

Sequence analysis and phylogenetic relationship of genes encoding heterodimeric phospholipases A2 from the venom of the scorpion *Anuroctonus phaiodactylus*

Norma A. Valdez-Cruz, Lorenzo Segovia, Miguel Corona¹, Lourival D. Possani*

Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Apartado Postal 510-3, Cuernavaca 622106221, Mexico

Received 12 December 2006; received in revised form 25 January 2007; accepted 12 March 2007

Available online 24 March 2007

Received by J.G. Zhang

Abstract

Some scorpion venom contain heterodimeric phospholipases A2. They were shown to be toxic to insects and to cause edema and/or hemolysis of mammalian erythrocytes. This manuscript describes the results of cDNA cloning of five different heterodimeric phospholipases from the venomous glands of the Mexican scorpion *Anuroctonus phaiodactylus*. The amino acid sequence deduced from the heterodimeric phospholipases open reading frames corresponds in each case to a different isoform. The nucleotide sequences corresponding to two of these genes were also obtained by directly sequencing genomic DNA. The cDNA isoforms show high similarity with the heterodimeric phospholipase Phaiodactylipin purified from the same scorpion. However, similar phospholipases were also found in scorpions from other species and the sequences available were used to construct a phylogenetic tree. In order to understand better the gene structure and phylogeny of these enzymes we analyzed their sequences and compared them with secretory phospholipases of other sources from groups I, II and III. The genomic DNA sequence of a similar phospholipase from bee venomous glands was also cloned. The information available on a *Drosophila* phospholipase was included in this analysis. The phospholipases of groups I and II contain a conserved exon–intron structure (four or five exons of the mature segment of the enzyme are separated by three or four introns). Also, the gene structure of the phospholipases from *A. phaiodactylus* and that of the bee venom, belonging to group III phospholipases, are interrupted by three introns. The mature peptide of the bee enzyme is a single polypeptide chain, coded by four exons, whereas those from the scorpion studied here although having four exons, showed the presence of two different polypeptides in its native state. The mature protein is processed after synthesis, producing the heterodimeric structure: a long and a short-peptide chain, linked by a disulfide bridge. The small subunit is the one coded by the fourth exon. The human phospholipase A2 and that of *Drosophila*, also classified into the group III phospholipases, have a considerably different exon–intron organization.

© 2007 Elsevier B.V. All rights reserved.

Keywords: cDNA clones; Gene structure; Phylogeny; Phospholipase; Scorpion

1. Introduction

Phospholipase A2 (PLA2), phosphatide 2-acylhydrolase, (EC 3.1.1.4) catalyzes the hydrolysis of the sn-2 acyl ester bond of phospholipids, resulting in the formation of lysophospholi-

pids and free fatty acids (Tischfield, 1997). PLA2s are ubiquitous in nature and especially abundant in snake venoms, mammalian pancreatic juices and some other sources (Valentin and Lambeau, 2000). Secretory PLA2s are classified into twelve groups based on the alignment of their primary sequences, the disulfide bond patterns and their biochemical properties (Six and Dennis, 2000; Rouault et al., 2003). Although the PLA2s share common catalytic properties and a high degree of similarity in primary, secondary and tertiary structures, they show a wide variety of diverse pharmacological and biological activities. Among these functions are: anti-inflammatory and myonecrotic action (Gutiérrez and Lomonte, 1995), blood platelet

Abbreviations: A, *Anuroctonus*; Pha1 to Pha5, phospholipase number 1 to 5; PLA2, phospholipase A2; Phai, phaiodactylipin.

* Corresponding author. Tel.: +52 77 73 171209; fax: +52 77 73172388.

E-mail address: possani@ibt.unam.mx (L.D. Possani).

¹ Present address: Department of Entomology, University of Illinois, 505 S. Goodwin Ave, Urbana, IL 61801.

aggregation (Gerrard et al., 1993), hemolysis (Condrea et al., 1980), neurotoxicity (Chang et al., 1977; Bon et al., 1979), cardiotoxicity (Fletcher et al., 1981) and ion channel blocking activity (Benishin, 1990). PLA2 purified from snake venoms have been classified into groups I and II. Phospholipases found in bees, scorpions, lizards, jellyfishes and one from a human source were classified into group III (Valentin et al., 2000). PLA2s of group III contain ten cysteines and are composed of one or two subunits (Valentin and Lambeau, 2000). They are phylogenetically distantly related to group I and II, but highly similar in the Ca^{2+} binding loop and the region of the active site. The bee phospholipase structure has a scaffold similar to that of groups I and II (Scott et al. 1990), but shows a different pharmacological activity (Fenard et al., 2001; Deregnacourt and Schrevel, 2000).

Only five different covalently linked heterodimeric phospholipases A2 from different venoms have been published. β -bungarotoxin was isolated from the snake *Bungarus multicinctus* and was classified in group II (Kondo et al., 1982; Delot and Bon, 1973); conodipin-m was purified from the cone snail *Conus magus*, classified into group IX (McIntosh et al., 1995); IpTxI and Phospholipin were isolated from the venom of the African scorpion *Pandinus imperator* (Zamudio et al., 1997; Conde et al., 1999) and Phaiodactylipin from the scorpion *Anuroctonus phaiodactylus* (Valdez-Cruz et al., 2004). The last three were placed into group III. Phaiodactylipin is toxic to arthropods, has a hemolytic effect on erythrocytes and causes inflammation in rat muscle (Valdez-Cruz et al., 2004). All five heterodimeric PLA2 contain a large subunit and a small subunit, linked by a disulfide bridge. There is additional information concerning a couple of other heterodimeric phospholipases from scorpion venoms, but they were obtained by cDNA cloning and the work is not published, only deposited in GenBank (accession nos. AAZ78243 and AY443497). It has also been reported that other venom sources contain complex phospholipase systems, mostly conformed by two subunits but not covalently linked. Among the heterodimeric PLA2 not covalently bound are: crotoxin from *Crotalus durrisus terrificus* (Bon et al., 1979), taipoxin from the Australian snake taipan *Oxyuranus s. scutellatus* (Lind and Eaker, 1982), mojave toxin from the snake *Crotalus scutulatus scutulatus* (Bieber et al., 1990), vipoxin from *Vipera ammodytes meridionalis* (Aleksiev and Shipolini, 1971; Mancheva et al., 1984), and its analogue from *Vipera russelli* (Rajashankar et al., 1999). But also known is taicatoxin, an oligomeric complex in which a PLA2 is the main component (Possani et al., 1992).

Here we report the isolation of cDNA clones encoding for five novel covalently linked heterodimeric phospholipases from the scorpion *A. phaiodactylus*, shown to be similar to Phaiodactylipin. Two of them were confirmed by directly sequencing the protein. We cloned and compared the gene structural organization of these novel heterodimeric PLA2 with that of the bee (this work), with one from *Drosophila* (GenBank accession number AAP45009) and another from a human source (GenBank accession number Q9NZ20).

