

# Comparative modeling and molecular docking of orphan human CYP4V2 protein with fatty acid substrates: Insights into substrate specificity

Suresh Kumar

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

Received December 05, 2011; Accepted December 06, 2011; Published December 10, 2011

## Abstract:

Cytochromes P450 (CYPs) are a super family of heme-containing enzymes well-known for their monooxygenase reaction. There are 57 CYP isoenzymes found in human which exhibit specific physiological functions. Thirteen members of this super family are classified as "orphan" CYP because of their unknown enzymatic functions. CYP4V2 is found to be a potential drug target for Bietti crystalline corneoretinal dystrophy (BCD). However, three-dimensional structure, the active site topology and substrate binding modes of CYP4V2 remain unclear. In this study, the three-dimensional model of CYP4V2 was constructed using the homology modeling method. Four possible fatty acid substrates namely, caprylic, lauric, myristic and palmitic acids were optimized and evaluated for drug likeness using Lipinski's rule of five. Further, these substrates were docked into active sites of CYP4V2 and several key residues responsible for substrate binding were identified. These findings will be helpful for the structure-based drug design and detailed characterization of the biological roles of CYP4V2.

**Keywords:** Human Cytochrome, P450 4V2, CYP4V2, homology modeling, molecular docking, fatty acid substrates, ligand binding site, caprylic acid, lauric acid, myristic acid and palmitic acid.

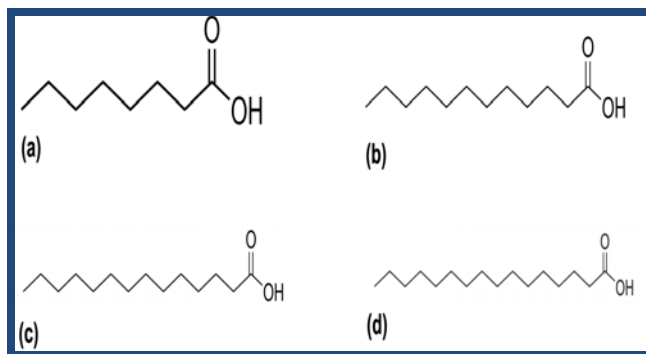
## Background:

Cytochrome P450s (CYPs) are important heme-containing proteins, well-known for their monooxygenase reaction. They are involved in the metabolism of xenobiotics and endogenous compounds, such as steroids and fatty acids. Human CYPs are primarily membrane-associated proteins located either in the inner membrane of mitochondria or in the endoplasmic reticulum of cells. Some CYPs metabolize only one (or a very few) substrate, while others may metabolize multiple substrates. Cytochrome P450 enzymes are present in most tissues of the human body, and play an important role in hormone synthesis and breakdown, cholesterol synthesis, and vitamin D metabolism. Cytochrome P450 enzymes also function to metabolize potentially toxic compounds, including drugs and products of endogenous metabolism [1]. In human there are fifty-seven functional cytochrome P450 isoenzymes with

well-established role in the metabolism of endogenous sterols, bile acids, vitamins and fatty acids. However, thirteen human P450s still remain "orphans" because their functional roles have yet to be elucidated [2]. Human CYP4 family consists of 6 subfamilies, 12 genes and 10 pseudo genes. Human CYP4V2, a relatively new member of human cytochrome P450 enzymes is termed as an orphan P450 because its substrate specificity and physiological roles are unknown. The CYP450 protein family is a group of enzymes that use iron to oxidise various substrates, including potentially harmful substances, thereby making them more water soluble and promoting their biological processing.

CYP4V2 gene encoded as a member of the cytochrome P450 hemethiolate protein super family is involved in oxidizing various substrates in the metabolic pathway. The gene contains 11 exons spanning 21.7 kb and is expressed in various tissues,





