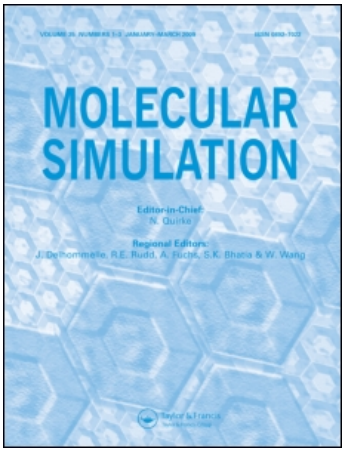


Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Structure prediction and R115866 binding study of human CYP26A1: homology modelling, fold recognition, molecular docking and MD simulations

J. H. Ren[a]; X. Q. Xiong[a]; Y. Sha[a]; M. C. Yan[a]; B. Lin[b]; J. Wang[a]; Y. K. Jing[c]; D. M. Zhao[a]; M. S. Cheng[a]
[a] School of Pharmaceutical Engineering, Shenyang Pharmaceutical University, Shenyang, P.R. China [b] Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX, USA [c] Department of Medicine, Mount Sinai School of Medicine, New York, USA

To cite this Article Ren, J. H. , Xiong, X. Q. , Sha, Y. , Yan, M. C. , Lin, B. , Wang, J. , Jing, Y. K. , Zhao, D. M. and Cheng, M. S.(2008) 'Structure prediction and R115866 binding study of human CYP26A1: homology modelling, fold recognition, molecular docking and MD simulations', Molecular Simulation, 34: 3, 337 — 346
To link to this Article: DOI: 10.1080/08927020801930562
URL: http://dx.doi.org/10.1080/08927020801930562

Molecular Simulation
Vol. 34, No. 3, March 2008, 337–346

Taylor & Francis
Taylor & Francis Group

Structure prediction and R115866 binding study of human CYP26A1: homology modelling, fold recognition, molecular docking and MD simulations

J.H. Ren[a], X.Q. Xiong[a], Y. Sha[a], M.C. Yan[a], B. Lin[b], J. Wang[a], Y.K. Jing[c], D.M. Zhao[a] and M.S. Cheng[a]*

[a]*School of Pharmaceutical Engineering, Shenyang Pharmaceutical University, Shenyang, P.R. China;* [b]*Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX, USA;* [c]*Department of Medicine, Mount Sinai School of Medicine, New York, USA*

(Received 31 March 2007; final version received 19 January 2008)

Two three-dimensional (3D) models of human cytochrome P450 26A1 (CYP26A1) were constructed using the programs Modeller and Sybyl-GeneFold, respectively. After refinement by molecular mechanics and molecular dynamics (MD) simulations, the two models were validated by structure analysis-validation online server. Subsequently, a flexible docking study was performed on the model constructed by GeneFold with the potent and specific inhibitor R115866 to examine the enzyme–inhibitor interactions. From the docking results, we can see R115866 interacts with amino acid residues at the active site by multiple hydrophobic interactions including the side chains of His111, Trp112, Ser115, Val116, Leu125, Ser126, Leu221, Phe222, Glu296, Phe299, Gly300, Glu303, Thr304, Pro371 and the cofactor heme. Trp112 and Thr304 form hydrogen bonds with R115866 and play important roles in stabilising the complex. This constructed CYP26A1 model may provide an opportunity to understand the action mode of the enzyme and could be useful in designing novel retinoic acid metabolism blocking agents (RAMBAs).

Keywords: molecular modelling; CYP26A1; molecular dynamics; docking; R115866

1. Introduction

Retinoids (vitamin A, its natural metabolites, and synthetic analogues) have a remarkably wide range of biological activities of regulating cell growth, differentiation, and morphogenesis [1]. All-*trans*-retinoic acid (ATRA), the most active metabolite of vitamin A, play important roles in cellular differentiation and proliferation of epithelial tissue and has been used for the treatment of acute promyelocytic leukaemia (APL) [2,3]. ATRA induces terminal differentiation of the malignant promyelocytes of APL patients with complete remission reported in 85–90% of the cases [4]. ATRA is metabolised into 4-hydroxyl-ATRA rapidly *in vivo* by cytochrome P450 enzymes and eventually into more polar metabolites [5]. The high ATRA metabolism has been thought to be a reason of ATRA resistance and disease relapse of APL. To inhibit the ATRA metabolism would be a feasible strategy to improve the therapeutic effect of ATRA and extend its use in the treatment of cancer. Several CYP isozymes have been shown to be capable of metabolising ATRA [6–8], but CYP26 appears to be the most dedicated ATRA 4-hydroxylase. CYP26 recognises only retinoic acid (RA) as its substrate and its expression is induced by ATRA both *in vitro* and *in vivo* at pharmacological concentrations [9]. Three subtypes of human CYP26 have been identified: CYP26A1 [10,11], CYP26B1 [12] and CYP26C1 [13]. CYP26A1 and CYP26B1 are responsible for the metabolism of ATRA both in the embryo and adults. CYP26C1 is a more recent discovered subtype that may play a specific role in the metabolism of both ATRA and 9-*cis*-retinoic acid. CYP26A1 was shown to be highly specific in the hydroxylation of ATRA and did not recognise the 9-*cis* and 13-*cis* retinoic acids [10]. Since CYP26A1 is induced in ATRA responsive leukemia and cancer cells and ATRA is clinically used for the treatment of cancer, it could used as a target to develop agents for inhibition of ATRA metabolism.

