Conserved domains and evolution of secreted phospholipases A₂

Timo J. Nevalainen¹, João C. R. Cardoso² and Pentti T. Riikonen³

¹ Department of Pathology, University of Turku and Turku University Hospital, Finland
² Comparative Molecular Endocrinology, Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, Faro, Portugal
³ Department of Information Technology, University of Turku, Finland

Keywords
conserved domains; eukaryotes; evolution; prokaryotes; secreted PLA₂s

Correspondence
Timo J. Nevalainen, Department of Pathology, University of Turku, Kiinamyllynkatu 10, FIN-20520 Turku, Finland
Fax: +358 2 3337456
Tel: +358 2 3337500
E-mail: timneva@utu.fi

(Received 1 September 2011, revised 8 December 2011, accepted 13 December 2011)


Secreted phospholipases A₂ (sPLA₂s) are lipolytic enzymes present in organisms ranging from prokaryotes to eukaryotes but their origin and emergence are poorly understood. We identified and compared the conserved domains of 333 sPLA₂s and proposed a model for their evolution. The conserved domains were grouped into seven categories according to the in silico annotated conserved domain collections of ‘cd00618: PLA₂-like’ and ‘pfam00068: Phospholip_A2_1’. PLA₂s containing the conserved domain cd04706 (plant-specific PLA₂) are present in bacteria and plants. Metazoan PLA₂s of the group (G) I/II/V/X PLA₂ collection exclusively contain the conserved domain cd00125. GIII PLA₂s of both vertebrates and invertebrates contain the conserved domain cd04704 (bee venom-like PLA₂), and mammalian GIII PLA₂s also contain the conserved domain cd04705 (similar to human GIII PLA₂). The sPLA₂s of bacteria, fungi and marine invertebrates contain the conserved domain pfam09056 (prokaryotic PLA₂) that is the only conserved domain identified in fungal sPLA₂s. Pfam06951 (GXII PLA₂) is present in bacteria and is widely distributed in eukaryotes. All conserved domains were present across mammalian sPLA₂s, with the exception of cd04706 and pfam09056. Notably, no sPLA₂s were found in Archaea. Phylogenetic analysis of sPLA₂ conserved domains reveals that two main clades, the cd- and the pfam-collection, exist, and that they have evolved via gene-duplication and gene-deletion events. These observations are consistent with the hypothesis that sPLA₂s in eukaryotes shared common origins with two types of bacterial sPLA₂s, and their persistence during evolution may be related to their role in phospholipid metabolism, which is fundamental for survival.

Introduction

Phospholipases A₂

Phospholipases A₂ (PLA₂s; EC 3.1.1.4) are a group of lipolytic enzymes that hydrolyze the sn-2 bond of phospholipids, such as phosphatidyleholine and phosphatidyethanolamine, resulting in the release of fatty acid and lysophospholipid. They have been isolated from organisms ranging from bacteria to mammals and are suggested to have emerged early in evolution. In general, PLA₂s are classified into three broad categories: Ca²⁺-dependent secreted PLA₂s (sPLA₂s); Ca²⁺-dependent cytosolic PLA₂s; and Ca²⁺-independent cytosolic PLA₂s. Cytosolic PLA₂s are known to play an important role in cellular signalling and

Abbreviations
EST, expressed sequence tag; G, group; ML, maximum likelihood; NJ, neighbour-joining; OtoC, N-terminal region of otoconin; OtoN, N-terminal region of otoconin; PLA₂, phospholipase A₂; sPLA₂, secreted phospholipase A₂.
prostanoid metabolism [1,2]. Secreted PLA2s are components of various body fluids, including blood plasma, pancreatic juice, tears, seminal fluid, and snake and other venoms, and they participate in diverse physiological and pathological functions, such as in the digestion of dietary phospholipids, in inflammatory reactions and in the defence against bacteria and other pathogens [1–7].

