWK (residues 140-144) seem to play a similar role in rCOMT [19]. These amino acid residues are located in the loop segment between β4 and α7 [19]. We chose to align these segments of rCOMT and rGAMT, and use them as a guidepost for aligning the other sequences relative to each other based on two facts; i) the two segments are both important parts of the AdoMet binding site, and ii) the two segments are both found in a equally flanking position relative to motif II, reported by [24] to be widely present in diverse AdoMet dependent methyl transferases.

All amino acid sequences included in the multiple alignments have an acidic residue 17-21 residues downstream to the C-terminal end of motif I, which also was used as a guidepost for aligning the sequences. The acidic residue separated by 20 residues from the C-terminal side of motif I has been shown to be hydrogen-bonded to the ribose 2'-hydroxyl in nucleotide binding enzymes [27], while the corresponding residue in rCOMT (Glu90) is hydrogen-bonded with the ribose hydroxyls of AdoMet [19]. Further, site directed mutagenesis studies of Glu89 located 17 residues from the C-terminal end of motif I in rGAMT indicated that this residue also was hydrogen bonded to AdoMet [25].

The structural superposition of rCOMT, *Hhai* and *TaqI* showed that these methyltransferases have an acidic residue at the C-terminal end of β6 [22]. In rCOMT this residue is a part of the substrate/inhibitor binding pocket. The crystal structure of rGNMT also contains an acidic residue at a similar position [20], indicating that this may be a general trend among the methyltransferases.

Construction of the model

Molecular mechanical energy minimizations and molecular dynamics simulations were performed with the AMBER 4.1 programs [28]. A distance dependent dielectric function (ε = r, where r = interatomic distance) was used in the calculations. Water molecules were not included. Energy minimizations were performed with an initial step length of 0.05,

the non-bonded pair list was updated after every 100 steps and the cut-off radius for non-bonded interactions was 15 Å. Molecular dynamics simulations were performed with a step length of 1 fs and a 12 Å cut-off radius for non-bonded interactions. The SHAKE option was used to constrain all bonds involving hydrogen atoms and the non-bonded pair list was updated after every 10 steps during the simulations.

Conserved α-helixes and β-sheet regions

The multiple sequence alignment was used to construct an initial model of the backbone atoms in structurally conserved regions of hTPMT. This was performed by substituting the residues of rCOMT into the corresponding residues of hTPMT using the MIDAS programs [29].

Loop regions

The initial backbone conformation of the loop segment 205-207 was constructed from the corresponding loop regions in rCOMT (Figure 1), while initial backbone conformation of other loop regions were constructed by searching for loop segments in the Brookhaven PDB-database. The 8 most preferred conformations of each segment were inspected visually for steric interactions with its local environment, and the conformation with most reasonable interaction and smallest structural deviation at the terminal ends was selected.

The selected loop segments were refined by 500 cycles of steepest descent minimization followed by 2000 cycles of conjugate gradient minimization. Loops of 5 or more amino acid residues were further refined by 50 cycles of simulated annealing molecular dynamics simulation. The loop was gradually heated to 1000 K during the first 5 ps of each cycle, kept at 1000 K between 5 and 10 ps, and gradually cooled to 0 K between 10 and 20 ps. The structure after 20 ps was used as the initial structure in the next cycle of simulated annealing molecular dynamics, which was performed in a similar manner as the previous cycle. A harmonic force of 5 kcal/Å^2 was used to constrain the terminal ends at an appropriate distance during the simulation. The structure after each cycle was energy minimized until convergence at 0.02 kcal/mol/Å r.m.s energy gradient difference between successive minimizations steps, resulting in 50 energy minimized structures of each loop. Potential energy and interactions with the local environments were then used as criteria for selecting between the energy minimized structures.

The loop regions were connected to the model using interactive computer graphics, and energy refined for 500 cycles of steepest descent minimization and 5000 cycles of conjugate gradient minimization. The loops were further refined by 50 ps of molecular dynamics simulation at 310 K, and thereafter by 500 cycles of steepest decent minimization and 2000 cycles of conjugate gradient minimization. α-helices and β-sheets were kept at fixed positions during the energy refinements of the loops. The two terminal segments (Figure 1) consisting of residues 1-4 and residue 231-245 were not included in the model.

