

Phylogenetic Analysis and *In Silico* Characterization of Cytochrome P450 1A (Cyp1A) Protein from the African Catfish, *Clarias gariepinus* (Burchell, 1822)

Muthuramalingam KARPAGAVALLI¹, Muthusamy THANGARAJ^{1*},
Duraismay ANNADURAI¹, Thangappan AJITHKUMAR²,
Mahapathra GYANAPRAKASH¹, Raman SURABI¹

¹Annamalai University, Faculty of Marine Sciences, Centre of Advanced study in Marine Biology, Parangipettai, Tamil Nadu 608 502, India; coralholder@yahoo.com (*corresponding author)

²PMFGR-ICAR-NBFGR Cochin Centre, Ernakulam, Kerala 682 018, India

Abstract

The CYP family enzymes are broadly used as biomarkers because of their pattern of expression. This study describes the application of *in silico* tools to predict the physico-chemical characters of CYP1A protein from the catfish, *Clarias gariepinus*. The nucleotide sequence analysis of *C. gariepinus* CYP1A gene showed higher similarity with *C. batrachus* and reflected in the phylogenetic tree. The comparative modelling results showed this CYP1A protein was highly similar with the 3-D crystal structure of human Cytochrome p450 1A1 (PDB: 1BE3). The prediction results depicted that most of the amino acids formed alpha helix. The predicted pI was 9.10, hydropathicity was -0.226, exposed and buried residues were 61.67, 38.33% respectively. Ramachandran plot analysis showed that most of the amino acids falling on the favoured region and exhibited right-handed alpha helices as the most stable secondary structure. Some amino acids were also found to form loops to interconnect different helices. The CYP1A protein was predicted to be localized in the mitochondrion of the eukaryotic cell.

Keywords: catfish; *Clarias gariepinus*; CYP1A; phylogeny; Ramachandran plot

Introduction

The Cytochrome p450 proteins (CYPs) are a family of haem proteins resulting from expression of a CYP gene super family. In fishes, 137 gene encoding p450s and they are classified into 18 CYP families: CYP1, CYP2, CYP3, CYP4, CYP5, CYP7, CYP8, CYP11, CYP17, CYP19, CYP20, CYP21, CYP24, CYP26, CYP27, CYP39, CYP46 and CYP51. The CYP1 family consists of four subfamilies, CYP1A, CYP1B, CYP1C, and CYP1D. Among them, CYP1A subfamily plays important roles in the metabolism and activation of carcinogenesis and is used as a biomarker to assess the contamination of the aquatic environment (Goksoyr, 1995; Nilsen *et al.*, 1998; Brammell *et al.*, 2010; Jung *et al.*, 2011). Understanding the functional evolution of these genes is essential to predicting and interpreting species differences in sensitivity to toxicity caused by some environmental pollutants (PAH, Dioxins etc.). Elucidating the evolutionary history of the CYP1 family may provide

insight into the origins of the physiological and toxicological functions of CYP1 enzymes.

In silico comparative modelling is a type of methods constructing an atomic resolution model of a protein from its amino acid sequences (Marti-Renom *et al.*, 2000). This method can produce high quality structural models when the target and templates are highly matched and give more information for structural genomics consortium dedicated to all class of experimental structures of proteins (Williamson, 2000). *Clarias gariepinus* is also called as North African Catfish. It is a freshwater fish and has been invaded to other parts of Africa, Europe and Asia. *Clarias gariepinus* is widely tolerant to extreme environmental conditions like highly polluted areas (de Moor and Bruton, 1988). Since the gene (CYP1A) is broadly used as biomarker and no structure was predicted and reported in this species as on date. So this study accomplishes the physico chemical characters and 3D structure of CYP1A protein in *C. gariepinus* using bioinformatics tools.