2. Materials and methods

2.1. Cloning genes from cDNA that code for heterodimeric phospholipases

RNA was isolated from the pair of venomous glands (telson) of only one scorpion of the species *A. phaiodactylus* from Baja California (Mexico), using the method of Chirgwin et al. (1979). Total RNA (500 ng) was used for the cDNA synthesis using the oligonucleotide poliT22NN (Corona et al., 2001). The cDNA was amplified by means of a set of sense and antisense primers. These primers were designed based on the peptide sequence of Phaiodactylipin (Valdez-Cruz et al., 2004). The 5' RACE-PCR experiments were carried out using the 5'RLM-RACE Kit (Ambion, Austin, TX USA). The 3' RACE-PCR were performed using the degenerate primer from the amino terminal region of Phaiodactylipin and the reverse primer poliT22NN (Corona et al., 2001). The amplification was obtained using the primers sense 5'-AATGGATGCTTTGGAGAG-3' or 5'-TTCCTCATCGTAC-CAGAT-3', and the antisense 5'-ACATCATCTCTCGCATACT-TATT-3'. PCR conditions were: 94 °C for 90 s, in the first cycle, 94 °C 40 s, 52 °C for 40 s, 72 °C for 50 s; repeated 30 cycles. The PCR products were analyzed on agarose gels. The appropriate bands were purified using the Qiaquick gel extraction kit (Quiagen, GmbH, Hilden, Deutschland), and subcloned into an EcoRV site of the phagemid pKS-(Stratagen, La Jolla, CA, USA). The ligated products were then transformed in *E. coli* strain DH5- α . The recombinants were selected on LB-ampicillin (50 $\mu\text{g/ml}$) plates supplemented with IPTG and X-Gal and subjected to Sanger dideoxy DNA sequencing, as previously described by our group (Corona et al., 2001). Approximately 50 positive clones were obtained. Under sequencing most of them (*Phai*, *Pha1* to *Pha4*), were obtained more than one time. *Pha5* was found only once, whereas *Phai* and *Pha4* were found about seven times each. Sequences were obtained from both sides 5'- and 3'-nucleotide end of the clones. The new sequences are deposited in GenBank accession nos. EF364039 to EF364043.

2.2. Isolation of DNA from scorpion *A. phaiodactylus* and *Apis mellifera*

The venomous glands of one *A. phaiodactylus* scorpion were frozen and pulverized into a fine powder in liquid nitrogen. The genomic DNA was extracted in the presence of SDS and Proteinase K (Boehringer Mannheim, Germany) using DNeasy Kit (Qiagen). The DNA pellet was suspended in TE buffer. The same protocol was applied to extract DNA from the venomous apparatus of one bee, collected in the field (Cuernavaca city, near the Institute of Biotechnology), and typified as a hybrid between *A. mellifera ligustica* and *A. mellifera adamsoni* (see acknowledgements). The sequences obtained are deposited in GenBank, accession nos. EF364044 for *Pha4* and EF373554 for the bee PLA2.

2.3. Amplification of genomic DNA

Oligonucleotide primers were designed based on the sequence of the highly conserved regions previously obtained

from the sequence of the cDNA of Phaiodactylipin. PCR was performed using two rounds of amplification. The first PCR condition was: 94 °C for 60 s, 47 °C for 90 s, and 72 °C for 120 s, for 30 cycles. The second PCR condition was: 94 °C for 60 s, 47 °C for 90 s and 72 °C for 120 s, for 28 cycles.

2.4. Phylogenetic analysis

A multiple alignment of 116 secretory PLA2 sequences, representative of each group, was performed using the program MUSCLE 3.5 (Edgar, 2004). The sequences selected were from snakes (*Agkistrodon contortrix*, *Agkistrodon acutus*, *Bothrops asper*, *Bothrops atrox*, *Bothrops jararacussu*, *Bothrops erythromelas*, *Crotalus durissus terrificus*, *B. multicinctus*, *Laticauda semifasciata*, *Laticauda laticaudata*, *Oxyuranus microlepidotus*, *Naja naja*, *Micrurus corallinus*, *Pseudechis porphyriacus*, *Pseudechis australis*, *Oxyuranus scutellatus*, *Trimeresurus gramineus*, *Tropidechis carinatus*, *Pseudonaja textiles*, *Vipera ammodytes*, *Vipera nikolskii*, *Vipera aspis aspis*, *Vipera palaestinae*, *V. russelli siamensis*, *Gloydus halys*, *Laticauda semifasciata*, *Trimeresurus mucrosquamatus*, *Trimeresurus jerdonii*, *Trimeresurus borneensis*, *Trimeresurus puniceus*, *Trimeresurus stejnegeri*, *Sistrurus catenatus tergeminus*, *Deinagkistrodon acutus*, *Crotalus durissus terrificus*, *Cerrophidion godmani*, *Trimeresurus flavoviridis*), scorpions (*A. phaiodactylus*, *Heterometrus fulvipes*, *Mesobuthus tamulus* and *P. imperator*) insects (*Drosophila melanogaster*, *Drosophila pseudoobscura*, *Anopheles gambiae*, *Bombus pennsylvanicus*, *Bombus terrestris* and *A. mellifera*), one snail (*C. magus*), mammals (*Homo sapiens* and *Mus musculus*), plants (*Arabidopsis thaliana*, *Oryza sativa*), fish *Danio rerio* and from the reptile *Heloderma suspectum*. The multiple alignment was generated using the default conditions of 15 iterations and was visually inspected for possible improvement. Some sequences were trimmed, conserving only the informative regions necessary for compatible analysis. Then, a Bayesian estimation of the phylogeny of these sequences was carried out, with the program mrBayes 3.1.2 (Huelsenbeck and

Ronquist, 2001) using 1,500,000 generations, sampling and printing frequencies of 500 and 8 chains for the Markov Chain Monte Carlo analysis. This procedure generated 3000 trees. Since stationarity was reached in around 500 cycles, we used a burning of 1500 trees so that only 3500 tree samples were used for the determination of a consensus tree and the determination of posterior probabilities for the trees. We repeated the Bayesian estimation starting from a different initial tree to verify whether it was possible to obtain the same phylogeny. The results were positive, i.e., a similar phylogeny was found. The unrooted tree diagrams are visualized with Tree-View (Page, 1996) from Glasgow University (internet reference: <http://taxonomy.zoology.gla.ac.uk/lod/treeview.htm/>).