**Secreted PLA2s**

Secreted PLA2s are the product of distinct genes and they have been classified, according to their molecular structure, into the following groups: IA, IB, IIA, IIB, IIC, IID, IIE, IIF, III, V, IX, X, XIA, XIB, XIIA, XIIIB, XIII and XIV [1]. Evidence of structural similarity among group (GI), GII, GV and GX members was taken to suggest that they may form a distinct GI/II/V/X PLA2 collection [8]. The GXIV PLA2s of bacteria and fungi differ in their primary structure and folding from the other sPLA2s [1].

Secreted PLA2s are small-molecular-mass proteins (14-18 kDa, 120-135 amino acid residues) that require the presence of Ca2+ at millimolar concentrations for their catalytic activity. They share a conserved 3D structure that is stabilized by five to eight disulfide bonds. The structure of the GI, GII and GX PLA2s consists of three α-helices, a two-stranded β-sheet (the β-wing) and a conserved Ca2+-binding loop [9–13]. The conserved catalytic network of GIB and GIIA PLA2s consists of hydrogen-bonded side-chains formed by histidine, which is localized in the long α-helix1, tyrosine and aspartic acid residues, as well as the hydrophobic wall that shields it [14]. The sPLA2s of plants have a different 3D structure. In the GXIB PLA2 of rice (Oryza sativa) the β-wing is absent and the C-terminal α-helix3 has a different orientation [15]. The sPLA2s of prokaryotes and fungi are characterized by a dominant α-helical fold [16]. The 3D structure of GIIA PLA2 differs from that of GI/II/V/X PLA2s but they share identical motifs [17].

**Conserved domains of sPLA2s and their evolution**

Functional motifs of sPLA2s that are well conserved include the Ca2+-binding and catalytic sites, and the conserved cysteine residues and disulfide bond pattern [1,2]. In silico database annotations identified two conserved signature patterns: the ‘PA2_HIS (PS0018, Phospholipase A2 histidine active site C-C-[P]-x-H-{LGY}-x-C’, where x represents a nonconserved amino acid, and amino acids within square brackets are not allowed) that contains the histidine residue of the sPLA2 active site, and the ‘PA2_ASP (PS00119, Phospholipase A2 aspartic acid active site [LIVMA]-C-{LIVMFYWPCST}-C-D-[GS]-[G]-[N]-x-{QS}-C’, where x represents a nonconserved amino acid, amino acids within curly brackets are not allowed and amino acids within square brackets are allowed) centred on the active site aspartic acid residue and localized towards the C-terminal portion of the molecule [18]. The Ca2+-binding motif, which, for example in mammalian GIB sPLA2s, is Y-x-G-x-G (where x represents a nonconserved amino acid) is localized before the histidine catalytic site towards the N-terminus.

Studies of sPLA2s have focussed mainly on eukaryotes [1–4]. However, the availability of molecular data from phylogenetically distant organisms, and the identification of new sPLA2s, challenges the current classification system, especially when applied to invertebrate and prokaryote sPLA2s [1,19–21]. Despite their potential evolutionary relationships [22,23], the origin and emergence of sPLA2s is still intriguing. Although all sPLA2s share the same catalytic mechanism involving the canonical histidine residue, there is considerable variation in their sequence identity. Based on their identity, members of the GI/II/V/X PLA2 collection have been designated ‘conventional sPLA2s’ and are believed to be evolutionally close, whereas GIII and GXII PLA2s are classified as ‘atypical sPLA2s’ and are evolutionally more distant [4,8]. In bacteria and fungi there are sPLA2s that show only limited sequence identity with other sPLA2s [1,16], and to date the evolutionary connection between these structurally distant proteins is lacking.