The human CYP26A1 consists of 497 amino acids [14], with a molecular weight of 56 163 Daltons. Cys442 connects to the cofactor heme directly through the sulfur atom and serves as the fifth ligand for it. However, a detailed analysis of the metabolic mechanism of ATRA at the molecular level has been hampered by the lack of a crystal structure for CYP26A1. Therefore it is necessary to determine the 3D structure which would pinpoint specific substrate contact residues and provide an opportunity to understand the interactions between the enzyme and the inhibitors. Such knowledge will also be helpful in designing novel blocking agents for ATRA metabolism. Gomaa et al. [15] reported a human CYP26 model constructed by homology modelling method

*Corresponding author. Email: mscheng@syphu.edu.cn

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020801930562
http://www.informaworld.com

based on CYP3A4 as template. We construct the model using homology modelling and fold recognition methods with improved stereochemical results.

2. Methods

2.1 Structure prediction

There are three methods for protein structure prediction including homology modelling, fold recognition and *ab initio* approach. *Ab initio* method is used for the prediction of small peptide structures, but not proteins such as CYP26A1 because the calculations required would be too expensive computationally. Therefore, we have attempted the other two methods in the current work.

Homology modelling is based on the assumption that the tertiary structures of two proteins may be similar if their sequences are related, and is an often used method when a clear homology template can be found. The amino acid sequence of human CYP26A1 obtained from the Swiss-Prot TrEMBL database (accession number O43174) was used to search for the templates using the program BLAST[1] [16] in the homology modelling method. *Mycobacterium tuberculosi*s CYP51 (MT-CYP51, PDB code: 1EA1 [17]) displayed the highest sequence identity to CYP26A1 (25%) in the Protein Data Bank [18] (PDB). Human CYP3A4 (HU-CYP3A4, PDB code 1TQN [19]) showed the highest sequence identity in mammalian to CYP26A1 (23%). Therefore, both MT-CYP51 and CYP3A4 were selected as the templates. The template sequences were extracted from the template PDB files. Using this two templates and target sequences, multiple sequence alignment was derived using CLUSTAL W 1.83 [20] package with some manual adjustment and default parameters, and a homology modelling procedure was performed by Modeller. Heme group in the two templates and fuconazole in 1EA1 were considered in our research, by adding parameter (set HETATM_IO = on) in the top file, using dots that represent the heme and fluconazole in the ali file, and numbering these groups after the last natural amino acid in the template pdb file.

In practice, the sequence alignment is very unstable when the sequence similarity is less than 30% of identical residues, and the quality abruptly drops [21]. Considering the relatively low sequence identity between CYP26A1 and all known protein structures in the PDB database, we also constructed a model using the fold recognition approach with GeneFold [22] program in the SYBYL6.9[2] molecular modelling package. GeneFold uses the probe sequence, target structure and a scoring function which measures the affinity of the probe sequence for the target structure to construct a model. With a representative library of protein folds, GeneFold finds the most appropriate fold for a given probe sequence. In this procedure, the numbers of top matches and sequence alignments between target and templates to be displayed were both set upto 20, and the crystal structure of *Bacillus megaterium* P450BM-3 (BM-CYP102, PDB code: 1BU7 [23]) showed high scores in the three scoring methods of GeneFold and was chosen as the template. Dai et al. [24] reported that P450BM-3 is the best choice when using a single P450 as a template, because it exhibits 20% sequence similarity with mammalian microsomal P450s and it is the only one interacting with NADPH cytochrome P450 reductase during the catalytic cycle as a class II P450. P450BM-3 has been used as a sole template for modelling of several P450s: 1A2 [25,26], 2B6 [27], 2D6 [25], 3A4 [25], etc. These also made us convinced of this template in fold recognition method. In our approach, the backbone of the structurally conserved regions (SCR) was constructed according to the template and the remaining parts of the backbone were constructed with Loop Search command in SYBYL. Side-chains were then added and the coordinates of heme were generated by superimposing the model to the template and extracting it from the counterpart of the template.

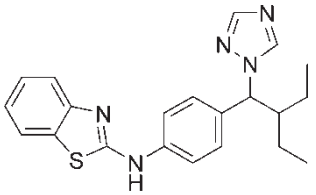
After model construction, energy minimisation and molecular dynamics (MD) simulations were performed in order to remove bad van der Waals contacts and to obtain stable conformations for all the constructed models of human CYP26A1. All hydrogen atoms were added and Amber7 charges were assigned to the two models. Then, energy minimisations were performed in SYBYL using steepest descent (SD) calculations with embedded Amber7 force field [28] for 5,000 iterations. In this procedure, heavy atoms were tethered to relieve possible steric clashes and overlaps of the hydrogen atoms. Conjugated gradient (CG) minimisations were then carried out until the systems converged to less than 0.05 kcal mol^{-1} $Å^{-1}$ for the gradient. Other miscellaneous parameters were assigned to their default values in SYBYL. In the CG minimisation procedure, heavy atoms and backbone were tethered successively, and then the whole structure was set to free. MD refinements were performed next using GROMACS [29] software. In this procedure, the simulations were carried out at 300 K using periodic boundary conditions as to avoid surface artifacts at the box edges. The two models were immersed in a cubic box ($0.65 \times 0.65 \times 0.65\,nm^3$) of extended SPC216 water molecules. The solvated system was neutralised by adding chloride ions whose positions were randomly chosen in the simulation. For the two constructed models, standard GROMACS amino acid residue topology and parameters based on the Gromacs Forcefield were used. Electrostatic energy was calculated using the particle mesh Ewald method [30]. Cutoff distances for the calculation of Coulomb and van der Waals interaction were 0.9 and 1.4 nm, respectively. After energy minimisation using a steepest decent

method, the system was subjected to equilibration at 300 K and normal pressure for 100 ps under the conditions of position restraints for heavy atoms. After that, 100-ps MD simulations with the protein main-chain and C_α atoms restrained were performed step by step. Finally, the two systems were subjected to free simulations under the same conditions for 3 ns. During the course of the simulation, the actual frame was stored every 4 ps. An average structure was refined further by a steepest decent and conjugate gradient energy minimisation.