Side chain conformations and refinements

All the side chains of the model were energy minimized by 500 cycles of steepest descent minimization and 4500 cycles of conjugate gradient minimization. After the energy minimization, a simulated annealing molecular dynamics procedure was used to generate the side chain conformation of each residue. The model was gradually heated to 750 K during 16 ps of simulation and kept at 750 K for 100 ps. Different coordinate sets observed during the simulation were gradually cooled from 750 K to 0 K during 23 ps of simulation and energy minimized. The main chain atoms were kept at fixed positions during the molecular dynamics simulating annealing and energy minimization procedure. Six different models obtained during the procedure were evaluated using the structure verification program CHECK [30] of the WHATIF program package.

Docking of ligands into the model

The structure with the best scoring values in the WHATIF structure verification test was considered as the most realistic model of hTPMT, and was used in the studies of putative hTPMT-ligand interactions. The structure of AdoMet from the crystal structure of rCOMT [19] was docked into the putative AdoMet binding site of hTPMT using the crystal structure of rCOMT as a guide. Further, we assumed that the sub-

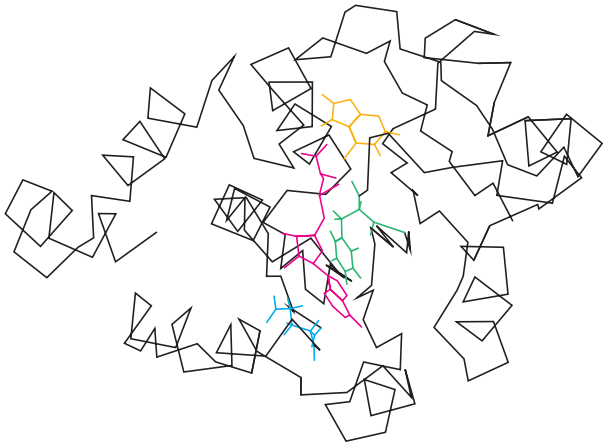


Figure 2 *The three dimensional Cα-trace model of hTPMT (black) complexed with AdoMet (magenta) and 6-MP (yellow). The putative catalytic center includes two residues suggested to be of special importance; Phe40 (green) and the conserved Glu91 (blue)*

strate binding site on hTPMT corresponded to the catechol binding site of rCOMT. The substrate 6-MP and different noncompetitive benzoic acid inhibitors for hTPMT [2, 14, 31] were docked into the site. The benzoic acid inhibitors used in the docking were: 3-chloro benzoic acid, 3-methyl benzoic acid, 3,5-dichlorobenzoic acid, 3,5-dimethyl benzoic acid and veratric acid.

Results

Evaluation of the model

The refined three dimensional model of hTPMT complexed with AdoMet is shown in Figure 2. The backbone conformation of the model was analyzed using the Procheck computer program [32]. The Ramachandran map (Figure 3) indicated that 11 out of the 226 non-glycine residues in the model were in disallowed regions. Comparing this result with the experimental detected template structures, 1 non-glycine residue in rCOMT (PDB-acquisition code: 1vid), 0 non-glycine residue in *Hhai* (PDB-acquisition code: 1hmy) and 3 non-glycine residues in *TaqI* (PDB-acquisition code: 2adm) are located in disallowed Φ/Ψ regions.

A criterion for a protein model either from experimental studies or from model building is that the energy associated with the interaction between each amino acid and the remainder of the protein model should be negative. The intermolecular potential energy of each amino acid residue calculated with the Analysis program of the AMBER package (Figure 4) showed that all residues had a negative energy peak. This indicates that there are no unfavourable van der Waals contacts or other bad local contacts in the the final energy refined model.

The Whatif/Quality Control method [30] was used to compare the packing environment of the residues with the average packing environment for the same residue type in good structures contained in the Protein Data bank. A residue in a structure with a quality score of -5.0 σ or worse indicates either poor packing, contacts with a cofactor, or that the residue is part of the active site. In the present model 10 residues out of the 226 residues included in the model were found to have a score < -5 σ. The corresponding values for rCOMT (216 residues), and the AdoMet binding domain of *Hhai* (residue 1-190 and residue 299 to 327) and *TaqI* (residue 21-245) were 4, 11 and 9, respectively. The average quality control value for the modeled structure, which evaluate the overall quality of the structure was -1.6. The corresponding values for rCOMT, *hhai*, and *Taq1* are -0.4, -1.5 and -0.8, respectively.

The putative binding site of AdoMet

Docking of AdoMet into the model (Figure 5) at a position corresponding to the position of AdoMet in rCOMT suggested the following possible interactions between the ligand and the surroundings:

i) H-bond from carboxyl oxygens on Glu91 sidechain to ribose 2'-hydroxyl. Glu91 corresponds to Glu90 in rCOMT which also interacts with the ribose hydroxyls. These residues correspond to the conserved acidic amino acid residue localized 17-21 residues downstream motif I.