The present study aimed to contribute to the understanding of the origin and evolution of sPLA2s based upon the identification of their conserved domains in a representative collection of prokaryotes and eukaryotes. Recently, based on published data and the collection of well-annotated multiple sequence alignment models of their conserved domains, the family hierarchy of sPLA2s was comprehensively classified in two general collections: ‘cd00618 PLA2_like: Phospholipase A2, a superfamily of secretory and cytosolic phospholipases A’ and ‘pfam00068: Phospholipid_A2_1, Phospholipase A2’ [24,25]. In the present study, secreted PLA2 protein sequences retrieved from public databases were classified according to their conserved domain structures. Their sequences were compared and specific motifs within each subfamily group were identified, and, based upon their sequence and structure similarity, a model for their origin and evolution was proposed.
Results

Repertoire of prokaryotic and eukaryotic sPLA2s

A total of 333 sPLA2 sequences from a wide range of taxa were retrieved from publicly available databases, and 358 conserved domains within the cd00618 (PLA2_−like) and pfam00068 (Phospholip_A2_1) collections were identified (Tables 1 and S1). Members of the cd-collection identified are grouped into five distinct subfamilies: the ‘cd04706 PLA2_plant: Plant-specific subfamily of Phospholipase A2’; ‘cd00125 PLA2c: Phospholipase A2, a family of secretory and cytosolic enzymes’; ‘cd04707 otoconin_90: Phospholipase A2-like domains present in otoconin-90 and otoconin-95’; ‘cd04704 PLA2_bee_venom_like: A subfamily of Phospholipase A2, similar to bee venom PLA2’ and ‘cd04705 PLA2_group_III_like: A subfamily of Phospholipase A2, similar to human group III PLA2’. The pfam-collection members are subdivided into ‘pfam06951: PLA2G12, group XII secretory phospholipase A2 precursor’ and ‘pfam09056 Phospholip_A2_3: Prokaryotic phospholipase A2 domain found in PLA2s of bacteria and fungi’ [24,25, http://www.ncbi.nlm.nih.gov/cdd].

All conserved domains of sPLA2s identified are present in representatives of the Kingdom Animalia with the exception of cd04706 of the sPLA2s of bacteria and plants (Table 1). Within this Kingdom, sPLA2s containing the conserved domain cd00125 are widespread in organisms ranging from basal metazoa (Porifera, Placozoa, Cnidaria and Rotifera), protostomes (Insects, Nematodes, Molluscs and Arthropods) and early deuterostomes (Echinodermata, Cephalochordata and Tunicata) to teleosts and tetrapods (Amphibia, Aves and Mammalia). The conserved domain cd00125 is typical of GIA, IB, IIA, IIB, ICC, IID, III, IIF, V and X sPLA2s. The conserved domain cd04707 was found in the N- and C-terminal regions of vertebrate otconins. The conserved domains cd04704 and cd04705 are present in metazoan GIII PLA2s, the latter domain exclusively in mammals. The sPLA2s of bacteria contain either cd04706 or pfam09056 (present in XIV PLA2s) and exceptionally pfam06951. In contrast, no sPLA2s were found in archaea. Pfam09056 is present in the sPLA2s of fungi. The majority of plant sPLA2s contained the conserved domain cd04706 (identified in XIA and XIB PLA2s). The conserved domain pfam06951 is present in the GXIIA PLA2s of unicellular and multicellular organisms, including marine algae and bacteria (Tables 1 and S1).

Conserved domain cd04706 of PLA2s of bacteria and plants

The sPLA2s of bacteria and plants share a number of conserved structural features for the histidine and aspartic acid catalytic motifs and Ca2+-binding domains (Fig. 1). The conserved domain cd04706 was identified in the sPLA2s of bacteria of the class Alphaproteobacteria of the phylum Proteobacteria and also in those of the phylum Firmicutes, which includes both Gram-negative and Gram-positive bacteria, such as the human pathogens Streptococcus pyogenes, Clostridium perfringens, Clostridium botulinum and Bacillus cereus [26] (Table S1). Secreted PLA2s containing cd04706 were also identified in numerous plants. The sPLA2s of O. sativa are well characterized and classified as GXI PLA2s [15,27]. In general, several GXI PLAS isoforms were identified in plants, which may have resulted from gene-duplication events, for example, four GXI PLA2s were identified in O. sativa (Table S1). The GXIB PLA2 of O. sativa is a 16.6 kDa protein in which the Ca2+-binding site contains tyrosine, glycine and aspartic acid residues, and the histidine residue of the active site is centred in the catalytic site motif (Fig. 1A).