The final refined models were evaluated by several programs: Procheck [31], Verify_3D [32,33], and Errat [34], which belong to the structure analysis-validation $(SAV)^3$ online server sponsored by the UCLA-DOE Institute for Genomics and Proteomics. Procheck was used to assess the 'quality' of the conformation of the polypeptide backbone and sidechains. Verify_3D and Errat were applied to evaluate the compatibility between the amino acid sequence and the environment of the amino acid sidechains in the model. The two approaches are complementary. Verify_3D assesses the environment of the sidechains based on the solvent accessibility of the sidechains and the fraction of the sidechains covered by polar atoms. Errat assesses the distribution of different types of atoms with respect to one another in the protein models [35].

2.2 Molecular docking study

Molecular docking involves fitting a protein and ligand together in a favourable conformation to form a complex. Structural information from such a complex may help to clarify the catalytic mechanism of the enzyme. The SiteID program in SYBYL was used to identify protein binding sites and functional residues. In this procedure, the FindPockets program was used to search the binding site by locating cavities in the protein structure. Combining the search results and the substrate recognition sites (SRS) [36] of other CYPs reported, the binding site of CYP26A1 can be predicted definitely. In order to understand the interactions between inhibitors and the enzyme, R115866 [37] (Figure 1), was docked into the model generated by GeneFold. The 3D structure of the R115866 was built with the SYBYL-Sketch module and the geometry was optimised. Computer-based automated docking of R115866 into the active site of CYP26A1 was performed using a grid based docking program, AUTODOCK 3.0.5 [38], with the Lamarckian genetic algorithm. A 52 × 52 × 52-point grid with a spacing of 0.375 Å centred at 47.703, 67.762, 38.717 Å was used for the protein model. Lennard-Jones parameters 12–10 and 12–6 were used for modelling H-bonds and van der Waals interactions, respectively.



R115866

Figure 1. Structure of R115866.

The maximum numbers of energy evaluation and generation were set to 250,000 and 27,000, respectively. A total of 50 separate docking runs were performed with the initial population of 50 individuals, and the lowest energy structure from each run was retained. Final docked conformations were clustered by using a tolerance of 1.0 Å root mean square deviation (RMSD). The ligand with the best receptor-ligand binding mode was selected.

3. Results and discussion

3.1 Template selection and sequence alignment

The BLAST searching results were shown in Table 1. In homology modelling approach, the crystal structure of 1EA1 showed the best sequence identity (25%) with CYP26A1. Furthermore, the co-crystallised ligand in 1EA1 was fluconazole which belongs to the early retinoic acid metabolism blocking agents (RAMBAs) [39]. Since substrate-induced conformational changes may occur, the constructed conformation of the active site with this template may be more approaching to the real binding mode of CYP26A1 with RAMBAs. Therefore, 1EA1 was chosen as template in the homology modelling method. Human CYP3A4 shows the highest sequence identity in mammalian to CYP26A1 (23%) and was also selected as a template in this method. The relatively low sequence identity between target and template proteins renders the CYP26A1 modelling a challenging task. However, it has been revealed that CYPs both in eukaryote and prokaryote share similar structural foldings [35].

The secondary structures of the temples were labelled as defined by Williams et al. [40] in Figure 2. From the alignment we observe that there is no insertion other than helix G in secondary structure elements. The conserved regions, such as heme binding domain PFGGGLRSCVG (underlined are the amino acids which are conserved in the heme binding domain), PERF domain (PDRF position 416–419) which forms an ERR triad [41], using Glu and Arg from the K-helix domain and Arg from the PERF domain, and ETLR conserved domain 50 amino acids upstream the PERF domain were matched correctly among our model and the templates.

Table 1. The P450 structures recovered by a BLAST search using the sequence of HU-CYP26A1.

Protein	PDB code	BLAST bit score	Sequence identity	% Sequence identity	Chain length
MT-CYP51	1EA1-A	101.252	112/443	25	455
BM-CYP102	1BU7-B	95.9237	93/363	25	455
HU-CYP3A4	1TQN-A	94.0232	103/434	23	486
HU-CYP2R1	2OJD-A	78.6192	96/437	21	479
RA[a]-CYP2B4	1PO5-A	75.5184	112/474	23	476
PS[b]-CYP108	1CPT	72.4176	62/224	27	412
HU-CYP2A6	1Z10-A	70.5171	108/480	22	476
HU-CYP2C9	1R90-A	68.2165	105/473	22	477
HU-CYP2A13	2P85-A	67.0162	108/477	22	476
HU-CYP2D6	2F9Q-A	65.5158	48/206	23	479
EC[c]-CYP107	2BVJ-A	63.9154	62/235	26	436
RA-CYP2C5	1DT6-A	63.9154	104/477	21	473
HU-CYP2C8	1PQ2-A	63.5153	112/477	23	476

[a] Rabbit. [b] Pseudomonas. [c] *Escherichia coli*.

3.2 Model construction and final model determination

In homology modelling method, ten CYP26A1 conformer models were generated by the Modeller program. The N-terminus (form Met1 to Val33) was discarded, because the coordinates of the corresponding region were not resolved in HU-CYP3A4 by X-ray diffraction and did not exist in MT-CYP51. This region is thought to be a membrane anchor in mammalian P450s, but it is not present in the soluble bacterial P450s [42]. The structural uncertainty of this region is not critical since it is not involved in substrate recognition which is the primary focus of our modelling study. Among the ten models produced by Modeller, CYP26A1.B99990005.pdb was selected due to its lower value of Modeller objective function and better performance in SAV validations (Table 2).