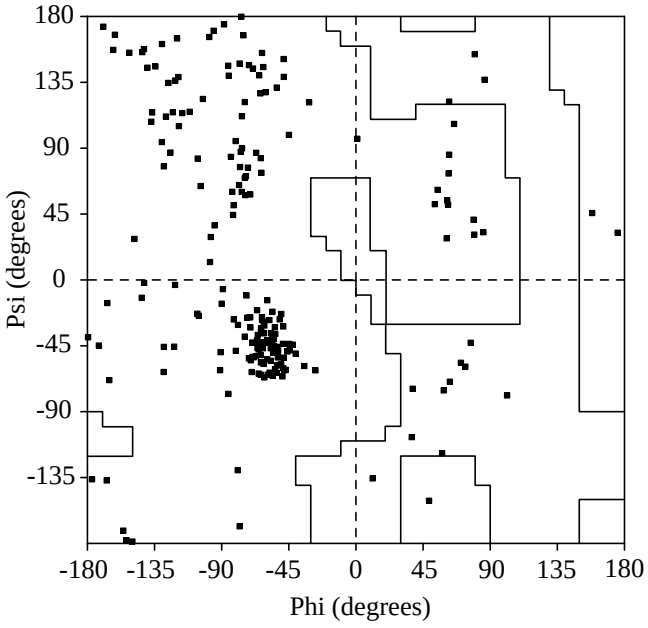


Figure 3 *Ramachandran map of the Φ/Ψ distribution of non-glycine residues for the hTPMT model*

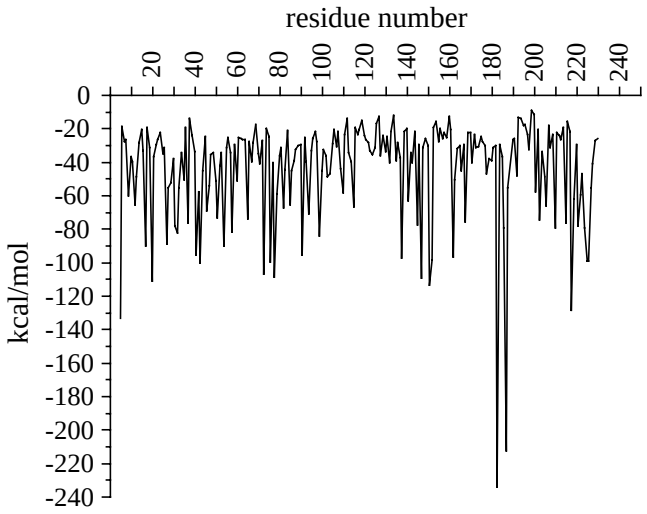


Figure 4 *The intermolecular interaction energy for each amino acid residue in the hTPMT model*

ii) H-bond from amino on Gln42 side chain to carboxyl on the methionyl part of AdoMet.

iii) H-bond from oxygen on Gln42 sidechain to amino on the methionyl part of AdoMet.

iv) Van der Waal interactions between Phe146 and adenine.

v) A possible cation-π-interaction between the positively charged sulphonium-methyl part of AdoMet and Phe40.

The putative binding site of substrates and inhibitors

Docking of 6-MP and the benzoic acid derivatives at the putative active site suggested that these ligands interact with several hydrophobic residues in hTPMT (Figure 5). The hydrophobic residues forming the substrate binding cleft were Phe40, Met148, Val184 and Val220. The chosen ligands seemed to be oriented inside the cleft guided by four charged groups (Lys145, Glu218, Lys219 and the sulphonium-methyl part of AdoMet) in the sorroundings. Negatively charged ligands were predicted to have the highest affinity for the enzyme due the net positive charge of the cleft. For aryl ligands like thiophenol substrates and benzoic acid inhibitors, the orientation of the ligand in the cleft also seemed to depend on substituents on the benzene ring other than the main functional group. Aryl ligands with hydrophobic meta-substituents with a negative electrostatic surface seemed to interact more intimately, which explains the differences in inhibition potency and the properties for benzoic acid derivatives as inhibitors.