3. Results

3.1. Nucleotide sequences of cloned gene

Five PLA2 isoforms of Phaiodactylipin were amplified and cloned from cDNA of the scorpion *A. phaiodactylus* (abbreviated *Pha1* to *Pha5* in Fig. 1). The region 5' preceding the N-terminal segment of the sequenced genes is conserved in all isoforms. An identity of approximately 90% was observed among the sequences of the mature proteins. The non-conserved amino acid sequences found are single point mutations situated in positions that do not correspond to the segment that binds calcium. Also the region corresponding to the catalytic site is highly conserved. For *Pha1* and *Pha5* the segment corresponding to the signal peptide was not obtained in the sequence reported here. For two of the isoforms cloned (*Pha2* and *Pha4* in Fig. 1) the corresponding proteins were isolated and the N-terminal amino acid sequence was directly determined by automatic Edman degradation. The first 35 residues (which includes the calcium binding site and part of the catalytic site (indicated in Fig. 1) were identical to the sequence deduced from the nucleotide sequence. For the isoform *Pha4*, the amino acid segment corresponding to the small subunit was also

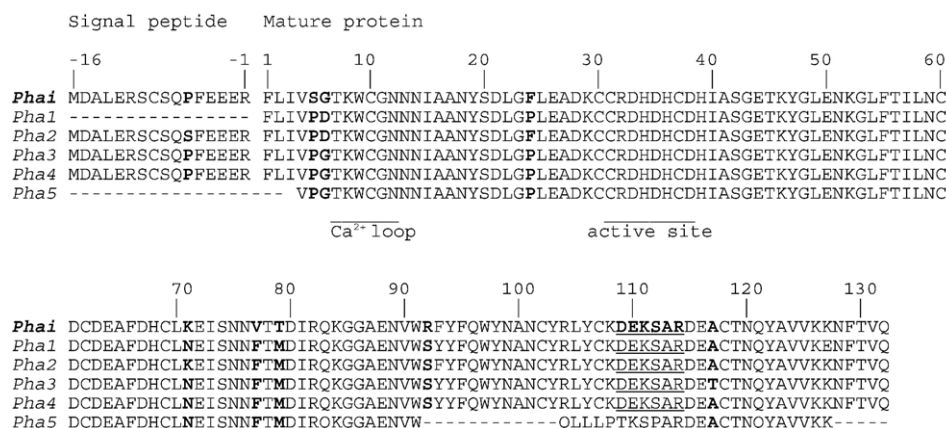


Fig. 1. Amino acid sequence alignment of five cDNA clones of heterodimeric phospholipases from the venomous gland of the scorpion *A. phaiodactylus*. The protein sequences were deduced from the nucleotide sequences. Single letter codes of amino acids were used. The putative signal peptide sequence is located from positions -16 to -1. The Ca^{2+} binding region and the active site are indicated below all sequences. The underlined segments correspond to the processed hexapeptides. For *Pha1* and *Pha5* the sequence corresponding to the signal peptide was not obtained, due to the protocol used for sequencing the cDNAs. *Phai* stands for phaiodactylipin. Dashes at the most C-terminal region of *Pha5* were included by the program used, in order to increase similarities. Bold letters were used to highlight the mutated amino acid residues.

sequenced up to residue number 7. Thus, *Pha3* although having an identical N-terminal sequence to *Pha4*, is a different isoform, because its C-terminal region that codes for the small subunit is different. This also supports the idea that the cloned proteins are indeed heterodimeric phospholipases. It is worth observing that in four isoforms reported in Fig. 1 the hexapeptides (underlined) are present and are assumed to be processed during maturation,

exactly as occurs with Phaiodactylipin (Valdez et al., 2004). An exception to this rule is *Pha5*, where the amino acid sequence of the hexapeptide is not conserved, and we do not know if this segment is also processed during maturation. Actually, *Pha5* has a missing segment (from residues 92 to 103), just before the position expected for the hexapeptide. In all cases, apart from *Pha5*, the preceding residue of the peptide normally eliminated

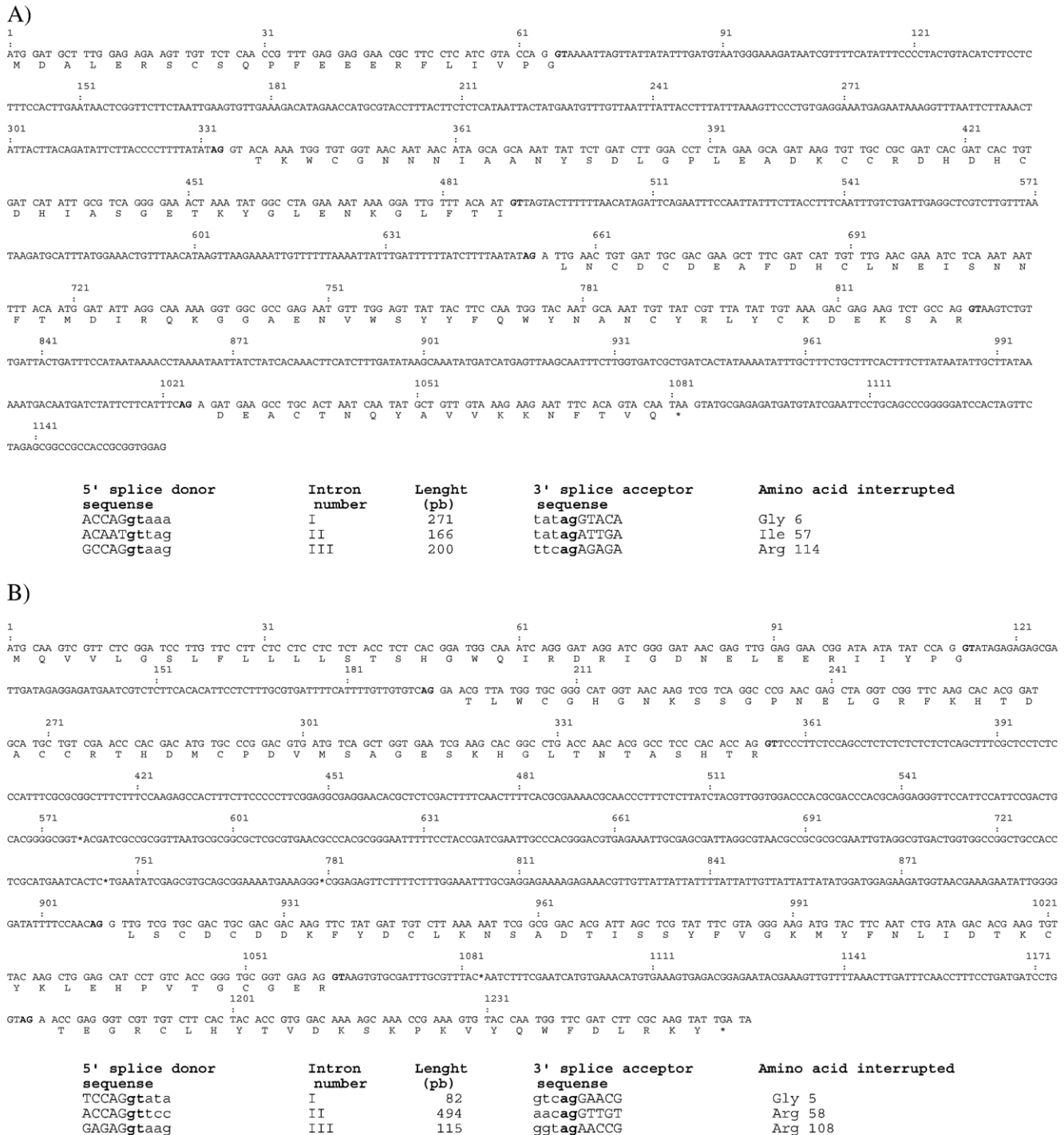


Fig. 2. Gene structure organization and nucleotide sequence of *A. phaiodactylus* and *Apis mellifera* sPLA2. A) Nucleotide and amino acid sequence of *Pha4*. The sequences underlined by their amino acid letters, correspond to the exons. Bold pair-bases indicate the donor and acceptor sequences positions. Below the sequences, the splicing regions (splice donor, splice acceptor, intron number, length and amino acids where the event occurred, are indicated. B) Contains the same information for the gene obtained from the bee PLA2.