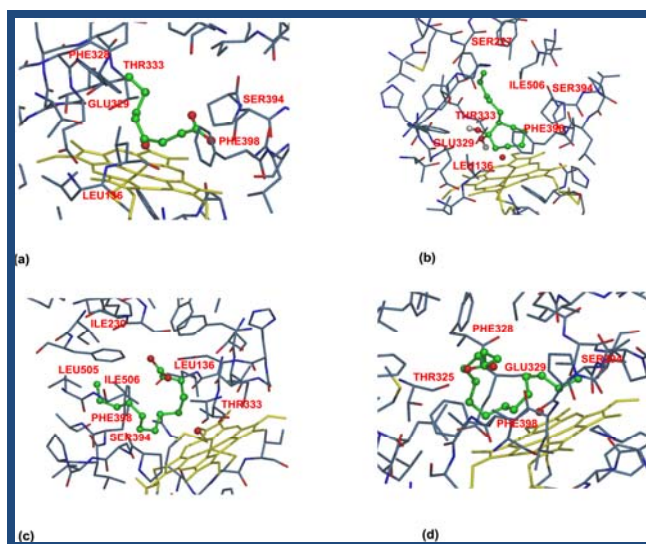
**Figure 4:** 2D structure of ligands; (a) Caprylic acid (b) lauric acid (c) Myristic acid (d) Palmitic acid.

### Prediction of binding site in homology model of CYP4V2

The homology model of CYP4V2 was submitted to ConSurf Server [11] for identification of functional regions by detecting the structurally conserved binding regions among their close sequence homologues.

### Ligand optimization and evaluation of drug likeness

The possible substrates like caprylic, lauric, myristic and palmitic acids were downloaded from the Pubchem in Structure Data Format (SDF). Conversion of SDF to Protein Data Bank (PDB) format was carried out using Open Babel program [12]. The MMFF94 force field was used for energy minimization of ligand molecules. Gasteiger partial charges were added to the ligand atoms, non-polar hydrogen atoms were merged and rotatable bonds were defined. Ligand geometries and electric properties were calculated using MOPAC2009. Further, ligands were evaluated for drug likeness using Lipinski's rule of five [13]. According to this rule, an ideal drug molecule should have a molecular weight of less than 500, total number of hydrogen bonds should not exceed 5, logP value should be less than 5 and the sum of N and O should not be more than 10. Further poor absorption or permeation is more likely when a ligand molecule violates this rule.



**Figure 5:** Close up View of CYP4V2 with (a) Caprylic acid (b) lauric acid (c) Myristic acid (d) Palmitic acid. Key residues are labeled. Substrates are shown in green stick.

### Molecular docking

Docking calculations were carried out using Docking Server [14] to compute the free energy of binding on protein model. Essential hydrogen atoms, Kollman united atom type charges and solvation parameters were added with the aid of AutoDock tools [15]. Affinity (grid) maps of 60X60X60Å grid points and 0.375 Å spacing were generated using the Autogrid program. AutoDock parameter set and distance dependent dielectric functions were used in the calculation of the Van der Waals and the electrostatic terms respectively. Docking simulations were performed using the Lamarckian genetic algorithm (LGA) and the Solis and Wets local search method. Initial position, orientation and torsions of the ligand molecules were set randomly and all rotatable torsions were released during docking. Each docking experiment was derived from 10 different runs that were set to terminate after a maximum of 250000 energy evaluations. The population size was set to 150. During the search, a translational step of 0.2 Å and quaternion and torsion steps of 5 were applied.

### Results and Discussion:

The protein sequence of human CYP4V2 was queried against the Protein Data Bank (PDB) using PSI-BLAST and the result showed less than 30% identity between target and the templates. The top six templates (1TQN, 3CZH, 3E4E, 1NR6, 1PO5 and 3EBS) were selected and their sequences were multiple aligned with CYP4V2 sequence. The majority of the structures were found to be conserved except for few gap inserts. The sequence conservation and the signature motifs of CYP4V2 were examined using multiple sequence alignment with templates (Figure 1). The structural motifs were in the I helix (xGxxT), a tetra peptide in the K helix (ExxR), a dodeca peptide prior to the L helix (ZxxPxxZxPxxZ) and a deca peptide between L helix and the dodeca peptide (FxxGxxxCxG), where, Z could be any aromatic amino acid and x could be any residue. Among these signature motifs, FxxGxxxCxG is the characteristic motif for the CYP super family, which includes a conserved cystine residue that ligates to the Fe of the heme. Several signature motifs were conserved for CYP4V2 as per previous findings [16, 17]. The conservation of sequence indicates that CYP4V2 model construction based on this alignment is reliable. The templates structures did not contain residues in N-terminal membrane-binding domain and thus the first 36 residues at the N-terminal were not included in model construction. The resulting alignment was used as input for Modeller to generate the initial 3D model of CYP4V2. The model of CYP4V2 was subjected to energy minimization using YAMBER force field implemented in YASARA. The total energy was 25283762.4 and -264806.2 kJ/mol for homology model of CYP4V2 before and after minimization. The energy minimized model was subjected to several structural quality assessment methods. The Procheck [18] evaluation showed that 98.5% of residues were in the allowed region and 1.5% in the disallowed region (Figure 2a). Prosa [19] evaluation showed Z score of -6.81 (Figure 2b) indicating that the predicted CYP4V2 model was a reliable one. The model contained several structurally conserved regions quite similar as those of other P450s such as helices D, E, I, J, K and L and the heme group sandwiched between helices (Figure 3a). Comparison of the model with the templates by superposition clearly showed that the overall CYP fold was preserved in CYP4V2 (Figure 3b).



