J Mol Model (2001) 7:324–333
DOI 10.1007/s008940100043

ORIGINAL PAPER

M. A. Thompson · R. M. Weinshilboum · J. El Yazal
T. C. Wood · Y.-P. Pang

Rabbit indolethylamine *N*-methyltransferase three-dimensional structure prediction: a model approach to bridge sequence to function in pharmacogenomic studies

Received: 9 March 2001 / Accepted: 19 June 2001 / Published online: 7 September 2001

Abstract Pharmacogenomics is the study of the genetic basis for individual variation in response to drugs and other xenobiotics. Successful prediction of effects of genetic variations that change encoded amino acid sequences on protein function and their consequent biomedical implications depends on three-dimensional (3D) structures of the encoded amino acid sequences. To bridge sequence to function, thus facilitating an in-depth pharmacogenomic study, we tested the feasibility of the use of a semi-computational approach to predict 3D structures of rabbit and human indolethylamine *N*-methyltransferases (INMTs) from their amino acid sequences, which share less than 26% sequence identity with known protein 3D structures. Herein, we report 3D models of INMTs predicted by using the crystal structure of rat catechol *O*-methyltransferase as a template, testing of the models both computationally and experimentally, and successful use of the models in retrospective prediction of the effects of genetic polymorphisms and in identification of residues that contribute to observed species-specific differences in substrate affinity. The results encourage the use of the semi-computational approach to predict 3D protein structures for use in pharmacogenomic studies when de novo prediction of protein 3D structures from their amino acid sequences is still not feasible and X-ray crystallography and/or solution nuclear magnetic resonance spectroscopy can only determine 3D structures for a small number of known amino acid sequences.

Keywords Single nucleotide polymorphisms · *S*-Adenosyl-*L*-methionine · Homology modeling · Molecular dynamics simulations · Molecular docking

Electronic supplementary material to this paper can be obtained by using the Springer LINK server located at http://dx.doi.org/10.1007/s008940100043

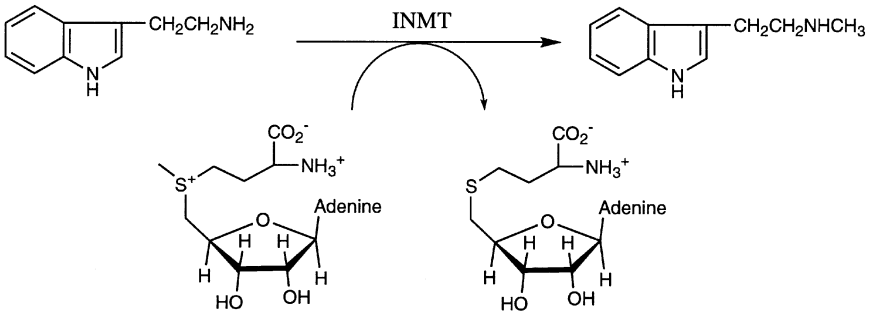
M.A. Thompson · R.M. Weinshilboum · J.E. Yazal · T.C. Wood
Y.-P. Pang (✉)
Department of Molecular Pharmacology
and Experimental Therapeutics, Mayo Clinic,
200 First Street SW, Rochester, MN 55905, USA
e-mail: pang@mayo.edu
Tel.: 507-284-7868, Fax: 507-284-9111

Y.-P. Pang · R.M. Weinshilboum
Mayo Clinic Cancer Center, 200 First Street SW,
Rochester, MN 55905, USA

Y.-P. Pang
Tumor Biology Program, Molecular Neuroscience Program,
Mayo Foundation for Medical Education and Research,
200 First Street SW, Rochester, MN 55905, USA

Introduction

Pharmacogenomics is the study of the genetic basis for individual variation in response to drugs and other xenobiotics. Successful prediction of the effects of genetic variations that change the encoded amino acid sequence on protein function and their consequent biomedical implications depends on three-dimensional (3D) structures of the encoded amino acid sequences. At present, de novo prediction of a protein 3D structure from its amino acid sequence is still not feasible; and only a small number of the known proteins have had their 3D structures determined by X-ray crystallography and/or solution nuclear magnetic resonance (NMR) spectroscopy. [1, 2] There is currently great interest in computational approaches that can help bridge sequence to function to facilitate in-depth pharmacogenomic studies. Reliable prediction of a protein 3D structure from its homologue with its 3D structure determined requires at least 30% sequence identity between the query protein and its homologue. [3] Unfortunately, a number of methyltransferases most relevant to our pharmacogenomic studies share less than 30% sequence identity with known protein 3D structures. This posed a tremendous challenge in predicting the 3D structures of methyltransferases of interest, because it is difficult to single out the correct sequence alignment from many possible alignments of the query protein over the sequence of the homologue as a template when low sequence identity is found between the two proteins. In this challenge, we investigated the



Scheme 1 *N*-Methylation of tryptamine catalyzed by INMT using AdoMet as a methyl donor

feasibility of a semi-computational approach [4, 5, 6, 7, 8, 9, 10] to the prediction of the 3D structures of methyltransferases such as indolethylamine *N*-methyltransferases (INMTs) and histamine *N*-methyltransferases (to be published). This approach identifies a handful of plausible sequence alignments, generates a 3D model according to each sequence alignment, triages the models derived from the incorrect alignments by the use of molecular dynamics simulations and docking studies, and experimentally tests the remaining 3D models.

Accordingly, we first set out to predict 3D structures of rabbit and human INMTs that catalyze the *N*-methylation of tryptamine, serotonin, and other structurally related amines (Scheme 1). [11, 12] INMT is one member of a large group of methyltransferases (MTs) that use *S*-adenosyl-*L*-methionine (AdoMet) as a methyl donor for the enzymatic methylation and share a conserved catalytic domain structure. [13, 14] The reasons that we chose to work with this enzyme were twofold. First, rabbit and human INMTs are 88% identical in amino acid sequence. [11, 12] Recombinant rabbit INMT catalyzes the methylation of a number of indoleamines such as tryptamine, whereas the corresponding enzyme in human has higher apparent K_M values for these substrates. [11, 12] Human INMT also has two single nucleotide polymorphisms (SNPs) that change encoded amino acids. Nucleotide 613 is either A or G, resulting in Met205 or Val205, while nucleotide 656 is either G or A, resulting in amino acid 219 being either Gly or Glu, respectively. [12] These two recombinantly expressed human INMT alleles have nearly identical apparent K_M values for tryptamine and AdoMet. From the pharmacogenomic standpoint, it would be interesting to determine whether the predicted 3D structures of human and rabbit INMTs could provide insight into the functional consequences of SNPs and the observed species-specific difference in substrate affinity. Second, on the basis of our survey of all MT 3D structures deposited in the Protein Data Bank, [15] the two INMT amino acid sequences are remote in sequence identity to known 3D structures of proteins (see later). From the computational biology standpoint, although low-homology sequences are challenging to model, successful predictions of such 3D models could serve to demonstrate the utility of computational approaches in pharmacogenomic studies.