is a lysine (position 108); for *Pha5* it is proline (also corresponding to position 108 in our alignment). The proteins assumed to be coded by the three other isoforms, not directly sequenced by Edman degradation, were not purified, but we assumed they are among the six different components of the venom previously identified and shown to display phospholipase activity, when separated by high performance chromatography (Valdez et al., 2004). For this reason we can conclude that the isoforms reported here are *bona fide* components expressed in the venomous gland. This is similar to what was described in snake venoms, where up to ten different isoforms of PLA2 were identified in a single species (Chen et al., 2004; Tsai et al., 2000, 2004). The same authors (Chen et al., 2004 and Tsai et al., 2000, 2004) have attributed this variability of isoforms to an accelerated process of evolution, being produced by point mutations of the parent gene.

3.2. Gene organization of the PLA2 isoforms

The complete nucleotide genomic sequence of *Pha4* (using DNA as template) was obtained, as shown in Fig. 2A. Since the sequences corresponding to the mature protein were identical to

that obtained by sequencing the cDNA it was possible to identify the segments corresponding to putative introns. In Fig. 2A the deduced amino acids corresponding to the signal peptide and to the mature protein are indicated below the nucleotide sequence. For the segments corresponding to the introns only the nucleotide sequence is shown. The intron/exon rule (GT/AG) is observed. Below each of the nucleotide sequences we have indicated the 5' splice donor, the intron number, the length of base pairs and the 3' splice acceptor sequences for the scorpion and for the bee phospholipases, respectively.

The signal peptide ranges from nucleotide 1 to 48 (Fig. 2A), coding for the amino acids from position -1 up to number -16. This sequence is immediately followed by the first six residues of the N-terminal region of the PLA2 (first exon). After this segment there is an intron of 271 nucleotides. The second exon has 51 amino acids (amino acids in position 6 to 57) and it is followed by 166 nucleotides from the second exon. The third exon is formed by amino acids from position 57 to 114, which is followed by the third intron (200 nucleotides). The last exon is from residues 114 to 133. Additionally, a 3' segment of 80 nucleotides was also obtained, corresponding to untranslated

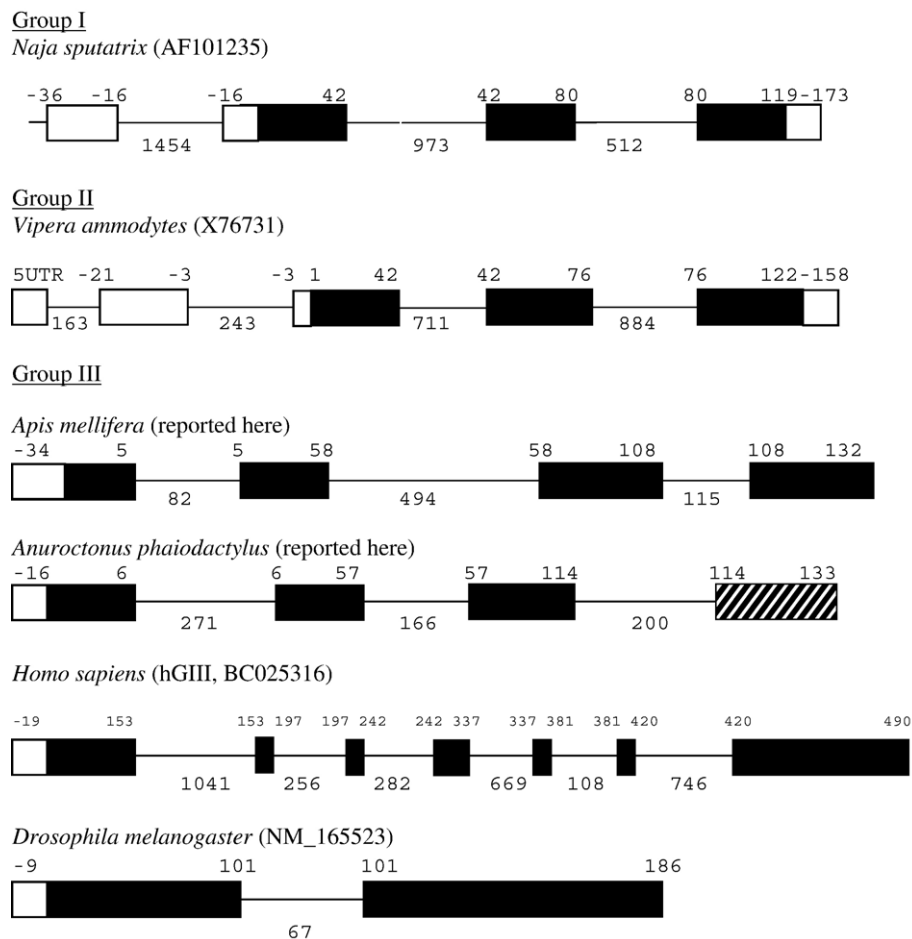
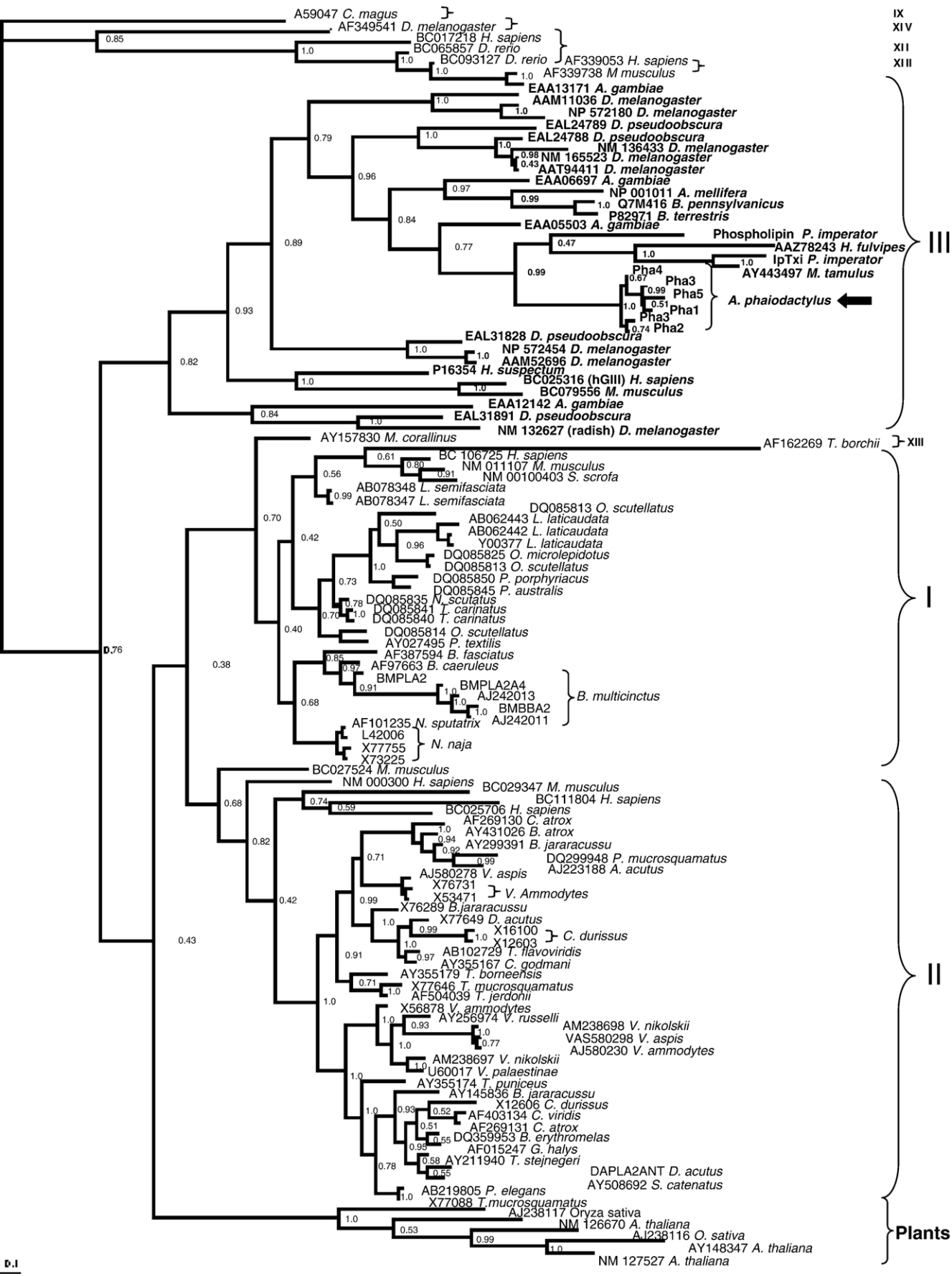


