J Mol Model (2008) 14:1101–1109
DOI 10.1007/s00894-008-0354-4

ORIGINAL PAPER

# Exploring CYP1A1 as anticancer target: homology modeling and in silico inhibitor design

**Abhay T. Sangamwar · Leena B. Labhsetwar · Sharad v. Kuberkar**

Received: 23 March 2008 / Accepted: 10 July 2008 / Published online: 30 July 2008


**Abstract** Microsomal cytochrome P450 family 1 enzymes has great importance in the bioactivation of mutagens. P450 catalyzed reactions involve a wide range of substrates, and this versatality is reflected in a structural diversity, evident in the active sites of available P450 structures. This structure offers a template to study further structure-function relationships of alternative substrates and other cytochrome P450 family 1 members. In this paper, we document a homology model of CYP P450 1A1 from Homo sapiens, developed on the basis of template crystal structure of human microsomal P450 1a2 in complex with inhibitor (PDB Id: 2HI4). Homology modeling is performed at the programs, both in the commercial and public realms. We tried to explore CYP1A1 as a potential target for anticancer chemotherapy. To gain an insight into the binding of ligands with enzyme, protein-ligand complex was developed by including information about the known ligand as spatial restraints and optimizing the mutual interactions between the ligand and the binding site. Active site characterization and the study for involvement of specific aminoacids in binding with ligand, facilitates structure based inhibitor design. This study should prove useful in the design and development of potential novel anticancer agents.

**Keywords** Active site · CYP1A1 · Drug design · Homology modeling · Protein-ligand complex

L. B. Labhsetwar · S. v. Kuberkar
Post Graduate Research Centre, Department of Chemistry,
Yeshwant Mahavidyalaya,
Nanded, 431605 MS, India

A. T. Sangamwar (✉)
Department of Pharmaceutics, Nanded Pharmacy College,
Shyamnagar, Nanded, 431605 MS, India
e-mail: abhaysangamwar@yahoo.co.in

## Introduction

Cancer is a group of diseases in which cells are aggressive, invasive and/or metastatic. These three malignant properties of cancer differentiate them from benign tumors, which are self-limited in their growth and do not invade or metastasize. Cancer may affect people at all ages, even fetuses, but risk for the more common varieties tends to increase with age [1]. Cancer causes about 13% of all deaths [2]. It is a leading cause of death globally.

Cancer can be treated by surgery, chemotherapy, radiation therapy, immunotherapy, monoclonal antibody therapy or other methods. The choice of therapy depends upon the location and grade of the tumor and the stage of the disease, as well as the general state of the patient. Complete removal of the cancer without damage to the rest of the body is the goal of treatment. Sometimes this can be accomplished by surgery, but the propensity of cancers to invade adjacent tissue or to spread to distant sites by microscopic metastasis often limits its effectiveness. Radiation also cause damage to normal tissues. Uptill now the effectiveness of chemotherapy is limited by toxicity to other tissues in the body. In current usage the term chemotherapy, usually refers to cytotoxic drugs which affect rapidly dividing cells in general, in contrast with targeted therapy. Chemotherapy drugs interfare with cell division in various possible ways like the duplication of DNA or the separation of newly formed chromosomes. It has the potential to harm healthy tissues, especially those that have a high replacement rate but these cells repair themselves after chemotherapy. This advantage of chemotherapy has created a scientific interest in developing novel therapeutic approaches to combat over cancer. It is evidenced that CYP1A1 is the covalently bound $m_r$ 50000 protein in sensitive cell lines [3]. It is the main oxidizing enzyme in phase 1 metabolism that catalyzes the

initial step in either detoxification or bioactivation of enviornmental toxins and xenobiotics [4–6]. CYP1 isoforms are capable of activating aromatic amines, therefore extrahepatic and cell line-specific regulation of CYP1A1 and CYP1B1 [7] are major determinants of chemosensitivity or resistance. The cytochrome P450 enzymes CYP1A1 and CYP1B1 have been reported to be the isoforms responsible for 6-hydroxylation of antitumor compounds. P450 catalyzed reactions involve a wide range of substrates, and this versatality is reflected in a structural diversity evident in the active sites of available P450 structure [8]. The novel antitumor agents modulates the expression and activity of CYP1A1 and CYP1B1 in sensitive cell lines but the critical involvement of CYP1A1 is well documented in sensitive ovarian cancer cells [9]. Treatment of sensitive cell lines with 10 μM resveratrol, an inhibitor of CYP1A1 induction, in combination with either 1 or 10 μM 5F-203 showed an ablation of the observed CYP1A1, but not CYP1B1 mRNA induction in parallel with a decreased sensitivity to 5F-203 [10]. A prior study demonstrated that coincubation of DF 203 with α-naphthoflavone, a CYP1A1 inhibitor, in sensitive cells can inhibit anticancer activity, which reveals the critical involvement of CYP1A1 in the anticancer activity of the benzothiazoles [11]. It is due to the fact that residues involved in inhibitor binding are different at CYP1A1 and CYP1B1. Therefore the generated model of CYP1A1 would furnish novel information on inhibitor binding for design of new scaffolds with improved therapeutic value.