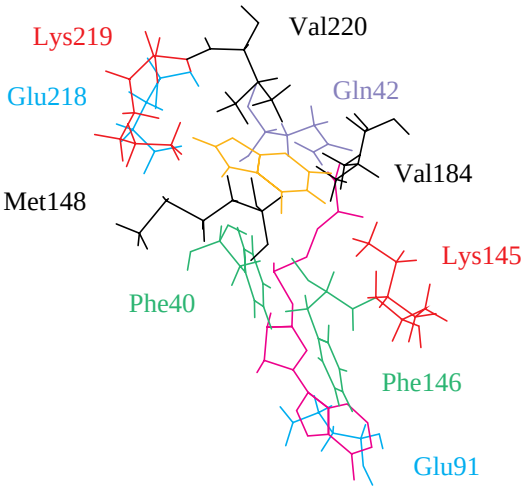


Figure 5 *A closer view at the putative ligand binding sites in the hTPMT model showing the residues predicted to interact with the ligands. Color coding; magenta: AdoMet, yellow: 6-MP, black: hydrophobic residues suggested to interact with 6-MP, red: positively charged residues, blue: negatively charged residues, green: aromatic residues. Phe40 is suggested to have a important catalytical role*

Favourable interactions were found between the substrate 6-MP and the surroundings (Figure 5). An electrostatic attraction was predicted between the electronegative sulphur on 6-MP and the positively charged sulphonium-methyl part of AdoMet. Possible H-bonds were found from N7 on 6-MP to Lys145 sidechain and from hydrogen at N2 to Glu218 sidechain.

Discussion

Resemblance between target and template proteins in both sequence identity, homology and functional properties is important in the construction of a model. An error in the sequence alignment affects all succeeding steps in the model construction, resulting in a model of low reliability. The most often used method in the construction of a homology model is based on an automated pairwise or multiple sequence alignment, including one structural template protein. However, the automated mode of sequence alignment may result in a model of relatively low significance when the homology between the target protein and the sequences in the alignment is low, a description that fits for hTPMT compared to the other AdoMet utilizing methyl transferases. The structural superposition of three AdoMet utilizing methyltransferases [22] gave the possibility to include structural information from three templates in the multiple sequence alignment, a strategy expected to give a more reliable model than just basing the model building on only one structural template.

In the crystal structure of rCOMT motif I (residues 62-70) forms a turn joining the first β-strand and α-helix of the β1-α4-β2 fold (Rossmann fold). The alignments procedure indicates that motif I seems to correspond to the segment VFFPLCGKA (residue 66-74) in hTPMT. The C-terminal end of this segments location 17 residues upstream the acidic Glu91 at the C-terminal end of β2 (Figure 1) strengthens this conclusion. All sequences included in the alignments have an acidic amino acid located 17-21 residues from the C-terminal end of motif I, which has been shown to be involved in binding to AdoMet both in rCOMT [19] and rGAMT [25]. The crystal structure of rCOMT shows that motif II (residue 133-140) forms β4. The alignment procedure suggests that motif II corresponds to the segment SIFDLPRTI (residues 135-142) in hTPMT. The loop segment flanking motif II has been shown to be important for binding of AdoMet both in rCOMT and rGAMT. Figure 1 indicates that the putative corresponding loop segment in hTPMT contains several aromatic and acid residues that may have a similar role as the DHWK segment in rCOMT and the DTYP segment in rGAMT. Figure 1 also indicates that this loop is longer in hTPMT than in rCOMT, and thereby may have a different structure. However, the modeling refinements procedure positioned Lys145 and Phe146 at positions similar to that of Lys144 and Trp143 in rCOMT, while Asp147 was positioned very close to a position similar to that of Asp141 in rCOMT, indicating that the GKFD segment may have a similar role as the DHWK segment of rCOMT and the DTYP segment in

rGAMT. Motif III (residue 159-168) is located in β5 in the crystal structure of rCOMT, and the alignments procedure suggests that motif III correspond to the segment LLGHHFQYLL (residue 173-182) in hTPMT.

The three sequence motifs widely present among the small molecule methyltransferases were not easily recognized in hTPMT. The DNA-metyltranserases lack sequence motifs II and III, but possess a three-dimensional fold of the AdoMet binding domain similar to that of the small molecule methyltransferases. In spite of the relative low overall amino acid sequence similarity (44.6 %) between rCOMT and rGNMT, their crystal structures show that they have a similar three-dimensional fold. The amino acid sequence similarity between hTPMT and rCOMT is 51.0 %, while the similarity between hTPMT and rGNMT is 44.3 %. Altogether, these observations indicate that hTPMT has a three-dimensional fold similar to rCOMT, rGNMT and the AdoMet binding domain of the DNA methyltransferases.