Fig. 3. Models indicating the gene structure of the PLA2 of groups I, II and III. A schematic representation of blocks for exons and lines for introns are used to show the gene structures of the various groups of PLA2 from *Naja naja sputatrix* (accession number AF101235), *Vipera ammodytes* (accession number X76731), *Apis mellifera* and *A. phaiodactylus* (this work), *Homo sapiens* hGIII (accession number BV025316) and *Drosophila* (accession number NM_165523). The empty boxes represent the regions that encode for signal peptides and the dashed block to the small subunit of *Pha4*. The numbers above the lines indicate the sequence positions of amino acids, whereas those below, the size of the introns.



sequences. A similar sequence organization was obtained for Phaiodactylipin (sequences not shown). Fig. 2B shows the genomic sequence obtained for the *A. mellifera* phospholipase. It is worth noting that the bee sequence (obtained in the present work) is different from that already reported by other authors for a putative phospholipase of *A. mellifera* strain DH4 (GenBank accession number AADG05004766.1). However the genomic organization is basically the same, with minor modifications at the level of the introns (about 95% identity between both sequences). The PLA2 described here belong to group III of the general classification of phospholipases. In order to compare our results with that available in the literature for other secretory PLA2 we analyzed the gene organization of PLA2 from groups I, II and III. Groups IV to VIII and groups X and XI are intracellular phospholipases, and were not used for comparison. Group IX was not included because the genomic organization is not known. Group V comprises PLA from mammals, and although composed by secretory enzymes was not included, because it is distantly located.

Fig. 3 shows the comparative results of the genomic organization of the secretory PLA2. As it can be observed groups I and II, represented by two different snake PLA2s, have a similar genomic organization in the region corresponding to the mature protein: three exons intercalated by two introns. The amino acid sequence of group III PLA2s includes different sources: bee (this work and the one deposited in the GenBank EF373554), scorpion, fruit fly, mosquito and one human; however the genomic organization is quite different. They have 4 exons (bee and scorpion), 7 exons (human) and from 2 to 6 for the mosquito and fruit fly. The bee venom PLA2 having four exons and three introns code for a single polypeptide chain, whereas the scorpion heterodimeric phospholipase codes for two mature polypeptide chains. In the last case, the fourth exon is the one that codes for the small subunit. The organization of the PLA2 gene (GenBank accession number NM_136433) from *Drosophila* has only two exons separated by one intron. The amino acid sequence of the *Drosophila* PLA2 reveals high similarity with the PLA2 of group III. The size of each intron/exon segment is indicated in the figure.

A comparative analysis of the introns of snake PLA2s shows some similarities in their nucleotide sequences (Fujimi et al., 2002; Guillemin et al., 2003), whereas in the Blast comparison we performed with the bee and scorpion PLA2 the similarities were practically non existent.

Comparison of the intron/exon organization of groups I and II with that of group III shows that the latter ones have an additional intron/exon. This seems to point in the same direction of that described by genes coding for snake phospholipases, where the segment situated at the most C-terminal region of the enzyme is the one presenting the highest degree of

variability (Kini and Evans, 1987). It is also important to observe that although the fly (*Drosophila*) is expected to be closely situated phylogenetically to insects (*Apis*) and scorpions (*Anuroctonus*), their most C-terminal located exon is the most dissimilar one (it is longer) among the genomic organization shown in Fig. 3.

3.3. Phylogenetic analysis

The program muscle 3.5 was used (Edgar, 2004) in order to generate sequence multiple alignments. One hundred and sixteen PLA2 sequences were selected from snakes of different genus as *Agkistrodon*, *Crotalus*, *Bothrops*, *Naja*, *Trimeresurus*, *Vipera*, *Oxyuranus*, *Bungarus*, *Notechis*, *Micrurus*, *Trimeresurus*, *Notechis*, *Gloydius* and *Laticauda*; from the snail (*C. magus*); from insects like *Drosophila*, *Anopheles*, *Bombus* and *Apis*, from the reptile (*Heloderma*) and fish (*Danio*), some from mammals (human and mouse) and a few ones from plants like *Arabidopsis* and *Oriza*. In this analysis the scorpion isoforms of Fig. 1 were included. The strategy described in Section 2.4. was repeated several times in order to obtain a robust phylogenetic result. Several independent groups were identified, as indicated in Fig. 4, specially group number III (highlighted), where all the isoforms of scorpion PLA2 are included, and *A. phaiodactylus* PLA2s are indicated with an arrow. The enzymes from the arthropods (scorpions and insects) are well clustered, but the ones from *Heloderma* (GenBank accession number P16354) and the human (GenBank accession number BC025316) are distantly located as well as some genomic sequences from *Drosophila* and *Anopheles*.