Rational inhibitor design relies both on mechanistic and structural information about the target enzyme. The three dimensional (3D) structure of a drug target, experimentally determined or theoretically modeled, particularly if it is complexed with a putative ligand, is vital to obtain a comprehensive understanding of the mode of action of a drug at the molecular level, to appreciate the strutural consquences of genetic variations and to construct accurate pharmacophores for drug design [12, 13]. In the absence of crystallographic structure, a variety of advanced homology modeling methods have been developed, which can provide reliable models of proteins that share 30% or more sequence identity with a known structure [14]. Generally, it is assumed that the tertiary structures of the two proteins will be similar if their sequences are related [15]. Protein models obtained with comparative modeling methods can be classified into three categories 1) models based on incorrect alignments between target and template; 2) models based on correct alignments are of course much better, but their accuracy can still be medium to low as the templates used during the modeling process have a medium to low sequence similarity with the target sequence. Such models are useful tools for the rational mutagenesis experiment design. They are however of very limited assistance in ligand binding studies; 3) model based on templates which share a high degree of sequence identity (>75%) with the target. Such models have proven useful during drug design projects and allowed the taking of key decisions in compound optimization and chemical synthesis [16].

We critically assessed the CYP1A1 model (A0N0X8) at MODBASE (http://www.salilab.org). The template used is 1z10A having sequence identity 33% (E value 0). It is evidenced that model reliability decreases as the sequence identity decreases. The target-template pairs sharing less than 50% sequence identity may often require manual adjustment of the alignment. High accuracy comparative models are based on >50% sequence identity to their templates [17]. We used open source software SWISS-MODEL to develop the initial homology model, due to certain advantages than other automated homology modeling servers; and BioMed CAChe (v 6.1.1), a commercial software program to generate protein-ligand complexes. SWISS-MODEL confirms that the common core of proteins sharing 50% sequence identity deviate by approximately 1 Å relative mean square deviation [18]. The protein-ligand complexes provides an insight into designing of new, potent inhibitors. The present work details the construction of a 3D model for CYP1A1 of Homo sapiens using the X-ray crystal structure of chain A of human microsomal P450 1a2, with inhibitor (PDB Id: 2HI4) as a template having 71% sequence identity and E value 1.92 $e^{-105}$. It is anticipated that the refined homology model along with protein-ligand complex provide a starting point for the structure based drug design.

## Results and discussion

To develop a 3-D structure of an enzyme, entire protein data bank at NCBI is searched and the sequence of cytochrome P450 1A1 from Homo sapiens is downloaded in FASTA format [19]. It is then submitted to BLAST (Basic Local Alignment Search Tool) [20–22] in search of the homologous protein.

To date the crystal structure of human microsomal P450 1A1 in complex with alpha-Naphthoflavone (PDB Id: 2HI41A) have reported 71% sequence identity with the raw sequence of cytochrome P450 1A1 from Homo sapiens. It has the score of 378 bits (97%) and positives 212/249 (85%) and gaps 3/249 (1%) with the raw sequence. Due to 71% sequence similarity automated mode at SWISS MODEL workspace server is selected to build a model of CYP1A1.

The structure of human P450 1a2 in complex with inhibitor alpha- Naphthoflavone is determined to a resolution of 1.95 Å. It is bound in the active site and about the distal surface of the heme prosthetic group. The structure

**Table 1** Multiple sequence alignment using SWISS-MODEL (The target sequence comprises 218 aminoacid residues while the template comprises 480 aminoacid residues, alignment starts from 264 numbered residue in the template with 1 number residue of the target

| | |
|---|---|
| TARGET 1 | MQKMVKEH YKTFEKGHIR |
| 2hi4A 234 | NPLDFFPILR YLPNPALQRF KAFNQRFLWF -LQKTVQEH YQDFDKNSVR |
| TARGET 19 | DITDSLIEHC QEKQLDENAN VQLSDEKIIN IVLDLFGAGF DTVTTAISWS |
| 2hi4A 282 | DITGALFKHS K KGPRASGN L-IPQEKIVN LVNDIFGAGF DTVTTAISWS |
| TARGET 69 | LMYLVMNPRV QRKIQEELDT VIGRSRRPRL SDRSHLPYME AFILETFRHS |
| 2hi4A 330 | LMYLVTKPEI QRKIQKELDT VIGRERRPRL SDRPQLPYLE AFILETFRHS |
| TARGET 119 | SFVPFTIPHS TTRDTSLKGF YIPKGRCVFV NQWQINHDQK LWVNPSEFLP |
| 2hi4A 380 | SFLPFTIPHS TTRDTTLNGF YIPKKCCVFV NQWQVNHDPE LWEDPSEFRP |
| TARGET 169 | ERFLTPD-GA IDKVLSEKVI IFGMGKRKCI GETIARWEVF LFLAILLQRV |
| 2hi4A 430 | ERFLTADGTS INKPLSEKMM LFGMGKRRCI GEVLAKWEIF LDLAILLQQL |
| TARGET 218 | EFSVPLGVKV DMTPIYGLTM KHACCEHFQM Q - |
| 2hi4A 480 | EFSVPPGVKV DLTPIYGLTM KHARCEHVQA RRFS |

reveals a compact, closed active site cavity that is highly adapted for the positioninig and oxidation of relatively large, planar substrates. This unique topology is clearly distinct from known active site architectures of P450 family 2 and 3 enzymes and demonstrates how P450 family 1 enzymes have evolved to catalyze efficiently polycyclic aromatic hydrocarbon oxidation. This has provided the first structure of a microsomal P450 from family 1 and offers a template to study further structure - function relationship of alternative substrates and other cytochrome P450 family 1 members.

