J Mol Model (2013) 19:3779–3791
DOI 10.1007/s00894-013-1913-x

ORIGINAL PAPER

Structural and phylogenetic basis for the classification of group III phospholipase A_2

Gururao Hariprasad · Alagiri Srinivasan · Reema Singh

Received: 10 April 2013 / Accepted: 6 June 2013 / Published online: 23 June 2013

Abstract Secretory phospholipase A_2 (PLA_2) catalyses the hydrolysis of the *sn-2* position of glycerophospholipids to liberate arachidonic acid, a precursor of eicosanoids, that are known mediators of inflammation. The group III PLA_2 enzymes are present in a wide array of organisms across many species with completely different functions. A detailed understanding of the structure and evolutionary proximity amongst the enzymes was carried out for a meaningful classification of this group. Fifty protein sequences from different species of the group were considered for a detailed sequence, structural and phylogenetic studies. In addition to the conservation of calcium binding motif and the catalytic histidine, the sequences exhibit specific 'amino acid signatures'. Structural analysis reveals that these enzymes have a conserved globular structure with species specific variations seen at the active site, calcium binding loop, hydrophobic channel, the C-terminal domain and the quaternary conformational state. Character and distance based phylogenetic analysis of these sequences are in accordance with the structural features. The outcomes of the structural and phylogenetic analysis lays a convincing platform for the classification the group III PLA_2s into (1A) venomous insects; (IB) non-venomous insects; (II) mammals; (IIIA) gila monsters; (IIIB) reptiles, amphibians, fishes, sea anemones and liver fluke, and (IV) scorpions. This classification also helps to understand structure-function relationship, enzyme-substrate specificity and designing of potent inhibitors against the drug target isoforms.

Keywords Amino acid signature · Classification · Group III phospholipase A_2 · Phylogeny · Structure

G. Hariprasad (✉) · A. Srinivasan
Department of Biophysics, All India Institute of Medical Sciences, Ansari nagar, New Delhi 110029, India
e-mail: g.hariprasad@rediffmail.com

R. Singh
Bioinformatics Centre, Indian Council of Medical Research, Ansari nagar, New Delhi 110029, India

Introduction

Phospholipases A_2 (PLA_2, EC 3.1.1.4) are upstream regulators of many inflammatory processes which hydrolyze phospholipids at the *sn*-2 position of the glycerol backbone, releasing lysophospholipids and fatty acids [1]. The enzymes are diverse and exist in various oraganisms from different species. The very first classification for 32 enzymes was based on the sequence, evolutionary origin and pharmacological implications [2]. Subsequently, in the beginning of this century, the entire PLA_2 family of enzymes was classified into 11 groups [3]. This particular classification was based on the following criteria: (i) catalytic function, (ii) knowledge of the amino acid composition, (iii) enzymes with identifiable sequence homology, (iv) grouping of catalytically active splice variants of the same gene and (v) assignment of subgroup letter to a paralog if more than one homologous PLA_2 gene existed within the same species. In a recent update, the PLA_2 superfamily classification consisted of 15 groups and many subgroups that includes five distinct types of enzymes, namely the secreted PLA_2s, the cytosolic PLA_2s, the calcium independent PLA_2s, the platelet-activating factor acetylhydrolases and the lysosomal PLA_2s [4]. With ever increasing number of PLA_2 enzymes that are discovered it has become important to include and update the classification. Accordingly, newer PLA_2s have been characterized and annotated under different groups and subgroups [5, 6]. The criteria for the classification too has widened with certain groups being sub-grouped based on the quality of the catalytic residues and structural determinants [7, 8].

Group III PLA_2s have been characterized from different sources like insects [9], parasites [10], scorpions [11–13], reptiles [14] and mammals including man [15]. Imperatoxin I, phaiodactylipin and phospholipin from the scorpions in South America have been characterized and because of their sequence homology with the bee venom PLA_2, a prototype enzyme in this group, they are studied as group III enzyme [16–18]. However, the reason for the varying nomenclatures for similar proteins is not exactly known. The group III bee venom PLA_2 is the only representative crystal structure that

is available for this group of enzymes and comprises a compact globular structure with a hydrophobic channel and active site with a catalytic histidine [9]. These enzymes are known to be calcium dependent [9–15]. Interestingly, up till now, there has been no attempt made to classify the group III PLA_2 despite the high number of available sequences from across the living organisms including man. It is obvious that though the functional outcome of the enzyme is initiating the inflammatory pathway, there are functional variations in terms of expression, regulation, duration and the purpose of the inflammation. This therefore necessitates a detailed comparative structural and phylogenetic analysis of the diverse enzymes under a single group to have a scientific rationale for the classification of the group III PLA_2. The classification will help to comprehend the diverse biophysical properties of the group, differentiate drug targets from the house keeping isoforms, relate the mediation of toxicity by enzymes in the venom and understand the evolutionary perspective of group constituents.

Materials and methods

Sequence analysis

Homology search for the group III PLA_2 was carried out using BLAST search tool at ExPASy proteomics server and 50 of the representative sequences from across different species were aligned using ClustalW [19]. Sequences chosen were based on a minimum of 40 % homology with bee venom PLA_2 which is a prototype representative of this group. The number of sequences under each phylum was proportional to total number of species sequences deposited in the data bank. Sequences of Imperatoxin I, phaiodactylipin and phospholipin from the scorpions were also taken for the analysis and will hence forth be referred to as PLA_2s of the respective scorpions.

Structural analysis

The sequences of the group III PLA_2 enzymes were submitted to the 3D prediction software tools such as Geno 3D [20], SWISS-MODEL [21], Jigsaw3D [22], ESyPred3D [23] and PHYRE [24] softwares. The coordinates of the structure obtained were viewed using the program PyMOL [25]. The analysis of the conformational correctness and reliability was carried out using PROCHECK [26]. The models of the enzymes with most residues in the core and allowed areas in the Ramachandran plot were taken for further studies.

The calcium coordinate was taken from the crystal structure complex of bee venom PLA_2 (PDB ID: 1POC) [9] and fitted into the validated enzyme models. The atomic position of Ca^{++} in the model and the side chain atoms of the model were optimized keeping the backbone of the model rigid. This was done in *Discovery Studio (DS) 2.0* (Accelrys Software Inc., San Diego) modeling environment. The hydrogen atoms were added and all atom CHARMm (version c33b1) forcefield [27, 28] parameterization was assigned to all the atoms of model including Ca^{++}. Then the optimization was performed in three gradual steps in the presence of distance dependent dielectric implicit solvent model. In the first step, it was minimized to remove the bad steric contacts using steepest descent algorithm with the convergence criteria of r.m.s gradient of <0.1 kcal mol^{-1} $Å^{-1}$ followed by minimization using conjugate gradient algorithm with the convergence criteria of r.m.s gradient of <0.05 kcal mol^{-1} $Å^{-1}$. Finally, the minimized model from above steps was further optimized using adopted basis-set Newton–Raphson (ABNR) algorithm with the convergence criteria of r.m.s gradient of <0.01 kcal mol^{-1} $Å^{-1}$. Theses optimized models were then taken for detailed structural analysis.

Phylogenetic analysis

The same 50 group III PLA_2 sequences from UniProt database that were taken for sequence and structural analysis were considered for the phylogenetic analysis. Multiple sequence alignment was performed using *M-coffee* [29] with default parameters. The consensus alignment was based on the result of eight multiple sequence alignment methods. Poorly aligned regions from the alignment were trimmed by using *trimAl v1.4* [30] with *Automated trimming heuristic: automated1* method (optimized for maximum likelihood phylogenetic tree construction) and the *Strict plus* method (optimized for neighbor-joining phylogenetic tree construction). The selected regions from the alignment were same after trimming. The resulting alignment were subjected to neighbor joining (NJ) [31] and maximum likelihood (ML) [32] phylogenetic analysis, performed by *PHYLIP 3.69* [33]. Maximum likelihood tree was calculated using *Jones-Taylor-Thornton* model of amino acid change [34]. The resulting tree was tested by bootstrapping with 1000 replicates [35]. Neighbor joining analyses using 1000 bootstrap replicates on default parameters were also performed to compare the grouping between two methods. The final trees were visualized and edited by tree figure drawing tool *FigTree v1.3.1* [36].