Additionally, a full likelihood analysis using the HYPHY package in the adaptive evolution server from <http://www.datamonkey.org/> was conducted with these sequences. The result of this analysis showed that the codons corresponding to the amino acids in positions: 5, 6, 24, 71, 77, 79, 92, 93, 117, and 127 are very likely under the effect of positive selection. When the amino acids corresponding to these positions are visualized in the three-dimensional structure of bee phospholipase (PDB accession number 1POC) it strongly supports the notion that this is the location of the greatest variability. Actually this region of the molecule is more C-terminally situated, in agreement with data reported by other authors (Kini and Chan, 1999; Chijiwa et al., 2003).

4. Discussion

In this communication we report the nucleotide sequences of different clones of PLA2, using as templates the cDNA and genomic materials from the scorpion *A. phaiodactylus*. The

Fig. 4. Phylogenetic tree of PLA2s. A multiple alignment of 116 secretory PLA2 sequences was generated, using data from the snakes of the genus: *Agkistrodon*, *Crotalus*, *Bothrops*, *Bungarus*, *Trimeresurus*, *Vipera*, and *Laticauda*. Additional sequences were obtained from the snail *Conus*, the insects: *Drosophila*, *Apis* and *Anopheles*, the mammals (*Homo* and *Mus*) and the reptile: *Heloderma* (see Materials and methods). Among the sequences from scorpion venoms are those of *Pandinus imperator*, *Heterometrus fulvipes*, *Mesobuthus tamulus* and *Anuroctonus phaiodactylus*. The consensus tree was visualized using the program TREEVIEW (Page, 1996). In order to avoid cumbersome names within the figure-tree, only group III indicates the accession numbers followed by the species name of the animal. The positions occupied by the PLA2 from this work are indicated by an arrow. All the other sequences are identified by the accession numbers, and they correspond to groups I, II and plants secretory PLA2 (major brackets followed by the Roman number). Also indicated are the accession numbers for the sequences of PLA2 of groups: IX, XII, XIII and XIV (see small brackets inside the tree). A complete tree containing the names of each sequence is available upon request to the authors.

cDNA sequences of five PLA2 are shown in Fig. 1, and the genomic sequence of *Pha4* is shown in Fig. 2, together with a PLA2 cloned from *A. mellifera*, for comparative purposes. *Pha1* to *Pha4* are assumed to be heterodimeric proteins, because they have the same hexaptide structure shown to be processed during maturation in the parent phospholipase named Phaiodactylipin (Valdez et al., 2004). *Pha4* was partially sequenced by automatic Edman degradation showing the presence of two amino acid sequences that correspond to the N-terminal region and to the small subunit sequences, exactly as expected from the nucleotide sequence, thereby confirming the assumption that it is in fact a heterodimeric protein in its mature form. A single gene at the level of the messenger is processed by excision of a hexapeptide, which is eliminated during the maturation process, exactly as earlier demonstrated for Phaiodactylipin and other heterodimeric phospholipases from scorpions, reported by our group (Valdez et al., 2004; Zamudio et al., 1997). However, *Pha5* probably is a single chain protein, because the segment corresponding to the putative processing site (hexapeptide) does not conform to the expected amino acid sequence.

Amino acid sequences obtained by direct Edman degradation of purified enzymes from the venom (HPLC profile not shown) do correspond to the amino acids expected for *Pha2* and *Pha4*. The first 28 amino acid residues (see Fig. 1) found for *Pha2* confirm the sequence. Position 5 is proline, position 6 is aspartic acid and position 24 shows phenylalanine, which are the sequences expected. For *Pha4* the first 23 amino acid residues were sequenced, showing proline and glycine for positions 5 and 6 respectively, whereas the additional sequence corresponding to the small subunit was aspartic acid followed by glutamic acid and alanine, which are the expected sequences for this isoform.

These direct sequencing data confirm the existence of isoforms of PLA2 both at the level of nucleotide sequences as authentic expressed proteins, although for *Pha1*, *Pha3* and *Pha5* direct evidence was not obtained. It is worth discussing the fact that most authors (Fan et al., 1999; Tsai et al., 2002; Fujimi et al., 2002; Rojnuckarin et al., 2006) dealing with similar situations (PLA2 obtained from snake sources) have not demonstrated the existence of the real amino acid sequences from protein directly purified from the venoms. Here, we have at least three isoforms (Phaiodactylipin, *Pha2* and *Pha4*) for which this has been clearly demonstrated.

The numbering system shown in Fig. 1 assumes that the sequence from positions -16 to -1 correspond to the signal peptides, and those from number 1 to 132 to the mature phospholipases, except for the underlined hexapeptides in positions 109 to 114, which are usually excised and eliminated during maturation of the heterodimeric PLA2s.

It is worth mentioning that the segment tentatively assumed to correspond to the signal peptides, starting at position -16 with a methionine, has several charged residues, usually not present in eukaryotic organisms. Similar results were reported for components of other venoms, as described elsewhere, for toxins Pn2-1 and Tx2-1 of the spider *Phoneutria nigriventer* (Kalapothakis et al., 1998). The amino acids preceding the first residue of the

mature PLA2 (FEEER) are similar to those of the propeptides of the spider toxins (VEELER). They are rich in glutamic acids and the last residue is the basic arginine (Kalapothakis et al., 1998).

The amino acids corresponding to positions 6 to 12 are the ones proposed to be the Ca^{2+} binding segment and those from positions 29 to 37 correspond to the catalytic site (segments indicated in Fig. 1). Both of these segments are highly conserved. Point mutations of the various isoforms are located in other positions of the sequences, but the degree of identity is very high (over 90%). Both the large and the small subunits show sequence similarities to that of Phaiodactylipin, and also with those of heterodimeric phospholipases purified from *P. imperator* venom: IpTxI (32%) and Phospholipin (49%), respectively (see Zamudio et al., 1997 and Conde et al., 1999). When these sequences (Fig. 1) are compared to those of phospholipases of bee and *Heloderma* venom the similarities are smaller. Only 28% and 26% identity were found, respectively.

