

Molecular modeling and identification of substrate binding site of orphan human cytochrome P450 4F22

Suresh Kumar*

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Coimbatore -641114, Tamil Nadu, India; Suresh Kumar - Email: sureshbio@gmail.com; *Corresponding author

Received October 08, 2011; Accepted October 11, 2011; Published October 14, 2011

Abstract:

Cytochrome P450s are superfamily of heme proteins which generally monooxygenate hydrophobic compounds. The human cytochrome P450 4F22 (CYP4F22) was categorized into "orphan" CYPs because of its unknown function. CYP4F22 is a potential drug target for cancer therapy. However, three-dimensional structure, the active site topology and substrate specificity of CYP4F22 remain unclear. In this study, a three-dimensional model of human P450 4F22 was constructed by comparative modeling using Modeller 9v5. The resulting model was refined by energy minimization, subjected to the quality assessment from both geometric and energetic aspects and was found to be of reasonable quality. Docking approach was employed to dock arachidonic acid into the active site of CYP4F22 in order to probe the ligand-binding modes. As a result, several key residues were identified to be responsible for the binding of arachidonic acid with CYP4F22. These findings provide useful information for understanding the biological roles of CYP4F22 and structure-based drug design.

Keywords: Human Cytochrome, P450 4F22, CYP4F22, comparative modeling, molecular docking, arachidonic acid, ligand binding site, 2D QSAR

Background:

Cytochromes P450 (CYP) which are a superfamily of heme containing enzymes play an important role in detoxification and metabolic activation of large number of xenobiotic chemicals such as drugs, carcinogens and many endogenous compounds like hormones and vitamins. There are 57 CYP isoenzymes found in humans which are well-known for their monooxygenase reaction. However, one quarter of the CYP family remains orphan, which means that their function, expression studies and substrate information are not yet clearly understood. Thirteen members of this superfamily are classified as orphans by Guengerich [1].

Human CYP4F22 was considered as one such orphan CYPs, amino acid sequence of which has been identified recently. The three-dimensional structure of this protein is not yet known. Experimental studies of CYP4 family members are previously

reported to metabolize arachidonic acid (AA) [2] like endogenous compounds to its hydroxyl product. CYP family members were found to share their substrates with one another. CYP4F22 gene as a part of a cluster of cytochrome P450 genes on chromosome 19 encodes an enzyme thought to play a role in the 12(R)-lipoxygenase pathway. Mutation in this gene is the cause of ichthyosis lamellar type 3. It was recently found that CYP4F22 was specifically expressed in human breast cancer tissues. For this reason, human CYP4F22 has been suggested to be a potential drug target for cancer therapy. However, to date, information regarding the structure and ligand binding site is not available for CYP4F22 and hence structural information might help us to understand the ligand interaction.

In this study, a 3D model of CYP4F22 was constructed using comparative modeling method, followed by energy minimization to refine the initial model. The quality of the

resulting 3D model of P450 4F22 was critically assessed through geometric and energetic assessments. After that, arachidonic acid (AA) was docked into the active site of the proposed 3D structure of CYP4F22 and certain key residues responsible for substrate specificity were identified. Further 2D QSAR study was used to find out the pharmacological characteristics of the ligand. The results obtained in this study will be useful for further elucidation of the biological role and the structure-based drug design of P450 4F22.

Methodology:

Sequence retrieval

The protein sequence (531 amino acids) of human CYP4F22 was obtained from the UniProtKB database (accession number: Q6NT55).

Template identification and sequence alignment

For template selection, BlastP [3] search was used against Protein DataBank (PDB). The templates having high similarity were selected for the model building. The templates were aligned and examined for conserved sequence with the target by ClustalW [4].

Comparative modeling of human CYP4F22

The alignment was used as the input and the homology model of P450 2F22 was generated using Modeller 9v5 [5]. Loops were modeled using ModLoop server [6]. The coordinates for heme were obtained from 1TQN and positioned as in the template.

Energy minimization and structural quality assessment

The resulting model was further refined by energy minimization using the Amber9 package [7] and subjected to quality assessment. The Procheck [8] and Verify 3D [9] were utilized for geometric evaluation. The Prosa [10] was employed to examine the energy of residue-residue interactions using a distance-based pair potential approach. The energy was transformed to a score called z-score where residues with negative z-score indicate reasonable side-chain interactions. The final model was then submitted to the protein modeling database [11] as PM0077605.

2D QSAR study:

The selected ligand arachidonic acid (AA) is a long-chain fatty acid with a hydrophobic tail. The arachidonic acid was downloaded from Pubchem in Structure Data Format (SDF). Conversion of SDF to Protein Data Bank (PDB) format was carried out using Swiss PDB viewer. The molecular structure of ligand was optimized using the Polak-Ribiere algorithm until the root mean square gradient was 0.01 kcal mol⁻¹ using Hyperchem V 8.09 [12]. The molecular descriptors were calculated for the ligand using QSAR properties of Hyperchem.

Prediction of key residues in homology model of CYP4F22:

The homology model of CYP4F22 was submitted to automated ConSurf web server [13] for the identification of functional regions of CYP4F22 by estimating the degree of conservation of the amino-acid sites among their close sequence homologues.