## Template selection and sequence alignment

X-ray crystal structure of chain A of human microsomal P450 1a2, with inhibitor (PDB Id: 2HI4) having 71% sequence identity with CYP1A1 is selected as a template. To establish a one-to-one connection between the amino-acids of the template and target protein sequence, alignment is the most prefered step in homology modeling [23]. The accuracy of the alignment is the most important factor that determines the quality of the model [24]. The alignment of the aminoacid sequence of CYP1A1 (Gene Accession No. CAJ84704.1) with the sequence from human microsomal P450 (PDB Id: 2HI4) is shown in Table 1. This alignment shows the model building by ab initio methds. ProModII was used for simple assignment of backbone. The loops of the model were built by ligating with anchor residues Pro174 and Ala177; Thr173 and Ala177. The number of ligations found was 82 and there were no clashes. In building CSP loop with anchor residues Gln29 and Gln32, the number of ligations found was 3 and there were no accepting loops and clashes. In building CSP loop with anchor residues Val39 and Ser42, the number of ligations found was 1 with no accepting loops and clashes. Side chains have been optimized by addition of hydrogen. The homology model of CYP1A1 was generated using SWISS-MODEL program [25–27]. The lowest energy rotamer is obtained by energy minimization.

## Homology model building and refinement

The homology model of CYP1A1 is build at automated mode of SWISS-MODEL. The detailed procedure has been discussed in the materials and methods section. Energy minimization of CYP1A1 and template, helped relieve any steric clashes or improper geometries in the protein structure to produce a model with correct bond lengths and bond angles and where individual atoms are not too close together. The crystal structure of human microsomal P450 1a2 in complex with inhibitor alpha-Naphthoflavone is shown in Fig. 1. The refined structure of homology model of CYP1A1 is shown in Fig. 2. The molecular mechanics calculation for energy minimization of the model is given in Table 2.

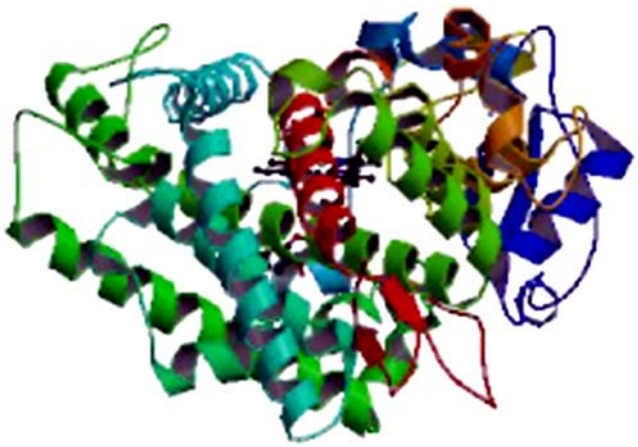

**Fig. 1** The crystal structure of human microsomal P450 1a2 in complex with inhibitor alpha-Naphthoflavone

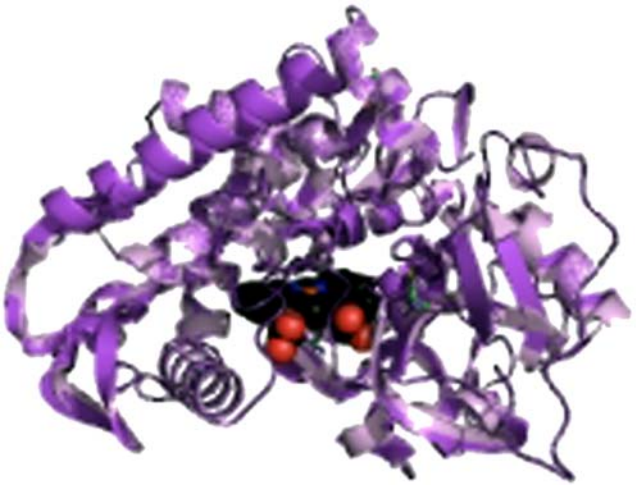

**Fig. 2** The refined structure of homology model of CYP1A1

## Validation of the model

The possible applications of protein models depend largely on the quality of the models [28–30]. The accuracy of a model can vary significantly, even within different regions of the same protein, usually highly-conserved core regions can be modeled much more reliably than varibale loop regions or surface residues. The models are built using the SWISS-MODEL server pipeline. The modeling log shows the individual steps during model building, especially which parts of the model have been built ab initio (i.e., insertions/deletions). Model is evaluated by the tools available at SWISS-MODEL. The ANOLEA results represent the y-axis of the plot, the energy for each amino acid of the protein chain. Negative energy values (in green) represent favorable energy environment whereas positive values (in red) unfavorable energy environment for a given amino acid. There are few aminoacids lying in an unfavorable energy enviornment, which can be ignored in overall consideration of the model structure. In Verify3D the vertical axis in the plot represents the average 3D-1D profile score for each residues in a 21-residue sliding window. The scores range from -1 (bad score) to +1 (good score). All aminoacds has +1 score, further validating the model. GROMOS has represented the y-axis of the plot, the energy for each amino acid of the protein chain. Negative energy values (in green) represent favorable energy environment whereas positive values (in red) unfavorable energy environment for a given amino acid. Very few aminoacids show red colour, confirming the homology model experimented. The force field energy [31] for the overall structure is −11103.370 KJ/mol. PROCHECK was used to assess the stereochemical quality of the model. This programme verifies the accuracy of parameters such as bond lengths, bond angles, torsion angles, and correctness of aminoacid chirality. The cut-off used was 10 standard deviations from the reference value. The stereochemical quality of the model is also judged by Ramchandran plot [32]. There were no spurious angle or bond length in our model. The reults are represented in Table 3 and shown in Fig. 3a and b. The distribution of main chain torsion angles phi and psi are examined in Ramchandran plot. The results clearly shows the vast majority of the aminoacids are in a phi-psi distribution consistent with right handed $\alpha$-helices. The remaining residues that fall into the random or beta configuration geometries are very short segments and are primarily in the loop regions of the protein. The 89.00% residues of the model CYP1A1 lie within the most favored region indicating good quality model. The backbone root mean square deviation (RMSD) between the final model and the template crystal structure is 1.56 Å, while that between the starting model and the template is only 1.72 Å; therefore, the adopted minimization and refinement procedures did not appear to cause a substantial distortion in the structure and accomplished the goal of relieving steric clashes and close contacts. Superposition of the modeled structure CYP1A1 with X-ray crystal structure of 2HI4 is shown in Fig. 4. The RMSD value of 1.56 Å strongly indicates that the two structures share common structural homology.