Based on Gothoh's proposal, predicted substrate binding sites (SRS) of the CYP4V2 such as SRS-1, SRS-2, SRS-3, SRS-4, SRS-5 and SRS-6 were observed respectively in B helix, F helix, G helix, I helix, K helix and between  $\beta$ 4 and  $\beta$ 5 (**Figure 3c**). The final model was deposited in Protein Model Database (PMDb) [20] and it is available under ID: PM0077758. The CYP4V2 model was further tested to predict functional regions using ConSurf server (Figure 3d) by identifying high conservational score (**Table 1, see supplementary material**). The optimized and energy minimized substrates namely caprylic (**Figure 4a**), lauric (**Figure 4b**), myristic (**Figure 4c**) and palmitic acids (**Figure 4d**) were subjected to ligand geometry and calculation of electric properties (**Table 2, see supplementary material**). Further, the ligand was also evaluated for drug ability assessment using Lipinski's rule of five. It was observed from the calculated ligand properties that all the substrates did not violate this rule.

To determine the key residues responsible for substrates interaction, the homology model of CYP4V2 was docked with selected fatty acid substrates. The calculated binding energies of CYP4V2 with lauric acid, myristic acid and palmitic acid were respectively -5.99, -5.72 and -6.14 kcal/mol. The lowest energy was found with caprylic acid as -4.30 kcal/mol (**Table 3, see supplementary material**). The negative and low value of free energy of binding indicates a strong favorable bond between CYP4V2 and the caprylic acid in most favorable conformations. Docking of CYP4V2 with caprylic acid revealed that major residues involved were Phe398, Leu136 and Phe329 which formed hydrophobic interaction, while Ser394 formed polar interaction (**Figure 5a**). With lauric acid, major residue Leu136 formed hydrophobic interaction while Phe398 formed both hydrophobic and cation interaction (**Figure 5b**). With myristic acid, the binding residues Ile 506 and Phe328 formed hydrophobic interaction, and Asp332 formed polar interaction (**Figure 5c**). With palmitic acid, Phe398 formed hydrophobic interaction (**Figure 5d**). Major residues involved in interaction with all selected fatty acid substrates were Phe328, Phe398, Leu136 and Ile506 (**Table 4, see supplementary material**).

## Conclusion:

The orphan human protein CYP4V2 was suggested to be associated with Bietti crystalline corneoretinal dystrophy (BCD) and a role in fatty acid and steroid metabolism in the eye. It was suggested to be potential drug target for BCD, but lack of structural information hinders the detailed characterization of its biological roles and it application in the structure based design. For this reason, the homology model of CYP4V2 was constructed using comparative modeling and assessed for geometric and energy aspects. To provide useful information and characterize the CYP4V2 function, four possible saturated fatty acid substrates such as caprylic, lauric, myristic and palmitic acids were docked into active site of CYP4V2 to

explore favourable binding modes and their substrate specificity. The interaction energy calculated between CYP4V2 and substrates revealed that caprylic acid had a lowest energy followed by myristic, lauric and palmitic acids. Several key residues Phe328, Phe398, Leu136 and Ile506 had predominant contribution to the substrate binding with CYP4V2. The key residues identified through ConSurf analysis (functional region identification) and docking studies were in concurrence with each other reflecting the importance of crucial residues involved in substrate binding. The key residues identified from the present study could be potential candidates for site directed mutagenesis studies. Furthermore, these residues might play a role in ligand-channeling and for identifying the reaction center of the protein. The structural information obtained from this study will also be useful for the structure-based drug design of CYP4V2.

