Delineation of the Structural Elements of Oriental Liver Fluke PLA₂ Isoforms for Potent Drug Designing

Gururao Hariprasad · Divya Kota ·
Sundararajan Baskar Singh · Alagiri Srinivasan ·
Souparno Adhikary

Received: 9 April 2013 / Accepted: 24 August 2013 / Published online: 11 September 2013
© Association of Clinical Biochemists of India 2013

Abstract *Clonorchis sinensis* or the Chinese liver fluke is one of the most prevalent parasites affecting a major population in the oriental countries. The parasite lacks lipid generating mechanisms but is exposed to fatty acid rich bile in the liver. A secretory phospholipase A₂, an enzyme that breaks down complex lipids, is important for the growth of the parasite. The enzyme is also implicated in the pathogenesis leading up to the hepatic fibrosis and its complications including cancer. The five isoforms of this particular enzyme from the parasite therefore qualify as potential drug targets. In this study, a detailed structural and ligand binding analysis of the isoforms has been done by modeling. The overall three dimensional structures of the isoforms are well conserved with three helices and a β -wing stabilized by four disulfide bonds. There are characteristic differences at the calcium binding loop, hydrophobic channel and the C-terminal domain that can potentially be exploited for drug binding. But the most significant feature pertains to the catalytic site where the isoforms exhibit three variations of either a histidine-aspartate-tyrosine or histidine-glutamate-tyrosine or histidine-aspartate-phenylalanine. Molecular docking studies show that isoform specific residues and their conformations in the substrate binding hydrophobic channel make unique interactions with certain inhibitor molecules resulting in a perfect tight fit. The proposed ligand molecules have a

predicted affinity in micro-molar to nano-molar range. Interestingly, few of the ligand binding interaction patterns is in accordance to the phylogenetic studies to thereby establish the usefulness of evolutionary mechanisms in aiding ligand design. The molecular diversity of the parasitic PLA₂ described in this study provides a platform for personalized medicine in the therapeutics of clonorchiasis.

Keywords *Clonorchis sinensis* · Structural analysis · Isoform specific potent drug designing · Phospholipase A₂

Introduction

Clonorchiasis, caused by the parasite *Clonorchis sinensis*, a liver fluke, is one of the major parasitic zoonoses in the oriental population affecting about 35 million people globally [1, 2]. The parasite causes a series of pathological, manifestations in the liver ranging from inflammation, fibrosis and cholangiocarcinoma. Life cycle of the parasite begins with human infection following ingestion of fish that harbor encysted younger forms of parasite called as metacercariae. These metacercariae are then released in the small intestine and migrate up the bile duct where they will develop into adult flukes over the course of several weeks. There has been a lot of interest to understand molecular mechanisms of the parasite. In a very recent development this year, the entire genomic assembly and transcriptome analysis was done to understand more about the migration, parasitism and pathogenic behavior of *C. sinensis* at the molecular level [3]. Some of the important results of the study were the fact that genes encoding the proteins for the facilitation and metabolism of lipid were highly expressed in the parasite. One of the enzymes identified was a

Electronic supplementary material The online version of this article (doi:10.1007/s12291-013-0377-1) contains supplementary material, which is available to authorized users.

G. Hariprasad (✉) · D. Kota · S. Baskar Singh · A. Srinivasan ·
S. Adhikary
Department of Biophysics, All India Institute of Medical Sciences,
Ansari Nagar, New Delhi 110029, India
e-mail: g.hariprasad@rediffmail.com

secretory phospholipase A₂ (CsIIIPLA₂) which breaks down phospholipids into lysophospholipids and fatty acids [4]. The importance of the enzyme lies in the fact that the metacercaria that harbor in the hepatic tissue feed on the bile that is rich in lipids. Interestingly, the genome of the liver fluke revealed that there are genes missing for the production of its own fatty acids. This makes the bile as the only source of lipid for its growth [5]. The CsIIIPLA₂ is therefore an important energy source enzyme of the adult parasite. Also, the enzyme has been shown to be crucial in the development of parasite mediated hepatic fibrosis [6, 7]. It therefore makes this enzyme an attractive drug target. There are a total of five isoforms of the enzyme that have been isolated so far. The amino acid sequences of these isoforms are available at the protein data bank. This makes it feasible to delineate the structural details that determine enzyme functionality, substrate specificity and drug binding. The rationale of structure based drug designing is an ideal platform to develop potent therapeutics for clonorchiasis [8].

We have in this study, delineated the structural elements of each of the five CsIIIPLA₂ enzyme isoforms that are necessary for drug binding and also propose isoform specific potent drug molecule candidates that can be taken up for clinical use.

Methodology

Sequence Analysis

Protein sequences of *C. sinensis* secretory PLA₂ (CsIIIPLA₂) isoforms were taken from National Centre for Biotechnology Information. The sequence accession numbers of the five enzyme isoforms are G7YK86, G7YK85, G7Y6W9, C9V3K9 and G7YK83. The sequences were aligned using ClustalW [9] available at ExPASy.