The secondary structure of the template and model is defined at BioMed Cache [33]. The residues in 4-helix, isolated beta bridge, extended strand, 3-helix and 5-helix are analysed. By this HELIX and SHEET record is verified and validated. By analyzing the validation results we come to the conclusion that the generated homology model is reasonable for further structure based drug design study.

**Table 2** Molecular mechanics calculation for energy minimization of the model at BioMed CAChe

| Number of iteration | 231 |
|---|---|
| Energy -total | 545.7444 |
| Delta E | 0.0004 |
| Average gradient | 0.49990 |
| Stretch | 302.612 |
| Angle | 787.486 |
| Stretch bend | 8.568 |
| Dihedral | 945.763 |
| Improper torsion | 60.509 |
| Torsion stretch | −37.473 |
| Bend bend | −1.189 |
| Van der walls | 975.166 |
| Electrostatistics | −1267.914 |
| Hydrogen bond | −1227.847 |

**Table 3** Results of protein structure check by PROCHECK

| No. | PROCHECK | 2HI4 | CYP1A1 |
|---|---|---|---|
| 1 | Most favored region | 377 (90.04%) | 169(89.00%) |
| 2 | Additional allowed region | 38 (9.1%) | 19(10.00%) |
| 3 | Generously allowed region | 0 (0.0%) | 0(0.00%) |
| 4 | Disallowed regions | 2 (0.5%) | 1(0.50%) |
| 5 | Non glycine and non proline residues | 417 | 189(100%) |
| 6 | End residues (excluding Gly and Pro) | 2 | 2 |
| 7 | Glycine residues (shown in triangle) | 28 | 12 |
| 8 | Protein residues | 33 | 15 |
| 9 | Total number of residues | 480 | 218 |
| 10 | Overall PROCHECK score[a] | 0.17 | 0.14 |

[a] Recommended value >0.50 and investigation is needed for <1.0

### Enzyme inhibitor complex

The binding of a ligand to the active site of a protein is typically associated with local and perhaps, also global structural rearrangement of the receptor. The protein-ligand complexes give greater insight in structure based drug design, so we develop a protein ligand complex. It gives a more detailed and accurate picture of the interactions and structural complementarities between the ligand and the active site. The representation of the binding pocket of ligand at 2HI4 is shown in Fig. 5.

The active site pocket of template reveals that the ligand is highly embeded in hydrogen bond donor region of the protein. The substitutions of highly electronegative groups at parent flavone skeleton may further stabilize the ligand at protein by formation of hydrogen bonds. The aromatic ring which protrudes from the parent skeleton is found to be within the hydrophobic cleft of protein exhibiting van der waals interactions.

Representation of the active site residues in 2HI4 within 5 Å, reveals that highly conserved residues Ala317, Val227, Leu382, Ile117, Leu497 shield the active site from the solvent, creating a hydrophobic enviornment. The residues Thr118, Ser122, Thr124, Thr223, Asn312, Thr321 and Thr498 remain hydrogen bond donors in the active site of the template.

The protein ligand complex in generated homology model is developed by sequence alignment and match selection procedures. Assuming that the ligand binding modes are similar in the target and the template protein structure, the coordinates of ligand was transferred from 2HI4 crystal structure. The active site contains the highly conserved residues Ile117(Ile115 in model), Thr118(Ser116), Ser122(Ser120),Thr124(Ser122),Phe125(Phe123),Thr223 (Asn221),Phe226(Phe224),Val227(Gly225),Phe256 (Leu254),Phe260(Phe258),Asn312(Ile310),Asp313(Val311), Gly316(Leu314), Ala317(Phe315), Asp320(Gly318), Thr321 (Phe319), Leu382(Ser380), Ile386(Phe384), Leu497(Tyr494) and Thr498(Gly495). The active site residues in CYP1A1, Ile310, Val311, Gly318, Phe319, Ser380, and Gly495 are different from those residues involved in the active site of CYP1A2.

Representation of the binding pocket of DF 203, anticancer agent, at the homology model is shown in Fig. 6. There are hydrophobic interactions of aromatic ring of benzthiazole with Ile310 and Val311 as they reside at a distance of 3.78 Å and 3.35 Å, respectively. Gly318, Gly495 and Phe318 has van der waals attractions with 2-substituted phenyl ring. Ser380 is involved in weak hydrogen bond with amino substitution at phenyl ring. It is found that these residues have a promising role in suggesting therapeutic activity of DF 203. The prosthetic group, heme is at a distance of 4.05 Å from the protruding aromatic ring of ligand. This binding has a vital role in potency of a lead compound and may prove effective in structure based drug design of new scaffolds.