Results and discussion

Sequence analysis

The PLA_2 sequences of the amino acids corresponding to the group III PLA_2 from different spectra of organisms were taken from UniProt for sequence analysis (Fig. 1). While the results of the sequence analysis are provided in this section, some of the discussions pertaining to the structural

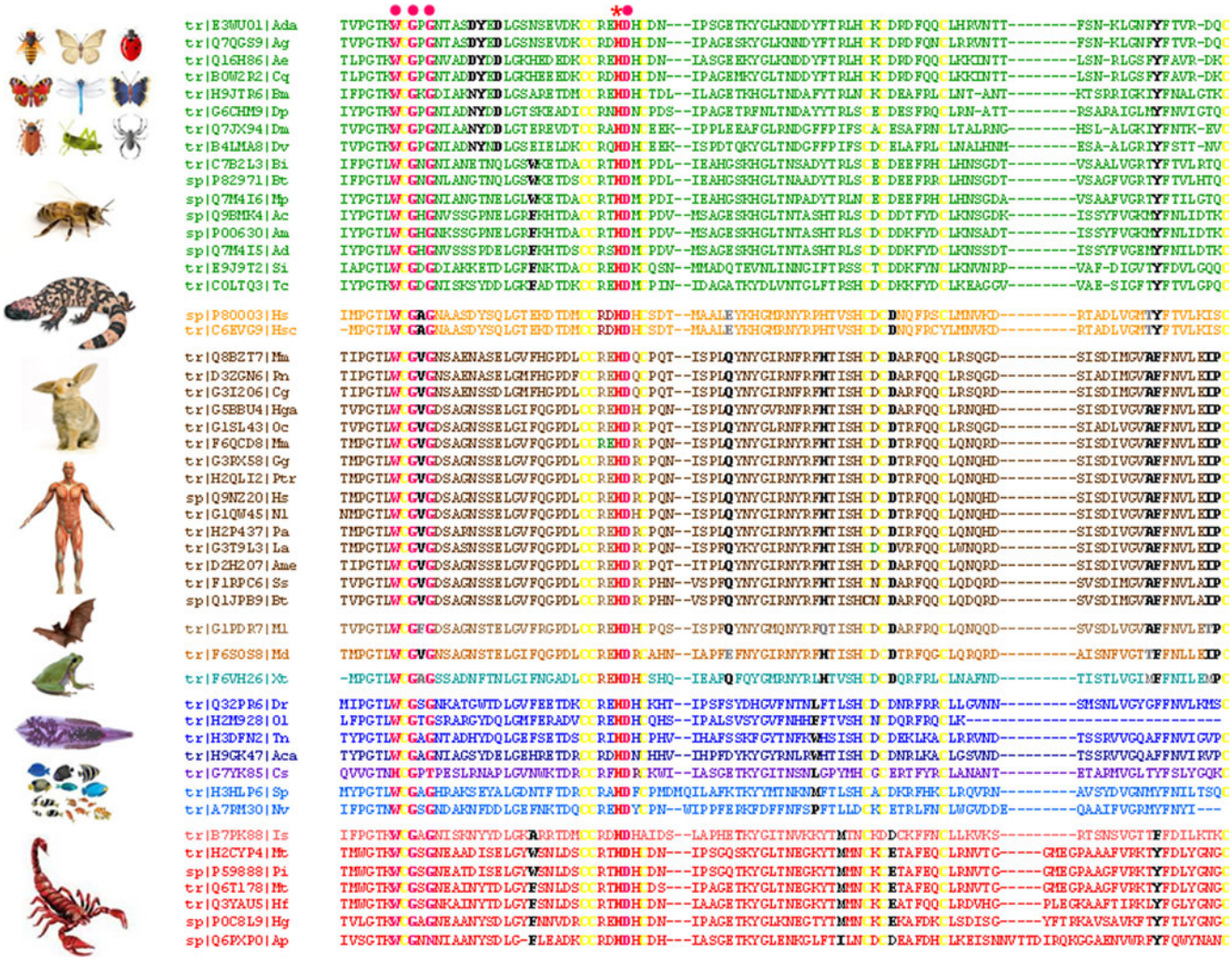

Fig. 1 Sequence homology studies of group III PLA_2 enzymes. The sequences are of Ada, *Anopheles darling*; Ag, *Anopheles gambiae*; Ae, *Aedes aegypti*; Cq, *Culex quinquefasciatus*; Bm, *Bombyx mori*; Dp, *Danaus plexippus*; Dm, *Drosophila melanogaster*; Dv, *Drosophila virilis*; Bi, *Bombus ignites*; Bt, *Bombus terrestris*; Mp, *Megabombus pennsylvanicus*; Ac, *Apis cerena*; Am, *Apis mellifera*; Ad, *Apis dorsata*; Si, *Selonopsis invicta*; Tc, *Tribolium castaneum*; Hs, *Heloderma suspectum*; Hsc, *Heloderma suspectum cinctum*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Cg, *Cricetulus griseus*; Hg, *Heterocephalus glabar*; Oc, *Oryclolagus cuniculus*; Mm, *Macaca mulatta*; Gg, *Gorilla gorilla*; Pt, *Pan troglodytes*; Hs, *Homo sapiens*; Nl, *Nomascus leucogenys*; Pa, *Pongo abeii*; La, *Loxodonta Africana*; Am, *Ailuropoda melanoleuca*; Ss, *Sus scrofa*; Bt, *Bos Taurus*; Ml, *Myotis lucifugus*; Md, *Monodelphins domestica*; Xt, *Xenops tropicalis*; Br, *Brachydanio rerio*; Ol, *Oryzias latipes*; Tn, *Tetradon nigroviridis*; Ac, *Anolis carolimonsis*; Cs, *Clorenchis sinensis*; Sp, *Strongylocentrotus purpuratus*; Nv, *Nematostella vectensis*; Is, *Ixodes scoapularis*; Ap, *Anuroctus phaiodoactylus*; Hg, *Hadrurus* gertschi; Mt, *Mesobuthus tamulus*; Hf, *Heterometrus fulvipes*; Pi, *Pandinus imperator* and Pc, *Pandinus cavimanus.* The conserved catalytic histidine is shown in red, cysteines are shown in yellow, calcium binding residues are shown in pink and the characteristic amino acid signatures in different species is shown in bold black. The UNIPROT accession numbers for the proteins are given preceding the sequences

and phylogenetic perspective are dealt with in the respective sections. All the enzymes in this group are secretory proteins and have a secretion signal peptide. The scorpion PLA_2s after the post-translational processing, form heterodimers by the association of the main enzymatic sub-unit with the non-enzymatic (C-terminal) unit by a disulfide bond and the latter is reversibly inhibited by a native peptide [13]. The rest of the enzymes are monomers with elongated C-terminal extension. The quaternary state conformation of the protein was not considered for the sequence analysis. Only the mature enzymatic unit sequences without the signal peptide and the C-terminal part distal to the last but one cysteine were used for the alignment.

There exists a very high identity of 90 % between sequences of the same species and a very low identity of 24 % between the higher order forms with the lower order ones. There are a total of ten conserved cysteines in the enzymes wherein the first nine are located in the main enzymatic sequence. The last cysteine (not shown) is located in the C-terminal region of the monomeric forms and the non-

enzymatic subunit of the heterodimeric forms of the scorpion enzymes. The catalytic histidine is conserved in all the sequences and is along side the conserved calcium binding aspartate making a 'His-Asp' dyad. The remaining three calcium binding residues are seen at the N-terminal end. In 48 sequences there is a calcium binding motif of WCGXG (X representing any amino acid) that is conserved, where Trp7, Gly9 and Gly11 are the calcium binding residues. However, there are two variations with the replacement of tryptophan by histidine in *Cs*PLA_2 and the replacement of glycine at position 11 by threonine and aspargine in *Cs*PLA_2 and *Ap*PLA_2 respectively.

There are two residues, aspartate and tyrosine that are usually conserved along with the catalytic histidine in the PLA_2 family [2]. However, in the group III PLA_2, while aspartate is conserved in 42 of the sequences, there is a conservative replacement of it to glutamic acid in the scorpions and in the insects-*Dm*PLA_2 *and Dv* PLA_2. On a similar note, while the tyrosine is conserved in 28 of the sequences it is replaced by phenylalanine in all the mammals, marine organisms and *Is*PLA_2.

An interesting feature of the alignment is that there are certain residues conserved in species or sub-species of organisms. They are: (a) a consensus sequence of Asp/Asn-Tyr-Asp/Glu-Asp after the calcium binding loop in the non-venomous insects; (b) an aromatic residue of Trp/Phe at position 23 in the venomous insects and scorpions; (c) hydrophobic stretch of Gly-Val-Ala-Phe at positions 83–86 and a valine in the calcium binding motif between the two glycines in the mammals; (d) a doublet of 'Iso-Pro' before the last cysteine in the mammals including the marsupial, gray short tailed opossum and (e) hydrophobic residue at position 54 in lower order organisms like fishes, sea anemones, liver fluke and chameleon. The *Ml*PLA_2 (bat), a flying mammal and *Md*PLA_2 (gray short tailed opossum), a marsupial have certain resemblance to the mammalian group III PLA_2 sequences. They are: (a) *Ml*PLA_2 has a Gln44, Ala-Phe doublet at 83–84 and a Pro90 and (b) *Md*PLA_2 has a Val10, His55 and an Iso-Pro doublet at 89–90 positions. These species specific 'amino acid signatures' are substrate dictated and are the evolutionary adaptations of the enzyme that best suit the essentials of the organisms in various environmental habitats [37–40].