Herein, we report the 3D models of rabbit and human INMTs predicted by using the crystal structure of rat catechol *O*-methyltransferase [16] as a template, testing of the hypothetical models both computationally and experimentally, and the successful use of the 3D models in pharmacogenomic studies. These results encourage the use of computer modeling to predict 3D protein structures in pharmacogenomic studies at a time when X-ray crystallographic and/or NMR spectroscopic methods can only determine 3D structures for a small number of known amino acid sequences.

Materials and methods

Materials

[^{14}C-CH$_3$]AdoMet (60 mCi mmol^{-1}) was obtained from NEN Life Science Products, Inc. (Boston, Mass.). Dulbecco's modified Eagle's medium and fetal calf serum were obtained from Life Technologies, Inc. (Gaithersburg, Md.). Bovine serum albumin, AdoMet HCl, dimethyl sulfoxide, *N*-methyltryptamine, and tryptamine HCl were obtained from Sigma (St. Louis, Mo.). α-Methyltryptamine was from Aldrich Chemical Co. (Milwaukee, Wisc.).

Site-directed mutagenesis

The PCR-based SDM method of Ho et al. [17] was used to generate PCR products for cloning into the expression vector pCR3.1 according to the manufacturer's instructions (Invitrogen, Carlsbad, Calif.). Inserts in expression vectors were sequenced on both strands to ensure that no alterations in sequence had been introduced during PCR amplification. Expression construct DNA (4 μg) was then used to transfect COS-1 cells using the DEAE-dextran method [18] as described elsewhere. [19] β-Galactosidase was co-transfected as a control, and INMT enzymatic activities were normalized to the β-galactosidase activity measured with the Promega β-galactosidase assay system (Madison, Wisc.). The COS-1 cells were harvested and a cytosol preparation was used for enzyme assays as previously described. [19]

INMT assay

The assay used to measure INMT activity for the recombinant rabbit or human cDNA expressed in COS-1 cells was as described elsewhere. [11, 12] This assay utilized [^{14}C-CH_3]AdoMet as a methyl donor. The formation of ^{14}C-methylated product was determined after incubation for 60 min at 37 °C in a total volume of 200 µl that contained 50 mM Tris-HCl, pH 8.5, 250 µg ml^{-1} BSA, 34 µM [^{14}C-CH_3]AdoMet (24 mCi mmol^{-1}) and methyl acceptor substrate. The reaction was terminated by the addition of 0.5 ml of 0.5 M potassium borate, pH 10, followed by extraction of the methylated product using 5 ml of toluene:isoamyl alcohol (97%:3%, v:v). The organic solution (3.5 ml, separated from the aqueous solution) was added to 5 ml of BioSafe II (Research Products International Corp., Mount Prospect, Ill.) prior to the determination of radioactivity. Assays were performed in duplicate, and values reported are averages of those duplicate determinations. Protein concentrations were measured with the dye binding assay of Bradford [20] with bovine serum albumin as a standard. Apparent K_M and V_{max} values were calculated by the method of Wilkinson [21] with a computer program written by Cleland. [22]

Molecular dynamics simulations

All molecular dynamics (MD) simulations were carried out on an SGI Octane computer (2xR10,000, 225 MHz) employing the AMBER 5.0 program [23] with the Cornell et al. force field and additional force field parameters provided in Supporting Information. [24] The values of keywords in uppercase letters used by the AMBER program are described in parentheses. All MD simulations used (1) the SHAKE procedure for all bonds of the system (NTC=3 and NTF=3); (2) a time step of 1.0 fs (DT=0.001); (3) a dielectric constant ε=1.0 (IDIEL=1.0); (4) the Berendsen coupling algorithm (NTT=1); (5) the Particle Mesh Ewald method for calculating the electrostatic interactions (BOXX=71.2007, BOXY=63.4124, BOXZ=53.5497, ALPHA=BETA=GAMMA=90, NFFTX=NFFTY=NFFTZ=64, SPLINE_ORDER=1, ISCHARGED=0, EXACT_EWALD=0, DSUM_TOL=0.00001); (6) the non-bonded atom pair list updated at every 20 steps (NSNB=20); (7) a distance cutoff of 8 Å to calculate the non-bonded steric interactions (CUT=8.0 Å); and (8) defaults of other keywords not mentioned here. The protein was simulated in a TIP3P water box with a periodic boundary condition at constant temperature (298 K) and pressure (1 atm) (NCUBE=20, QH=0.4170, DISO=2.20, DISH=2.00, CUTX=CUTY=CUTZ=8.2, NTB=2, TEMP0=298, PRES0=1, TAUTP=0.2, TAUTS=0.2, TAUP=0.2, NPSCAL=0, and NTP=1). The resulting system consisting of 21,892 atoms was first energy minimized for 500 steps in order to remove close van der Waals contacts in the system. The minimized system was then slowly heated to 298 K (10 K/ps, NTX=1) and equilibrated for 100 ps before a 1.0 ns simulation.

Substrate docking studies

The Restrained ElectroStatic Potential (RESP) charges [24] of AdoMet, tryptamine, *N*-methyltryptamine, α-methyltryptamine, and 5-hydroxytryptamine were calculated by employing the Gaussian98 program with the HF/6-31G*//HF/6-31G* method. [25]

All thermodynamically accessible conformations of tryptamine and its analogs were generated by the conformational search program CONSER (devised by Y.-P. Pang). This program first generates conformations by specifying all discrete possibilities at 60° of arc increment in a range of 0–360° for all rotatable torsions. The program then optimizes such conformers through energy minimizations with the RESP charges and the Cornell et al. force field. It thereafter performs a cluster analysis to delete duplicates. In the cluster analysis, two conformers were judged different if at least one of the defined torsions differed by more than 30° of arc.