In the GeneFold method, all of the three scoring methods indicated that chain B of 1BU7 was the best template. The alignment of target sequence and template sequence was accomplished automatically with the BLAST procedure embedded in GeneFold. And the sequence alignment of 1BU7 with CYP26A1 in GeneFold was integrated in Figure 2. Using this template, a 3D model containing the residues from Arg36 to Ala488 was constructed. The N- and C-termini were not constructed because their counterparts were lacking in the template. However, these missing residues are far from the active site and are expected to have minimal effects on the results.

After energy minimisation, MD simulations were performed on the two models constructed by Modeller and GeneFold. The dynamic behaviours and the structural changes of the models during the course of the simulation period were analysed. The RMSD from the starting structure is an important criterion for the convergence of the protein system. The RMSD values of all C_α atoms of the GeneFold and Modeller models showed that the two systems reached convergence after 2 ns equilibrations in Figure 3. The average protein structures (Figure 4) between 2 and 3 ns were calculated as the final results for the two methods.

Protein backbone and sidechain conformations of the refined two models were evaluated by Procheck which includes a Ramachandran plot parameter for the assessment of protein quality based on the ϕ/ψ-angle distribution. The Ramachandran plots indicated that more than 98% of the residues in each model were located in favoured and allowed regions. Cys256 (Figure 5(a)) in the Modeller model and Phe77, His129 and Phe430 (Figure 5(b)) in the GeneFold model were found in the disallowed regions, but all these residues occurred on the surface of the protein were far away from the active site and seemed not to affect the most important part of the structure. The Verify_3D and Errat results of the two models are listed in Table 3. From the results of the SAV validation, we conclude that the two models have satisfactory stereochemistry and good amino acid sidechain environment properties. It seems that the model constructed by GeneFold is slightly reasonable by its better performance in Verify_3D and Errat.

The helices F and G in the structure made by Modeller (Figure 4(a)) are shorter comparing with those regions in other P450s. The helices F and G have been thought to comprise part of the substrate pocket or the substrate access channel in bacterial and mammalian P450s [40,44,45] and F–G loop might be the membrane anchor point for mammalian P450s. In 2001, Haines et al. [46] reported that helix F, helix G and F–G loop composed a 'lid domain' as a substrate access channel in studying of P450BM-3 and N-palmitoylglycine complex. The composed channel is involved in a clam shell-like movement to trap substrate and exhibits a rocking motion with helix I as a fulcrum. A number of structural studies have suggested that the C-terminal portion of helix F may be important in forming the ceiling of the substrate

Figure 2. Multiple sequence alignment of human CYP26A1 to MT-CYP51 and HU-CYP3A4. Identical residues in the three CYPs are depicted by red letters. Magenta, yellow and blue boxes on templates represent α-helix, β-strand and Cys-pocket, respectively.

Table 2. The value of objective function and results of the Ramachandran checks, Verify_3D, Errat for the ten CYP26A1 conformer models produced by Modeller.

Conformer model	Objective function	Ramachandran plot[a] (%)				Verify_3D[b] (%)	Errat[c] (%)
		Most favoured	Additional	Generously	Disallowed		
CYP26A1.B99990001.pdb	49674.0508	83.5	13.0	2.5	1.0	77.42	44.298
CYP26A1.B99990002.pdb	34271.0273	82.5	14.2	2.5	0.8	62.37	46.930
CYP26A1.B99990003.pdb	34918.0313	83.2	12.8	3.0	1.0	67.74	42.325
CYP26A1.B99990004.pdb	47381.6797	82.2	15.8	1.5	0.5	73.55	36.889
CYP26A1.B99990005.pdb	38891.5195	84.2	12.8	1.5	1.5	72.90	41.009
CYP26A1.B99990006.pdb	38144.8867	81.8	14.0	3.0	1.2	67.10	38.377
CYP26A1.B99990007.pdb	46660.1797	80.8	15.0	3.2	1.0	72.26	41.667
CYP26A1.B99990008.pdb	36889.1992	83.5	13.2	2.8	0.5	67.74	51.974
CYP26A1.B99990009.pdb	48327.8438	84.0	12.5	2.8	0.8	64.73	45.175
CYP26A1.B99990010.pdb	43213.5625	82.2	14.0	3.0	0.8	58.92	49.561

[a]Percentage of residues with Φ, Ψ conformation in the corresponding regions of the Ramachandran plot. [b]Percentage of residues with an average 3D-1D score >0.20. [c]Overall quality factor.

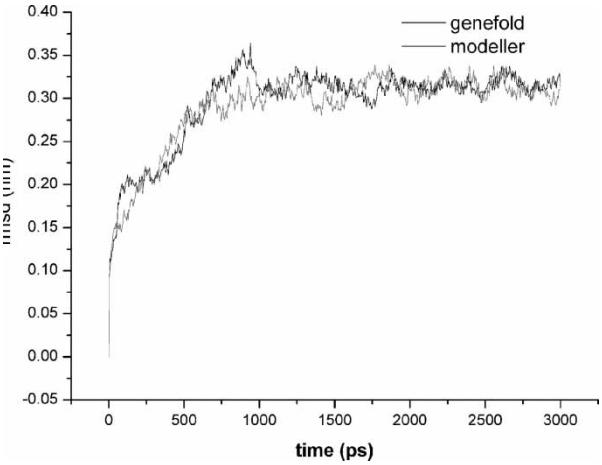


Figure 3. RMSD values for the GeneFold model (gray) and Modeller model (black).

binding pocket in P450s [47]. However, helix F in the model constructed by Modeller was so short that the C-terminal potion was far from the active site and could not exert this function. These regions may be improper for their anomalistic structure which is significantly different from other P450s reported so far.

Considering the structure information mentioned above and the SAV validation, we chose the GeneFold model for further studies.