Model building and validation

The models of the group III PLA_2s were built based on the only available crystal structure of bee venom PLA_2. The model of *Ae*PLA_2 whose structure is not known has been shown and explained as a demonstrative example (Fig. 2). The Ramachandran plot using PROCHECK software [26] judged the reliability of all the models obtained from different softwares. Ramachandran plot for *Ae*PLA_2 structure of 136 residues (Fig. 3) shows a total of 95.0 % of the residues in the core and allowed regions, 4.2 % in the generously allowed region and just 0.8 % in the disallowed region of the plot. The G factor is −0.15 for the dihedrals and −0.09 for the covalent bonds with an overall G factor of −0.10. A total of 97.9 % of the main chain bond lengths and 94.6 % of the main chain bond angles are within limits. Validation of the model on the *Verify_3D* shows a total of 72 % of the residues to be having an averaged 3D-1D score of >0.2. The overall quality factor at *Errat* is 61.4. The RMSD for the main chain between the *Ae*PLA_2 and the template bee venom PLA_2 is 0.4 Å. Based on these validation results and the stability provided by the five disulfide bonds, the models obtained by SWISS MODEL were considered reasonably good to be taken up for further structural analysis. In the *Ae*PLA_2 there is one residue Thr75 that is seen just outside the allowed confines of the Ramachandran plot. This residue is present in the loop connecting the second helix to the third helix. The main chain oxygen of the residue makes a hydrogen bond (3.2 Å) with the main chain nitrogen of Arg79 on the third helix thereby stabilizing its conformation. As this residue is a part of the flexible loop region and not integral to the functionality of the enzyme, the disallowed conformations of its main chain dihedral angles have be overlooked.

Structural analysis of group III PLA_2

The over all structure of a monomeric group III PLA_2 essentially comprises three α-helices, two β-wings, a calcium binding loop at the N-terminus and an extended C-terminus (Fig. 2). The calcium binging loop extends from Lys6-Asn12 with the calcium ion at its center. The first helix extends from Glu24-His35, second helix extends from Cys59-Lys71, third helix extends from Leu76-Val87, the two β-sheets in the first wing extend from Gly42-Lys45 and Leu48-Asn50 and the two β-sheets in the second wing at the C-terminal region extend from Cys125-Gln129 and Gln159-Phe163 respectively. The entire structure is stabilized by five disulfide bonds between the ten cysteines: Cys8-Cys30, Cys29-Cys68, Cys35-Cys61; Cys59-Cys91 and Cys101-Cys113.

The enzyme has a substrate binding hydrophobic channel whose left superior wall is formed by the calcium binding channel, the roof and the right wall is formed by the second and third helices and the floor is formed by the loops that traverse between the first β-wing and the respective helices (Fig. 4). The catalytic histidine is seen nested at the deep end of the channel. Though the entire enzyme takes up a compact globular structure, the C-terminal region is seen to form an additional functional domain away from the main catalytic part of the enzyme. This C-terminal region, distal to the end of third helix, comprises 52 residues that take up an anti-parallel β-strand and a loop connecting the two strands (Fig. 2). Thirty five residues that form the initial part of the first strand and the major part of the loop comprises 11 basic residues and 13 non-polar residues. This constitutes a

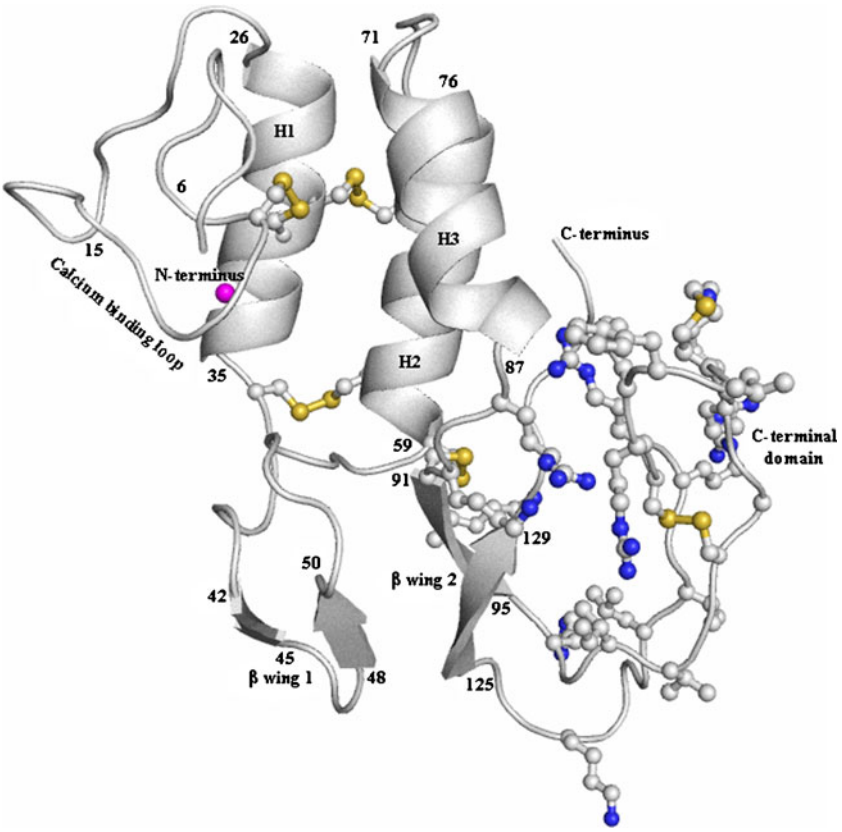

Fig. 2 Ribbon diagram showing the overall structure of human group III PLA_2. The three main helices: H1, H2 and H3, β-wings 1 and 2, calcium binding loop and C-terminal extension are shown. The catalytic histidine and the five disulfide bonds between the ten cysteines are also shown

'cationic-hydrophobic' patch suggestive of a functional role that is independent of the catalytic activity. These types of cationic-hydrophobic stretches enable strong electrostatic-interactions and bilayer penetration of the anionic lipid [41] to induce skeletal muscle necrosis and lyse endothelial cells [42]. The patch is also reminiscent of the myotoxic site on the group I PLA_2 [43]. It may be noted that from a clinical view point, that the group III PLA_2 along with milittin in the insects, inserts into the phospholipids layer of the cell membrane to cause muscle necrosis and disruption of the vascular endothelium [44]. This is in contrast to the non-enzymatic subunit of the scorpions which have an 'acidic-hydrophobic' patch in an anti-parallel β-sheet conformation [13] and the mammalian C-terminal extension which has a hydrophobic patch of residues in an elongated loop conformation [15]. The structural variations between the species could have possibly resulted from adaptation to different prey ecology and substrate exposure.

We have previously described the structure of a heterodimeric variant of group III *Mt*PLA_2 from the scorpions [13]. Briefly, the enzyme transcript from the scorpions codes for three distinct products that include a large enzymatic subunit, a pentameric peptide and a small non-enzymatic subunit. The enzymatic subunit comprises three helices, the calcium binding loop and a substrate binding hydrophobic channel where the structure is stabilized by four disulfide bonds. The non-enzymatic subunit comprises extensive hydrophobic residues with a conformation of an anti-parallel β-sheets making it ideal for tissue specific targeting. The free cysteine on both these units suggests a potential disulfide bond for a quaternary structural association. Also, functional studies on other scorpions PLA_2s have established the existence of the enzyme as heterodimers. Lastly, the native pentapeptide is seen to have a physio-chemical complimenatrity to the substrate binding channel making it a potential reversible inhibitor of the enzyme preventing any non-specific interactions.

Calcium binding site analysis

The structural analysis reveals all the PLA_2 in this group to be calcium dependent. The calcium binding loop is in close proximity to the N-terminus unlike the group I and group II PLA_2s. It extends from Trp7 to Ala14. The main chain carbonyl oxygens from Trp7, Gly 9 and Gly 11 along with the side chain carboxyl oxygens of aspartate that precedes the catalytic histidine form the pentagonal coordinate cage with the calcium ion (Fig. 5a). While this arrangement is consistently present in most of the group III family, there are subtle variations to the calcium binding in some of the enzymes. In the *Ap*PLA_2, there is an aspargine in place of glycine in the 11 position, with a conformation such as to orient the main chain carbonyl oxygen toward the calcium to form a coordinate bond of 2.4 Å (Fig. 5b). In another instance, the parasite, *Cs*PLA_2 has a histidine replacing the bulky tryptophan at