Model evaluation and quality

The model was subjected to a series of structural tests for its internal consistency and reliability. The Ramachandran map (Figure 3) indicates that the backbone conformations of the model structure are of nearly as good reliability as those of the template structures. A Whatif/Quality Control overall value of -1.6 may be considered as relatively low. However, water molecules were not included in the structural refinements procedure, and the average quality control value of the hTPMT model is close to the value for the template structures *hhai*, and *Taq1*. Thus, the backbone conformations (Figure 2), the packing environment of side chains, and the energy profile of each amino acid residue (Figure 4) indicate a reliable model.

AdoMet binding

The structure of AdoMet in the rCOMT-AdoMet complex structure [19] was docked into the hTPMT model at a site (cavity) corresponding spatially to the position of AdoMet in the rCOMT-AdoMet complex. The purpose was to look for interactions between functional groups on the AdoMet molecule and the surrounding residues constituting the AdoMet binding site on hTPMT. Interestingly, Phe40 was found in the vicinity of the sulphonium-methyl part of AdoMet, implying that this aromatic residue has a catalytic role in the transmethylation by interactions with the positively charged donor group of AdoMet, analogous to the cation-π interaction mechanism suggested for Tyr137 in rGAMT [25]. An aromatic residue is found at a position identical to Phe40 in 5 of the 20 sequences included in the alignment, while the other sequences have several aromatic residues in the putative loop joining β4 and α7, suggesting the necessity of an aromatic residue in the catalytic centre, performing the same role as proposed above.

TPMT inhibitors and substrates

Different inhibitors for TPMT and the TPMT substrate 6-MP [2, 14, 31] were docked into the putative substrate binding site. The inhibitors, briefly described as consisting of a benzoic acid core with different meta, ortho and para substituents, has been predicted to exert their inhibitory effect situated in a hypothetical hydrophobic cleft in the enzyme [31]. Based on experimental observations the authors also concluded that the most important factor in the binding of benzoic acid derivatives to the enzyme was the hydrophobicity of one of the meta substituents. We docked some potent inhibitors into the substrate binding cleft, even though some benzoic acid derivatives may be noncompetitive inhibitors and may exert their inhibitory effect by binding elsewhere on both the free enzyme and the enzyme-substrate complex, a phenomenon associated with noncompetitive inhibition. By visual inspection of the model of the protein-ligand complex the following predictions were made:

i) Some of the residues (Met148, Val184, Val220 and Phe40) surrounding the substrate or inhibitor are hydrophobic and favour the enclosure of generally hydrophobic ligands through Van der Waal interactions.

ii) A net positive charge of the cleft, due to charges in the surroundings (Lys145, Glu218, Lys219 and the sulphonium-methyl part of AdoMet), should favour hydrophobic negatively charged or neutral ligands to bind to the cleft.

iii) The spatial mutual orientation of the positive charges around the cleft is nearly identical to the mutual orientation

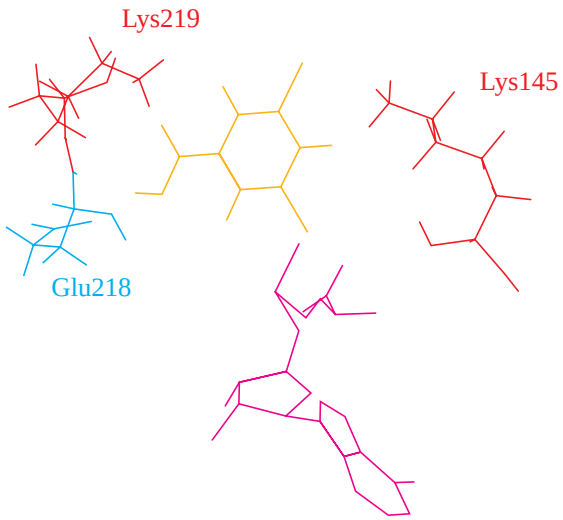


Figure 6 *Charged residues in the substrate/inhibitor binding cleft are suggested to influence on ligand orientation and affinity. Color coding; magenta: AdoMet, yellow: 3,5-dichlorobenzoic acid,a hTPMT inhibitor, with the chlorides pointing in the direction of the sulphonium-methyl part of AdoMet and the Lys145 sidechain, red: positively charged residues, blue: negatively charged residue*

of the sustituents of 1,3,5-trisubstituted benzene ring (Figure 6). Therefore, a ligand of this category with hydrophobic substituents with a negative electrostatic surface should bind with relatively high affinity and in a preferred orientation promoted by the attractive forces between the positive charges in the surroundings and the electron-rich substituents.