Modeling Studies and Validation

Homology modeling of CsPLA₂ isoforms were done based on the crystal structure of bee venom PLA₂ (PDB ID: 1POC) using SWISS-MODEL [10] software. The coordinates obtained were visualized using program PyMOL [11]. Analysis of conformational correctness and reliability was carried out using PROCHECK [12]. Calcium coordinate was taken from the crystal structure complex of 1POC and fitted into the validated model. Hydrogen atoms were added and the model was minimized in the presence of calcium in Molegro Virtual Docker (MVD) [13]. These models with the calcium ion were taken for structural analysis and docking studies.

Molecular Docking

Eight small molecule PLA₂ inhibitors: aspirin (2-(acetyloxy)benzoic acid), diclofenac(2-[2,6-dichlorophenyl]amino]benzeneacetic acid), indomethacin(2-[1-(4-chlorobenzoyl)-5-methoxy-2-methylindol-3-yl]acetic acid), atropine {(1R,5S)-8-methyl-8-azabicyclo[3.2.1]oct-3-yl(2r)-3-hydroxy-2-phenylpropanoate}, anisic acid(4-methoxy benzoic acid), indole 2-carbamoylmethyl-5-propyl-octahydro-indol-7-yl)-acetic acid), atenolol (2-(4-(2-hydroxy-3-(isopropylamino)propoxy)phenyl)ethanamide) and LY311727 (3-[3-(2-amino-2-oxoethyl)-1-benzyl-2-ethylindol-5-yl]oxypropyl phosphonic acid) were chosen for binding studies with CsIIIPLA₂ isoforms. The ligands were docked to each of the five CsIIIPLA₂ isoforms using Molegro Virtual Docker (MVD) [13].

Ligand and Receptor Preparation

The coordinates for ligands aspirin, diclofenac, indomethacin, atropine, anisic acid, indole, atenolol and LY311727 were obtained from their respective crystal structure complexes with Protein Data Bank Id: 1OXR, 2B17, 3H1X, 1TH6, 2QUE, 1OXL, 2OTF and 1DCY. These molecules and the validated CsIIIPLA₂ isoform structures were prepared by assigning any missing bonds, bond orders, hybridization and charges.

Ligand Binding Site Prediction

The possible ligand binding cavity was predicted by MVD by using the following settings: minimum cavity volume: 10 Å, probe size: 1.40 Å, maximum number of ray checks: 16, minimum number of ray hits: 12 and grid resolution: 0.80 Å. Based on these parameters the best cavity was chosen for docking with the ligands.

Molecular Docking

The docking calculations were done using MolDock SE (simplex evolution) with MolDock Score [GRID] scoring function.

Search Algorithm by MolDock Simplex Evolution

The simplex evolution parameters were set with 1500 maximum iterations, population size of 50 at 300 steps with neighbor distance factor of 1. The energy threshold value was set below 100 with 10 minimum torsions/translations/rotations, 10 maximum positions and 10 quick try values.

Pose Clustering and Generation

Pose Clustering was used to reduce the number of poses found during the docking run and identify the best pose. After docking, energy minimization was performed by a short *Nelder–Mead Simplex* algorithm for minimization of the translation, orientation and flexible dihedral angles of the found poses. Based on a RMSD threshold of 1.0, five poses were generated after 10 runs following the one with the lowest energy was chosen as the best pose.

Scoring Function by MolDock Score (GRID)

In the MolDock Score [GRID] a grid resolution of 0.3 Å was set to initiate the docking process. The ignore-distant-atoms option was used to ignore atoms far away from the binding site. The docking scoring function MolDock Score used was an adaptation of the Differential Evolution (DE) algorithm [13]. The MolDock score is based on a Piecewise-Linear Potential (PLP) introduced by Gehlhaar et al. [14] and further extended in Generic Evolutionary Method for molecular DOCKing (GEMDOCK) [15]. In MolDock, the docking scoring function is extended with a new term, taking hydrogen bond directionality into account. The MolDock score energy (E_{score}) [13] is the summation of the internal energy of the ligand (E_{intra}) and the ligand–protein interaction energy (E_{inter}). The E_{inter} comprises of the PLP which is an approximation of the potential for hydrogen bonds, steric interactions between atoms and electrostatic interactions between charged atoms.

$$E_{\text{score}} = E_{\text{inter}} + E_{\text{intra}}$$

However, in the MolDock Score [Grid] the hydrogen bond directionality is not taken into account. The grid-based scoring function facilitates speed-up by precalculating potential-energy values on an evenly spaced cubic grid. Therefore, the hydrogen bonding is now determined solely on distance and hydrogen bonding capabilities. The energy potential is evaluated by using trilinear interpolation between relevant grid points.

Prediction of Binding Affinities

The explicit hydrogens were created and their hydrogen bonding patterns were determined. The candidates with the best conformational and energetic results were taken as the best fit. Interaction energies between the ligands and the proteins were estimated as the MolDock score. The binding affinities were calculated based on the MolDock Score using the binding affinity module of MVD. Binding affinity (K_i) is calculated by: $E_{\text{binding}} = -5.68 \text{ pKi}$.