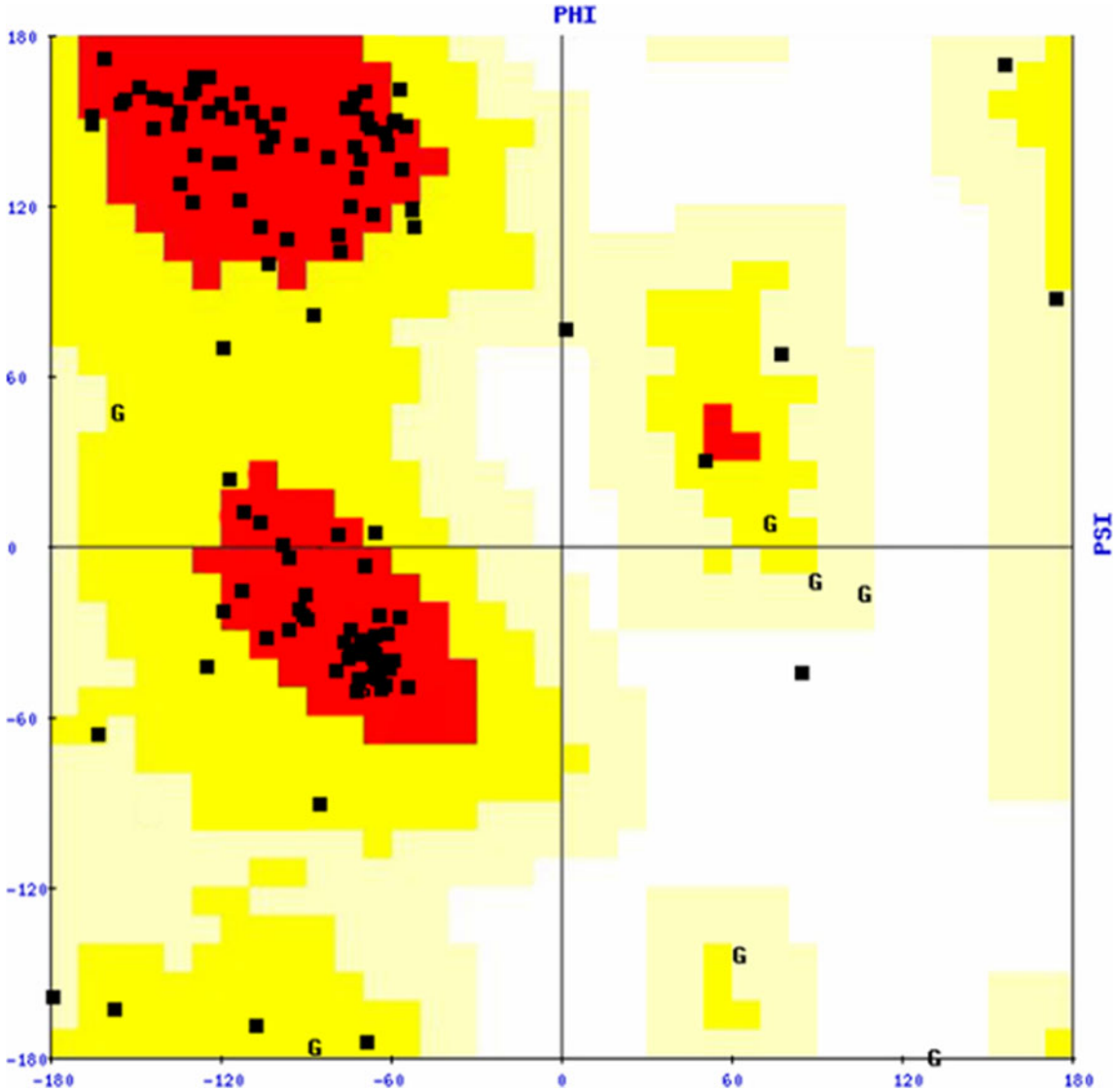

Fig. 3 Ramachandran plot for the structural model of PLA_2 from *Aedes aegypti* built using NIH structural analysis and verification software. The plot is for a total of 136 residues which are shown by black squares or by the letter 'G' that stands for the residue glycine

position 7 and a threonine replacing glycine at position 11. Here too, the side chains are oriented outward while the carbonyl oxygen atom interacts with the calcium (Fig. 5c). The glycines which contribute their carbonyl oxygens to the calcium ion cage are generally thought of as 'universally conserved components' of the PLA_2 enzyme family. So the replacements of these calcium binding-glycine residues in the group III PLA_2 are itself unprecedented.

The analysis of the calcium binding loop helps to derive the following conclusions: (a) the loops in the enzymes are stabilized by the calcium binding and the conserved disulfide bond between the cysteine in the loop with the cysteine in the first helix which comes to lie in close proximity to the loop; (b) provides the structural basis for the *In-vitro* assays that showed enzymes from *Mt, Ap, Cs, Dm* and *Hs* to be having calcium dependent properties [12, 17, 45–47]; (c) the loop forms the left superior wall of the substrate binding channel (Fig. 4) and it is likely that calcium coordination orders the local protein structure in a manner that optimizes the interactions with the substrate, just as in the case of bee venom PLA_2 [9]; (d) the loop offers variability in terms of the hydrophobicity offered for the substrate binding. The variability is due to non-conserved residues that are present between the two conserved calcium binding-glycines in the loop. While this position is occupied by a non-polar residue in all the mammalian enzymes, it is occupied by a polar residue in the insects and scorpions.

Active site analysis

This stereochemical mechanism of $sPLA_2$ is reminiscent of the catalytic system of the serine proteases [48]. A conserved

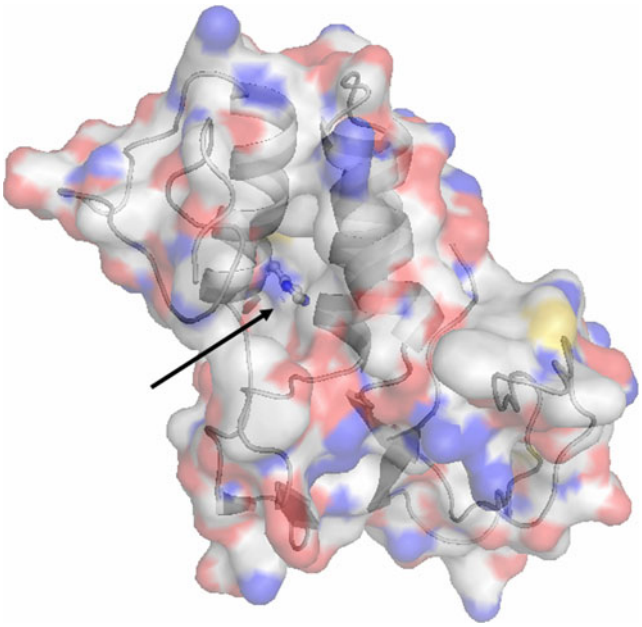

Fig. 4 Hydrophobic channel wall composition. The overall GRASP representation (transparent) of the *Ad*PLA$_2$ model is shown in conventional white-blue-red. The cartoon diagram of the structure is shown within the confines of the GRASP. The entry of the hydrophobic channel is indicated by black arrow. The catalytic histidine is shown nested at the very end of the channel

water molecule acts as the attacking nucleophile and attacks the *sn-2* bond. A conserved histidine in the active site of PLA$_2$ abstracts a proton from the water molecule at the Nδ1 position. The positive charge acquired by the histidine is stabilized through an extended hydrogen-bonded network that includes the carboxylate group of aspartic acid and the phenolic group of tyrosine residues. It may be noted that the invariant aspartate and tyrosines at the active site are from non-analogous backbone positions in different groups of the sPLA$_2$ enzymes. The histidine, aspartate and the single/dual tyrosine residues are rightly called as catalytic residues of the sPLA$_2$. It is noteworthy that this pattern of catalytic machinery is consistent in most groups of the PLA$_2$ ranging from group I, II IV, V, X and XII [49–52].

In comparison, the group III PLA$_2$s, show a total of nine variations at the active site of the enzyme (Fig. 6). In many of the insects, gila monsters and in the aquatic based organisms the classical PLA$_2$ triad of His-Asp-Tyr arrangement is maintained (Fig. 6a). In three of these, the hydroxyl of tyrosine makes a lone hydrogen bond with the Nδ2 of catalytic histidine, while the glutamate in the invariant position of aspartate is oriented away from the histidine. This is seen in most of the scorpions (Fig. 6b), *Dm*PLA$_2$ (fruit fly) (Fig. 6c), *Cs*PLA$_2$ (liver fluke), *Dv*PLA$_2$ (wine fly) and *Nv*PLA$_2$ (sea anemone) (Fig. 6d). In the mammalian enzymes including the *Hs*PLA$_2$ (human) (Fig. 6e), there is a phenylalanine at the active site where its aromatic ring fails to make any hydrogen bonded interactions with either the histidine or the aspartate which are in its vicinity. In these enzymes, the stabilization of aspartate is either by another histidine from that same helix as in *Md*PLA$_2$ (gray short tailed opossum) (Fig. 6f) or by a glycine or a serine present on the loop below as seen in *Ag*PLA$_2$ (house fly) (Fig. 6g) and *Br*PLA$_2$ (zebra fish) (Fig. 6h), respectively. In the last variant, the tyrosine that interacts with the catalytic histidine is seen to be stabilized by a glutamine residue as in the case of *Cs*PLA$_2$ (liver fluke) (Fig. 6i). The active site analysis helps to draw three important implications: (a) there is a more polar environment at the active site of venomous and parasitic enzyme isoforms as against the void of polar residues in the human isoform, (b) the analysis complements the functional studies that establish that the hydrogen bonding network between tyrosine, aspartate and histidine is not crucial for catalytic function [53] and (c) the loss of the hydrogen bonded network at the site has been known to cause non-localized changes in the structure conformations that affect the substrate binding and the enzyme kinetics