All docking studies were performed by using the EUDOC program. [26] This program systematically translates and rotates a ligand in a putative binding pocket of a receptor to search for energetically favorable orientations and positions of the ligand. A box is defined within the binding pocket to confine the translation of a ligand. The resolution of a docking study was determined by the translational and rotational increments and the size of the box employed. In this work, the translational and rotational increments were set at 1.0 Å and 10° of arc, respectively. The box size was 9×3×5 Å^3. The energy used to judge the preferred orientation and position of the ligand is termed intermolecular interaction energy, and is defined as the potential energy of the complex relative to the potential energies of the ligand and receptor in their free states. The potential energies are calculated with the additive, all atom force field by Cornell et al. [24] A distance-dependent dielectric function was used to calculate the electrostatic interactions. The cutoff for steric and electrostatic interactions was set at 8.0 Å in this study. The estimated binding affinity for each ligand was calculated by subtracting the ligand solvation energy from the intermolecular interaction energy. The ligand solvation energy was calculated with the HF/6-31G*//HF/6-31G* method using the CPCM model implemented in the Gaussian98 program. [25]

Results

Semi-computational approach

To determine the 3D structures of INMTs, we used a four-step semi-computational approach. First, we aligned the INMT amino acid sequence with a related sequence of a known MT 3D structure (see later). Second, we gen-

erated and refined the INMT 3D structure according to the related MT of known 3D structure. Third, we computationally tested the 3D model for self-consistency via MD simulations and docking studies. If the model had been inconsistent with the assumptions used in the first two steps, we would have iterated the process from the first step using a different alignment or using a different sequence of a known 3D structure until the model was consistent with the assumptions. Fourth, we experimentally tested the 3D model with site-directed mutagenesis experiments. If the model had been inconsistent with the experimental observations, we would have iterated the four-step procedure until the model was consistent with the experimental observations.

Sequence alignment

We surveyed the 3D structures in the 1999 release of the Protein Data Bank [15] and found 33 X-ray structures for AdoMet-dependent MTs, which represented 13 distinct enzymes. All the MTs showed a similar folding pattern in the AdoMet-binding domain, with a central β-sheet surrounded by α-helices. [14, 27, 28, 29, 30, 31, 32, 33] Consistent with the report that nine AdoMet-dependent MTs shared a common catalytic domain structure, [13] our visual inspection revealed that the foldings of the catalytic, AdoMet-containing domains of the 33 crystal structures (13 AdoMet-dependent MTs) were similar despite low amino acid sequence identity and wide variation in methyl acceptor substrates. The crystal structures of rat catechol *O*-methyltransferase (COMT), [16] which was the only monomeric, small-molecule MT for which the crystal structure was determined, was thus considered as a template structure for our computer homology modeling of rabbit INMT, since INMT was a monomeric, small-molecule MT. Rat COMT was reportedly used as a template for predicting a 3D model of human thiopurine *S*-methyltransferase. [34] Rabbit INMT was first modeled rather than the human enzyme because the rabbit cDNA [11, 12] was cloned and functionally characterized prior to the human INMT cDNA. [11, 12] Unfortunately, the sequence identity between rabbit INMT and rat COMT was less than 26% calculated by using various sequence alignment algorithms, which discouraged the use of any computer-assisted sequence alignment.

A search for protein-folding propensity of the rabbit INMT sequence using the 3DPSSC program [35] suggested that the folding of rabbit INMT might be similar to those of rat glycine *N*-methyltransferase (GNMT) [30] and rat COMT [16]. The search algorithm of the 3DPSSC program is based on 1D and 3D sequence profiles coupled with secondary structure and solvation potential information. [35] Visual inspection of the two 3D structures of the alpha carbon atoms of rabbit INMT derived from rat COMT and rat GNMT by the 3DPSSC program revealed a discontinuity problem in the hypothetical active site of the two 3D INMT models, namely, a large separation of two adjacent alpha carbons located in the active site of INMT.

Given the above-mentioned problems, we had to resort to manual alignment of the rabbit INMT amino acid sequence over that of rat COMT. It has been reported that MT enzymes that utilize AdoMet as a co-substrate often contain three regions of high sequence homology, termed regions I–III, respectively. [36] A subsequent study of 84 non-DNA MT sequences revealed region I in 69 sequences, region II in 46 sequences, and region III in 61 sequences. [37] All three regions were present in 45 of the 84 sequences, and 60 sequences had motifs I and III. [38, 39] Region II was relatively varied [39] according to studies that emphasized the motif position as well as sequence. [37, 38, 39, 40, 41] In rat COMT, region I was located in close proximity to AdoMet, whereas region III formed a β-strand on the exterior of the enzyme located away from the active site. [16] Although region III was not involved in AdoMet binding, it was the second most common MT sequence motif [37] and a sequence that was unique to mammalian MTs. By visual inspection, we noticed similarity in the consensus sequences of regions I and III between the sequences of rabbit INMT and rat COMT. Therefore, we manually aligned regions I and III of rabbit INMT with the corresponding regions of the rat COMT template (Fig. 1). This manual alignment was consistent with the secondary structures of rabbit INMT and rat COMT predicted by the PHDsec program (Fig. 1). [42] Because the amino acid sequence of rabbit INMT was longer than that of rat COMT and because a number of carboxyl- and amino-terminal residues of rat COMT had not been determined in the X-ray crystal structure, we excised the last 20 amino acids of the carboxyl-terminus and a fragment extending from Pro156 to Ser185 in rabbit INMT, assuming that these fragments were located away from the active site like those fragments located away from the active site of rat GNMT. This assumption was legitimate as we were using the a posteriori approach to predict the 3D structure of rabbit INMT. If the joints of the excised fragments had later been found to be located in or near the active site of the homology-derived 3D model, this model would have been discarded and alternative sequence alignments with COMT or with other MTs of known 3D structures would have been pursued.