The 3,5-disubstituted benzoic acid inhibitors docked into the cleft were oriented with the carboxyl group towards Glu218 and Lys219. This favoured H-bonds from the Lys218 sidechain to the carbonyl oxygen of the inhibitor and from the carboxyl oxygen at the Glu218 sidechain to the hydrogen at the carboxyl of the inhibitor.

iv) The 3,5-substituted benzoic acid should have higher inhibition potency compared to its 3-substituted analogue due to one extra coordination with the surroundings.

The assumtions (i-iv) above are verified by experimental results reported in the literature [2, 14, 31]:

i) The name 'aryl thiol methyltransferase' has been proposed for TPMT because of its diverse specificity for aromatic substrates.

ii) The great majority of the aromatic/heterocyclic thiol substrates and the benzoic acid inhibitors tend to be negatively charged due to their deprotonated sulfhydryl group and carboxyl group, respectively.

iii) The majority of the most potent inhibitors reported for TPMT are 3,5-disubstituted or 3,4,5-trisubstituted benzoic acids. The compounds substituted in 3- and 5-position with two identical hydrophobic substituents with a negative electrostatic surface seem to be better inhibitors compared to compounds with two identical hydrophilic (polar) meta-substituents.

iv) Benzoic acids with the same substituent in both 3- and 5- position seem to be better inhibitors than benzoic acids with the same substituent in the 3- position only. An eksample is 3,5-dichlorobenzoic acid versus 3-chlorobenzoic acid.

Favourable interactions were also predicted between the substrate 6-MP and the surroundings. Hydrophobic interactions are described above. In addition, when 6-MP is positioned with its electronegative sulphur towards the positively charged sulphonium-methyl part of AdoMet, possible H-bonds are seen from N7 on 6-MP to Lys145 sidechain and from hydrogen at N2 to Glu218 sidechain. A large substituent attached to N9 does not seem to interfere sterically with the protein as this substituent points out of the cleft to the outside of the protein. This may be the explanation why large ligands such as 6-MP riboside diphosphate serve as substrates for TPMT.

The crystal structure of rCOMT complexed with one of its competitive inhibitors [19] shows that the ligand binding groove in rCOMT has functional/structural similarities with the substrate binding cleft in the present model.

The ligand binds to the groove in rCOMT through interactions with the charged sulphonium-methyl part of AdoMet and the Mg^{2+}-ion, the charged sidechains of Lys144, Glu199 and the hydrophobic residues Trp38 and Trp143. The interacting residues in the present model are the charged sulphonium-methyl part of AdoMet, Lys145, Glu218, Lys219 and the hydrophobic residues Met148, Val184, Val220 and Phe40. The structural similarities between the ligand binding groove in rCOMT and the hTPMT model are further confirmed by the fact that some of the benzoic acid derivatives are inhibitors of rCOMT [33].

Consequences of TPMT mutations

Most of the mutations naturally occurring in the hTPMT gene are associated with substitution of residues in the structure and loss of enzyme activity [7, 8, 34]. The model of hTPMT was used to predict possible structural consequences of these substitutions.

Nucleotide transition 460G-A results in an Ala154-Thr substitution. Ala154 is situated in the loop between β4 and α7 on the surface of the protein and is connected to a part of the AdoMet binding site. The Ala-Thr substitution may introduce structural alterations into the loop affecting the affinity for AdoMet and the activity of the enzyme. It has been shown [35] that the intrinsic stability of wild type hTPMT enzyme is enhanced by the presence of its cosubstrate AdoMet, a known phenomenon for enzymes in general. However, the intrinsic stability of the Ala154-Thr mutated protein was not stabilized by AdoMet [35]. This observation verifies our prediction by indicating that the mutated protein has lost its affinity for AdoMet.

Transition 719A-G results in a Tyr240-Cys substitution. Tyr240 is situated in the outer C-terminal part of the peptide chain. This part and the outer N-terminal part of the structure was not included in our model since the structure of these domains could not be predicted from the structural templates. However, inspection of the model showed that the terminal domains, and therefore the substitution Tyr240-Cys most probably is situated at the surface of the protein with no apparent intrinsic structural consequences. Experimental studies showed that the intrinsic stability of the Tyr240-Cys mutated protein was enhanced remarkably by the presence of AdoMet [35], which indicates that the mutation has minor structural influences on the binding site of AdoMet since the affinity for AdoMet is preserved.