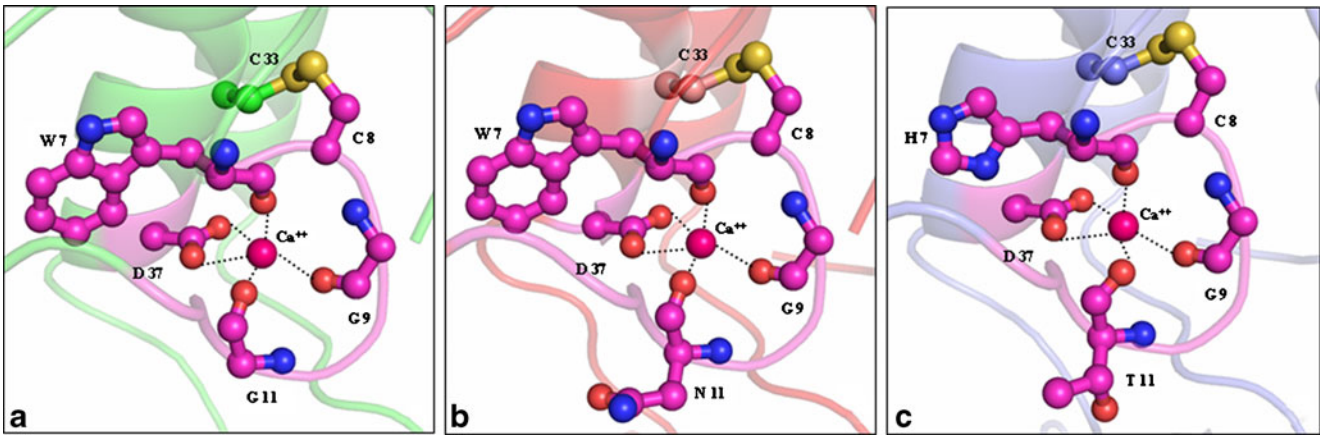

Fig. 5 Calcium binding analysis. The calcium binding loop residues and the calcium ion are shown for the group III PLA$_2$ enzymes in **a**. human; **b**. scorpion and **c**. liver fluke. The loop, residues and the calcium are shown in pink

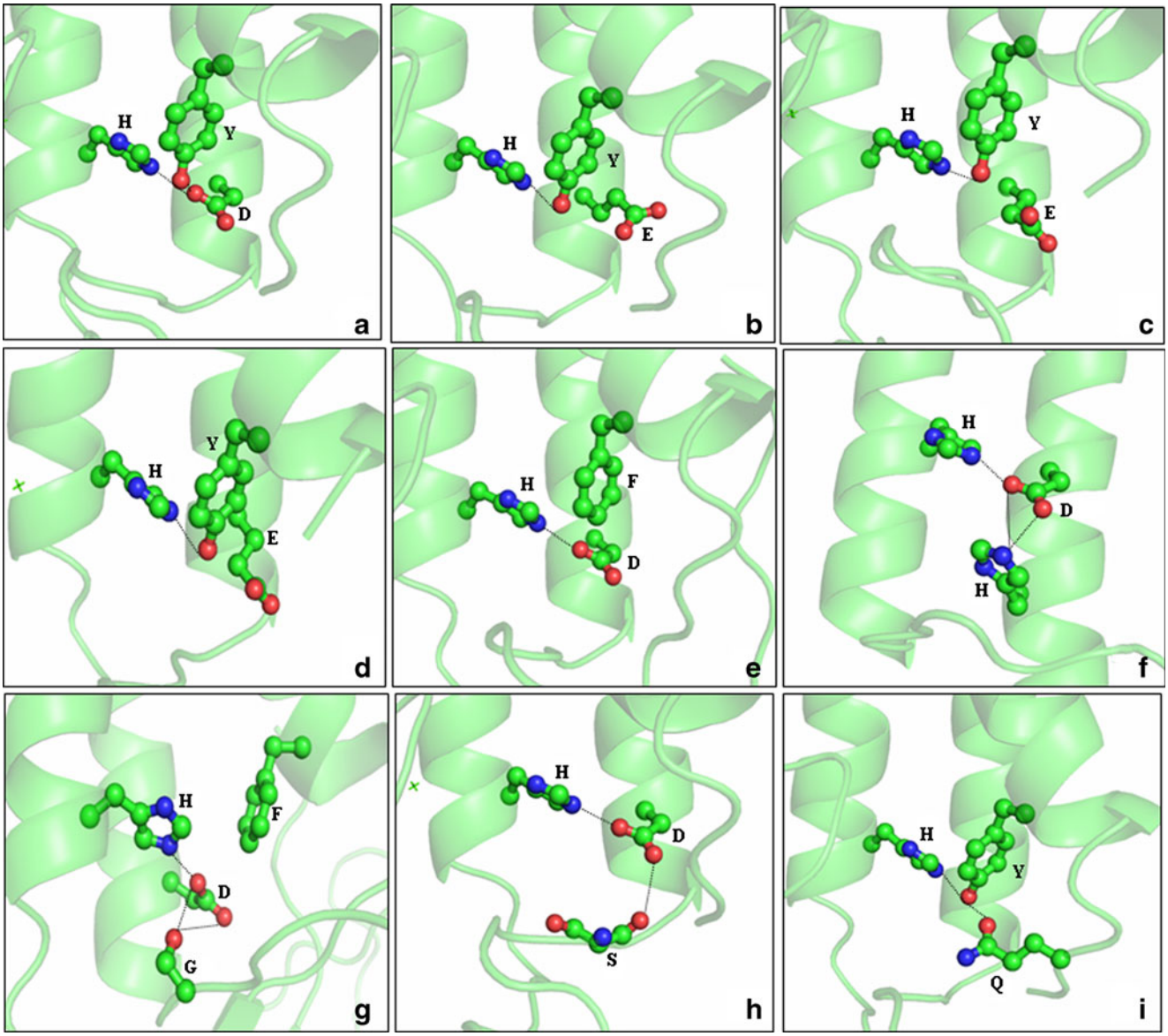

Fig. 6 Active site analysis. The residues present at the active site are shown for the enzymes of **a**. gila monster; **b**. scorpion; **c**. fruit fly; **d**. human; **e**. frog; **f**. house fly; **g**. zebra fish; **h**. liver fluke and **i**. bumble bee. The hydrogen bonded interactions (2.5 Å-3.5 Å) between the residues are shown in black dotted lines

[54]. The differences observed at the catalytic site of group III PLA_2 helps to provide a basis for the house keeping functions of the mammalian species and the differences in the severity of toxicity amongst the venomous species.

Hydrophobic channel analysis

During the binding of the substrate to the enzyme the fully acyl chains of the lipid occupy the core space in the enzyme called the hydrophobic channel. At the interior end of the channel lies the Nδ1 of the catalytic histidine that initiates the catalysis. The first two long helices are riveted together by a disulfide bridge to create a rigid backbone brace that forms the back wall of the channel. This channel comprises a cluster of hydrophobic residues that could make potential interactions with the substrate. Along side these residues there are patches of hydroxyl baring residues that are positioned lining the channel. These hydrophobic and hydroxyl residues play a significant role in the binding of the substrate [55]. The scorpion has the residues Thr87, Tyr88, Tyr54 and Ser10 lining the channel (Fig. 7a). The fruit fly and the human isoforms have a relatively more non-polar environment with the former comprising the residues Gly9, Phe55, Phe66, Leu81, Phe 86 (Fig. 7b) and the later comprising Val10, Phe54, Phe84 and Ala85 (Fig. 7c). The parasite, liver fluke, in contrast shows the presence of Ser10, Phe55 and