Construction of the 3D model of apoenzyme

A 3D structure of rabbit INMT was generated by employing the PROTEIN module of the Quanta 97 program [43] using the manual alignment shown in Fig. 1. Visual inspection of the 3D structure of rabbit INMT refined by energy minimizations in vacuo revealed that the model contained a pocket that was similar to the common AdoMet-binding pockets of other MTs. More importantly, we observed that the excluded segments were not near the AdoMet binding region, encouraging further refinement and validation of the a posteriori model

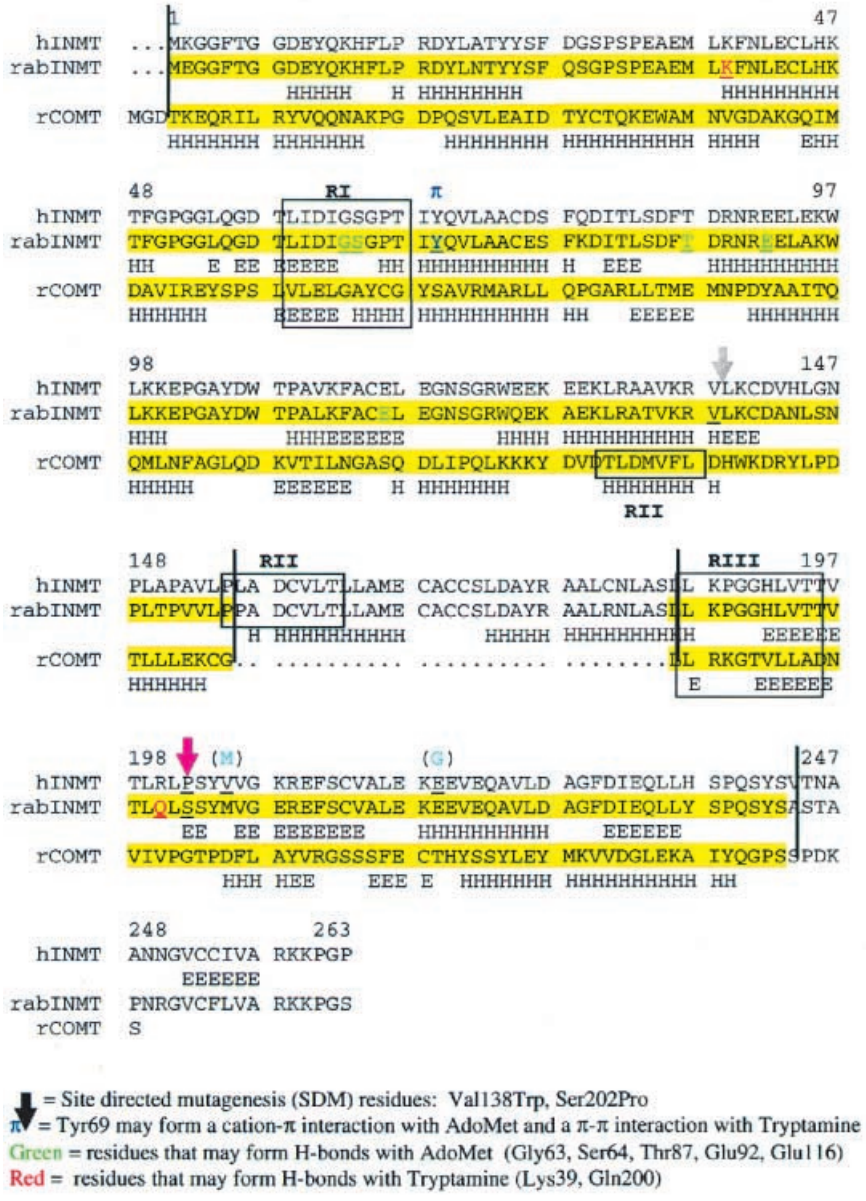


Fig. 1 Human (h) and rabbit (rab) INMT alignments with rat (r) COMT. Numbering is based on the INMT sequence. The residues highlighted in yellow were used in homology modeling. Methyltransferase signature sequence motif regions I, II, and III (RI, RII, and RIII, respectively) are boxed. Regions I and III were used in the homology model sequence alignment. Amino acids of interest are underlined

(Fig. 2). The 3D structure of rabbit INMT was then refined by a 1.0 ns MD simulation in water at room temperature. By examining instantaneous structures at 100 ps intervals during this 1.0 ns simulation as well as the time-average structure of 1000 instantaneous structures at 1 ps intervals, we observed by visual inspection that the 3D structure of apo-INMT did not diverge from the structures of the AdoMet binding domain for the 33 reported 3D structures of MTs, despite an observed loop conformational change during the 1.0 ns MD simulation, which is depicted in Fig. 3.

Construction of the 3D model of the binary complex

Docking the AdoMet molecule with the RESP charges and the pharmacophore identified in the crystal structure of the rat COMT complex into the 3D structure of rabbit

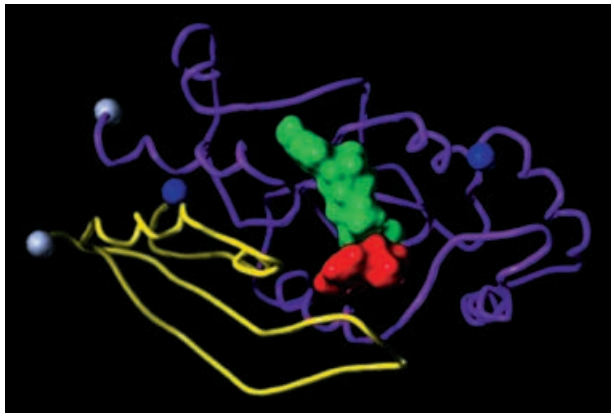


Fig. 2 Tube model of rabbit INMT liganded with AdoMet (*green*) and tryptamine (*red*) showing the joints (*blue ball*: the N-terminal; *white ball*: the C-terminal) of the internal "excluded" segment located away from the active site

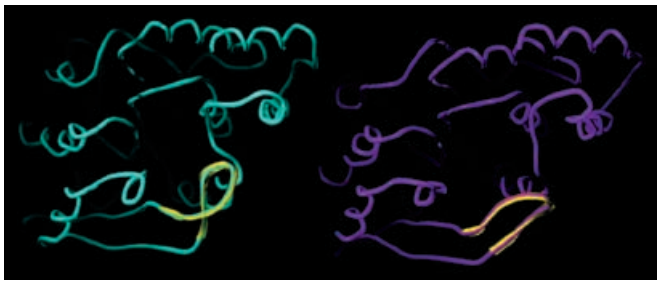


Fig. 3 Loop (*yellow*) conformational change resulted from a 1.0 ns MD simulation of rabbit INMT 3D structure in water (*left*: energy-minimized structure before the MD simulation; *right*: time-average structure of the MD simulation)