## Materials and methods

### Software

Molecular modeling was carried out on BioMed CACHe (v6.1.1) [34] workstation running on Pentium IV. For sequence alignment and homology modeling studies, the benchmark used was SWISS-MODEL [35–47]. Energy minimization and molecular dynamic simulations were carried out at BioMed CAChe. Validation of the model is done by PROCHEK (v.3.0) [48], Verify 3D and Ramchandran plot.

### Template selection

The amino acid sequence of the target proetein CYP1A1 (Gene bank Accession No.CAJ84704.1) is extracted from the NCBI protein sequence database (http://www.ncbi.nlm.

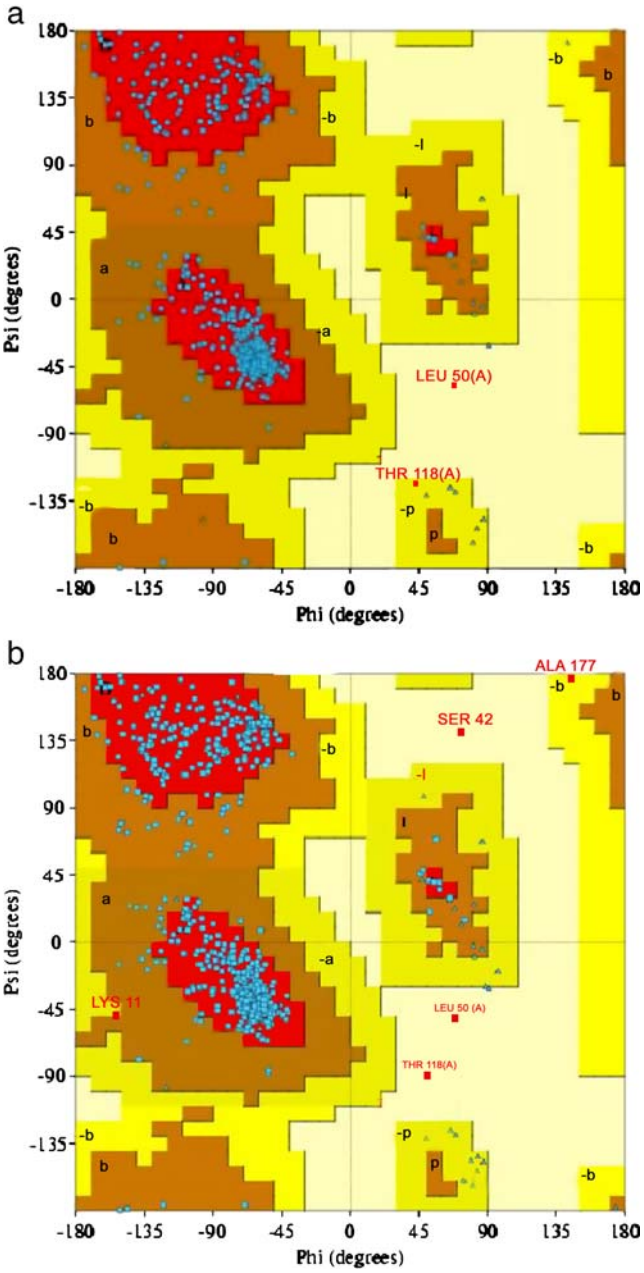


**Fig. 3** **a**. Ramchandran plot of 2HI4 (refer Table 3). **b**. Ramchandran plot of homology modeled structure of CYP1A1 (refer Table 3)

nih.gov). The sequence is composed of 251 residues. To select templates for a given protein, sequences of the template structure library is searched. The pipeline has automatically selected a template (PDB Id:2HI4) based on Blast having E-value $1.92e^{-105}$.

## Alignment

To align a target sequence CYP1A1 with template 2HI4 at SWISS-MODEL the five template structures are superposed using an iterative least squares algorithm. A structural alignment is generated after removing incompatible templates. A local pair-wise alignment of the target sequence to the main template structures is calculated [49], followed by a heuristic step to improve the alignment for modeling purposes. The placement of insertions and deletions is optimized. The isolated residues in the alignment are moved to the flanks to facilitate the loop building process.

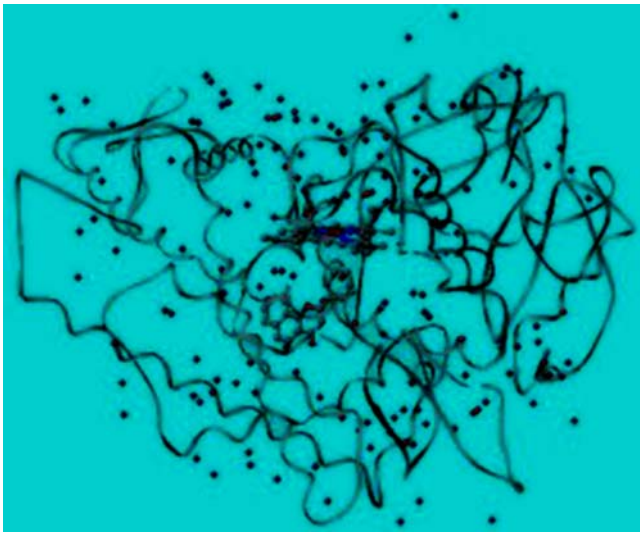

**Fig. 4** Superposition of the modeled structure CYP1A1 with X-ray crystal structure of 2HI4

## Building homology model

Modeling requests are computed by the SWISS-MODEL server homology modeling pipeline. It is a fully automated