Transition 238G-C gives an Ala80-Pro substitution which may result in a kink in helix α4. A kink in α4 further affects the conformation of the loop between β1 and α4 which is in very close proximity to AdoMet and probably interacts with the ligand. A significant structural alteration is indicated by the experimental results [35] showing that the Ala80-Pro mutation causes the protein to be degraded rapidly.

Transition 146T-C causes the substitution Leu49-Thr in a loop on the surface of the protein with no intrinsic structural consequences. But an introduction of a hydroxyl in this region of the chain may establish a H-bond to a spatially close Arg in a sequentially distant part of the chain, affecting the dynamics of the protein by introducing rigidity to the system, with loss of activity as a result.

Transition 539A-T causes the substitution Tyr180-Phe in β6. In the model, the hydroxyl group of Tyr180 has a strong H-bond with Lys228 in β7. Tyr180 is located four amino acid residues upstream of Val184, which is a part of the putative

substrate binding pocket (Figure 5). Removing the H-bond possibility by introducing a phenylalanine may increase the structural flexibility of the region around Val184 and thereby decrease the activity of the enzyme.

Heterologous yeast expression of wild type hTPMT and hTPMT carrying mutations resulting from both nucleotide transitions 460G-A and 719A-G showed that mutant hTPMT mRNA levels were comparable to wild type, indicating that the mutations had no effect on transcription or mRNA stability [36]. hTPMT protein levels were found to be 400-fold less in yeast producing the mutant protein compared to yeast producing the wild type protein, indicating a post-transcriptional mechanism for the loss of hTPMT activity in yeast. Based on our model, some of the most frequent mutations associated with experimentally measured loss of enzyme activity, could be classified as mutations introducing an alteration of the surface of the protein with no major changes in the tertiary structure. A post-transcriptional mechanism considering a correlation between mutational changes on the protein surface and protein instability involves proteases, protein degrading enzymes detecting molecular determinants on proteins that target them for proteolysis. A substitution of a surface residue might generate or contribute to the exposure of a marker making the protein vulnerable to proteolytic activity. This mechanism has been experimentally proved [35] for some of the mutations discussed above.

Conclusions

The amino acid sequence similarities of hTPMT with rCOMT and rGAMT indicates that hTPMT has a three dimensional fold similar to that of rCOMT and rGAMT and the AdoMet binding domain of the DNA-methyltransferases. In the present study we have constructed a three dimensional model of hTPMT by molecular modeling, based on a multiple amino acid sequence alignment of AdoMet dependent methyltransferases. Comparison of the conformation and packing properties of the model with those of rCOMT, *HhaI* and *TaqI* and structures in the PDB-database indicated a reliable overall model structure. The structural and functional properties of the model agreed with experimental results from the literature and indicated a structural explanation for inhibitor/substrate preferences of hTPMT. However, in some regions of the model the structure needs further verification, and experimental work validating the structure of the model are in progress in our lab. The present model provide a useful approach for further experimental studies of hTPMT and its ligand interactions, and may contribute to a better understanding of the structural effects of TPMT mutations. However, a further and definitive assessment of the three dimensional structure of hTPMT and its ligand interactions awaits an x-ray crystal structure.

Acknowledgements This work was supported by The Norwegian Cancer Society and The Norwegian Supercomputing Committee.