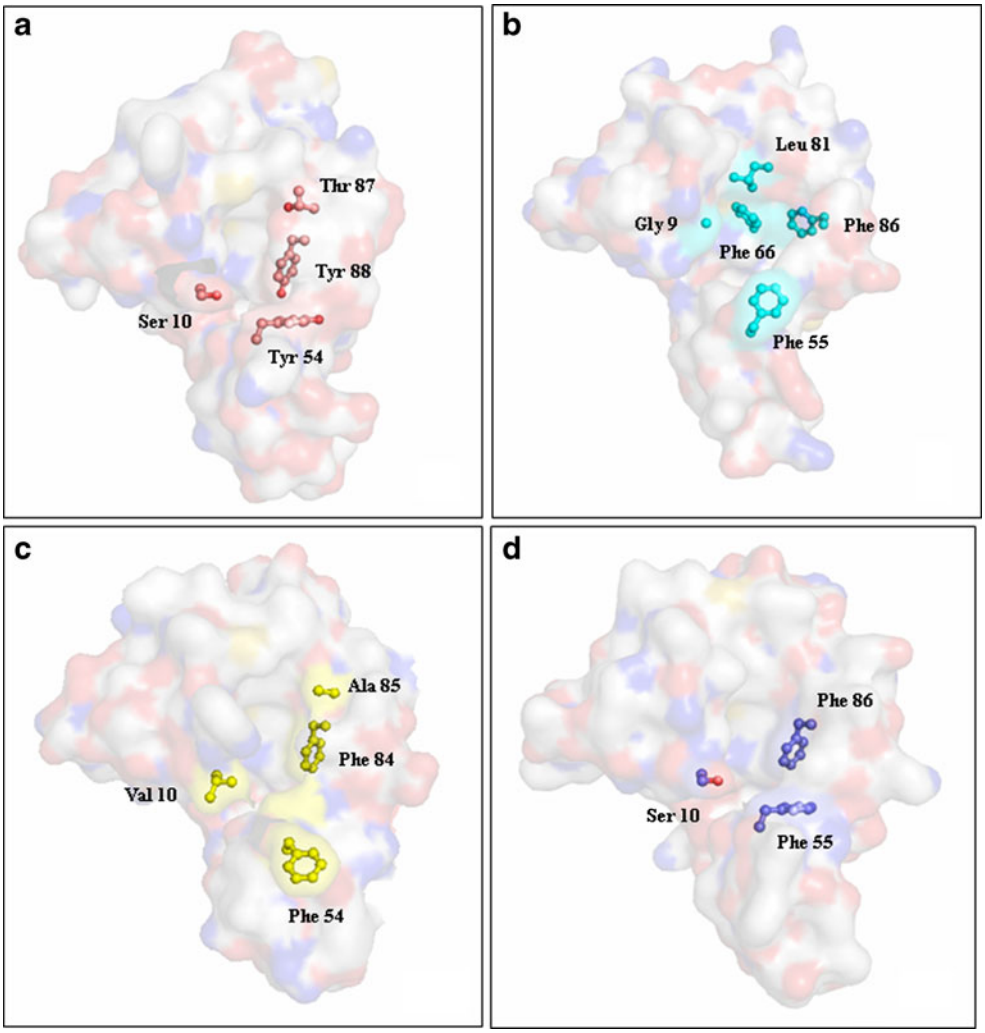

Fig. 7 Hydrophobic channel residue composition analysis. **a**. scorpion; **b**. fruit fly; **c**. human and **d**. liver fluke. The hydrophobic and the hydroxyl residues in the hydrophobic channel are shown

Phe86 (Fig. 6d). These channel lining residues are important in terms of aromaticity, hydrophobicity, and hydrophilicity conferred to the binding with the substrate and affects the kinetic parameters toward various substrates [56]. This explanation helps to put the following observations in the right perspective: (a) cause of diverse functional outcomes of toxicity, neuronal growth, hepatic fibrosis and cancer by these enzymes; (b) non-implication of the human isoform is in the causation of hepatic fibrosis despite its high expression in the liver tissue [47] and (c) the offensive and defensive role adopted by the enzyme in the venom of scorpions and insects respectively.

These studies show that the quality of residues and their conformations dictate the enzyme kinetics and thereby the functional outcomes. From the detailed structural studies the group III PLA_2 can be grouped into venomous insects, non-venomous insects, mammals, scorpions and aquatic forms. The analysis so far sets the platform for a systematic phylogenetic study to put the structural studies of the enzymes in the right perspective.

Phylogenetic analysis

In order to understand the molecular evolution of group III PLA_2 genes, we compiled the 50 group-III PLA_2 protein from scorpions, insects, mammals, reptiles, liver fluke and other sea living organisms including fishes and inferred evolutionary relatedness in the group III PLA_2 using maximum likelihood (ML) and neighbor joining (NJ) phylogenetics. Phylogenetic tree obtained using both ML (Fig. 8) and NJ (Fig. 9) yield similar tree topologies showing three primary divergent groups which are the insects, the scorpions and mammals including the bird, *Ml*PLA_2 (bat) and the marsupial, *Md*PLA_2 (gray short tailed opossum). While the enzymes in insects, mammalian and the aquatic organisms have followed a common pathway of evolution to attain a monomeric form, the scorpion group may have evolved separately after species diversification to attain the heterodimeric forms. The expression of distinct quaternary conformations of group III PLA_2 in different venom has probably resulted from natural selection in the PLA_2 paralogs through

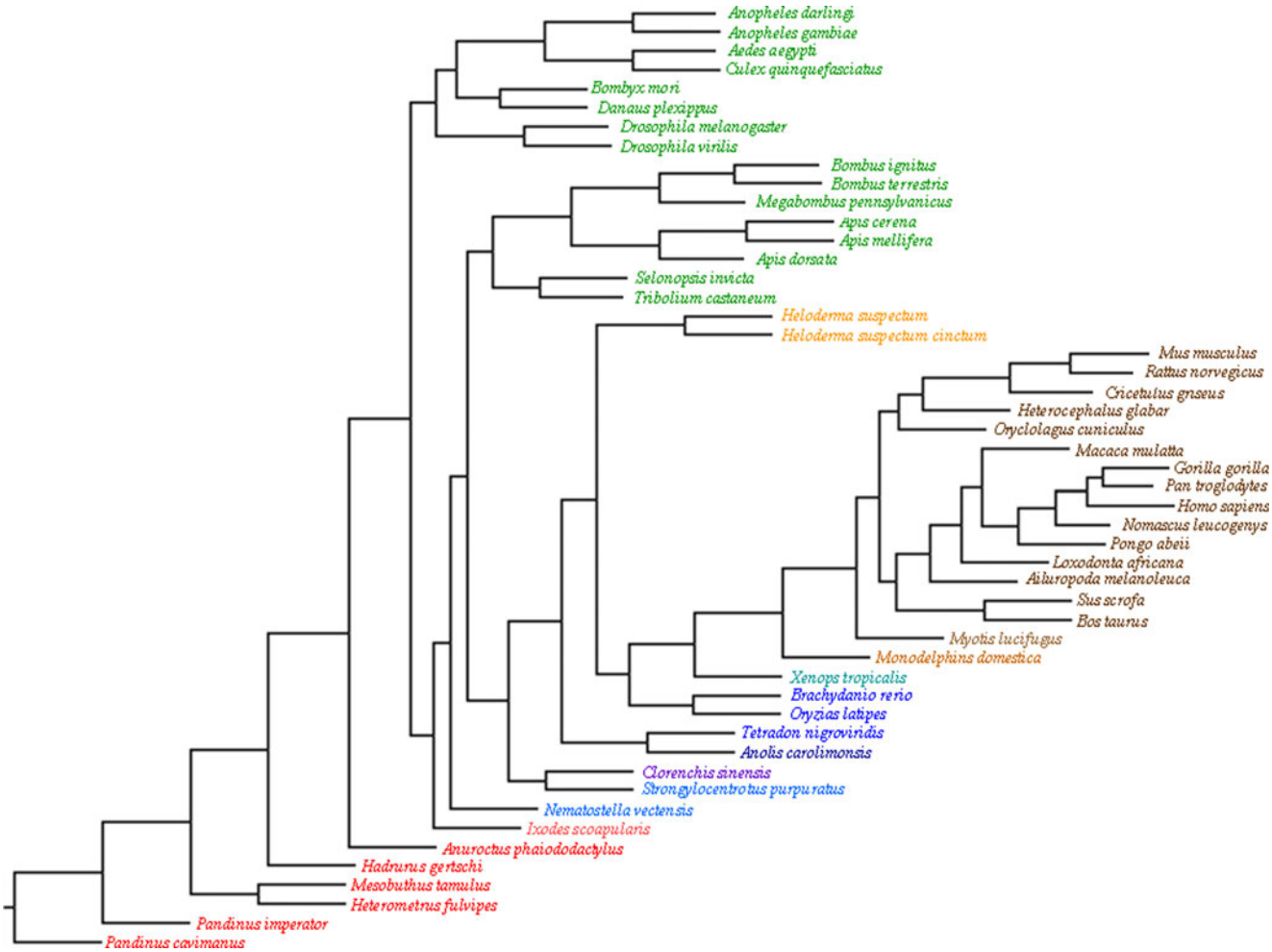

Fig. 8 Phylogenetic analysis by neighbor joining method

evolution, and there is every possibility that the monomeric and dimeric forms were present in their ancestral predecessors. The last distinct group to emerge from the two phylogenies is of the lower order organisms that include the liver fluke, sea anemones, fishes and frogs.