Materials and Methods

Sample collection

Live *Clarias gariepinus* (n=15) were collected from a commercial fish pond from Sirkazhi, Tamilnadu. The species was identified based on its morphological characters described in previous reports (Jayaram, 1966; Nelson, 1994). Caudal fin clips were dissected out and stored in 95% ethanol for DNA isolation. The voucher specimen were stored in 10% formalin and kept in the laboratory.

DNA isolation and PCR

About 50 mg of tissue was used for DNA isolation as per standard phenol/chloroform method (Sambrook *et al.*, 1989). Quality of the DNA was assessed by 0.8% agarose gel electrophoresis. The quantitative estimation was done by UV spectrophotometer at 260 nm.

CYP1A gene was amplified using the universal primers *hm-F* (5'-CGAGGGTGAGAGTTCTGAGT-3') and *hm-R* (5'-CAGCTTCCTGTCCTCACAGT-3'). The 50 µL PCR reaction volume included 1.5 µL PCR buffer (10×), 1.5 µL dNTPs (2.5 mM each), 2.0 µL MgCl₂ (30 mM), each primer 0.5 µL (10 µM), *Taq* polymerase 0.5 µL (3U/µL) and template DNA 1.0 µL (50 ng/µL). Polymerase chain reactions were performed in the following temperature and timing condition programmed in a thermal cycler (Genei). The thermal regime for CYP1A consists of initial denaturation of 2 min at 95 °C, 35 cycles of 60 sec at 94 °C, 45 sec at 52 °C, 60 sec at 72 °C, followed by final extension of 5 min at 72 °C and hold at 4 °C. The amplified products were checked on 1.5% agarose gel electrophoresis and the molecular weight was assessed using molecular weight marker (100 bp ladder) in a gel documentation system (Genei).

Sequencing and characterization CYP1A gene

The amplified CYP1A gene products (3 samples) were sequenced by Dideoxy method in an automated ABI sequencer in a sequencing facility centre (Eurofins, Bangalore). All the CYP1A sequences of *Clarias gariepinus* were conceptually translated into proteins using vertebrate genomic genetic code translation pattern. NCBI's BankIt protocol was properly followed for submitting the sequences. In addition to those sequences, *Clarias batrachus* (KP893713), *Platycephalus bassensis* (EU107276), *Astronotus ocellatus* (EU553595), *Gambusia affinis* (AB371607) and *Siganus canaliculatus* (KP336485) sequences were retrieved from NCBI and used for phylogenetic analysis. The nucleotides composition, K2P genetic variation, phylogenetic tree using neighbour joining algorithm were analysed in MEGA 4.0 software (Tamura *et al.*, 2011). The phylogeny test was performed by 1000 bootstrap replications and K2P nucleotide substitution model (Kimura, 1980).

Prediction of physical and chemical parameters

All the nucleotides were converted into amino acid sequences by choosing definite reading frame in translate tool (web.expasy.org/translate/). Thus the translated CYP1A protein consists of 180 amino acids. For the

prediction of molecular weight, hydropathicity, theoretical isoelectric point (pI), charges of amino acids, Expasy's ProtParam server was used (<http://web.expasy.org/protparam/>). For Ramachandran plot prediction, Rampage server was used (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). Secondary structure, subcellular localization and transmembrane helices were predicted by predict protein server was used (<https://www.predictprotein.org/>).

Template for target CYP1A gene

To find the best template for high quality structural model, Swiss Model (<https://swissmodel.expasy.org/>) was used and it was validated through PDBSum server (www.ebi.ac.uk/pdbsum). By this comparative modelling, 3-D crystal structure of human Cytochrome p450 1A1 was available at 2.6 Å resolution (PDB: 1BE3) and it was selected as a template structure to generate the 3-D model of the target CYP1A. The crystal 3-D structure of template was searched from protein data bank (<http://www.rcsb.org/pdb/>). Homology modelling was used to generate the 3-D structure of CYP1A with the help of Swiss Model (<https://swissmodel.expasy.org/>) and visualized the 3-D structure in RASMOL (Sayle and Milner-White, 1995).