References

1. Remy, C. N. *J. Biol. Chem.* **1963,** *239*, 1078.
2. Woodson, L. C.; Ames, M. M.; Selassie, C. D.; Hansch, C.; Weinshilboum, R. M. *Mol. Pharmacol.* **1983,** *24*, 471.
3. Tidd, D. M.; Paterson, A. R. *Cancer. Res.* **1974,** *34*, 738.
4. Lennard, L.; Rees, C. A.; Lilleyman, J. S.; Maddocks, J. L. *Br. J. Clin. Pharmacol.* **1983,** *16*, 359.
5. Weinshilboum, R. M.; Sladek, S. L. *Am. J. Hum. Genet.* **1980,** *32*, 651.
6. Weinshilboum, R. M. *Xenobiotica* **1992,** *22*, 1055.
7. Szumlanski, C.; Otterness, D.; Her, C.; Lee, D.; Brandriff, B.; Kelsell, D.; Spurr, N.; Lennard, L.; Wieben, E.; Weinshilboum, R. *DNA. Cell. Biol.* **1996,** *15*, 17.
8. Tai, H. L.; Krynetski, E. Y.; Yates, C. R.; Loennechen, T.; Fessing, M. Y.; Krynetskaia, N. F.; Evans, W. E. *Am. J. Hum. Genet.* **1996,** *58*, 694.
9. Lennard, L.; Van Loon, J. A.; Weinshilboum, R. M. *Clin. Pharmacol. Ther.* **1989,** *46*, 149.
10. Lennard, L.; Gibson, B. E.; Nicole, T.; Lilleyman, J. S. *Arch. Dis. Child*. **1993,** *69*, 577.
11. Evans, W. E.; Horner, M.; Chu, Y. Q.; Kalwinsky, D.; Roberts, W. M. *J. Pediatr.* **1991,** *119*, 985.
12. Aarbakke, J.; Janka Schaub, G.; Elion, G. B. *Trends. Pharmacol. Sci.* **1997,** *18*, 3.
13. Lysaa, R. A.; Giverhaug, T.; Wold, H. L.; Aarbakke, J. *Eur. J. Clin. Pharmacol.* **1996,** *49*, 393.
14. Deininger, M.; Szumlanski, C. L.; Otterness, D. M.; Van Loon, J.; Ferber, W.; Weinshilboum, R. M. *Biochem. Pharmacol.* **1994,** *48*, 2135.
15. Szumlanski, C. L.; Weinshilboum, R. M. *Br. J. Clin. Pharmacol.* **1995,** *39*, 456.
16. Cheng, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J. *Cell* **1993,** *74*, 299.
17. Labahn, J.; Granzin, J.; Schluckebier, G.; Robinson, D. P.; Jack, W. E.; Schildkraut, I.; Saenger, W. *Proc. Natl. Acad. Sci. U.S.A.* **1994,** *91*, 10957.
18. Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995,** *82*, 143.
19. Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994,** *368*, 354.
20. Fu, Z.; Hu, Y.; Konishi, K.; Takata, Y.; Ogawa, H.; Gomi, T.; Fujioka, M.; Takusagawa, F. *Biochemistry* **1996,** *35*, 11985.
21. Honchel, R.; Aksoy, I. A.; Szumlanski, C.; Wood, T. C.; Otterness, D. M.; Wieben, E. D.; Weinshilboum, R. M. *Mol. Pharmacol.* **1993,** *43*, 878.
22. O'Gara, M.; McCloy, K.; Malone, T.; Cheng, X. *Gene* **1995,** *157*, 135.
23. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994,** *22*, 4673.
24. Kagan, R. M.; Clarke, S. *Arch. Biochem. Biophys.* **1994,** *310*, 417.
25. Hamahata, A.; Takata, Y.; Gomi, T.; Fujioka, M. *Biochem. J.* **1996,** *317*, 141.
26. Takata, Y.; Konishi, K.; Gomi, T.; Fujioka, M. *J. Biol. Chem.* **1994,** *269*, 5537.

27. Wierenga, R. K.; Hol, W. G. J. *Nature* **1983,** *302*, 842.
28. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. M.; Kollmann, P. A. *J. Am. Chem. Soc*. **1995,** *117*, 5179.
29. Ferrin, T. E.; Huang, C. C.; Langridge, R. *J. Mol. Graphics* **1988,** *6*, 1.
30. Vriend, G.; Sander, C. *J. Appl. Cryst.* **1993,** *26*, 47.
31. Ames, M. M.; Selassie, C. D.; Woodson, L. C.; Van Loon, J. A.; Hansch, C.; Weinshilboum, R. M. *J. Med. Chem.* **1986,** *29*, 354.
32. Laskowski, R. A. *J. Appl. Cryst.* **1993,** *26*, 283.
33. Borchardt, R. T.; Huber, J. A.; Houston, M. *J. Med. Chem.* **1982,** *25*, 258.
34. Otterness, D.; Szumlanski, C.; Wood, T.; Lennard, L.; Klemetsdal, B.; Aarbakke, J.; ParkHah, J.; Iven, H.; Branum, E.; OBrien, J.; Weinshilboum, R. *Clin. Pharmacol. & Ther.* **1997,** *61*, OIB4.
35. Tai, H. L.; Krynetski, E. Y.; Schuetz, E. G.; Yanishevski, Y.; Evans, W. E. *Proc. Natl. Acad .Sci. U.S.A* **1997,** *94*, 6444.
36. Krynetski, E. Y.; Tai, H. L.; Yates, C. R.; Fessing, M. Y.; Loennechen, T.; Schuetz, J. D.; Relling, M. V.; Evans, W. E. *Pharmacogenetics* **1996,** *6*, 279.