In the insect clades, it is seen that there are two sub-clades in both type of NJ and ML type of cladistic analysis. While one sub-clade comprises the non-venomous flies and mosquitoes that belong to the class insecta and order diptera, the other sub-clade comprises venomous ants, bees and beetles which belong to class insecta and order hymenoptera. While the *Hs*PLA$_2$ (gila monster) belonging to phylum chordate and class reptilia have been grouped alongside the aquatic forms in the ML, they seem to have evolved as a separate group from the former in NJ. Three dimensional structure of the *Hs*PLA$_2$ shows an alanine at position 10 contributing hydrophobicity to the substrate binding channel. This substrate binding function of the residue is the same as seen in the case of *Xt*PLA$_2$ (frog), *Tn*PLA$_2$ (spotted green puffy fish), *Sp*PLA$_2$ (sea urchin) and *Ac*PLA$_2$ (American chameleon), a class reptilia animal. It is therefore structurally justified that the gila monster variants be grouped with the aquatic isoforms. A notable and interesting aspect of the NJ method of phylogeny analysis is the proximity of the *Is*PLA$_2$ (deer tick) belonging to the class arachnida, to the scorpion group which belongs to the same class under the phylum arthropoda. This result is structurally justified by the presence of a hydrophobic residue, metheonine and an aromatic residue at position 23, 54 and 84 just as in the scorpion family. The *Is*PLA$_2$ (deer tick) also has a conserved catalytic aspartate just as the scorpion, *Ap*PLA$_2$ (mafia scorpion), and therefore rightly is adjacent to it on the phylogeny tree. Though *Xt*PLA$_2$ (frog) belongs to phylum chordate, it lacks the amino acid signatures of either the mammals or the fishes and is rightly seen wedged between the two. The primary structure of *Cs*PLA$_2$ (liver fluke) belonging to phylum platyhelminthes has the residues Thr26, Lys37, Ser42, Asn53, Leu54 and Arg68 just as seen in *Br*PLA$_2$ (Zebra fish) thereby justifying its grouping with the aquatic organisms in both the NJ and ML trees. The topology of the trees also shows taxonomic affinity between different types of mammals and are in agreement with their sequence analysis

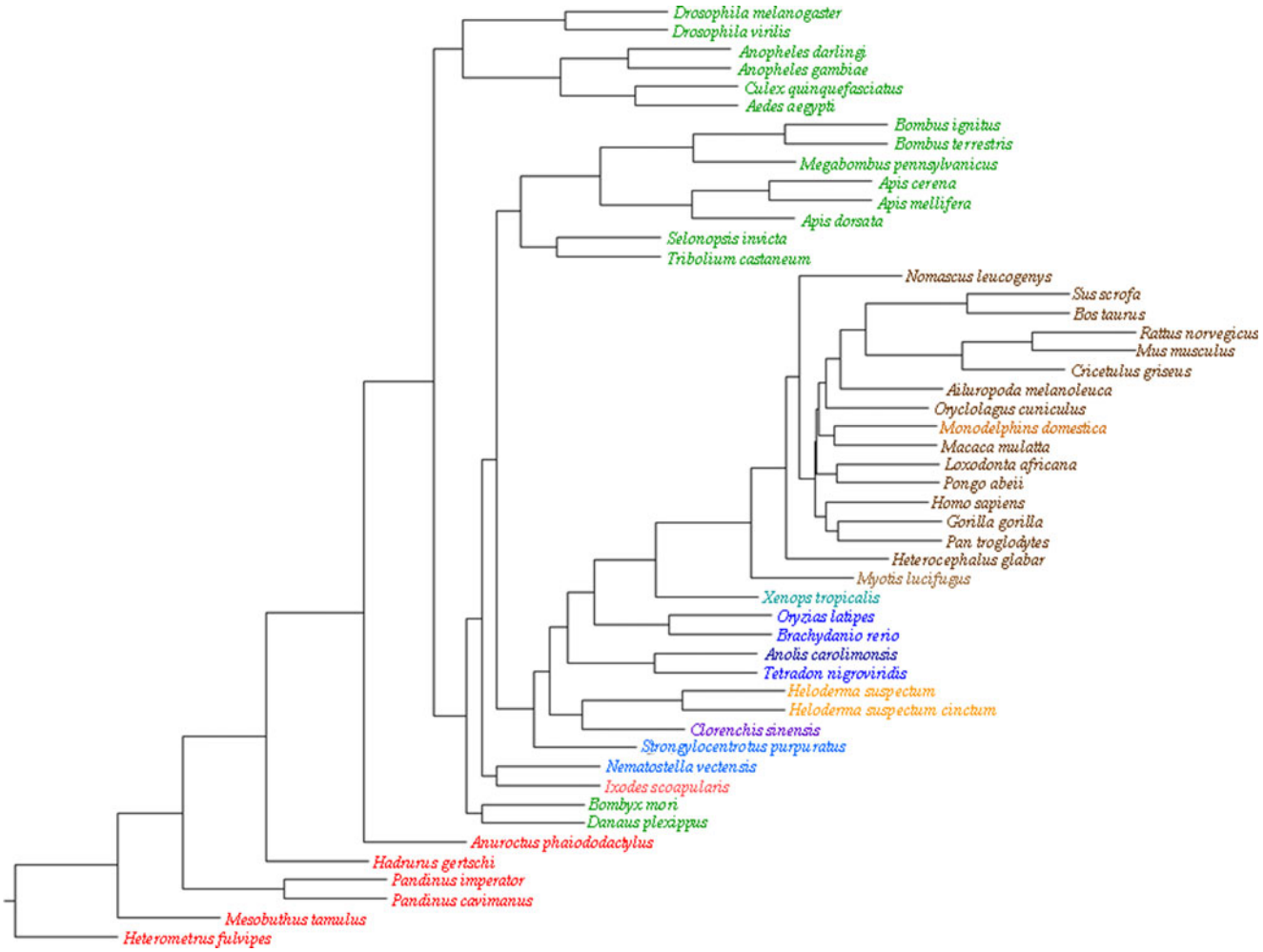

Fig. 9 Phylogenetic analysis by maximum likelihood method

of the conventional mammals, flying mammal (bat) and marsupial mammal (gray short tailed opossum). It therefore justifies the presence of all these different types of mammals under one group.

In summary, there is a convergent evolution for the calcium binding, histidine mediated catalysis and the overall structure of the enzyme and a divergent type of evolution that is seen for the quality of substrate binding channel, C-terminal region composition and the residues present alongside histidine at the active site. The interactions of different species with their respective environment may have been responsible for the adaptive evolutionary mechanisms such as gene duplication, diversification and segregation. The phylogenetic analysis of the group III PLA_2 paralogs can therefore be an added dimension for the group classification.

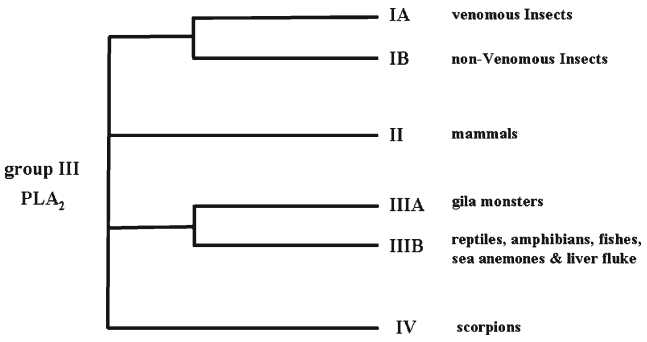

Fig. 10 Classification of group III PLA_2

Conclusions

Classification of group III PL A_2

Group III PLA_2 enzymes that are present across organisms in different species are collectively grouped under a single category in the PLA_2 enzyme classification. Though all the isoforms in this family share similar enzymatic activity, they are structurally different from each other for diverse functional outcomes across different species of organisms making the classification of the

group very relevant. Based on the detailed structural and phylogenetic analysis of group III PLA_2, the enzyme family can be classified into four sub-groups (Fig. 10): (1A) venomous insects; (IB) non-venomous insects; (II) mammals; (IIIA) gila monsters; (IIIB) reptiles, amphibians, fishes, sea anemones and liver fluke and (IV) scorpions. This classification is necessary for the following outcomes: (1) to understand the structure-function relationship in this group of enzymes; (2) to differentiate toxic, house keeping and the drug target forms of the enzyme; (3) to understand the modes of PLA_2 toxicity and (4) to comprehend the evolutionary perspective of the enzyme keeping in mind the survival of the organism.

Acknowledgments This work is funded by a FAST TRACK project award (SR/FT/18) to GH from Department of Science and Technology, Government of India.