Results and Discussion

CYP1A gene was successfully amplified using the primers *hm-F* and *hm-R* in the study animal, *C. gariepinus* and the amplicon size was 575 bp. All the sequences got the accession number as MH760775, MH760776, MH760777 in NCBI GenBank. The K2P genetic distance between species is shown in Table 1. The maximum K2P distance (1.268) was found between the *Clarias gariepinus* and *Siganus canaliculatus*, whereas the minimum distance (0.010) was observed between *Clarias gariepinus* and *Clarias batrachus*.

Fig. 1 shows the phylogenetic relationship based on CYP1A gene sequences using Neighbor - Joining Algorithm with distance scale. The tree shows a significant phylogenetic relationship among the studied fish species. Here the catfish species, *Clarias gariepinus* and *Clarias batrachus* was is related therefore, they formed a monophyletic group. The rest of the four species formed a separate clade showing more similarities among themselves.

Hydrophobic amino acids such as Valine, Phenylalanine, Isoleucine, hydrophilic amino acids such as Serine, Glutamic acid, Theronine were found and Tryptophan was absent in the CYP1A protein. Charged polar amino acids were less in number as compared to non-polar amino acids. Most of the amino acids formed alpha helix and rest of them formed loops. The predicted physico chemical properties are given in Table 2. More residues (61.67%) were predicted as exposed residues and made it a solvent accessible protein. Therefore, the net hydropathicity value was negative (-0.226). This protein was predicted to be present in the mitochondria of the eukaryotic cell. Since it is a transmembrane protein, it may localize in the inner membrane of the mitochondrion.

Ramachandran plot analysis of CYP1A showed most of the amino acids felled on favoured region and exhibited

right-handed alpha helices as the most stable secondary structure of the protein. Some amino acids were also found to form loops to interconnect different helices (Fig. 2). The 3-D structure of protein showed most of the amino acids formed alpha helices and some of the amino acids formed loops (Fig. 3). Lack of the complete CYP1A gene amplification, the complete protein structure could not be predicted so beta strands were not seen in the predicted structure.

In the present study, genetic distances were calculated for these species and on that basis phylogenetic tree was constructed in which *Clarias gariepinus* and *Clarias batrachus* were having the least genetic distance of 0.010 and formed monophyletic clade and showed 95% sequence similarity. The rest of the species showed higher genetic distances (1.0) The CYP1A gene of the studied teleost fishes showed more similarity to human CYP1A1 gene. Earlier studies on teleost CYP1A proteins also confirmed this result where several properties, including low spin haem iron, catalytic functions and patterns of induction, particularly the strong induction in extrahepatic organs are same as in human CYP1A1 (Stegeman and Hahn, 1994). Antigenic similarities also suggested a closer relationship between teleost and mammalian CYP1A1 proteins (Stegeman, 1989).

Examination of the amino acid alignment revealed regions with a span of 10 or more amino acids of fish CYP1A proteins that resembled the corresponding domains in the mammalian CYP1A1s (Edwards *et al.*, 1991). In the present study CYP1A gene sequence of *Clarias gariepinus* was found to be 575 bp in length partial at both ends. Pairwise sequence alignment of the target CYP1A protein with the human CYP1A1 protein showed 56.64% amino acid identity. Further analysis of the target protein for the nature of the amino acids showed most of the amino acids were hydrophilic in nature having charged and

uncharged amino acids with the isoelectric point was found to be 9.10. Hydrophobicity of value -0.226 showed CYP1A was hydrophilic in nature and interact with water effectively. Total number of charged amino acids was 30.

CYP1A protein was found to be a transmembrane protein localized in the membrane of endoplasmic reticulum and mitochondria (inner mitochondrial membrane). Human CYP1A1 was also found to be localized in the membrane of the endoplasmic reticulum and the mitochondria. Ramachandran plot analysis of the target CYP1A protein showed most of the amino acids exhibited right-handed alpha helices as the most favourable conformation for the protein and some amino acids formed loops which interconnected the alpha-helices. The secondary structure of the protein consisted of right-handed alpha helices and loops as predicted in the Ramachandran plot analysis. Human CYP1A1 protein also exhibited secondary structure consisted of right-handed alpha helices and loops similar to the *Clarias gariepinus* CYP1A protein.