The final structure of human CYP26A1 constructed by GeneFold is shown in Figure 4(b). The overall fold can be divided into two domains: a helix-rich domain containing the cofactor heme and a β-sheet domain which mostly consists of antiparallel β-strands. The heme of P450 is sandwiched between helix L on the side of the heme proximal to the surface of the protein and helix I in the interior of the protein. Hydrogen bond interactions between heme and amino acids near the active site including the propionic acid sidechain of heme with Arg440 in the β bulge before helix L and the porphyrin nitro atoms with Thr304 in helix I contribute to the stability of the active site. The spatial organisation of the heme binding sites both in the soluble microbial P450s and in mammalian enzymes are highly conserved even though the proteins exhibit very low sequence identities [48]. The axial ligand at the fifth coordination site of the heme iron is provided by the sulfide of Cys442 in CYP26A1. The entrance to the substrate access channel between the F–G loop and the N-terminal β sheet system is located in this membrane attachment surface.

And comparing with the model constructed by Gomaa et al. (MOE-Homology), our model (GeneFold) has relatively better stereochemical quality and amino acid environment (Table 4).

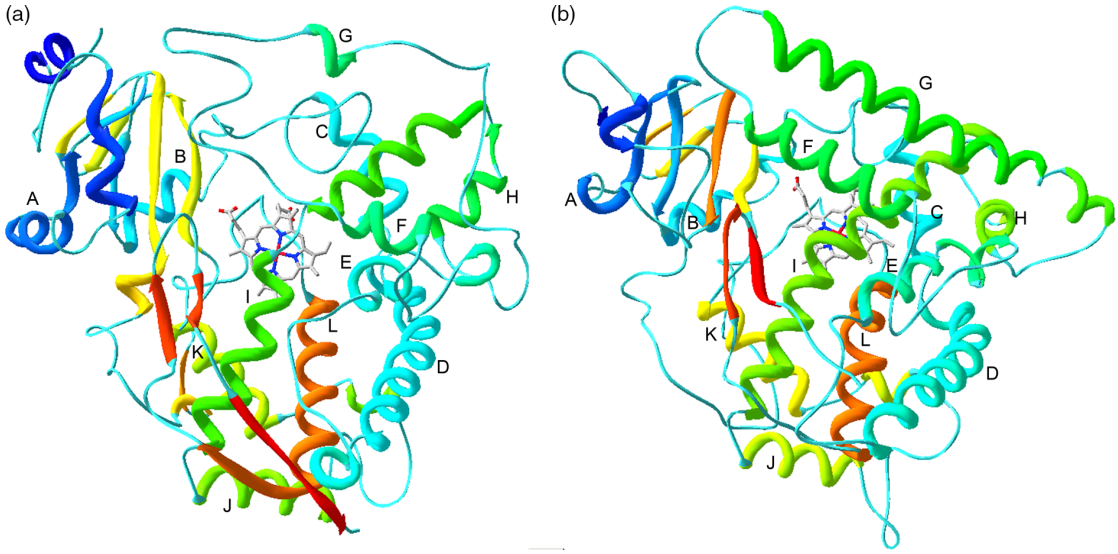


Figure 4. 3D model of CYP26A1 generated by Modeller (a) and by GeneFold (b) after MD simulations and displayed by SPDBV software [43].

3.3 Interactions between the constructed model and inhibitor

SiteID program gave us a possible binding pocket in the CYP26A1 model in an area corresponding to the substrate binding pocket of other CYP crystal structures, which was above the heme group and located in the core of the protein. Amino acids in this predicted binding area included His111, Trp112, Ala114, Ser115, Val116, Leu125, Ser126, Thr218, Leu221, Phe222, Glu296, Phe299, Gly300, Glu303, Thr304, Pro371, Phe374 and heme. The key residues involved in the ligand binding in this model are mainly consistent with the results reported by Gomaa et al. However, it is not straightforward to interpret whether an amino acid is directly involved in ligand binding or structurally important for maintaining the conformation of the binding pocket. Therefore, a molecular docking study was performed to confirm the binding residues with typical substrates. R115866, the most potent and selective CYP26A1 inhibitor reported