Molecular Dynamics

The binding of the small molecules in the gas phase was analyzed by molecular dynamics (MD) simulation in Accelrys Discovery Studio (DS 2.5). The minimization of the protein–ligand complex was carried out in ABNR algorithm which incorporates the initial ‘Steepest descent’ minimization steps with convergent criteria of $<0.1 \text{ kcal/mol/Å}$ r.m.s gradient for 1,000 steps. The production run was carried out for another 5 ns for canonical ensemble (NVT) using Leapfrog Verlet dynamics integrator. In both the stages (minimization and production) of simulation, a time step of 1 fs was used and only the backbone of the protein was kept rigid.

Validation of Docking Methodology

The docking methodology was validated on a crystal structure complex of human group IIA PLA_2 with LY311727, one of the ligands used in the binding analysis in this study. The molecule, LY311727 was extracted from the complex (PDB Id: 1DCY), re-built, energy optimized and the new coordinates were used as ligand input for docking to the same protein. The conformation, total interaction energies and the quality of interactions were determined for the docked complex and compared with that of the crystal structure complex.

Phylogenetic Analysis

The aligned sequence of the CsIII PLA_2 isoforms in ClustalW format was taken. The sequences were trimmed to the length of the isoform with the shortest primary structure. This data was submitted to the software MEGA 5.1 [16] for phylogenetic analysis by maximum likelihood and maximum parsimony methods were carried out. The respective trees were constructed after validation by bootstrapping.

Results and Discussion

Sequence Analysis

Sequences of amino acids corresponding to five CsIII- PLA_2 s enzymes were aligned for sequence analysis (Fig. 1). CsIII PLA_2 isoforms exist as monomers in the native state unlike few other enzymes in the group III PLA_2 family [17, 18]. The lengths of sequences and molecular weights of the isoforms vary from one another. Isoform 3 is the longest with 197 residues and isoform 4 is the shortest with 115 residues. It is difficult to tell from the protein annotations in the data bank whether the lengths of

sequences are primer dictated or not. The sequences will therefore be considered as native forms in this study.

Alignment of the sequences shows conservation of eight cysteines for a possible four disulfide bonds. There is a high homology of 41 % among the sequences from N-terminus up to the conserved phenylalanine at position 96 with no significant homology in C-terminal region. This implies that functional motifs required for the catalytic function are contained within the conserved stretch. Catalytic histidine is conserved in all isoforms confirming histidine dependent hydrolysis of the enzyme. Comparison with bee venom PLA₂ shows variations at positions of two other residues that participate along side histidine in catalysis. Isoform 2 has a glutamic acid instead of aspartic acid at position 64 and isoforms 4 and 5 have phenylalanine instead of tyrosine at position 87. Calcium binding motif is at N-terminal region of the enzymes with only glycine residues conserved at position 11. While the ninth position has either tryptophan, histidine or tyrosine, the thirteenth position has either glycine, threonine or asparagine. Calcium binding motif therefore shows a convergent type of evolution through possible random mutations that have been acquired by the isoforms with the same ancestor with exposure to similar environments within the bile duct. The fourth calcium binding residue, aspartate is conserved along side catalytic His35 to form the classical ‘His-Asp dyad’. Based on structural characterization of bee venom PLA₂ with a substrate binding analogue [17], the substrate binding residues on CsIIPLA₂ isoforms were identified to reveal interesting features. They are: (1) valine at position 83 and phenylalanine at position 67 are the only non-polar residues that are identical; (2) non-polar residues are conserved at the positions 4, 57, 87 and 91 and (3) cavity residues at positions 12, 41, 43, 58, 60, 61, 86 and 93 are not even semi-conserved. This highlights the fact that there

are physio-chemical differences in the channel for possible substrate recognition function and specificity of isoforms. This may partly explain the functional outcomes of inflammation, hepatic fibrosis and cancer caused by infection of the parasite, *C. sinensis* [5–7].

Model Building and Validation

Models of CsIII PLA₂ were built based on crystal structure of bee venom PLA₂ using SWISS MODEL software (Fig. 2). SWISS MODEL has been known to build the best quality structures for this group of enzymes [18]. All five isoform structures were analyzed for their conformational correctness. Ramachandran plot using PROCHECK software along with other structure validating parameters judged the reliability of all models as shown in Table 1. Based on results of Ramachandran plot, the total number of residues in the model and the RMSD of the model from the template main chain, isoform models built by SWISS MODEL were found to be correct. These isoform models were taken for structural analysis and subsequent docking studies. A representative Ramachandran plot for CsIIPLA₂ isoform 1 has been shown in Fig. 3.