Fig. 3 shows a comparative analysis of the PLA2 found for the scorpion and bee venoms (see Fig. 2) with those of snakes, human and *Drosophila* genes. In Fig. 3 the entire gene structure is clearly indicated, highlighting the sections of each intron and exon of the corresponding genes. For the PLA2 of the scorpion and that of the bee, there are three introns and four exons. However, the gene organization of those from *Drosophila* and human hGIII, showed a distinct structure (Fig. 3). The structure arrangement for the genes starting at the propeptide to the C-terminal ends of groups I and II from snakes is also different; they show only two introns within the sequence corresponding to the mature protein.

It is worth noticing that the first intron separates the calcium binding domain and the last intron separates the small subunits of the heterodimeric PLA2s. For both introns (the first and the last) of scorpion and bee PLA2 the same amino acids separate the introns from exons (Gly and Arg, respectively) as also indicated in Fig. 2.

For the phylogenetic analysis we have selected only representative sequences from the PLA2, for which the gene structures were deposited in data banks at the time of the analysis. A complete analysis of the chosen sequences is shown in Fig. 4. The most variable region of the phospholipases is located at the amino- and carboxyl-terminal ends, whereas the most conserved segments correspond to the central part, mainly at the α -helices, where the catalytic site and the calcium binding region are located.

For the scorpions, bees, wasps, flies, mosquitoes and *Heloderma* the most variable region is located at the C-terminal region. All these phospholipases belong to group III.

Concerning the genomic structure of the genes coding for the mature protein PLAs, groups I and II in general have three exons separating this region (Kordis et al., 1998; Jeyaseelan et al., 2000; Fujimi et al., 2002; Guillemain et al., 2003; Chijiwa et al., 2003). On the contrary, group III is highly variable, showing from two to six exons. In group III, the genomic structure of the scorpion *A. phaiodactylus* and the bee *A. mellifera* are closely related. The genes for both arthropods contain four exons.

It is worth mentioning that concerning the known functions of the PLAs included in group III, there is a wide diversity of activities reported. For example, the phospholipase “radish” found in *Drosophila* defines a neural circuit involved in anesthesia-resistant memory (Chiang et al, 2004; Folkers et al, 2006). The gene structure of this phospholipase has six exons, and has an ortholog in the mosquito *A. gambiae* (see Fig. 4). The bee phospholipases are allergenic to humans (Jilek et al., 2001), whereas those from the scorpion reported here are toxic to insects.

Thus while there is a wide variability within the genomic structure, and they play distinct functions, they all converge to group III, when analyzed according to their phylogeny, as shown in Fig. 4.

Acknowledgements

The authors wish to acknowledge the assistance received from Shirley Ainsworth B.A., during the writing of this manuscript.

Dr. Ricardo Ayala is kindly acknowledged for having identified the bee species used in this publication. This work was partially supported by grants from DGAPA (*Dirección General de Asuntos del Personal Académico*) of UNAM and the National Council of Science and Technology (Mexican Government).

References

- Aleksiev, B., Shipolini, R., 1971. Further studies on the fractionation and purification of the toxic components from the venom of the Bulgarian viper (*Vipera ammodytes ammodytes*). Hoppe Seylers Z. Physiol. Chem. 352, 1183–1188.
- Benishin, C.G., 1990. Potassium channel blockade by the B subunit of beta-bungarotoxin. Mol. Pharmacol. 38, 164–169.
- Bieber, A.L., et al., 1990. The complete sequence of the acidic subunit from Mojave toxin determined by Edman degradation and mass spectrometry. Biochim. Biophys. Acta 1037, 413–421.
- Bon, C., Changeux, J.P., Jeng, T.W., Fraenkel-Conrat, H., 1979. Postsynaptic effects of crotoxin and of its isolated subunits. Eur. J. Biochem. 99, 471–481.
- Chang, C.C., Lee, C.C., Eaker, D., Fohlman, J., 1977. Short communications the presynaptic neuromuscular blocking action of taipoxin. A comparison with beta-bungarotoxin and crotoxin. Toxicon 15, 571–576.
- Chen, Y.H., Wang, Y.M., Hseu, M.J., Tsai, I.H., 2004. Molecular evolution and structure-function relationships of crotoxin-like and asparagine-6-containing phospholipases A2 in pit viper venoms. Biochem. J. 381, 25–34.
- Chiang, A.S., et al., 2004. Radish encodes a phospholipase-A2 and defines a neural circuit involved in anesthesia-resistant memory. Curr. Biol. 14, 263–272.
- Chijiwa, T., et al., 2003. Interisland mutation of a novel phospholipase A2 from *Trimeresurus flavoviridis* venom and evolution of Crotalinae group II phospholipases A2. J. Mol. Evol. 57, 546–554.
- Chirgwin, J.M., Przyvyla, A.E., McDonald, R.J., Rutter, W.J., 1979. Isolation of biologically active ribonucleic acid from source enriched in ribonuclease. Biochem. 18, 5294–5299.
- Conde, R., Zamudio, F., Becerril, B., Possani, L.D., 1999. Phospholipin, a novel heterodimeric phospholipase A2 from *Pandinus imperator* scorpion venom. FEBS Lett. 460, 447–450.
- Condrea, E., Yang, C.C., Rosenberg, P., 1980. Comparison of a relatively toxic phospholipase A2 from *Naja nigricollis* snake venom with that of a relatively non-toxic phospholipase A2 from *Hemachatus haemachatus* snake venom—I. Enzymatic activity on free and membrane bound substrates. Biochem. Pharmacol. 29, 1555–1563.
- Corona, M., Valdez-Cruz, N.A., Merino, M., Possani, L.D., 2001. Genes and peptides from the scorpion *Centruroides sculpturatus* Ewing, that recognize Na⁺-channels. Toxicon 39, 1893–1898.
- Delot, E., Bon, C., 1993. Model for the interaction of crotoxin, a phospholipase A2 neurotoxin, with presynaptic membranes. Biochemistry 32, 10708–10713.
- Deregnacourt, C., Schrevel, J., 2000. Bee venom phospholipase A2 induces stage-specific growth arrest of the intraerythrocytic *Plasmodium falciparum* via modifications of human serum components. J. Biol. Chem. 275, 39973–39980.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
- Fan, C., Qian, Y., Yang, S., Gong, Y., 1999. cDNA cloning and sequence analysis of Lys-49 Phospholipase A2 from *Agkistrodom acutus*. Gen. Anal. Biomol. eng. 15, 15–18.
- Fenard, D., Lambeau, G., Maurin, T., Lefebvre, J.C., Doglio, A., 2001. A peptide derived from bee venom-secreted phospholipase A2 inhibits replication of T-cell tropic HIV-1 strains via interaction with the CXCR4 chemokine receptor. Mol. Pharmacol. 60, 341–347.
- Fletcher, J.E., Rapuano, B.E., Condrea, E., Yang, C.C., Rosenberg, P., 1981. Relationship between catalysis and toxicological properties of three phospholipases A2 from elapid snake venoms. Toxic. Appl. Pharmacol. 59, 375–382.
- Folkers, E., Waddell, S., Quinn, W.G., 2006. The *Drosophila* radish gene encodes a protein required for anesthesia-resistant memory. Proc. Natl. Acad. Sci. U.S.A. 103, 17496–17500.
- Fujimi, T.J., Kariya, Y., Tsuchiya, T., Tamiya, T., 2002. Nucleotide sequence of phospholipase A(2) gene expressed in snake pancreas reveals the molecular evolution of toxic phospholipase A(2) genes. Gene 292, 225–231.
- Gerrard, J.M., Robinson, P., Narvey, M., McNicol, A., 1993. Increased phosphatidic acid and decreased lysophosphatidic acid in response to thrombin is associated with inhibition of platelet aggregation. Biochem. Cell. Biol. 71, 423–439.
- Guillemin, I., Bouchier, C., Garrigues, T., Wisner, A., Choumet, V., 2003. Sequences and structural organization of phospholipase A2 genes from *Vipera aspis aspis*, *V. aspis zinnikeri* and *Vipera berus berus* venom. Identification of the origin of a new viper population based on ammodytin II heterogeneity. Eur. J. Biochem. 270, 2697–2706.
- Gutiérrez, J.M., Lomonte, B., 1995. Phospholipase A2 myotoxins from Bothrops snake. Toxicon 33, 1405–1424.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
- Jeyaseelan, K., Armugam, A., Donghui, M., Tan, N.H., 2000. Structure and phylogeny of the venom group I phospholipase A(2) gene. Mol. Biol. Evol. 17, 1010–1021.
- Jilek, S., Barbey, C., Spertini, F., Cortesby, B., 2001. Antigen-independent suppression of the allergic immune response to bee venom phospholipase A (2) by DNA vaccination in CBA/J mice. J. Immunol. 166, 3612–3621.
- Kalopothakis, E., et al., 1998. Cloning of cDNAs encoding neurotoxic peptides from the spider *Phoneutria nigriventer*. Toxicon 36, 1843–1850.
- Kini, R.M., Evans, H.J., 1987. Structure-function relationships of phospholipases. The anticoagulant region of phospholipases A2. J. Biol. Chem. 262, 14402–14407.
- Kini, R.M., Chan, Y.M., 1999. Accelerated evolution and molecular surface of venom phospholipase A2 enzymes. J. Mol. Evol. 48, 125–132.
- Kondo, K., Toda, H., Narita, K., Lee, C.Y., 1982. Amino acid sequences of three beta-bungarotoxins (beta 3-, beta 4-, and beta 5-bungarotoxins) from *Bungarus multicinctus* venom. Amino acid substitutions in the A chains. J. Biochem. (Tokyo) 91, 1531–1548.
- Kordis, D., Bdojlah, A., Gubensek, F., 1998. Positive Darwinian selection in *Vipera palaestinae* phospholipase A2 genes is unexpectedly limited to the third exon. Biochem. Biophys. Res. Commun. 251, 613–619.
- Lind, P., Eaker, D., 1982. Amino-acid sequence of the alpha-subunit of taipoxin, an extremely potent presynaptic neurotoxin from the Australian snake taipan (*Oxyuranus s. scutellatus*). Eur. J. Biochem. 124, 441–447.
- Mancheva, I., Kleinschmidt, T., Aleksiev, B., Braunitzer, G., 1984. Sequence homology between phospholipase and its inhibitor in snake venom. The primary structure of the inhibitor of vipoxin from the venom of the Bulgarian viper (*Vipera ammodytes ammodytes*, Serpentes). Hoppe Seylers Z. Physiol. Chem. 365, 885–894.
- McIntosh, J.M., et al., 1995. Conodipine-M, a novel phospholipase A2 isolated from the venom of the marine snail *Conus magus*. J. Biol. Chem. 270, 3518–3526.
- Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