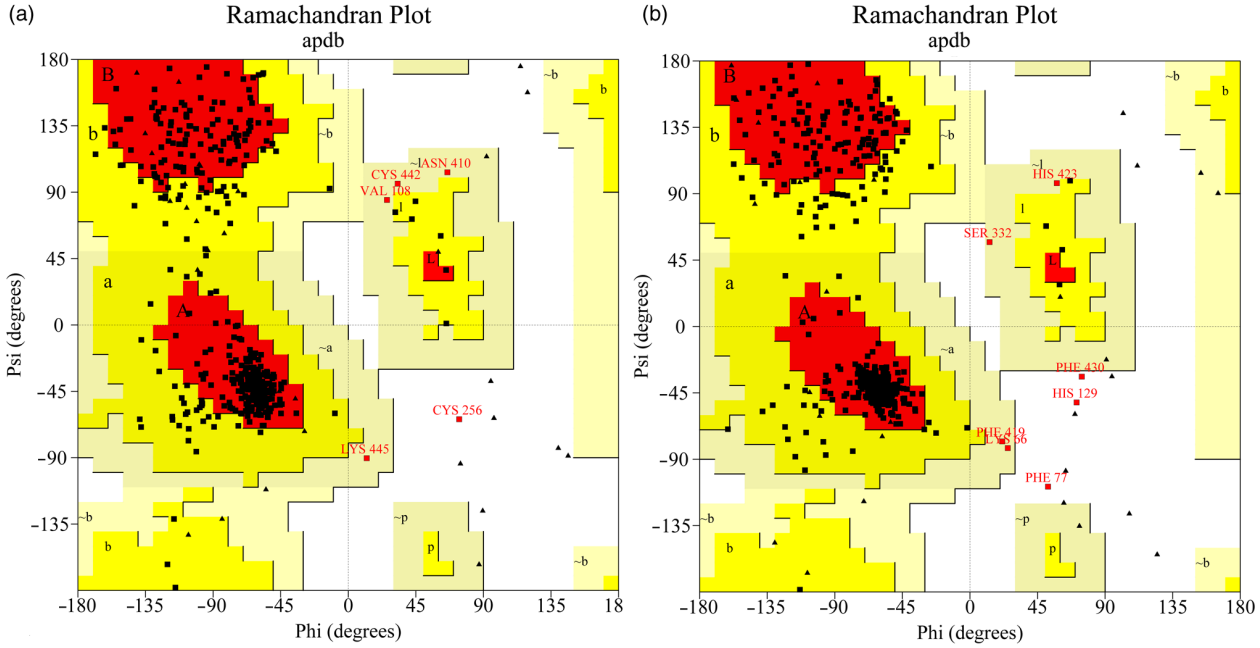


Figure 5. Ramachandran plots of the Φ, Ψ distribution of Modeller (a) and GeneGold (b) models.

Table 3. The results of the Ramachandran checks, Verify_3D and Errat performed on the models generated by Modeller and GeneFold methods and the templates.

	Ramachandran plot (%)					
	Most favoured	Additional	Generously	Disallowed	Verify_3D (%)	Errat (%)
The templates						
1EA1	88.9	10.3	0.5	0.3	99.33	94.050
1TQN	84.7	12.6	2.0	0.7	95.74	93.696
1BU7	93.1	6.7	0.2	0.0	99.34	93.960
Model after MM						
Modeller	84.2	12.8	1.5	1.5	67.10	75.000
GeneFold	81.6	15.3	1.0	2.0	81.94	46.742
Model after MD						
Modeller	80.0	18.8	1.0	0.2	81.29	82.018
GeneFold	83.9	14.3	1.0	0.8	87.89	93.679

to date, was selected for this study. Fifty conformations of R115866 produced by AutoDock3.0 have been analysed visually. Considering the significance of the azole group of many drugs to inhibit P450 enzyme [49] and the fluconazole binding model in the MT-CYP51 template, the triazole in R115866 which was in a perpendicular position with respect to the heme iron in the enzyme pocket was selected as the best binding mode. It was detected with an estimated binding free energy of $-11.92\,\mathrm{kcal\,mol^{-1}}$ and an estimated inhibition constant Ki of $1.84\,\mathrm{nmol\,l^{-1}}$. The magnitude order of Ki is approximately consist with its experimental activity ($IC_{50} = 4\,\mathrm{nmol\,l^{-1}}$) [50] which could be considered as an indirect proof of the rationality of the model established. From the docking results, it was observed that R115866 interacted with amino acid residues at the active site by multiple hydrophobic interactions, including the side chains of His111, Trp112, Ser115, Val116, Leu125, Ser126, Leu221, Phe222, Glu296, Phe299, Gly300, Glu303, Thr304, Pro371 and heme (Figure 6). Additional factors involved in the binding of the substrate including hydrogen bonding and $\pi-\pi$ stacking interactions were shown in Figure 7. Trp112 formed hydrogen bonding interactions with the nitrogens both in benzothiazole and aniline of R115866, and Thr304 formed same interaction with the nitrogen in triazole. The complementary aromatic residues in the active site of CYP26A1, such as the indole ring of Trp112 and the phenyl ring of Phe222 formed $\pi-\pi$ interactions with the benzothiazole ring of R115866. This is similar to Gomaa's research which shown Phe84, Trp112, Phe222, Phe299, Pro371

Table 4. Validation results comparison between Gomaa's model and our GeneFold model.

Method	Ramachandran Plot (%) most favoured	Verify_3D[a]	Errat (%)
MOE-Homology	81.4	121	84.5
GeneFold	83.9	170	93.7

[a] Total Verify_3D score summed over all the residues.

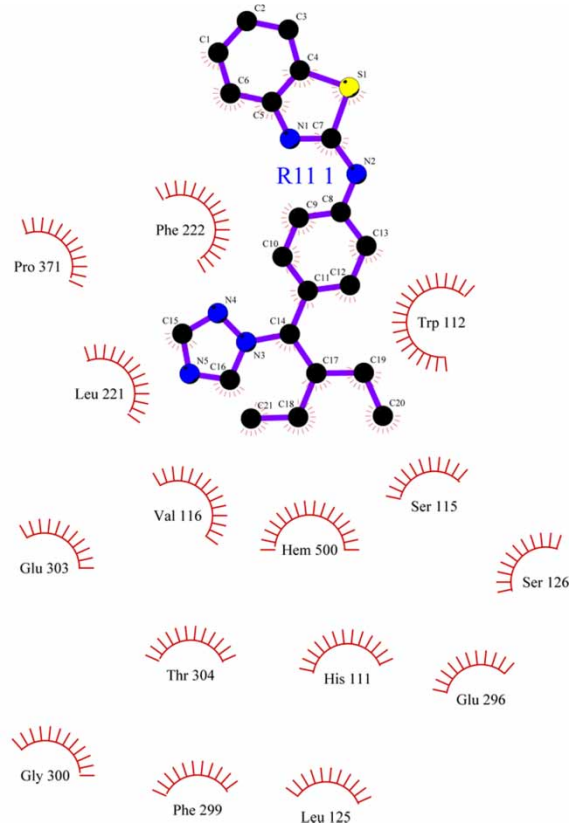


Figure 6. Hydrophobic interactions between CYP26A1 and R115866.