Overall Structure

Over all tertiary folds of all five CsIIPLA₂ isoforms essentially comprise of three α -helices, a β -wing, a calcium binding loop at N-terminus (Fig. 2). The entire structure is stabilized by four disulfide bonds between first eight conserved cysteines. There are variations pertaining to the presence and length of C-terminal domain. While the fourth isoform ends at end of the third helix, the first, second and the fifth isoforms extend to different lengths taking up a loop conformation. C-terminal region is the longest for the third

Fig. 1 Sequence homology studies of *Clonorchis sinensis* secretory PLA₂ isoforms. The conserved residues are shown in bold. The calcium binding residues are shown in bright pink, cysteines are shown in yellow, catalytic residues are shown in red and residues

lining the hydrophobic channel are shown in green, cyan, purple, blue and orange in each of the five isoforms. ':', '*' and '.' indicate identical, conserved and semi-conserved residues respectively

Fig. 2 Superimposed ribbon diagrams of the CsPLA₂ isoforms. The five isoforms are shown in green, cyan, purple, blue and orange. The three main helices: H1, H2 and H3, β -wings, calcium binding loop and C-terminal extension are shown. The four conserved disulfide bonds are shown in yellow. The numbering used is for CsPLA₂ isoform 1

Table 1 Validation of CsIIPLA₂ isoform models

Structure validating parameters	CsIIIsPLA ₂ isoforms				
	1	2	3	4	5
Residues in the core and allowed region (%)	99.1	97.4	94.7	96.7	93.5
Residues in the generously allowed region (%)	0.9	1.3	4.3	3.3	5.6
Main chain bond lengths (%)	97.7	100	98.9	99.6	99.5
Main chain bond angles (%)	93.5	97.6	95.3	95.5	94.4
3D–1D > 2 (%)	74.1	74.2	53.2	63.3	58.1
Overall quality factor	59.9	92.9	75.0	86.5	61.7
G-factor for dihedral	−0.2	−0.1	−0.2	−0.1	−0.2
G-factor for covalent	−0.2	0.3	−0.1	0.2	0.1
Overall G factor	−0.2	0.1	−0.1	0.03	−0.07
r.m.s.d of the main chain from the template (Å)	0.06	0.06	0.06	0.06	0.05
Number of disulfide bonds	4	4	4	4	4

isoform and a portion of the length takes up an anti parallel β -sheet conformation which is the second β -wing for this isoform, while rest form an elongated loop. β -Wing with the elongated loop is here designated as C-terminal domain. Enzyme–substrate interaction and their catalysis are structurally independent of the length and conformation of C-terminal domain [17].

Calcium Dependency of the Enzymes

Calcium binding loop is completely conserved in all CsIIPLA₂ isoforms. While three of the calcium binding residues come from the loop at the N-terminal region, oxygen atoms that complete the coordinate cage around calcium arise from the side chain of aspartate in the

Fig. 3 A representative Ramachandran plot for the structure of C5PLA₂ isoform 4. The plot was built using NIH structural analysis and verification software. 96.7 % of the residues in either the core region or allowed region and remaining 3.3 % of the residues in the generously allowed region with no residue in the disallowed region

invariant position next to catalytic histidine (Fig. 4). A five coordinate bonded calcium cage is therefore seen for all isoforms establishing the structural basis for calcium dependent nature of these enzymes. The backbone carbonyl oxygen of glycine at position 11 is present in all isoforms. However, there are variations in the remaining two residues of the loop that participate in calcium binding. Carbonyl oxygen of tryptophan in isoforms 1, histidine in isoforms 2 and 4, and tyrosine in isoforms 3 and 5, at ninth position form coordinate bond with calcium. Likewise, carbonyl oxygen of glycine and threonine, in isoforms 1 and 2, and asparagine at thirteenth position in rest of the three isoforms participate in calcium binding. In addition to coordinate bonds, calcium binding loop is stabilized by a disulfide bond between Cys10 on the loop with Cys32 on the adjacent first helix. Though calcium binding loop forms the superior wall of the hydrophobic channel, the side chains of calcium binding residues are oriented away from the cavity. Therefore the only possible interactions of the loop with substrate would be with C α main chain atoms. Conversely, there exists a high possibility of a similar pattern of interactions from back bone atoms of the loop with competitively binding ligand molecules at the cavity.

Catalytic Site

Phospholipase A₂ enzyme hydrolyses the *sn*-2 ester bond of phospholipids by catalytic histidine and a nucleophile

Fig. 4 Calcium binding loop. The loop is shown in white and pink. The calcium binding residues in the five isoforms are shown in green, cyan, pink, white and orange. The calcium ion is shown as an asterisk in pink

water molecule. The conformational state of histidine is therefore important for hydrolysis of the phospholipids to release arachidonic acid, a precursor for inflammatory eicosanoids. Conformation of catalytic histidine at active site is same in all five isoforms. The participation of this residue makes the residue's pK_a the optimum pH of enzyme [6]. Conformational orientation and stability of histidine has been classically known to be through its interactions with aspartate and tyrosine residues, which are in the vicinity of catalytic site [17, 19]. In some of enzymes there is more than one tyrosine involved in hydrogen bonded interaction [20, 21]. These aspartates and tyrosines arise from non-analogous positions of the enzyme. This pattern of arrangement is seen in C5HPLA₂ isoforms 1 and 3 (Fig. 5a) and is reminiscent of group X PLA₂ [19]. In isoform 2, there is a replacement of aspartate to a glutamate at position 65. Though this is a classical conservative mutation, the oxygen atoms of the carboxyl side chain are oriented away from catalytic histidine and do not form any hydrogen bonds with tyrosine (Fig. 5b). Inter-atomic distance between carboxyl oxygen of aspartate and hydroxyl oxygen of tyrosine is 4 Å. Conformation of glutamate is stabilized by van der Waals interactions with His60, Cys61, Cys62, Tyr86, Phe87 and Lys94. Orientation of glutamate also rules out any possibility of a water mediated hydrogen bond formation. This type of conformational arrangement of residues at the active site is seen in PLA₂ enzymes in venom of scorpion [22]. In isoforms 4 and 5, there is a replacement of hydrophilic tyrosine to a