Table 1. Distribution of the conserved domains of sPLA2 in the Kingdoms of Life. +, presence; −, absence.

<table>
<thead>
<tr>
<th>sPLA2 conserved domain</th>
<th>Group number</th>
<th>Archaea</th>
<th>Bacteria</th>
<th>Viridiplantae</th>
<th>Fungi</th>
<th>Animalia</th>
</tr>
</thead>
<tbody>
<tr>
<td>cd00125</td>
<td>IA, IB, IIA, IIB, IIC, IID, IIE, IIF, V, X</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>cd04704</td>
<td>III</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>cd04705a</td>
<td>III</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>cd04706</td>
<td>XIA, XIB</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>cd04707</td>
<td>Otoconin</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>pfam06951</td>
<td>XIA, XIB</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>pfam09056</td>
<td>XIV</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

a Group number according to the classification in Schaloske and Dennis [1]. b cd04705 is found only in mammalian GIII PLA2s.
aspartic acid residue. In the GXIB PLA2 of function [15], instead of the more commonly occurring molecule, which probably contributes to the catalytic an asparagine residue in the C-terminal portion of the amino acid sequence identity. In both proteins, there is sPLA2 contains five cysteine residues, four of which bonds [15], while the conserved domain of there are 12 cysteine residues that form six disulfide (C1-C2 and C3-C5), align with the plant GXIB PLA2 cysteines (C4-C8 and C9-C10) and form conserved disulfide bonds, and are suggestive of structural similarity (Fig. 1B).

Conserved domain cd00125 of sPLA2 of animals

The conserved domain cd00125 was identified in the basal invertebrates Placozoa up to mammals and was prevalent in the vertebrate GI/II/V/X PLA2 collection where the functional sites and cysteine residues are highly conserved (Fig. 2). Members of this collection share 26–50% amino acid sequence identity within the conserved domain region. Among the members of the GI/II/V/X PLA2 collection, the catalytic histidine is centred in the active site motif and the aspartic acid residue of the active site is present in the C-terminal portion of the molecule. In the Ca2+-binding site, Ca2+ is bound by tyrosine and two glycines and an aspartic acid adjacent to the catalytic histidine (Fig. 2).

Homologues of GI/II/V/X PLA2 collection members are present also in invertebrate genomes, such as the sea anemone Nematostella vectensis, where six genes were identified (Table S1) [28]. The sPLA2 of the sea anemone Adamsia carciiniopados (also called Adamsia palliata) contains the pancreatic loop characteristic of vertebrate GI PLA2s and lacks the C-terminal extension found in GII PLA2s but shares homology for both GI and GII PLA2s [19]. Comparison of the disulfide bond positions of the A. carciiniopados sPLA2 with those of the human GI/II/V/X PLA2 collection indicates that the sea anemone sPLA2 shares five conserved disulfide bonds with the vertebrate homologues, suggesting structural, and possibly functional, conservation across the phylogenetically distant organisms of Cnidaria and Mammalia (Fig. 3).