- Possani, L.D., et al., 1992. Isolation and physiological characterization of taicatoxin, a complex toxin with specific effects on calcium channels. *Toxicon* 30, 1343–1364.
- Rajashankar, K.R., Inn-Ho Tsai, I.H., Betzela, C., 1999. Crystallization and preliminary X-ray diffraction studies of the neurotoxic, heterodimeric phospholipase A2 from the Taiwan viper (*Vipera russelli formosensis*). *Acta Crystallogr., D Biol. Crystallogr.* 55, 1064–1065.
- Rojnuckarin, P., Muanpasitporn, C., Chanhom, L., Arpijuntarakoon, J., Intragumtornchai, T., 2006. Molecular cloning of novel serine proteases and phospholipases A2 from green pit viper (*Trimeresurus albolabris*) venom gland cDNA library. *Toxicon* 47, 279–287.
- Rouault, M., Bollinger, J.G., Lazdunski, M., Gelb, M.H., Lambeau, G., 2003. Novel mammalian group XII secreted phospholipase A2 lacking enzymatic activity. *Biochemistry* 42, 494–503.
- Scott, D.L., Otwinowski, Z., Gelb, M.H., Sigler, P.B., 1990. Crystal structure of bee-venom phospholipase A2 in a complex with a transition-state analogue. *Science* 250, 1563–1566 Erratum in: *Science* 1991, 252, 764.
- Six, D.A., Dennis, E.A., 2000. The expanding superfamily of phospholipase A (2) enzymes: classification and characterization. *Biochim. Biophys. Acta* 1488, 1–19.
- Tischfield, J.A., 1997. A reassessment of the low molecular weight phospholipase A2 gene family in mammals. *J. Biol. Chem.* 272, 17247–17250.
- Tsai, I.H., Wang, Y.M., Au, L.C., Ko, T.P., Chen, Y.H., Chu, Y.F., 2000. Phospholipases A2 from *Callosellasma rhodostoma* venom gland. *Eur. J. Biochem.* 267, 6684–6691.
- Tsai, I.H., Hsu, H.Y., Wang, Y.M., 2002. A novel phospholipase A(2) from the venom glands of *Bungarus candidus*: cloning and sequence-comparison. *Toxicon* 40, 1363–1367.
- Tsai, I.H., Wang, Y.M., Chen, Y.H., Tsai, T.S., Tu, M.C., 2004. Venom phospholipases A2 of Bamboo viper (*Trimeresurus stejnegeri*): molecular characterization, geographic variations, and evidence of multiple ancestries. *Biochem. J.* 377, 215–223.
- Valdez-Cruz, N.A., Batista, C.V., Possani, L.D., 2004. Phaiodactylipin, a glycosylated heterodimeric phospholipase A from the venom of the scorpion *Anuroctonus phaiodactylus*. *Eur. J. Biochem.* 271, 1453–1464.
- Valentin, E., Lambeau, G., 2000. Increasing molecular diversity of secreted phospholipases A(2) and their receptors and binding proteins. *Biochim. Biophys. Acta* 1488, 59–70.
- Valentin, E., Ghomashchi, F., Gelb, M.H., 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.
- Zamudio, F., 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.