hydrophobic phenylalanine at position 87 (Fig. 5c). Loss of hydroxyl group in these isoforms results in a void of hydrogen bond with aspartate residue and is similar to the human isoform [23]. The conformation of Phe87 is stabilized by an orthogonal ‘side-face’ hydrophobic interaction with aromatic rings of Phe67 and Phe88. It may be noted that the latter two phenylalanine residues are conserved in all five isoforms. Functionality of *CsIIIPLA₂* enzyme isoforms despite the catalytic site variations are supported by functionality of scorpion and human *PLA₂* enzymes [24, 25]. Analogous residue mutations of D64E and Y52F *PLA₂* enzymes too show activity with a substrate binding role of these residues [26, 27]. These interesting variations at catalytic site offer ample scope for design of physiochemical specific drug molecules towards *CsIIIPLA₂* enzyme isoforms.

Hydrophobic Channel

Hydrophobic channel of the *PLA₂* enzyme is the site of binding of substrate phospholipids. Catalytic site with the catalytic His35 is nested deep in this cavity (Fig. 6). The channel is lined by a number of non-polar residues that can make potential interactions with hydrocarbon tails of lipid [17]. The parasite liver fluke is dependent on lipids in bile as a source of energy. *PLA₂* isoforms are therefore important for the parasite for uptake and breakdown of lipid source for its metabolism. The isoforms however have variations to the quantity and quality of hydrophobicity in the channel. While primary structure gave an idea of the possible residues at the channel, three dimensional analysis helps to define the precise physical orientations and chemical moieties lining the channel. These unique conformations of residues in each of the isoforms not only reiterate the possibility of substrate specificity but also provide information for precise design and binding of drug candidates.

Ligand Binding Studies

Based on binding studies of small molecule inhibitors of group III *PLA₂* enzyme family, a total of eight molecules were chosen for docking analysis. Docking studies show the molecules to be occupying the catalytic site within the substrate binding channel thereby making them competitive inhibitors of the pathogenic enzymes as shown in Fig. 6. Volumes of the cavity that are chosen by default for binding in the five isoforms are 118.8, 131.6, 143.8, 156.2 and 146.4 Å³, respectively. The bulky side chain of Gln43, Val57 and Iso93 is one of the factors for the relatively small binding volume in isoform 1 as compared to others. Details of the interacting residues, hydrogen bonding pattern and predicted binding affinity of the eight ligands to all five isoforms is provided in the Supplementary Material. The best inhibitors were those with reasonable number of interactions, optimum hydrogen bonds and good predicted binding affinity. There is no linear correlation between these parameters. The reasons for this discrepancy are two fold. Firstly, conformational stability of the ligand is not accounted for in the parameters and secondly, the predicted binding is based upon averaged multiple linear regression of a set of 200 ligand–protein model complexes.

Structure Based Rational for Isoform Specific Drug Potency

Most potent ligands against the five isoforms of *CsIIIPLA₂* are atenolol, indomethacin, indole, LY311727 and indomethacin, respectively (Table 2). All these five molecules make interactions with catalytic histidine, have coordinate bonded interactions with calcium ion and have affinities in low micro-molar to nano-molar range. Ligand conformations are such that they occupy a significant space in channel as per the principle of lock and key mechanism for a tight fit. There are interesting binding interactions made

Fig. 5 Catalytic sites of *CsPLA₂* isoforms. Residues from analogous positions are shown in: **a** histidine, aspartate and tyrosine; **b** histidine, glutamate and tyrosine and **c** histidine, aspartate and phenylalanine.

The hydrogen bonded interactions are shown as black dotted line and the distance between the atoms is shown in grey. The distances in Å are shown adjacent to the lines

Fig. 6 A GRASP model of CsPLA₂ isoforms1 with the ligand atenolol. The catalytic site with its constituent residues is shown in a mesh. The ligand, atenolol is shown in white-blue-red rendering. The interacting residues are shown in green-blue-red rendering and the calcium ion is shown as a green asterix. The hydrogen bonded interactions are shown as dotted lines