The CYP1A protein sequence of *Clarias gariepinus* also showed more similarity to the human CYP1A1 protein sequence. In future, the study of the complete protein structure of the CYP1A of the *Clarias gariepinus* can help us to decipher substrate specificity of the enzyme as well as the flexibility of the active site for different ligands. The CYP1A enzyme metabolism of the polycyclic aromatic hydrocarbons (PAH) into carcinogen can also be learned by studying the complete protein structure of the CYP1A protein. Alignment of the complete CYP1A gene sequence with the human CYP1A gene sequence can tell us about the differences in the gene. These differences can lead to different catalytic specificity of the enzymes for different substrates. Differences in the gene sequence can lead to different conformation of the active site of the enzyme and hence differences in the catalytic function of the enzyme.

Table 1. K2P genetic distance between selected fishes based on CYP1A gene sequence

Species	<i>Clarias gariepinus</i>	<i>Clarias batrachus</i>	<i>Platycephalus bassensis</i>	<i>Astronotus ocellatus</i>	<i>Gambusia affinis</i>	<i>Siganus canaliculatus</i>
<i>C. gariepinus</i>	****					
<i>C. batrachus</i>	0.010	****				
<i>P. bassensis</i>	1.215	1.215	****			
<i>A. ocellatus</i>	1.176	1.176	0.195	****		
<i>G. affinis</i>	1.135	1.135	0.221	0.215	****	
<i>S. canaliculatus</i>	1.268	1.268	0.239	0.276	0.268	****
Overall mean	0.733					

Table 2. Physico-chemical properties of CYP1A

Properties	Value
Number of amino acids	180
Molecular weight	20.814 KDa
Isoelectric point	9.10
Number of negatively charged residues	17
Number of positively charged residues	22
Estimated half life	30 hours
Instability index	43.41
Hydrophobicity	-0.226
Exposed residues	61.67 %
Buried residues	38.33 %