Molecular docking of CYP4F22-ligand complex:

Both the modeled protein and ligand were optimized using "dock prepare" of the USCF chimera [14]. The optimized ligand

molecule was docked into the refined CYP4F22 protein model using Hex software [15]. The molecular graphical representation was prepared using Pymol program [16].

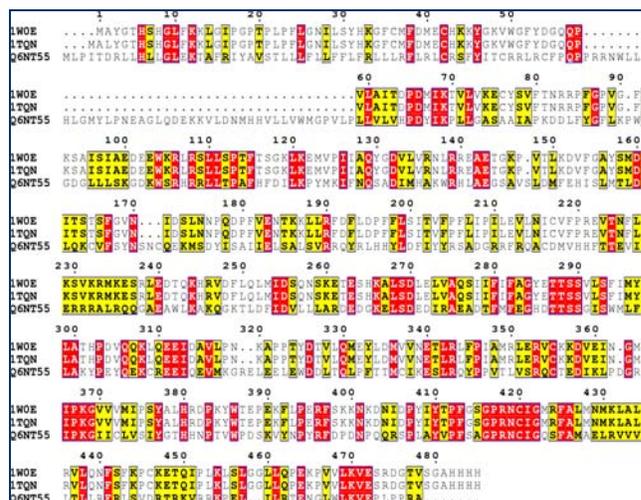


Figure 1: Multiple sequence alignment of target (CYP4F22:Q6NT55) and two templates (1W0E and 1TQN) represented using ClustalW Program. The alignment figure was prepared using Esript 2.2. The red colour indicates the residues conserved in all three sequences.

Results and Discussion:

The first step in homology modeling is to detect appropriate template proteins whose structures are known that are expected to be similar to the target protein. A BlastP search confirmed that there are several mammalian CYPs structures which could serve as potential templates for P450 4F22 modeling. The top-ranked two templates are 1W0E [17] and 1TQN [18]. The sequence conservation and the signature motifs of CYP4F22 were examined using multiple sequence alignment with templates (Figure 1). The essential signature motif FxGxxxCxG is characteristic motif for the CYP superfamily which includes a conserved cystine residue that ligates to the Fe of the heme. The amino acid sequence identity between target and the templates was above 40% which is high enough to construct a 3D model. The initial 3D model of CYP4F22 was energy-minimized and assessed for both geometric and energy aspects. The Procheck evaluation for good stereochemistry of the refined model resulted in 97.8% in the allowed regions and 2.1% in disallowed regions. Verify 3D analysis resulted in a model with about 81% of residues with a score over 0.2, indicating a good model. The Prosa analysis showed that z-score in the CYP4F22 model was negative in most residues, further indicating a reliable model. The 3D structure model of CYP4F22 exhibits a fold similar to those of other P450s and contains the heme group sandwiched between helices (Figure 2a). CYP4F22 model of this study also reveals that a highly conserved threonine is located in the middle of helix I which is similar to other reported human CYP structures. This conserved residue has been suggested to participate in the protein delivery and play an important role in the dioxygen bond cleavage during the catalytic cycle [19]. The CYP4F22 model was further tested using ConSurf server for the prediction of active residues (Figure 2b) that are expected to be accessible, either at the surface or in a pocket of the protein. The

functional residues of CYP4F22 model with high conservation scores were identified **Table 1** (see supplementary material). The energy minimized model of arachidonic acid was also subjected to 2D QSAR study using Hyperchem. Topological (surface area of 603.66 Å²) and constitutional descriptors (LogP of 10.10, molecular weight of 303.46 amu, polarizability of 36.83 Å³ and refractivity of 62.82 Å³) were computed and found to be desirable pharmacological characteristics for docking. To determine the key residues responsible for ligand interaction, the homology model of CYP4F22 was docked with arachidonic acid (**Figure 2c**). The free energy of binding (ΔG_{bind} kcal/mol) for the designed ligand was found to be -3.53 kcal/mol. The negative and low value of ΔG_{bind} indicates a strong favorable bond between CYP4F22 protein and the AA ligand in most favorable conformations. Several important residues were detected in homology model of P450 4F22. The residue ARG 156 appears to play an important role in substrate specificity. The corresponding TRP 152 and THR 160 pairs were conserved and they were directly in contact with ligands. In addition, the rest of the ligand-interacting residues GLY 471, CYS 475, ILE 476 and GLN 478 were mostly located near the key residues (**Figure 2d**).