by these ligands with the channel lining residues at the end of molecular dynamics at the end of 5 ns. The key interactions of these five ligands with their respective isoforms are explained: (A) Atenolol binds at the catalytic site of CsIIIPLA₂ isoform 1 enzyme and makes hydrogen bonded interactions with all three active site residues- His35, Asp65 and Tyr87 (Figs. 6, 7a). The main chain nitrogen of Gly11 makes a hydrogen bond with the oxygen of the ligand. Thr58 makes a polar contact through its main chain with the oxygen atom of the ligand acetamide. The ligand makes coordinate bond with the calcium and an array of hydrophobic interactions with Val83 (minimum interacting distance of 4 Å), Val57 (4 Å) and Leu86 (3 Å). The main chain oxygen atom of Lys12 makes van der Waals contact of 3.9 Å with the ligand. It may be noted that residues Lys12, Val57 and Leu86 are unique to isoform 1. (B) Indomethacin is most potent in terms of occupying the cavity volume as well as having chemical complementarity to isoform 2. Carboxyl oxygen atom at one end of the ligand makes two hydrogen bonds each with the residues side chains of Asp36 (3.4 Å) and main chain oxygen atom of His9 (2.8 Å). The chlorobenzoyl ring and the indole ring of the ligand make a number of hydrophobic interactions with Tyr92. The ligand also makes hydrophobic contacts with Gly57, Pro58, Tyr59, Met60, Tyr87, Tyr92 and Gln94 (Fig. 7b). These residues along with His9 are unique to

isoform 2. (C) The natural molecule indole is the best fit for the isoform 3 enzyme and makes a total of seven hydrogen bonds, saturating all the polar moieties on the ligand thereby stabilizing its binding conformation. Hydrogen bonded interactions are with the residues Gly11, His35, Asp36, Phe60, His61 and Tyr87 (Fig. 7c). The ligand also makes important hydrophobic interactions with Phe60 (3.2 Å) and van der Waals interaction with Tyr9 (3.6 Å). These two aromatic residues are unique to this isoform. (D) The synthetic molecule LY311727 binds to isoform 4 enzyme at catalytic site and the carboxyl oxygens make two hydrogen bonded interactions with Gly11 (3.4 Å) and His35 (2.6 Å). The etheryl oxygen interacts by a hydrogen bond with main chain nitrogen of Asn13 (Fig. 7d). The ligand also makes hydrophobic contacts with the residues Thr58 (3.4 Å), Phe68 (3.9 Å) and Phe87 (3.5 Å) which are unique to this isoform. (E) Indomethacin molecule at the catalytic site of isoform 5 makes five hydrogen bonded interactions. Two of these are from carboxyl oxygen with Gly11 and His35, two are from carbonyl oxygen atom with Thr58 and Asn13 and the last one is between methoxy oxygen atom of the ligand and the side chain amino group of Asn41. The ligand also makes hydrophobic interactions with Tyr43 (3.9 Å) and Val57 (3.8 Å).

In summary, every ligand binds in a unique and a definitive way making certain ligands more isoform

Table 2 Docking results

CsIIPLA ₂	Ligand	Docking parameters			Contacting residues of CsIIPLA ₂ (up to 4.0 Å) in the final docked position (hydrogen bonded residues are highlighted in bold)	Predicted K _i
		MolDock score (kcal/mol)	H bonded score	H bonds		
Isoform 1	atenolol (synthetic)	−121.0	−4.5	6	Trp9, Cys10, Gly11 , Cys32, His35 , Asp36, Asn41, Val57, Thr58, Ile59, Ser60 , Asp65 , Phe68, Leu86, Tyr87 , Ca ⁺⁺	0.2 μM
Isoform 2	Indomethacin (NSAID)	−148.7	−4.7	2	His9 , Cys10, Gly11 , Cys32, His35 , Asp36, Asn55, Gly57, Pro 58, Tyr59, Phe68, Thr86, Tyr87, Tyr91, Gln93, Ca ⁺⁺	4.1 nM
Isoform 3	Indole (natural)	−141.8	−8.6	5	Tyr9, Cys10, Gly11 , Asn13, His35 , Asp36, Cys38, Asn41 , Thr58, Met59, Phe60, His61 , Cys64, Asp65, Tyr87, Ca ⁺⁺	97.2 nM
Isoform 4	LY311727 (synthetic)	−146.0	−4.7	4	Met4, His9, Cys10, Gly11 , Pro12, Cys32, His35 , Asp36 , Ala57, Thr58 , Phe68, Val83, Leu86, Phe87, Tyr91,	0.2 μM
Isoform 5	Indomethacin (NSAID)	−134.4	−6.3	5	Tyr9, Cys10, Gly11 , Pro12, Asn13 , His35 , Asp36, Cys38, Asn41 , Tyr43, Val57, Thr58 , Ser60, Asp65, Phe68, Phe87, Ca ⁺⁺	3.2 nM

Fig. 7 Molecular dynamics trajectory shown as ball and stick representations for the CsIIPLA₂ enzyme **a** isoform 1 with atenolol; **b** isoform 2 with indomethacin, **c** isoform 3 with indole and **d** isoform 4 with LY311727. Snapshots of the molecules and the channel residue conformers extracted from the production dynamics trajectory at the time intervals of 1, 2, 3, 4 and 5 nano-seconds are shown. The

molecules are shown in thick stick rendering and the interacting residues of the channel are shown as thin sticks. Residues making hydrogen bonded interactions are labeled in red and those interacting by hydrophobic or van der Waals interactions are labeled by blue. The calcium ion is labeled in pink