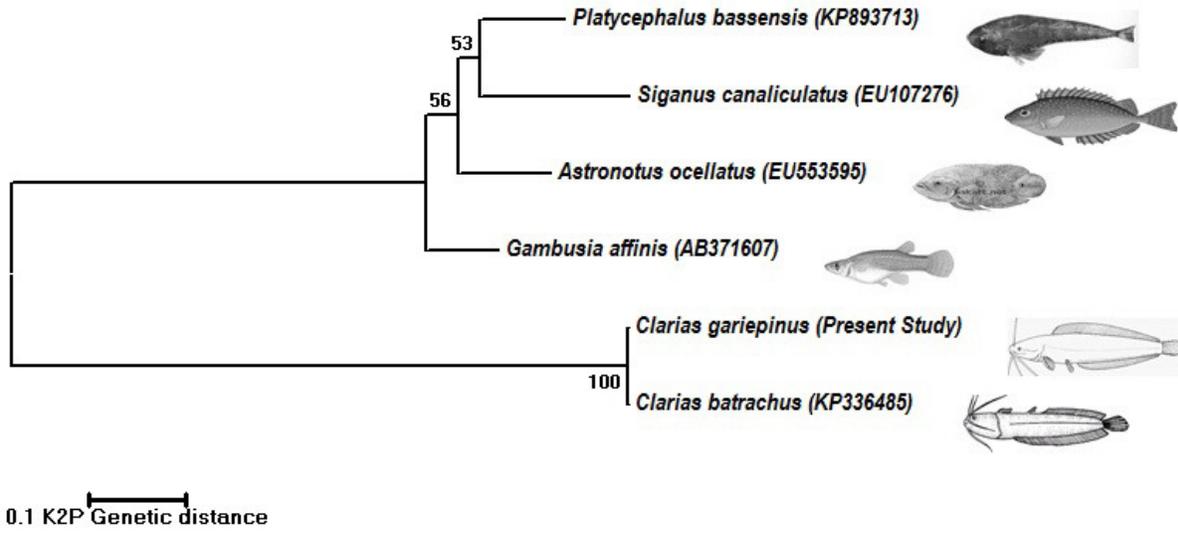


Fig. 1. NJ tree of selected fishes based on CYP1A gene sequence

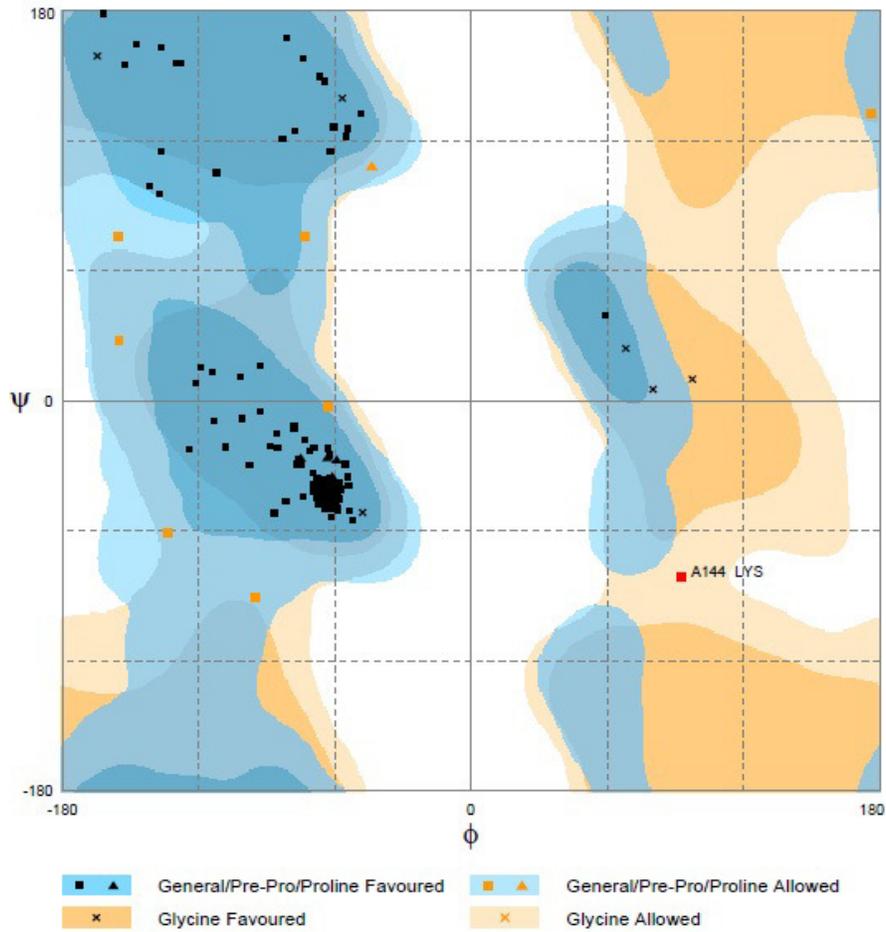


Fig. 2. Ramachandran plot of *Clarias gariepinus* CYP1A protein

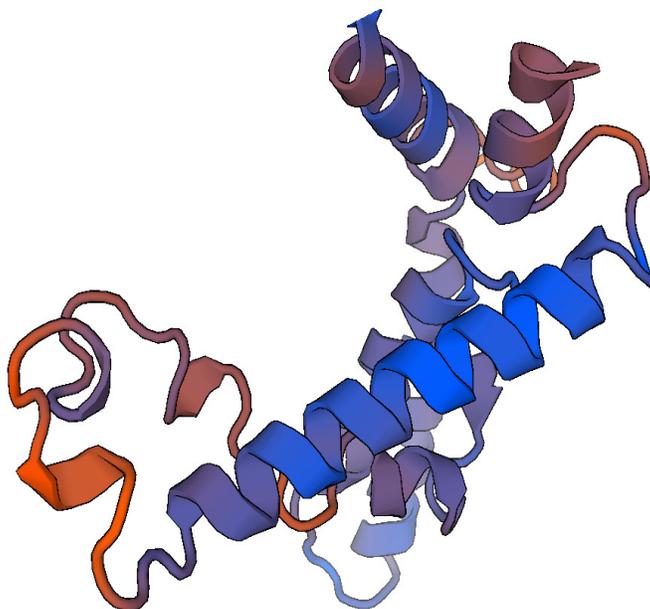


Fig. 3. The 3D structure of *Clarias gariepinus* CYP1A protein

Acknowledgements

The authors gratefully acknowledge the Dean and Director, of the Faculty of Marine Sciences and the authorities of Annamalai University for the facilities provided.

Conflict of Interest

The authors declare that there are no conflicts of interest related to this article.

References

- Brammell BF, McClain JS, Oris JT, Price DJ, Birge WJ, Elskus AA (2010). CYP1A expression in caged rainbow trout discriminates among sites with various degrees of polychlorinated biphenyl contamination. *Archives of Environmental Contamination and Toxicology* 58(3):772-782.
- de Moor IJ, Bruton MN (1988). Atlas of alien and translocated indigenous aquatic animals in southern Africa. National Scientific Programmes Unit: CSIR. Report No. 144. Port Elizabeth, South Africa pp 310.
- Edwards RJ, Singleton AM, Murray BP, Murray S, Boobis AR, Davies DS (1991). Identification of a functionally conserved surface region of rat cytochromes P4501A. *Journal of Biochemistry* 278:249-757.
- Goksoyr A (1995). Use of cytochrome P450 1A (CYP1A) in fish as a biomarker of aquatic pollution. *Archives of Toxicology* 17:80-95.
- Jayaram KC (1966). Contributions to the study of the fishes of the family Bragidae. *Zoological Survey of India* 207:1-136.