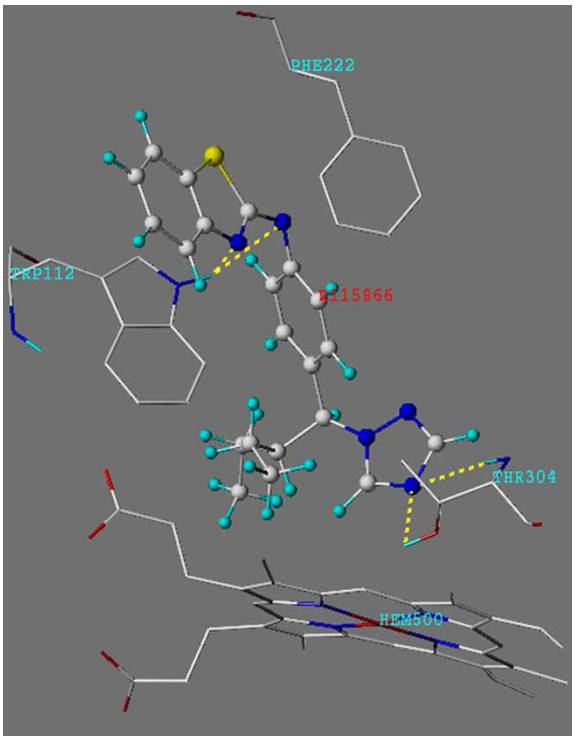


Figure 7. Hydrogen bonding and π−π stacking interactions between CYP26A1 and R115866.

and Phe374 had hydrophobic interactions with R115866. However, they showed that Ser115 but not Trp112 and Thr304 formed hydrogen bonded interactions with R115866 in their model.

4. Conclusion

The three-dimensional models of human CYP26A1 were constructed with homology modelling and fold recognition methods respectively. From the SAV validation results, it seems that the model generated by fold recognition method is a bit more reasonable. Moreover, the structure of helices F and G in Modeller model was improper for their short length. Therefore the model constructed by GeneFold was selected as the final human CYP26A1 structure. And this model performs better in terms of mainchain stereochemistry and amino acid environment with the previous reported model. The inhibitor R115866 docked into the active site of this confirmed model and revealed the essential residues involved in hydrogen bonds and hydrophobic interactions with CYP26A1. This comparative model of human CYP26A1 is useful in the absence of experimentally determined structure to provide an opportunity to understand the action mode of the enzyme and to present the possibility of using structure-based design to develop inhibitors that are effective against the metabolism of CYP26A1. Further work in this area is currently in progress to discover novel RAMBAs by virtual screening.

Acknowledgement

The MD simulations were carried out on the Sunway supercomputer platform at Beijing Institute of Pharmacology and Toxicology. We also acknowledge Professor Song Li for his support.

Notes

1. http://www.expasy.org/tools/blast
2. SYBYL 6.91; Tripos Inc. (1699 South Hanley Road, St Louis, MO 63144, USA); http://www.tripos.com
3. http://nihserver.mbi.ucla.edu/SAVS/

References

[1] V.C.O. Njar, *Cytochrome p450 retinoic acid 4-hydroxylase inhibitors: Potential agents for cancer therapy*, Mini. Rev. Med. Chem. 2 (2002), p. 261.
[2] S. Castaigne et al., *Retinoic acids in the treatment of acute promyelocytic leukemia*, Nouv. Rev. Fr. Hematol. 32 (1990), p. 36.
[3] L. Degos and Z.Y. Wang, *All* trans *retinoic acid in acute promyelocytic leukemia*, Oncogene 20 (2001), p. 7140.
[4] W.L. Zhao et al., *Hemostatic abnormalities associated with acute promyelocytic leukemia and corrective effects of all*-trans-*retinoic acid or arsenic trioxide treatment*, Chin. Med. J. 113 (2000), p. 236.
[5] J.L. Napoli, *Retinoic acid biosynthesis and metabolism*, FASEB J. 10 (1996), p. 993.
[6] M.A. Leo, S. Iida, and C.S. Lieber, *Retinoic acid metabolism by a system reconstituted with cytochrome P-450*, Arch. Biochem. Biophys. 234 (1984), p. 305.
[7] R. Martini and M. Murray, *Kinetic evidence for the involvement of a common enzyme in the microsomal reduction of retinal and androstenedione in rat liver*, J. Steroid. Biochem. Mol. Biol. 45 (1993), p. 581.
[8] L.C. McSorley and A.K. Daly, *Identification of human cytochrome P450 isoforms that contribute to all*-trans-*retinoic acid 4-hydroxylation*, Biochem. Pharmacol. 60 (2000), p. 517.
[9] W.J. Ray et al., *CYP26, a novel mammalian cytochrome P450, is induced by retinoic acid and defines a new family*, J. Biol. Chem. 272 (1997), p. 18702.
[10] E. Sonneveld et al., *Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for all*-trans-*RA and can be induced through RA receptors in human breast and colon carcinoma cells*, Cell Growth Differ. 9 (1998), p. 629.
[11] M.E. Trofimova-Griffin and M.R. Juchau, *Expression of cytochrome P450RAI (CYP26) in human fetal hepatic and cephalic tissues*, Biochem. Biophys. Res. Commun. 252 (1998), p. 487.
[12] D.R. Nelson, *A second CYP26 P450 in humans and zebrafish: CYP26B1*, Arch. Biochem. Biophys. 371 (1999), p. 345.
[13] M. Taimi et al., *A novel human cytochrome P450, CYP26C1, involved in metabolism of 9-cis and all-trans isomers of retinoic acid*, J. Biol. Chem. 279 (2004), p. 77.
[14] J.A. White et al., *cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RAI) identifies a novel family of cytochromes P450*, J. Biol. Chem. 272 (1997), p. 18538.
[15] M.S. Gomaa et al., *Homology model of human retinoic acid metabolising enzyme cytochrome P450 26A1 (CYP26A1): Active site architecture and ligand binding*, J. Enzyme Inhib. Med. Chem. 21 (2006), p. 361.