References

1. van Deenen L, de Haas G, Heemskerk CH (1963) Hydrolysis of synthetic mixed-acid phosphatides by phospholipase A from human pancreas. Biochim Biophys Acta 67:295–304
2. Dufton MJ, Hider RC (1983) Classification of phospholipases A2 according to sequence: evolutionary and pharmacological implications. Eur J Biochem 137:545–551
3. Six DA, Dennis EA (2000) The expanding superfamily of phospholipase A(2) enzymes: classification and characterization. Biochim Biophys Acta 1488:1–19
4. Schaloske RH, Dennis EA (2006) The phospholipase A2 superfamily and its group numbering system. Biochim Biophys Acta 1761:1246–1259
5. Lindbom J, Ljungman AG, Lindahl M, Tagesson C (2002) Increased gene expression of novel cytosolic and secretory phospholipase A(2) types in human airway epithelial cells induced by tumor necrosis factor-alpha and IFN-gamma. J Interferon Cytokine Res 22:947–955
6. Kini RM (2003) Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon 42:827–840
7. Heinrikson RL, Krueger ET, Keim PS (1977) Amino acid sequence of phospholipase A2-alpha from the venom of *Crotalus adamanteus*. A new classification of phospholipases A2 based upon structural determinants. J Biol Chem 252:4913–4921
8. Tanaka H, Minakami R, Kanaya H, Sumimoto H (2004) Catalytic residues of group VIB calcium-independent phospholipase A2 (iPLA2gamma). Biochem Biophys Res Commun 320:1284–1290
9. Scott DL, Otwinowski Z, Gelb MH, Sigler PB (1990) Crystal structure of bee-venom phospholipase A2 in a complex with a transition-state analogue. Science 250:1563–1566
10. Hariprasad G, Kaur P, Srinivasan A, Singh TP, Kumar M (2012) Structural analysis of secretory phospholipase A_2 from *Clonorchis sinensis*: therapeutic implications for hepatic fibrosis. J Mol Model 18:3139–3145
11. Hariprasad G, Baskar S, Das U, Ethayathulla AS, Kaur P et al. (2007) Cloning, sequence analysis and homology modeling of a novel phospholipase A_2 from *Heterometrus fulvipes* (Indian black scorpion). DNA Seq 18:242–246
12. Hariprasad G, Saravanan K, Baskar S, Das U, Sharma S et al. (2009) Group III PLA2 from the scorpion, *Mesobuthus tamulus*: cloning and recombinant expression in *E. coli*. Electro J Biotech 12:3
13. Hariprasad G, Kumar M, Srinivasan A, Kaur P, Singh TP et al. (2011) Group III phospholipase A_2 from the scorpion, *Mesobuthus tamulus*: targeting and reversible inhibition by native peptides. Int J Biol Macr 48:423–431
14. Sosa BP, Alagón AC, Martin BM, Possani LD (1986) Biochemical characterization of the phospholipase A2 purified from the venom of the Mexican beaded lizard (*Heloderma horridum horridum Wiegmann*). Biochemistry 25:2927–2933
15. Hariprasad G, Kumar M, Kaur P, Singh TP, Kumar RP (2010) Human group III PLA_2 as a drug target: structural analysis and inhibitor binding studies. Int J Biol Macr 47:496–501
16. Zamudio FZ, Conde R, Arévalo C, Becerril B, Martin BM et al. (1997) The mechanism of inhibition of ryanodine receptor channels by imperatoxin I, a heterodimeric protein from the scorpion *Pandinus imperator*. J Biol Chem 272:11886–11894
17. Valdez-Cruz NA, Batista CV, Possani LD (2004) Phaiodactylipin, a glycosylated heterodimeric phospholipase A2 from the venom of the scorpion *Anuroctonus phaiodactylus*. Eur J Biochem 27:1453–1464
18. Conde R, Zamudio FZ, Becerril B, Possani LD (1999) Phospholipin, a novel heterodimeric phospholipase A2 from *Pandinus imperator* scorpion venom. FEBS Lett 460:447–450
19. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinfo 2:2–3
20. Combet C, Jambon M, Deléage G, Geourjon C (2002) Geno3D: automatic comparative molecular modelling of protein. Bioinformatics 18:213–214
21. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics 22:195–201
22. Bates PA, Kelley LA, MacCallum RM, Sternberg MJE (2001) Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 5:39–46
23. Lambert C, Léonard N, De Bolle X, Depiereux E (2002) ESyPred3D: prediction of proteins 3D structures. Bioinformatics 18:s1250–s1256
24. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371
25. De Lano WL (2002) The PyMOL molecular graphics system. San Carlos, CA, USA
26. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S et al. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 4(1983):187–217
28. Momany FA, Rone R (1992) Validation of the general purpose QUANTA® 3.2/CHARMm®/force field/. J Comp Chem 13:888–900
29. Wallace IM, O'Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 34:1692–1699
30. Capella-Gutiérrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973
31. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
32. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376
33. Tuimala J (2005) A Primer to Phylogenetic Analysis Using the *PHYLIP* Package, 4th edn. Centre for Scientific Computing Ltd, Espoo, Finland
34. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282
35. Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–192

36. Rambaut A (2010) Tree Figure Drawing Tool, Version 1.3.1. [http://tree.bio.ed.ac.uk/software/figtree], 1-3-2010. Ref Type: Electronic Citation
37. Lugtigheid RB, Otten-Kuipers MA, Verheij HM, De Haas GH (1993) Arginine 53 is involved in head-group specificity of the active site of porcine pancreatic phospholipase A2. Eur J Biochem 213:517–522
38. Lugtigheid RB, Nicolaes GA, Veldhuizen EJ, Slotboom AJ, Verheij HM et al. (1993) Acylation of porcine pancreatic phospholipase A2 influences penetration and substrate head-group binding, depending on the position of the acylated lysine in the enzyme molecule. Eur J Biochem 216:519–525
39. Kuipers OP, Dekker N, Verheij HM, de Haas GH (1990) Activities of native and tyrosine-69 mutant phospholipases A2 on phospholipid analogues. A reevaluation of the minimal substrate requirements. Biochemistry 29:6094–6102
40. Han SK, Kim KP, Koduri R, Bittova L, Munoz NM et al. (1999) Roles of Trp31 in high membrane binding and proinflammatory activity of human group V phospholipase A2. J Biol Chem 274:11881–11888
41. Santamaría C, Larios S, Angulo Y, Pizarro-Cerda J, Gorvel JP et al. (2005) Antimicrobial activity of myotoxic phospholipases A2 from crotalid snake venoms and synthetic peptide variants derived from their C-terminal region. Toxicon 45:807–815
42. Lomonte B, Angulo Y, Calderón L (2003) An overview of lysine-49 phospholipase A2 myotoxins from crotalid snake venoms and their structural determinants of myotoxic action. Toxicon 42:885–901
43. Kini RM, Iwanaga S (1986) Structure-function relationships of phospholipases. II: Charge density distribution and the myotoxicity of presynaptically neurotoxic phospholipases. Toxicon 24:895–905
44. Sing-Yi F, Collin G (2008) Hymenoptera envenomation: bees, wasps and ant. Ped Emer Med Pract 5:6
45. Hu F, Hu X, Ma C, Zhao J, Xu J, Yu X (2009) Molecular characterization of a novel *Clonorchis sinensis* secretory phospholipase A2 and investigation of its potential contribution to hepatic fibrosis. Mol Biochem Parasitol 167:127–134
46. Ryu Y, Oh Y, Yoon J, Cho W, Baek K (2003) Molecular characterization of a gene encoding the *Drosophila melanogaster* phospholipase A_2. Biochim Biophys Acta 25:206–210
47. Valentin E, Ghomashchi F, Gelb MH, Lazdunski M, Lambeau G (2000) Novel human secreted phospholipase A(2) with homology to the group III bee venom enzyme. J Biol Chem 275:7492–7496
48. Annand RK, Kontoyianni M, Penzotti JE, Dudler T, Lybrand TP et al. (1996) Active site of bee venom phospholipase A_2: the role of histidine-34, aspartate-64 and tyrosine-87. Biochemistry 35:4591–4601
49. Jabeen T, Singh N, Singh RK, Ethayathulla AS, Sharma S et al. Crystal structure of a novel phospholipase A2 from *Naja naja sagittifera* with a strong anti-coagulant activity. Toxicon 46: 865–875
50. Jasti J, Paramasivam M, Srinivasan A, Singh TP (2004) 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: Biol Crystallogr 60:66–72
51. Pan YH, Yu BZ, Singer AG, Ghomashchi F, Lameau G et al. (2002) Crystal structure of human group X secreted phospholipase A2 electrostatically neutral interfacial surface targets zwitterionic membrane. J Biol Chem 277:29086–29093
52. Six DA, Barbayianni E, Loukas V, Constantinou-Kokotou V, Hadjipavlou-Litina D et al. (2007) Structure-activity relationship of 2-oxoamide inhibition of group IVA cytosolic phospholipase A2 and group V secreted phospholipase A2. J Med Chem 50:4222–4235
53. Masuda S, Yamamoto K, Hirabayashi T, Ishikawa Y, Ishii T et al. (2008) Human group III secreted phospholipase A2 promotes neuronal outgrowth and survival. Biochem J 409:429–438
54. Nicolas JP, Lambeau YG, Ghomashchi F, Lazdunski M, Gelb MH et al. (1997) Localization of structural elements of bee venom phospholipase A2 involved in N-type receptor binding and neurotoxicity. J Biol Chem 272:7173–7181
55. Chandra V, Jasti J, Kaur P, Dey S, Srinivasan A et al. (2002) Design of specific peptide inhibitors of phospholipase A2: structure of a complex formed between Russell's viper phospholipase A2 and a designed peptide Leu-Ala-Ile-Tyr-Ser (LAIYS). Acta Crystallogr D: Biol Crystallogr 58:1813–189
56. Dupureur CM, Yu BZ, Mamone JA, Jain MK, Tsai MD (1992) Phospholipase A2 engineering. The structural and functional roles of aromaticity and hydrophobicity in the conserved phenylalanine-22 and phenylalanine-106 aromatic sandwich. Biochemistry 31:10576–10583