Table 3 Validation of the docking methodology

No.	Structure	Methodology used	Total energy (kcal/mol)			r.m.s.d of docked complex with the crystal structure complex (Å)	Contacting residues of the protein with the ligand (residues making H bonded interactions are shown in bold)
			E _{inter}	E _{intra}	E _{score}		
1	Human group IIA PLA ₂ complex with inhibitor LY311727	Crystallization (PDB ID: 1DCY)	−144.0	−1.3	−145.3	–	Leu2, Phe5, His6, Ala17, Ala18, Tyr21, Gly22, His27, Cys28, Gly29 , Val30, Cys44, His47 , Asp48, Tyr51, Lys62, Ca ⁺⁺
2	Human group IIA PLA ₂ complex with inhibitor LY311727	Docking validation (This study)	−138.5	−2.1	−140.6	0.8	Leu2, Phe5, His6, Ile9, Ala17, Ala18, Tyr21, Gly22, His27, Cys28, Gly29 , Val30, Cys44, His47 , Asp48, Tyr51, Lys62, Phe98, Ca ⁺⁺

Fig. 8 Validation of the docking methodology. Crystal structure and the docked conformations of the ligand LY311727 in the hydrophobic channel of the human group IIA PLA₂. The crystal structure complex is shown in green while the docked structure complex is shown in white

specific. It would therefore be useful to know the parasitic PLA₂ isoform by gene sequencing in patients with liver fluke disease. This would be a meaningful exercise in the direction of pharmacogenomics of clonorchiasis from a clinical perspective.

Fig. 9 Phylogenetic analysis. **a** Phylogeny tree by maximum likelihood method and **b** phylogeny tree by parsimony method

Validation of Docking Methodology

As there is no crystal structure of a group III PLA₂ with a small molecule, the docking methodology was validated by redocking LY311727 to the human group IIA PLA₂ and interaction parameters were compared to crystal structure complex PDB ID: 1ILY. Total interaction energy of docked complex is comparable to that of crystal structure complex (Table 3). The r.m.s.d conformations of the two complexes are less than 1 Å (Fig. 8). Lastly, the docked complex has same residue interactions as its crystal counterpart. Hydrogen bonded interactions of Gly29 and His47 are common to both complexes. Ile9 and Phe98 are the only two additional interacting residues that do not feature in the crystal complex interactions. Predicted binding affinity of 17.3 nM is comparable to experimentally determined inhibitory constant of ligand LY311727 with the same enzyme.

Evolutionary Perspectives for Isoform Specific Drug Potency

Though sequence analysis helps to understand the extent of identity of isoforms with any one of the sequences, phylogeny helps to evaluate inter sequence relationship

amongst each other. Phylogenetic studies were done to evaluate evolutionary proximity between the five CsIII-PLA₂ isoforms. As isoforms were of different lengths, sequences were trimmed to a uniform length of hundred residues. Sequences therefore essentially include calcium binding motif, active site residues and hydrophobic channel residues that play an important role in substrate identification and binding. Both the maximum likelihood method and the Maximum parsimony method show similar pattern of trees (Fig. 9). The results show that isoform 3 and isoform 5 clustered together in a single clade, isoform 1 and isoform 2 are closely related and isoform 4 is all by itself in the two trees. It may be noted that ligand binding studies pertain to the molecules that competitively bind at substrate binding channel with an increased probability of making interactions with the same substrate binding residues. It therefore may be implied that understanding of the evolutionary proximity amongst the CsPLA₂ isoforms have a direct bearing on ligand binding. From ligand binding analysis carried out it is seen that there are consistent binding patterns between: (1) first and second isoforms, such as participation of Gly11 and His35 in hydrogen bonded interactions with indomethacin; (2) third and fifth isoforms, such as participation of Asn13 and Asn41 in interactions with most potent ligands and (3) uniqueness of the isoform 4 with participation of residues Met4, His35, Asp36 and Leu86 in interactions with LY311727, and also absence of a coordinate bond with the calcium. It is likely that studies on a larger number of complexes will bring out statistical significance of correlation between phylogenetic studies and ligand binding fingerprints. Evolutionary relations of proteins therefore have scope to complement structural analysis as a rational for isoform based drug designing as in case of CsIIIPLA₂.

Conclusion

CsIIIPLA₂ isoforms are potential drug targets for infection by parasite *C. sinensis*. All isoforms have a well conserved three dimensional structure but have residue variations at calcium binding site, catalytic site and at substrate binding hydrophobic channel. Detailed structural analyses combined with ligand binding studies help to understand factors that dictate isoform specific drug potency. Some of the drug molecules proposed may either be further developed or considered as a basis for personalized medicine for hepatic fibrosis caused by the parasite.

Acknowledgments This research was supported by a Fast Track project Grant to GH from the Department of Science and Technology, Government of India.