[16] S.F. Altschul et al., *Basic local alignment search tool*, J. Mol. Biol. 215 (1990), p. 403.
[17] L.M. Podust, T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors*, Proc. Natl Acad. Sci. USA 98 (2001), p. 3068.
[18] F.C. Bernstein et al., *The Protein Data Bank: a computer-based archival file for macromolecular structures*, J. Mol. Biol. 112 (1977), p. 535.
[19] J.K. Yano et al., *The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution*, J. Biol. Chem. 279 (2004), p. 38091.
[20] J.D. Thompson, D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice*, Nucleic Acids Res. 22 (1994), p. 4673.
[21] G. Vogt, T. Etzold, and P. Argos, *An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited*, J. Mol. Biol. 249 (1995), p. 816.
[22] L. Jaroszewski et al., *Fold prediction by a hierarchy of sequence, threading, and modeling methods*, Protein sci. 7 (1998), p. 1431.
[23] I.F. Sevrioukova et al., *Structure of a cytochrome P450-redox partner electron-transfer complex*, Proc. Natl Acad. Sci. USA 96 (1999), p. 1863.
[24] R. Dai, M.R. Pincus, and F.K. Friedman, *Molecular modeling of mammalian cytochrome P450s*, Cell. Mol. Life Sci. 57 (2000), p. 487.
[25] F.D. Rienzo et al., *Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4*, J. Comput. Aid. Mol. Des. 14 (2000), p. 93.
[26] R. Dai et al., *Inhibition of human cytochrome P450 1A2 by flavones: a molecular modeling study*, J. Protein Chem. 17 (1998), p. 643.
[27] D.F. Lewis et al., *Molecular modeling of CYP2B6, the human CYP2B isoform, by homology with the substrate-bound CYP102 crystal structure: evaluation of CYP2B6 substrate characteristics, the cytochrome b5 binding site and comparisons with CYP2B1 and CYP2B4*, Xenobiotica 29 (1999), p. 361.
[28] S.J. Weiner et al., *A new force field for molecular mechanical simulation of nucleic acids and proteins*, J. Am. Chem. Soc. 106 (1984), p. 765.
[29] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. 91 (1995), p. 43.
[30] U. Essmann et al., *A smooth particle mesh Ewald method*, J. Chem. Phys. 103 (1995), p. 8577.
[31] R.A. Laskowski et al., *PROCHECK: a program to check the stereochemical quality of protein structures*, J. Appl. Cryst. 26 (1993), p. 283.
[32] J.U. Bowie, R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure*, Science 253 (1991), p. 164.
[33] R. Luthy, J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles*, Nature 356 (1992), p. 83.
[34] C. Colovos and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*, Protein Sci. 2 (1993), p. 1511.
[35] S.B. Kirton, C.A. Baxter, and M.J. Sutcliffe, *Comparative modeling of cytochromes P450*, Adv. Drug Deliv. Rev. 54 (2002), p. 385.
[36] O. Gotoh, *Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences*, J. Biol. Chem. 267 (1992), p. 83.
[37] P. Stoppie et al., *R115866 inhibits all*-trans-*retinoic acid metabolism and exerts retinoidal effects in rodents*, J. Pharmacol. Exp. Ther. 293 (2000), p. 304.
[38] G.M. Morris et al., *Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function*, J. Comput. Chem. 19 (1998), p. 1639.
[39] E.L. Schwartz et al., *Inhibition of all*-trans-*retinoic acid metabolism by fluconazole* in vitro *and in patients with acute promyelocytic leukemia*, Biochem. Pharmacol. 50 (1995), p. 923.
[40] P.A. Williams et al., *Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity*, Mol.Cell 5 (2000), p. 121.
[41] Y.M. Zheng et al., *Identification of a meander region praline residue critical for heme binding to cytochrome P450: implications for the catalytic function of human CYP4B1*, Biochemistry 37 (1998), p. 12847.
[42] T.L. Poulos, *Cytochrome P450: Molecular architecture, mechanism, and prospects for rational inhibitor design*, Pharm. Res. 5 (1988), p. 67.
[43] N. Guex and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling*, Electrophoresis 18 (1997), p. 2714.
[44] H.L. Lin et al., *Threonine-205 in the F helix of P450 2B1 contributes to androgen 16β-hydroxylation activity and mechanism-based inactivation*, J. Pharmacol. Exp. Ther. 306 (2003), p. 744.
[45] C.A. Hasemann et al., *Structure and function of cytochromes P450: a comparative analysis of three crystal structures*, Structure 3 (1995), p. 41.
[46] D.C. Haines et al., *Pivotal role of water in the mechanism of P450BM-3*, Biochemistry 40 (2001), p. 13456.
[47] I.A. Pikuleva, A. Puchkaev, and I. Bjorkhem, *Putative helix F contributes to regioselectivity of hydroxylation in mitochondrial cytochrome P450 27A1*, Biochemistry 40 (2001), p. 7621.
[48] C.A. Hasemann et al., *Structure and function of cytochromes P450: a comparative analysis of three crystal structures*, Structure 3 (1995), p. 41.
[49] F.C. Odds and H.V. Bossche, *Antifungal activity of itraconazole compared with hydroxy-itraconazole* in vitro, J. Antimicrob. Chemother. 45 (2000), p. 371.
[50] P. Stoppie et al., *R115866 Inhibits All*-trans-*Retinoic Acid metabolism and exerts retinoidal effects in rodents*, J. Pharmacol. Exp. Ther. 293 (2000), p. 304.