- Jung JH, Kim M, Yim UH, Ha SY, An JG, Won JH, ... Shim WJ (2011). Biomarker responses in pelagic and benthic fish over 1 year following the Hebei Spirit oil spill (Taean, Korea). *Marine Pollution Bulletin* 62(8):1859-1866.

- Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.
- Marti-Renom MA, Stuart AS, Fiser A, Sanchez R, Melo F (2000). Comparative protein structure modelling of genes and genome. *Annual Review of Biophysical and Biomolecular Structure* 29:291-325.
- Nelson JS (1994). *Fishes of the world*. 3rd Ed. John Wiley and sons. New York pp 622.
- Nilsen BM, Berg K, Goksoyr A (1998). Induction of cytochrome P450 1A (CYP1A) in fish. A biomarker for environmental pollution. *Methods of Molecular Biology* 107:423-438.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Sayle R, Milner-White EJ (1995). RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences* 20(9):374-376.
- Stegeman JJ (1989). Cytochrome P450 forms in fish: Catalytic, immunological and sequence similarities. *Xenobiotica* 19(10):1093-1110.
- Stegeman JJ, Hahn ME (1994). Biochemistry and molecular biology of monooxygenases- current directions in forms, functions, and regulation of cytochrome P450 in aquatic species. In: Malins DC, Ostrander GK (Eds). *Aquatic toxicology cellular molecular and biochemical perspectives*. CRC Press, Boca Raton, Florida pp 87-206.
- Tamura K, Dudley J, Nei M, Kumar M (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software v. 4.0. *Molecular Biology and Evolution* 24:1596-1599.
- Williamson AR (2000). Creating a structural genomic consortium. *Nature Structural Biology* 7(11):953.