References

1. Lun ZR, Gasser RB, Lai DH, Li AX, Zhu XQ, Yu XB, et al. Clonorchiasis: a key foodborne zoonosis in China. *Lancet Infect Dis*. 2005;5:31–41.
2. Li T, He S, Zhao H, Zhao G, Zhu XQ. Major trends in human parasitic diseases in China. *Trends Parasitol*. 2010;26:264–70.
3. Huang Y, Chen W, Wang X, Liu H, Chen Y, Guo L, et al. The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, re-annotation and analysis of the genome and characterization of tissue transcriptomes. *PLoS One*. 2013;8:e54732.
4. Murakami M, Sato H, Taketomi Y, Yamamoto K. Integrated lipidomics in the secreted phospholipase A₂ biology. *Int J Mol Sci*. 2011;12:1474–95.
5. Wang X, Chen W, Huang Y, Sun J, Men J, Liu H, et al. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. *Genome Biol*. 2011;12:R107.
6. Hu F, Hu X, Ma C, Zhao J, Xu J, Yu X. Molecular characterization of a novel *Clonorchis sinensis* secretory phospholipase A₂ and investigation of its potential contribution to hepatic fibrosis. *Mol Biochem Parasitol*. 2009;167:127–34.
7. Zhang F, Liang P, Chen W, Wang X, Hu Y, Liang C, et al. Stage-specific expression, immunolocalization of *Clonorchis sinensis* lysophospholipase and its potential role in hepatic fibrosis. *Parasitol Res*. 2013;12:737–49.
8. Hariprasad G, Kaur P, Srinivasan A, Singh TP, Kumar M. Structural analysis of secretory phospholipase A₂ from *Clonorchis sinensis*: therapeutic implications for hepatic fibrosis. *J Mol Model*. 2012;18:3139–45.
9. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2002;2(2):3.
10. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics*. 2006;22:195–201.
11. De Lano WL. The PyMOL molecular graphics system. San Carlos: DeLano Scientific; 2002.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26:283–91.
13. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*. 2006;49:3315–21.
14. Gehlhaar DK, Verkhivker G, Rejto PA, Fogel DB, Fogel LJ, Freer ST. Docking conformationally flexible small molecules into a protein binding site through evolutionary programming. *Proceedings of the fourth international conference on evolutionary programming*. 1995;615–627.
15. Yang JM, Chen CC. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins*. 2004;55:288–304.
16. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
17. Scott DL, Otwinowski Z, Gelb MH, Sigler PB. Crystal structure of bee-venom phospholipase A₂ in a complex with a transition-state analogue. *Science*. 1990;250:1563–6.
18. Hariprasad G, Baskar S, Das U, Ethayathulla AS, Kaur P, Singh TP, et al. Cloning, sequence analysis and homology modeling of a novel phospholipase A₂ from *Heterometrus fulvipes* (Indian black scorpion). *DNA Seq*. 2007;18:242–6.
19. Pan YH, Yu BZ, Singer AG, Ghomashchi F, Lameau G, Gelb MH, et al. Crystal structure of human group $\times$ secreted phospholipase A₂ electrostatically neutral interfacial surface targets zwitterionic membrane. *J Biol Chem*. 2002;277:29086–93.

20. Jabeen T, Singh N, Singh RK, Ethayathulla AS, Sharma S, Srinivasan A, et al. Crystal structure of a novel phospholipase A2 from *Naja naja sagittifera* with a strong anti-coagulant activity. *Toxicon*. 2005;46:865–75.
21. Jasti J, Paramasivam M, Srinivasan A, Singh TP. Structure of an acidic phospholipase A2 from Indian saw-scaled viper (*Echis carinatus*) at 2.6 Å resolution reveals a novel intermolecular interaction. *Acta Crystallogr D*. 2004;60:66–72.
22. Hariprasad G, Kumar M, Srinivasan A, Kaur P, Singh TP, Jithesh O. Group III phospholipase A₂ from the scorpion, *Mesobuthus tamulus*: targeting and reversible inhibition by native peptides. *Int J Biol Macromol*. 2011;48:423–31.
23. Hariprasad G, Kumar M, Kaur P, Singh TP, Kumar RP. Human group III PLA₂ as a drug target: structural analysis and inhibitor binding studies. *Int J Biol Macromol*. 2010;47:496–501.
24. Hariprasad G, Saravanan K, Baskar S, Das U, Sharma S, Kaur P, et al. Group III PLA2 from the scorpion, *Mesobuthus tamulus*: cloning and recombinant expression in *E. coli*. *Electron J Biotechnol*. 2009;12:3.
25. Valentin E, Ghomashchi F, Gelb MH, Lazdunski M, Lambeau G. Novel human secreted phospholipase A(2) with homology to the group III bee venom enzyme. *J Biol Chem*. 2000;275:7492–6.
26. Nicolas JP, Lin Y, Lambeau G, Ghomashchi F, Lazdunski M, Gelb MH. Localization of structural elements of bee venom phospholipase A2 involved in N-type receptor binding and neurotoxicity. *J Biol Chem*. 1997;272:7173–81.
27. Dupureur CM, Yu BZ, Jain MK, Noel JP, Deng T, Li Y, et al. Phospholipase A2 engineering Structural and functional roles of highly conserved active site residues tyrosine-52 and tyrosine-73. *Biochemistry*. 1992;31:6402–13.