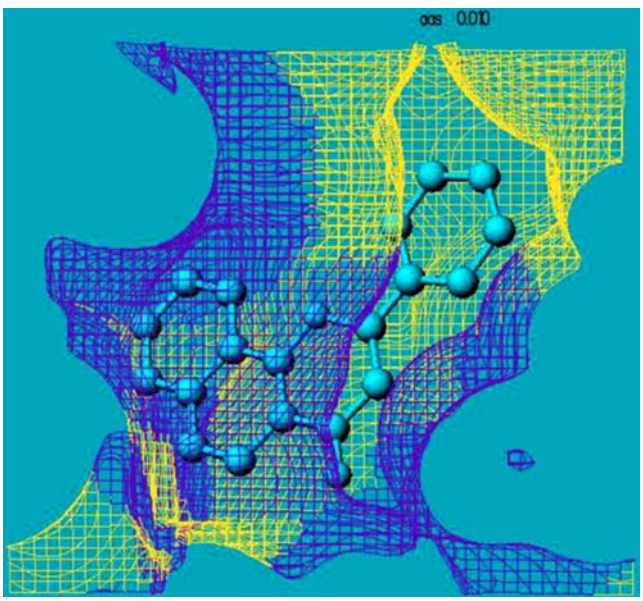

**Fig. 5** A representation of the binding pocket (active site) in the template 2HI4 at BioMed CAChe

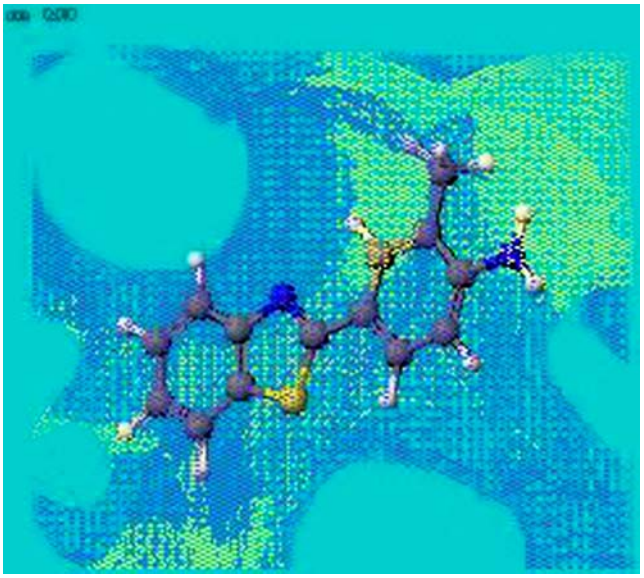

**Fig. 6** A representaion of the binding pocket (active site) in the template CYP1A1 at BioMed CAChe

protein structure homology modeling server accessible via the ExPASy Web server (http://swissmodel/expasy.org). As the target and template share 71% sequence identity automated sequence alignment mode is selected.It takes as input a sequence alignment and a PDB file for the template. These are submitted over a server, and the knowledge-based homology model is constructed using the ProModII program. To generate the core of the model, the backbone atom portions of the template structure are averaged. To generate insertions or deletions in the target-template alignment, an ensemble of fragments compatible with the neighbouring stems is constructed using constraint space programming. The best loop is selected using a scoring scheme, which accounts for force field energy, steric hindrance and favorable interactions like hydrogen bond formation.

The reconstruction of the model side chains is based on the weighted portion of corresponding residues in the template structure [50]. A scoring function assessing favorable interactions and unfavorably close contacts is applied to select the most likely confirmation. The steepest descent energy minimization using the GROMOS 96 force field is done to regularize the protein structure geometry.

## Refining the model

To refine the homology model, charge in the protein is balanced to 1 and hydrogen atoms, electrons and hybridizations are addedd to the protein and water molecules under beautification of valency. The model is then refined by performing an optimize geometry calculation in mechanics using augmented MM3 parameter at BioMed CaChe/ workspace module. Cache molecular mechanics includes energy terms for bond stretch, bond angle, dihedral angle, improper torsion, torsion stretch, bend bend, van der waals electrostatistics, and hydrogen bond interactions. Conjugate gradient is used to locate the energy minimum. All atoms are moved at once during minimization. The normal movements of atoms are scaled by 1.0000. The dielectric is 1.50. Van der waals interactions between atoms separated by greater than 9.00 Å is excluded. The van der waals interactions list is updated every 50 iterations. The optimization continues until the energy change is less than 0.00100000 kcal/mol.