INMT averaged over a 1.0 ns MD simulation employing the EUDOC program with the conditions described above, we found that the AdoMet molecule docked in the same region of rabbit INMT as it did in the rat COMT complex, but the orientation of AdoMet was different from that in the rat COMT complex. This observation suggested that the structure of rabbit INMT was reasonable on the basis of the computational study. Otherwise, AdoMet could not be docked in the active site. The change of AdoMet orientation was probably due to the MD simulation of rabbit INMT in the absence of the co-substrate AdoMet, resulting in a slightly decreased binding pocket size in which AdoMet could not fit unless its orientation is changed. We then manually placed AdoMet in the active site of rabbit INMT based on the orientation identified in the rat COMT–AdoMet complex and refined the rabbit INMT–AdoMet complex with an MD simulation in water for 0.5 ns under the same conditions as those used during the simulation of the apoenzyme. The average structure of the rabbit INMT-AdoMet complex derived from 500 instantaneous structures at 1 ps intervals was then used as an MD-refined INMT–AdoMet binary complex in subsequent docking studies of prototypic substrates. In the most energetically stable rabbit INMT–AdoMet binary complex depicted in Fig. 4, the amino group of AdoMet formed a cation–π interaction with the aromatic ring of Tyr69 and hydrogen bonds with Gly63, Ser64, Thr87, Glu92, and Glu116 of rabbit INMT.

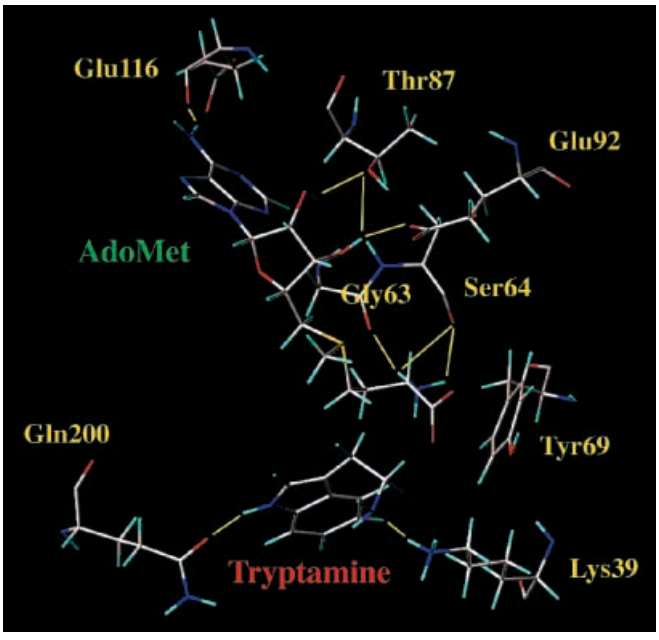


Fig. 4 Interaction map of the rabbit INMT–AdoMet–tryptamine complex

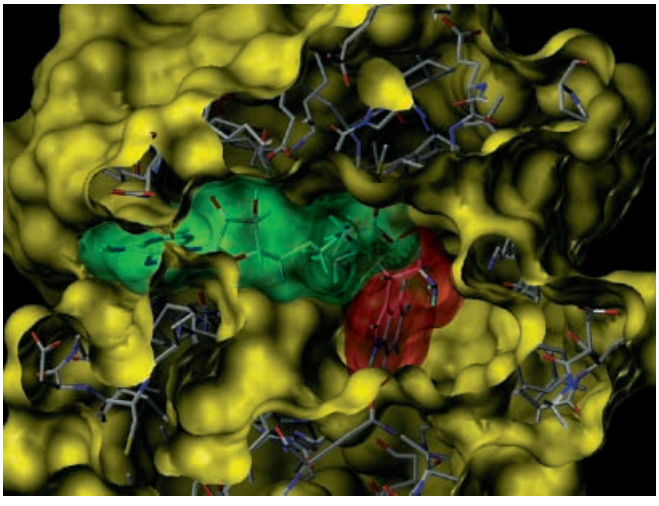


Fig. 5 Close-up of rabbit INMT liganded with AdoMet and tryptamine. (INMT, AdoMet, and tryptamine are wrapped with *opaque*, *green transparent*, and *red transparent* van der Waals surfaces, respectively)

Construction of the 3D model of the ternary complex

The prototypic INMT substrate tryptamine and its analogs with the RESP charges in different conformations generated by the CONSER program were docked into the rabbit INMT–AdoMet binary complex by employing the same methods as used above. It was encouraging that tryptamine was able to dock in the active site of the INMT–AdoMet binary complex on the basis of the intermolecular interaction energy (Fig. 5). This result was consistent with the experimental observation that tryptamine was a prototypic substrate for rabbit INMT, and it supported the model of the AdoMet–INMT complex. Otherwise tryptamine would not be able to dock in the active site of the binary complex. In the most energetically stable rabbit INMT–AdoMet–tryptamine ternary complex shown in Fig. 4, tryptamine formed hydrogen bonds with Lys39 and Gln200 and van der Waals interactions with the methylene group of Glu43, as well as with Val138 and Ile202. There was also a tilted T-shaped π–π interaction between tryptamine and Tyr69. By further examining the ternary complex depicted in Fig. 5, we found no steric hindrance that could prevent the ethylamino group of tryptamine from reacting with the methyl

group of AdoMet. This observation suggested that tryptamine was a potential substrate from the modeling standpoint.

Computational testing of the 3D INMT model

The program WHAT CHECK [44] was used to determine if changes in amino acid chirality took place during the model generation and refinement process. The energy-minimized rabbit INMT structure, and the rabbit INMT structure optimized by a 1.0 ns MD simulation in water were evaluated by WHAT CHECK and no changes in chirality were observed for the two protein structures. Therefore, the energy minimization process and the 1.0 ns MD simulation of INMT had not changed amino acid chirality.