## References:

- [1] Strausberg RL *et al. Proc Natl Acad Sci USA*. 2002 **99**: 16899 [PMID:12477932]
- [2] Guengerich FP *et al. Biochem Biophys Res Commun*. 2005 **338**: 465 [PMID: 16126164]
- [3] Li A *et al. Am J Hum Genet*. 2004 **74**: 817 [PMID: 15042513]
- [4] Wada Y *et al. Am J Ophthalmol*. 2005 **139**: 894 [PMID:15860296]
- [5] Lai TY *et al. Invest Ophthalmol Vis Sci*. 2007 **48**: 5212 [PMID: 17962476]
- [6] Wilson DJ *et al. Arch Ophthalmol*.1989 **197**: 213 [PMID: 2783846]
- [7] Nakano M *et al. Drug Metab Dispos*. 2009 **37**: 2119 [PMID: 19661213]
- [8] Altschul SF *et al. J Mol Biol*. 1990 **215**: 403 [PMID: 2231712]
- [9] Sali A *et al. Proteins*. 1995 **23**: 318 [PMID: 8710825]
- [10] Krieger E *et al. Proteins*. 2009 **77**: 114 [PMID: 19768677]
- [11] Glaser F *et al. Bioinformatics*. 2003 **19**: 163 [PMID: 12499312]
- [12] O'Boyle NM *et al. J Cheminform*. 2011 **3**: 33 [PMID: 21982300]
- [13] Lipinski CA *et al. Adv Drug Del Rev*. 2001 **46**: 3 [PMID: 11259830]
- [14] Bikadi Z & Hazai E. *J Cheminform*. 2009 **11**: 1 [PMID: 20150996]
- [15] Morris GM *et al. J Comput Chem*. 2009 **30**: 2785 [PMID: 19399780]
- [16] Lewis DF *et al. Mutat Res*. 1998 **410**: 245 [PMID: 9630657]
- [17] Gajendrarao P *et al. J Mol Graph Model*. 2010 **28**: 524 [PMID: 20079672]
- [18] Morris AL. *Proteins*. 1992 **12**: 345 [PMID: 1579569]
- [19] Wiederstein M & Sippl MJ, *Nucleic Acids Res*. 2007 **35**: W407 [PMID: 17517781]
- [20] Castrignanò T *et al. Nucleic Acids Res*. 2006 **34**: D306 [PMID: 16381873]

Edited by P Kanguane

Citation: Kumar. Bioinformation 7(7): 360-365 (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 CYP4V2 predicted using ConSurf server with high amino acid conservation score (9 - conserved, 1 - variable).

Residue	Conservation score	Buried (B)/Exposed (E)	Residual variety
W130	9	B	W,G
L131	9	B	M,I,L
G132	9	E	S,A,N,R,K,G
L133	5	E	A,F,S,T,N,K,E,Q,C,D,R,I,L
G134	9	B	S,C,G
L135	9	E	M,P,I,L
L136	8	B	F,M,I,L,V
T137	8	E	T,N,R,I,L,V
S138	9	E	F,A,S,T,N,C,D,G,L
L150	8	B	A,F,M,I,L
T151	9	E	A,W,T,D,N,I,G
P152	9	E	S,Q,M,P
T153	8	E	S,A,T,C
F154	9	B	F,I
H155	9	B	H,N,P
F156	9	B	H,S,F,L,Y
R320	9	E	Q,C,I,K,R,L
E321	9	E	A,Q,N,D,K,E
E322	9	E	Q,E,V
V323	9	B	I,V
D324	9	E	H,N,D,E
T325	9	B	T,I,LF
F326	9	B	F,I,V
M327	9	B	T,M,I,L,V
F328	9	B	F,I,L
E329	8	E	A,R,G,E
G330	9	B	A,G,BV
H331	9	E	H,Q,N
D332	9	E	D,E
T333	9	E	S,T
F398	5	B	F,T,Y,V,H,Q,M,I,L
I506	8	B	T,I,L,V
L507	8	B	T,N,V,H,M,I,R,L
R508	8	E	S,M,T,I,K,R,L