## Evaluation of the model

Possible applications of protein models depend largly on the quality of the models. The correctness of a model is essentially dictated by the quality of the sequence alignment used to guide the modeling process. The quality of the final refined model was assessed by subjecting it to a series of tests for its internal consistancy and reliability. We used the automated SWISS-MODEL [51] pipeline which is continuosly evaluated by the EVA project [52]. To determine the quality of protein model and template structure, we used graphical plots of ANOLEA mean force potential [53], GROMOS empirical force field energy [54] Verify 3 D profile evaluation [55] and PROCHECK. ANOLEA is used to assess packing quality of the models. The program performs energy calculations on a protein chain, evaluating the “non- local environment” (NLE) of each heavy atom in the molecule. The Verify3D method assess protein structures using three-dimensional profiles. This program analyzes the compatibilty of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, apolar etc). Then a database generated from good structures is used to obtain a score for each of the 20 amino acids in this structural class. GROMOS is a general-purpose molecular dynamics computer simulation package for the study of biomolecular systems and can be applied to the analysis of conformations obtained by experiment or by computer simulation. The stereochemical quality of the model is additionally evaluated by Ramchandran plot. While analyzing the results we could confirm the observation that the common core of proteins sharing 50% sequence identity deviate by approximately 1 Å relative mean square deviation. RMSD values were calculated for the minimized homology model and compared to the template structures. RMSDs were also calculated for the structures before and after minimization to show there were no gross movements, indicating the suitability of the template used.

Enzyme-inhibitor complex formation

The crystal structure of human microsomal P450 1a2 in complex with inhibitor (PDB Id: 2HI4) is submitted for energy minimization and refined for hydrogen atoms, atom hybridization, and correct bond types. Active site is located by selecting inhibitor, embeded in a 5 Å shell of residues, water and HETs.

*Displaying the active site pocket*

This is the surface of the protein adjacent to the ligand and the surface that forms a pocket around the ligand. To generate this surface in BioMed CAChe, the ligand is selected first and then the accessible surface of the protein contained in a box enclosing the ligand is drawn. To design a new candidate drug, the interaction of ligand molecules when bound within the protein, with those of known active compounds is analyzed.

*Discovering the active site in the homology model by sequence alignment*

Molecular modeling represents a relatively rapid approach to determine the likelihood that the lead compound will exhibit binding selectivity. We aligned the sequences of the homology model with the sequence of the homologous 2HI4 for which the active site is known. We identified the active site residues in homology model from the alignment. The gaps appear in the homology model and 2HI4 is aligned in the sequences according to the maximum scoring alignment using BLOSUM 50 substitution matrix in the Needlemham-Wunsch alignment algorithm [56–58]. The matching residues in the homology model are discovered by match selection by choosing 2HI4 as the master sequence.

*Docking ligands into homology model*

The inhibitor DF 203 is selected and refined for lowest energy rotamer by molecular mechanics procedure at BioMed CAChe. It is then docked over the homology model. CAChe automates the docking of ligand into the active site by using a genetic algorithm with a fast, simplified potential of mean force (PMF) [59]. PMF has been demonstrated to show a significant correlation between experimental binding affinities and its computed score for diverse protein-ligand complexes [60–64].

## Conclusions

A 3-D structure of CYP1A1 has been developed using homology modeling. The model was constructed using the template structure 2HI4 having resolution 1.95 Å. The generated model was validated by various bioinformatics techniques. After analyzing the stereochemical parameters of the model at PROCHECK and structural similarity by RMSD with the template, we found the model is a reasonable one for further study. The enzyme-inhibitor complexes were developed by newly improved procedures at BioMed CAChe software. It is examined that the active site site pocket is highly conserved within hydrogen bond donors. The interaction of the ligand with prosthetic group heme has provided evidence that it has a vital role in establishing activity of inhibitor at the active site.

Furthermore, the improvization of the interaction with the active site residues by structural rearrangement over the ligand is suggested. Thus the generated model may prove potential in structure based drug design to make a major impact on anticancer chemotherapy.

## References


1. Cancer Research UK (2007) Internet references. Retrieved from http://info.cancerresearch.org/cancerstats/incidence/age/ecitations.html 25/06/2007
2. WHO World Health Organization (2006) Internet references. Retrieved from http://www.who.int/mediacetre/factsheets/fs 297/en/ecitations.html 25/06/2007
3. Hutchinson I, Chua MS, Browne HL, Trapani V, Bradshaw TD, Westwell AD, Stevens MF (2001) J Med Chem 44:1446–1455
4. Gyengerich FP (1992) Drug Metab Dispos 21:1–5
5. Stefaan S, Jason KY, Rosamund LR, Guillaume AS, Keith JG, David CS, Eric FJ (2007) J Biol Chem 19:14348–14355
6. Crofts FG, Sutter TR, Strickland PT (1998) Carcinogenesis 19: 1969–1973
7. Kress S, Greenlee WF (1997) Cancer Res 57:1264–1268
8. Somsen S, Yano JK, Reynold RL, Schoch GA, Griffin KJ, Stout CD, Johnson EF (2007) J Bio Chem 282:14348–14352
9. Eileen B, Valentina T, Michael CA, Curtis DH, Tracey DB, Malcolm PG, Edward ES, Sherman FS (2004) Drug Metab Dispos 32:1392–1401
10. Curtis DH, Melinda H, Edward AS, Anne M (2003) Mol Cancer Ther 2:1265–1272
11. Chua MS, Kashiyama E, Bradshaw TD, Stinson SF, Brantley E, Sausville EA, Stevens MF (2000) Cancer Res 60:5196–5203
12. White PJ, Blundell TL (1994) Annu Rev Biophys Biomol Struct 23:349–352
13. Blundell TL (1996) Nature 23:384–392
14. Burley SK (2000) Nat Struct Biol Supp 17:932–938
15. Krochser RT, Doughty SW, Robinson AJ, Richards WG (1996) Protein Eng 9:493–497
16. Peitsch MC (1996) Biochem Soc Trans 24:274–279
17. Ursula P, Narayanan E, Ashley CT, Valentin AI, Andrej S (2002) Nucleic Acids Res 30:255–259
18. Chotia C, Lesk AM (1986) EMBO J 5:823–827
19. http://WWW.ncbi.nlm.nih.gov
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1996) J Mol Biol 215:403–410
21. Altshul SF, Thomon LM, Alejandro AS, Zhang J, Zang Z, Miller W, David JL (1997) Nucleic Acids Res 25:3389–3402