The MD-optimized rabbit INMT 3D structure passed the following self-consistency tests: (1) The segments excluded in the homology model generation were not near the substrate binding site of the MD-optimized rabbit INMT 3D structure (Fig. 2), an observation consistent with the assumption made during the sequence alignment that the excised regions were probably located away from the substrate binding site. (2) The MD-optimized rabbit INMT 3D structure did not diverge from the structural motif of the AdoMet binding domain observed in the 33 reported crystal structures of MT enzymes, which agreed with the assumption that INMTs had a similar fold to that of rat COMT. (3) The co-substrate AdoMet was able to dock in the substrate binding region of the 3D model and formed a stable rabbit INMT–AdoMet binary complex. (4) The two Asp and one Asn residues that constitute the Mg^{2+} binding site in rat COMT were changed by evolution to two Val and one Thr residues in rabbit INMT, suggesting that Mg^{2+} was not required for activity. This observation is consistent with the experimental observation of a lack of Mg^{2+} requirement for INMT enzymatic activity, in contrast with COMT. [11] (5) The intermolecular interactions of AdoMet in rabbit INMT resembled those in rat COMT but not in rat GNMT. In the crystal structure of COMT, a Glu residue formed a "bidendate" hydrogen bond network with the two hydroxyl groups of the sugar ring of AdoMet; the adenine ring was in contact with a Met residue via *d*–π interaction (the π electrons of the adenine ring occupy the vacant *d* orbitals of the sulfur atom); and the amino group of AdoMet formed a salt bridge with a Glu. In the crystal structure of rat GNMT whose folding of the AdoMet binding domain was similar to that of rat COMT, one hydroxyl group of AdoMet formed a hydrogen bond with one His residue and the other hydroxyl group formed a hydrogen bond with one Tyr residue; the sulfur–adenine interaction was missing; and the amino group of AdoMet had a salt bridge with an Asp residue. In the proposed 3D structure of INMT, a Thr group was forming a "bidendate" hydrogen bond network with the two hydroxyl groups; the adenine ring was in contact with a Cys residue via *d*–π interaction; and the amino group of AdoMet formed a cation–π interaction with a Tyr residue. (6) The retrospective prediction of tryptamine as a feasible substrate was consistent with our experimental identification of tryptamine as a prototypic substrate. (7) Further docking studies performed with tryptamine analogs, using the same procedure as in the tryptamine docking, revealed that the estimated binding affinities of tryptamine, *N*-methyltryptamine, α-methyltryptamine, and 5-hydroxytryptamine qualitatively agreed with experimentally determined apparent K_M values (Table 1). Tryptamine, *N*-methyltryptamine, and α-methyltryptamine had relative binding energies of less than –20 kcal mol^{-1} (higher affinity) while that of 5-hydroxytryptamine was –14 kcal mol^{-1} (lower affinity). Thus the predicted apparent K_M value for 5-hydroxytryptamine was higher than that for the other indoleamines tested. This prediction was experimentally validated by the relatively high apparent K_M value for 5-hydroxytryptamine when compared with those of tryptamine, *N*-methyltryptamine and α-methyltryptamine.

Table 1 Experimental and computationally estimated binding affinities of indoleamines

INMT substrate	Calculated relative binding energy (kcal mol^{-1})	K_M (mM)	Relative activity (%)
Tryptamine	–24.5	0.27	100
N-Methyltryptamine	–24.2	0.086	44
α-Methyltryptamine	–23.2	0.49	131
5-Hydroxytryptamine	–14.0	1.38	11

Experimental testing of the 3D INMT model

To experimentally test the INMT 3D structure, we generated an expression construct by SDM and conducted substrate kinetic studies to confirm the predicted interactions of tryptamine with residue Val138 identified in the rabbit INMT–AdoMet–tryptamine 3D model. The Val138Trp mutant was designed to determine if the substitution of Val138 with a bulky amino acid residue in the putative tryptamine binding site would sterically disrupt tryptamine binding, since Val138 is located in the center of the putative active site of INMT (Fig. 6a) and this amino acid is conserved in rabbit and human INMT, mouse thioether *S*-MT, human and mouse nicotinamide *N*-MT, as well as rat, mouse, and human phenylethanolamine *N*-MT. [12] The Val138Trp mutant had a V_{max}/K_M ratio that was only 2% of that of the wild-type enzyme (Table 2). The steric hindrance to tryptamine binding caused by the Trp substitution is clearly visible in Fig. 6b. Although more SDM experiments would be needed for rigorous validation of the 3D INMT model, the confirmatory result of the Val138Trp mutant encouraged us to proceed with testing of the utility of the 3D model in pharmacogenomic studies, because the objective of the present work was not to use mutagenesis-coupled homology modeling to generate 3D INMT struc-

Table 2 Site directed mutagenesis study

Rabbit INMT	K_M±SE (mM)	V_{max}±SE [pmol/(min*mU β-gal)]	Rel. V_{max} (%)	V_{max}/K_M] [pmol/(min*mU β-gal)	Rel. V_{max}/K_M (%)
Wild-type	0.43±0.04	0.68±0.04	100	1.59	100
Val138Trp	3.50±0.76	0.09±0.02	13	0.03	2
Ser202Pro	1.27±0.17	0.21±0.02	30	0.16	10

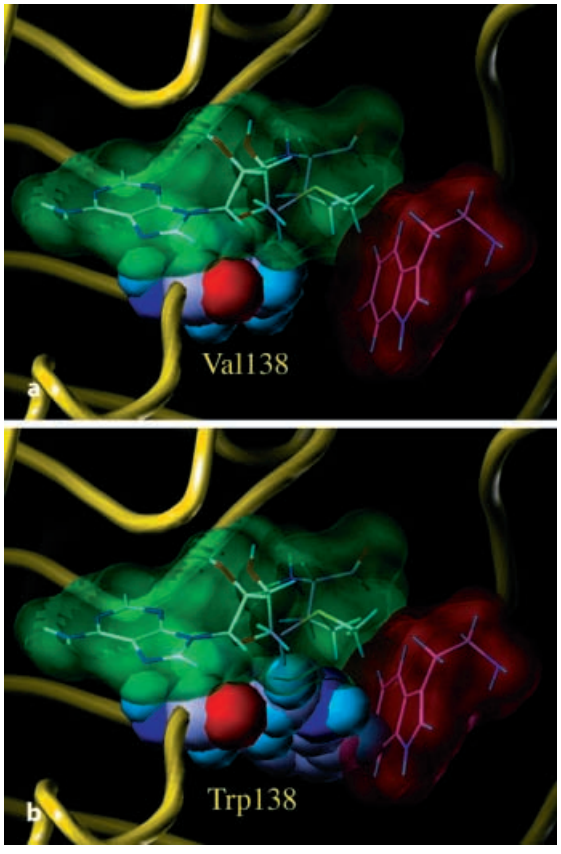


Fig. 6 Close-up of the tryptamine binding site in the ternary complex of wild-type rabbit INMT (**a**) and Val138Trp mutant rabbit INMT (**b**) showing steric hindrance in the center of the active site caused by Trp138

tures as accurate as those derived from X-ray and/or NMR experiments, but rather to determine whether this approach might have predictive value for phamacogenomic studies.