**Table 2:** Ligand properties

Ligand	Molar weight	LogP	H acceptor	H donor	pKa	pKb	psa	logD	logK	Total charge
Caprylic acid	143.204	2.44	2	1	5.19	0	37.30	0.63	1.417	0.001
Lauric acid	200.318	3.21	2	2	11.88	4.03	40.46	3.21	-5.2	0.002
Myristic acid	227.363	4.82	2	1	4.95	0	37.30	2.78	-8.031	0.001
Palmitic acid	255.416	5.61	2	1	4.95	0	37.30	3.57	-13.55	0.001

**Table 3:** Docking analysis of CYP4V2 with selected ligands

Ligand	Free Energy of Binding (kcal/mol)	Inhibition Constant, Ki (uM)	vdW + Hbond + desolv Energy (kcal/mol)	Electrostatic Energy (kcal/mol)	Total Intermolec. Energy (kcal/mol)	Interact. Surface
Caprylic acid	-4.30	707.14	-5.85	-0.05	-5.89	353.002
Lauric acid	-5.99	40.85	-7.64	-0.05	-7.69	427.342
Myristic acid	-5.72	64.20	-8.70	0.07	-8.63	480.982
Palmitic acid	-6.14	31.58	-9.60	-0.03	-9.63	534.563

**Table 4:** Details of intermolecular interactions in the binding site of CYP4V2 with ligands

Ligand complex	Hydrophobic*	Polar	Cation	Other
Caprylic acid	C2-LEU136 (3.25), C1-LEU136 (3.13), C7-PHE328 (3.62), C6-PHE398 (3.76)	O2 -SER394 (3.38)	-	C1-GLU329(3.51), C3-GLU329(3.58), C3-THR333(3.63) C5-THR333(3.38) C7-THR333(3.33) C8-SER394(3.89) O1-SER394 (3.61) O2-SER394(3.48) O2-PHE398(3.43)
Lauric acid	C10-LEU136 (3.87) C11-ILE230 (3.51) C11-PHE231(3.87) C5-PHE328(3.68) C7-PHE328(3.84) C9-PHE328(3.48) C11-PHE328(3.77) C6-PHE398(3.59) C8-PHE398(3.82) C7-ILE506(3.69) C9-ILE506(3.89) C11-ILE506(3.50)	-	H1-PHE328(3.35)	O1-LEU136(3.52) C9-SER227(3.36) C11-SER227(3.78) H1-GLU329(3.67) O2-GLU329(3.53) H2-GLU329(3.53) C7-ASP332(3.86) C9-ASP332(3.10) O2-THR333(3.31) H2-THR333(3.36) C4-SER394(3.76)
Myristic Acid	C6-LEU136(3.53) C13-ILE230(3.84) C10-PHE328 C12-PHE329(3.71) C14-PHE328(3.62) C2-PHE398(3.78) C1-PHW398(3.65) C3-PHE398(3.69) C9-PHE398(3.51) C11-LEU505(3.40) C13-LEU505(3.71) C7-ILE506(3.76) C14-ILE506(3.90)	O2-ASP332(3.14)	-	O1-PHE328(3.42) C14-ASP332(3.90) C8-THR333(3.64) C10-THR333(3.87) O2-THR333(3.71) C3-SER394(3.32) C5-SER394(3.26) O2-ILE506(3.80) O1-ILE506(3.77)
Palmitic acid	C2-LEU136 (3.11) C1-LEU136 (3.61) C4-LEU136 (3.64) C12-ILE230 (3.80) C15-PRO393 (3.37) C5-PHE398 (3.45) C2-PHE398 (3.20) C4-PHE398 (3.76) C6-PHE398 (3.30) C8-PHE398 (3.88) C10-PHE398 (3.14) C12-LEU505(3.14) C14-LEU505(3.17)	-	-	C11-THR333(3.33) C13-THR333(3.84) C15-THR(3.63) C7-SER394(3.36) C9-SER394(3.32) C13-SER394(3.32) C15-SER394(3.58) O2-PHE398(3.84)

\*C(X) -R(X) (D):

C(X) - atom name, R(X) - residue name, number within 4 Angstrom from the given ligand atom, (D) - lowest inter atomic distance between the given ligand atom and the residue. For example: C2-LEU136 (3.11), here C2 is C(X), LEU136 is R(X) and 3.11 is (D)