22. Schaffer AA, Aravind L, Thomon LM, Shavirin S, John LS, Yuri IW, Eugene VK, Stephen IA (2001) Nucleic Acids Res 29:2994–2999
23. Lee KW, Briggs JM (2004) Protein 54:693–698
24. Krieger E, Nabuurs SB, Vriend G (2003) Methods Biochem Anal 44:509–514
25. Schwede T, Kopp J, Guex N, Peitsch MC (2003) Nucleic Acids Res 31:3381–3385
26. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
27. Amold K, Bordoli L, Kopp J Schwede T (2006) Bioinformatics 22:195–202
28. Sali A, Kuriyan J (1999) Trends Cell Biol 9:M20–M24
29. Peitsch MC (2002) Bioinformatics 18:934–938
30. van Gunsteren WF, Billeter SR, Eising A, Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG (1996) Bimolecular simulations: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ, Zurich
31. van Gunsteren WF, Billeter SR, Eising A, Hunenberger PN, Krruger P, Mark AE, Scott WRP, Tironi IG (1996) Vdf Hochschulverlay ETHZ, Zurich (1963) J Mol Biol 7:95–98
32. Ramchandran GN, Ramkrishnan C, Sasisekharan V (1963) J Mol Biol 7:95–99
33. Kabsch W, Sander C (1983) Biopolymers 22:2577–2582
34. BioMed CAche 6.1 Molecular Modelling Programme Package, Fujitsu Limited, Life Science and Material Science Division 9–3 Nakase 1- Chame,Avihama- ku Chiba city, Chiba 261–8588 Japan (2006)
35. Peitsch MC, Jongeneel V (1993) Int Immunol 5:233–238
36. Peitsch MC (1995) PDB Qua Newsletter 72:4–8
37. Peitsch MC (1995) Bio/Technology 13:658–660
38. Peitsch MC (1996) Biochem Soc Trans 24:274–279
39. Peitsch MC, Herzyk P (1996) Molecular modeling of G-protein coupled receptors. In:Mulford N, Savage LM (eds) Molecular modelling of G- Protein coupled rreceptors. IBC Biomedical Library Series pp 6.29–6.37
40. Peitsch MC, Herzyk P, Wells TNC, Hubbard RE (1996) Recept Channels 4:161–164
41. Peitsch MC, Wilkim MR, Tonella L, Sanchez J-C, Appel RD, Hochstrasser DF (1997) Electrophoresis 18:498–501
42. Peitsch MC (1997) Large scale protein modeling and model repository. In: Gaasterland T, Karp P, Karplus K, Ouzounis C, Sander C, Valenica A (eds) Large scaleprotein modelling and model repository. AAAI Press, pp 234–236
43. Peitsch MC, Guex N (1997) Large-scale comparative protein modeling. In: Wilkins MR, Williams KL, Appel RO, Hochstrasser DF (eds) Large scale comparative protein modelling. Springer, Berlin, pp 177–186
44. Guex N, Peitsch MC (1999) Immunol News 6:132–134
45. Guex N, Diemand A, Peitsch MC (1999) TiBS 24:364–367
46. Peitsch MC, Schwede T, Guex N (2000) Pharmacogenomics 1:257–266
47. Schwede T, Diemand A, Guex N, Peitsch MC (2000) Res Microbiol 151:107–112
48. Laskowski RA, Chistyakov VV, Thornton JM (2005) Nucleic Acids Res 33:D266–D268
49. Huang X, Miller W (1999) Adv Appl Math 12:337–357
50. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) Proteins 40:389–408
51. Schwede T, Kopp J et al (2003) Nucleic Acids Res 31(13):3381–3385
52. Koh IY, Eyrich VA et al (2003) Nucleic Acids Res 31(13):3311–3315
53. Melo F, Feytmans E (1998) J Mol Biol 277(5):1141–1152
54. van Gunsternen WF, Billeter SR et al (1996) The GROMOS 96 manual and user guide, Zyrich, vdf Hochschulverlag ETHZ
55. Eisenberg D, Luthy R et al (1997) Methods Enzymol 277:396–404
56. Needleman SB, Wunsch CD (1970) J Mol Biol 48:443–453
57. Gotoh O (1996) J Mol Biol 264:823–828
58. Durbin R, Eddy SR, Krough A, Mitchinson G (1998) Cambridge University Press, p 19
59. Muegge I, Mrtin Y (1999) J Med Chem 42:791–804
60. Mugge I (1999) Med Chem Res 9:490–500
61. Muegge I, Mrtin Y, Hajduk PJ, Fesik SW (1999) J Med Chem 42:2498–2503
62. Ha S, Andreani R, Robbins A, Muegge I (2000) J Comp Aided Mol Design 14:435–448
63. Muegge I, Rarey M (2001) Rev in Comp Chem 17:1–60
64. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz Jr KM, Forguson DM, Spellmeyer DC, Fox T, Caldwell JM, Kollman PA (1995) J Am Chem Soc 117:5179–5197