Utility of the 3D models

In human INMT two SNPs that change the encoded amino acids were observed. These amino acids were Met/Val205 and Gly/Glu219. [12] Since rabbit INMT shared 88% amino acid identity with the human enzyme, it was plausible to use the rabbit INMT 3D model to predict the effects of residues 205 and 219 on apparent K_M values of the human enzyme for AdoMet and tryptamine. Visual inspection of the 3D rabbit INMT model revealed that residues 205 and 219 were located away from the AdoMet and tryptamine binding sites. Therefore, it was predicted that the two identified human INMT SNPs would not significantly alter the apparent K_M values for tryptamine and AdoMet. This retrospective prediction agreed with the study of human INMT alleles expressed in COS-1 cells, which showed that both alleles had similar apparent K_M values for tryptamine and AdoMet. [12]

Among the 31 or 33 amino acid differences (depending on the human INMT allele compared) between rabbit and human INMT, the rabbit INMT 3D model suggested that amino acid 202 (Ser in rabbit INMT and Pro in human INMT) was the most likely residue, of those that differed between species, to result in a change in K_M value. The prediction was based on the fact that residue 202 was located in the active site of the 3D models of both human and rabbit enzymes. Therefore, we used the SDM construct rabbit INMT Ser202Pro to test the hypothesis that the species-specific difference in substrate affinity might be due to the species difference in residue 202. We found that rabbit INMT Ser202Pro had a higher apparent K_M value (1.27 mM) and a lower V_{max} value than the wild-type rabbit protein (Table 2), observations consistent with the "humanization" of rabbit INMT by this alteration in amino acid. The change in K_M value from 0.43 mM to 1.27 mM explained approximately 34% of the difference in apparent K_M values of INMT for tryptamine between the two species (tryptamine K_M value of 2.9 mM in human [11]).

Discussion

In this study, we used rigorous Particle Mesh Ewald Method-based nanosecond-length MD simulations in water. The purpose of these simulations was twofold. First, such nanosecond-length MD simulations could remedy some "stress" of backbone conformations inherited from "alien" proteins that were used as 3D templates in generating 3D models of the proteins in question. As is evident from Fig. 3, the loop marked in yellow preferred a twisted conformation in rat COMT, but an untwisted conformation in rabbit INMT. Inherited from the rat COMT structure, the loop conformation in the energy-minimized 3D rabbit INMT structure was twisted. Only after a 1.0 ns MD simulation was the twisted, energetically unstable loop conformation refined to an untwisted, energetically stable loop conformation. Second, such simulations were also used to test self-consistency,

since if the model were correct, it would not have unfolded during the MD simulation. Although a 1.0 ns MD simulation was not long enough to observe folding or unfolding of a protein, it was, however, long enough to observe regional backbone conformation changes involving a partial unfolding of a 3D model that was incorrectly folded by homology modeling. A model that could sustain the 1.0 ns MD simulation was not necessarily correct, but a model that failed in the 1.0 ns MD simulation would be sufficient to call for a new iteration of the four-step procedure described above for generating a new model.

In conventional homology modeling, excessive energy minimizations and nanosecond-length MD simulations are not recommended, since such calculations made the 3D homology structure diverge from the template of the known 3D structure. [45] According to our studies, such divergences were often the results of performing the energy calculations without consideration of long-range electrostatic interactions and without placing solvent molecules in the cavity of the active site. As long as long-range electrostatic interactions were calculated and explicit solvent molecules were included, according to our experience, the time-averaged structures of nanosecond-length MD simulations employing the AMBER 95 force field were identical, within experimental error, to the protein structures determined by X-ray crystallographic analysis. [46, 47] Accordingly, in the present work, we emphasized the use of the Particle Mesh Ewald Method-based nanosecond-length MD simulations in water.

In this report, we present what is probably a "best-case" scenario for the prediction of a 3D model from its amino acid sequence employing the semi-computational approach since no iteration of the four-step procedure was needed. However, it should be noted that iteration is likely to be required for predictions of other low-homology proteins and that process can be time-consuming. Nevertheless, even in a "worst-case" scenario, the semi-computational approach proposed here can still be useful because experimental elucidation of 3D structures of proteins by X-ray crystallography or NMR experiment can be hampered by difficulties in obtaining sufficient quantities of purified proteins and/or well diffracting crystals as well as the limitation that NMR techniques are currently applicable only to proteins with molecular weights less than 20 kDa.

Current estimates suggest that there are, on average, about eight SNPs in a protein of about 500 amino acid residues, approximately half of which would be nonsynonymous. [48, 49] Evolution leads to the creation of protein clusters in terms of amino acid sequence homology and folding pattern. SNP effects within a highly homologous protein family can be computationally predicted and confirmed with a small number of experimental studies if the 3D structure of one member is determined by an experimental or theoretical approach. In the present case of INMT, there are 31 or 33 amino acid differences (depending on the human INMT allele compared) between rabbit and human enzymes, a situation that would require at least 31 SDM experiments to single out the residues that contribute to the species-specific difference in tryptamine binding. However, using the semi-computational approach, a single SDM experiment has identified Ser202 of rabbit INMT as being responsible for approximately 50% of the difference between human and rabbit enzyme in apparent K_M values for tryptamine. These results demonstrate the utility of the predicted 3D INMT structures for application in pharmacogenomic studies and encourage the use of computer modeling to predict 3D protein structures for the purpose of pharmacogenomic studies at a time when X-ray crystallography and/or solution NMR spectroscopy can determine 3D structures of only a small number of known amino acid sequences.