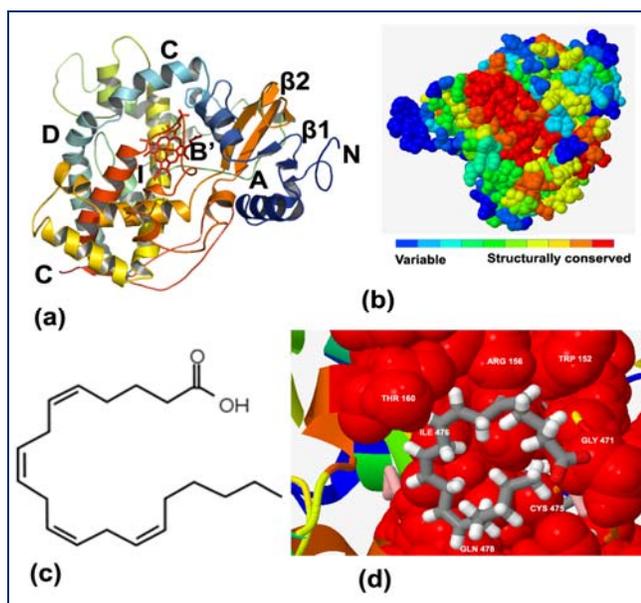


Figure 2: (a) Ribbon schematic representation of the homology model of CYP4F22. The heme is shown in the center of the figure in red stick. Various visible regions of secondary structure elements are labeled in the figure, (b) Predicted active sites of CYP4F22, (c) 2D structure of arachidonic acid and (d) a close view of CYP4F22 with arachidonic acid. Key residues are shown in red spheres and labeled. The substrate arachidonic acid is shown as grey stick.

Conclusion:

The human P450 4F22 theoretical 3D model was constructed using homology modeling, refined with energy minimization and assessed for stereochemical and energetic aspects. Docking approaches were employed to dock arachidonic acid ligand into the active site of the proposed 3D model in order to probe ligand binding modes. The docking studies as well as the ConSurf analysis revealed the important residues involved in the substrate specificity. The docking and ConSurf analysis are in concurrence with each other reflecting the importance of crucial residues. Interacting key residues gave information on the substrate specificity of the CYP4F22 active site when they interact with arachidonic acid ligand of hydrophobicity. Also these residues might act as ligand-channeling and could help the reaction center of the protein. The current results might be very helpful for the determination of biological role to provide insights into arachidonic acid.

References:

- [1] Guengerich FP *et al. Biochem Biophys Res Commun.* 2005 **338**: 465 [PMID: 16126164]
- [2] Hsu MH *et al. Drug Metab Rev.* 2007 **39**: 515 [PMID: 17786636]
- [3] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]
- [4] Chenna R *et al. Nucleic Acids Res.* 2003 **31**: 3497 [PMID: 12824352]
- [5] Sali A *et al. Proteins* 1995 **23**: 318 [PMID: 8710825]
- [6] Fiser A & Sali A. *Bioinformatics* 2003 **19**: 2500 [PMID: 14668246]
- [7] Case DA *et al. J Comput Chem.* 2005 **26**: 1668 [PMID: 16200636]
- [8] Morris AL. *Proteins* 1992 **12**: 345 [PMID: 1579569]
- [9] Bowie JU *et al. Science* 1991. **253**: 164 [PMID: 1853201]
- [10] Wiederstein M & Sippl MJ, *Nucleic Acids Res.* 2007 **35**: W407 [PMID: 17517781]
- [11] Castrignanò T *et al. Nucleic Acids Res.* 2006 **34**: D306 [PMID: 16381873]
- [12] <http://www.hyper.com/>
- [13] Glaser F *et al. Bioinformatics* 2003 **19**: 163 [PMID: 12499312]
- [14] Pettersen EF *et al. J Comput Chem.* 2004 **25**: 1605 [PMID: 15264254]
- [15] Ritchie DW *et al. Bioinformatics* 2008 **24**: 1865 [PMID: 18591193]
- [16] <http://www.pymol.org/>
- [17] Williams PA *et al. Science* 2004 **305** : 683 [PMID: 15256616]
- [18] Yano JK *et al. J Biol Chem.* 2004 **279**: 38091 [PMID: 15258162]
- [19] Denisov IG *et al. Chem Rev.* 2005 **105**: 2253 [PMID: 15941214]

Edited by P Kanguane

Citation: Kumar. *Bioinformation* 7(4): 207-210 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Function regions of CYP4F22 predicted using ConSurf server with high amino acid conservation score (9 - conserved, 1 - variable).

| Residue | Conservation score | Buried (B)/Exposed (E) | Residual variety |
|----------------|---------------------------|-------------------------------|-------------------------|
| G149 | 9 | E | G |
| D150 | 5 | E | N, D, E |
| K151 | 9 | E | K |
| W152 | 8 | E | W |
| S153 | 4 | E | S, N, G |
| R154 | 3 | E | H,S,Q,R |
| H155 | 9 | B | H |
| R156 | 9 | B | R |
| R157 | 9 | E | R |
| L158 | 4 | B | M, L |
| L159 | 9 | B | L |
| T160 | 9 | B | T |
| S469 | 9 | E | S |
| A470 | 9 | E | A |
| G471 | 9 | B | G |
| P472 | 9 | E | P |
| R473 | 9 | E | R |
| N474 | 9 | E | N |
| C475 | 9 | B | C |
| I476 | 9 | B | I |
| G477 | 9 | B | G |
| Q478 | 9 | E | Q |
| S479 | 3 | B | A,S,T,N |
| F480 | 9 | B | F |