Conclusion

We have predicted 3D models of human and rabbit INMTs using a semi-computational approach on the basis of the X-ray structure of rat COMT and confirmed the two models both computationally and experimentally. Retrospective prediction of the effects of genetic polymorphisms using the 3D models of INMTs was in agreement with experimental study. Most importantly, such 3D models offered spatial information with regard to species-specific residues of INMTs and facilitated our experimental identification of active-site residues that contribute to observed species-specific differences in substrate affinity. The results thus encourage the use of mutagenesis-coupled computer modeling to predict 3D protein structures for use in pharmacogenomic studies when de novo prediction of protein 3D structures from their amino acid sequences is still not feasible and X-ray crystallography and/or solution NMR spectroscopy can only determine 3D structures for a small number of known amino acid sequences.

Supplementary material. The AMBER atom types and RESP charges of AdoMet, tryptamine, 5-hydroxytryptamine, α-methyltryptamine, and *N*-methyltryptamine.

Acknowledgements We thank Thomas M. Kollmeyer of the Mayo Clinic for discussion and Luanne Wussow for assistance in preparing this manuscript. This work was supported by the Mayo Foundation for Medical Education and Research and in part by the NIH R01 GM28157 (RMW), R01 GM35720 (RMW) and U01 GM61388-01B (RMW and YPP).

References

1. Burley SK, Almo SC, Bonanno JB, Capel M, Chance MR, Gaasterland T, Lin DW, Sali A, Studier FW, Swaminathan S (1999) Nat Genet 23:151–157
2. Service RF (2000) Science 287:1954–1956
3. Orengo CA, Jones DT, Thornton JM (1996) Protein folds and their recognition from sequence. In: Sternberg MJE (ed) Protein structure prediction. Oirl Press, Oxford, pp 173–206
4. Blundell TL, Sibanda BL, Sternberg MJ, Thornton JM (1987) Nature 326:347–352

5. Strader CD, Fong TM, Tota MR, Underwood D, Dixon RA (1994) Annu Rev Biochem 63:101–132
6. Herzyk P, Hubbard RE (1998) J Mol Biol 281:741–754
7. Szklarz GD, Halpert JR (1997) Life Sci 61:2507–2520
8. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325
9. Lazaridis T, Karplus M (2000) Curr Opin Struct Biol 10:139–145
10. Strahs D, Weinstein H (1997) Protein Eng 10:1019–1038
11. Thompson MA, Weinshilboum RM (1998) J Biol Chem 273: 34502–34510
12. Thompson MA, Moon E, Kim U-J, Xu J, Siciliano MJ, Weinshilboum RM (1999) Genomics 61:285–297
13. Schluckebier G, O'Gara M, Saenger W, Cheng X (1995) J Mol Biol 247:16–20
14. Djordjevic S, Stock AM (1997) Structure 5:545–558
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
16. Vidgren J, Svensson LA, Liljas A (1994) Nature 368:354–358
17. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Gene 77:51–59
18. McCutchan JH, Pagano JS (1968) J Natl Cancer Inst 41: 351–357
19. Aksoy S, Szumlanski CL, Weinshilboum RM (1994) J Biol Chem 269:14835–14840
20. Bradford MM (1976) Anal Biochem 72:248–254
21. Wilkinson GN (1961) Biochem J 80:324–332
22. Cleland WW (1963) Nature 198:463–465
23. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, Debolt S, Ferguson D, Seibel G, Kollman PA (1995) Comput Phys Commun 91:1–41
24. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz Jr KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
25. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Hohnson BG, Robb MA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzales C, Pople JA (1999) Gaussian 98. Gaussian, Pittsburgh, Pa.
26. Pang Y-P, Perola E, Xu K, Prendergast FG (2001) J Comput Chem, in press
27. Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ (1993) Cell 74:299–307
28. Labahn J, Granzin J, Schluckebier G, Robinson DP, Jack WE, Schildkraut I, Saenger W (1994) Proc Natl Acad Sci USA 91: 10957–10961
29. Reinisch KM, Chen L, Verdine GL, Lipscomb WN (1995) Cell 82:143–153
30. Fu Z, Hu Y, Konishi K, Takata Y, Ogawa H, Gomi T, Fujioka M, Takusagawa F (1996) Biochemistry 35:11985–11993
31. Hodel AE, Gershon PD, Shi X, Quiocho FA (1996) Cell 85: 247–256
32. Yu L, Petros AM, Schnuchel A, Zhong P, Severin JM, Walter K, Holzman TF, Fesik SW (1997) Nat Struct Biol 4:483–489
33. Bussiere DE, Muchmore SW, Dealwis CG, Schluckebier G, Nienaber VL, Edalji RP, Walter KA, Ladror US, Holzman TF, Abad-Zapatero C (1998) Biochemistry 37:7103–7112
34. Lysaa RA, Sylte I, Aarbakke J (1998) J Mol Model 4:211–220
35. Kelley LA, MacCallum RM, Sternberg MJE (2000) J Mol Biol 299:499–520
36. Ingrosso D, Fowler AV, Bleibaum J, Clarke S (1989) J Biol Chem 264:20131–20139
37. Kagan RM, Clarke S (1994) Arch Biochem Biophys 310: 417–427
38. Fujioka M (1992) Int J Biochem 24:1917–1924
39. Gomi T, Tanihara K, Date T, Fujioka M (1992) Int J Biochem 24:1639–1649
40. Wu G, Williams HD, Zamanian M, Gibson F, Poole RK (1992) J Gen Microbiol 138:2101–2112
41. Niewmierzycka A, Clarke S (1999) J Biol Chem 274:814–824
42. Rost B, Sander C (1994) Proteins 19:55–72
43. Accelrys Inc (1997) QUANTA/CHARMm San Diego, CA
44. Hooft RW, Vriend G, Sander C, Abola EE (1996) Nature 381: 272
45. Peitsch MC, Guex N (2000) http://www.expasy.ch/swissmod/course/text/chapter6.htm
46. Pang YP (1999) J Mol Model 5:196–202
47. Pang YP, Xu K, El Yazal J, Prendergast FG (2000) Protein Sci 9:1857–1865
48. Halushka MK, Fan JB, Bentley K, Hsie L, Shen NP, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Nat Genet 22: 239–247
49. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Nat Genet 22:231–238