Conserved domain cd04707 of otoconins

Otoconins containing the conserved domain cd04707 were identified in a number of vertebrates, from fish to mammals, and two conserved domains within the N-terminal (OtoN) and C-terminal (OtoC) regions of the same collection were identified within the mature protein sequence (Tables 1 and S1). The human otoconin-90 is a 53 kDa protein, and the OtoN and OtoC domains are 37% identical and structurally closely related to the members of the GI/II/V/X PLA2 collection, suggesting common ancestry (Fig. 2). The human OtoN and OtoC domains are 36% and 34% identical, respectively, when compared with GIB PLA2. Moreover, the disulfide bond patterns of human otoconin and GIB PLA2 are identical (Fig. 3). Both OtoN and OtoC domains contain a histidine residue within the conserved catalytic site. However, they are believed to be catalytically inactive as a result of mutations in the Ca2+-binding sites (Fig. 2), leading to loss of the usual Ca2+-binding residues in the OtoN domain, and in the OtoC domain the conserved second glycine of the mammalian GIB sPLA2s (Y-x-G-x-G) is replaced with a glutamic acid, although PLA2 activity remains to be investigated [29].

Conserved domain cd04704 of group III PLA2s

Secreted PLA2s, containing the conserved domain cd04704 (bee venom-like PLA2), were identified in arthropods, reptiles and vertebrates, including humans.
The honeybee *Apis mellifera* venom GIII PLA$_2$ shares considerable sequence identity with the members of the GI/II/V/X PLA$_2$ collection (e.g. 44% identity with the human GIV PLA$_2$) and is related in its 3D structure and catalytic mechanism to GI and GII PLA$_2$s [11,17]. The backbone of the GIII PLA$_2$ molecule contains the conserved Ca$^{2+}$-binding loop with tryptophan, glycine and aspartic acid residues and the conserved catalytic histidine and aspartic acid residues (Fig. 4). Human GIII PLA$_2$ is a 57 kDa protein in which cd04704 is localized in the middle part of the molecule and flanked by N- and C-terminal extensions [30]. The cd04704 of human GIII PLA$_2$ displays features similar to the arthropod GIII PLA$_2$s, including the Ca$^{2+}$-binding and catalytic site residues and also the 10 conserved cysteines that form five disulfide bonds at similar locations, and shares 44% sequence identity with honeybee venom GIII PLA$_2$.

**Conserved domains of sPLA$_2$s**

*Fig. 2.* Multiple sequence alignment of the conserved domains of human sPLA$_2$s. HaspiB$_{25}$ (GIB PLA$_2$), HaspiI$_{25}$ (GIIA), HaspiII$_{25}$ (GIIC), HaspiID$_{25}$ (GIID), HaspiIE$_{25}$ (GIE), HaspiIF$_{25}$ (GIIF), HaspiV$_{25}$ (GV) and HaspiX$_{25}$ (GX) PLA$_2$s contain the conserved domain cd00125 and belong to the structurally related GI/II/V/X PLA$_2$ collection. HaspiII$_{04}$ and HaspiII$_{05}$ (GIIA PLA$_2$) represent cd04704 and cd04705, respectively. HaspiGXIIA$_{51}$ and HaspiGXIIB$_{51}$ contain pfam06951. HaspiOtoN$_{07}$ and HaspiOtoC$_{07}$ contain the N- and C-terminal sections of cd04707 of otoconin-90, respectively. The catalytic histidine (H), aspartic acid (D) and asparagine (N) residues are highlighted in red, and the Ca$^{2+}$-binding amino-acid residues tyrosine (Y), glycine (G), histidine (H), phenylalanine (F), tryptophan (W), proline (P), leucine (L) and aspartic acid (D) are highlighted in blue.

The conserved domain cd04705 of group III PLA$_2$.

Proteins containing the conserved domain cd04705 (similar to human group III PLA$_2$) were found only in mammals (Table 1). The conserved domain cd04705 is localized in the C-terminal part of the GII PLA$_2$ molecule and is structurally unrelated to cd04704 (Fig. 2). In contrast to cd04704, cd04705 is considered to be catalytically inactive and its function is unknown [30].
different in their molecular structure from the bacterial sPLA2s containing cd04706. Comparison of the conserved domains of *Streptomyces violaceoruber* (pfam09056) and *S. pyogenes* (cd04706) sPLA2s indicates that they share only 8% sequence identity. The catalytic motif of the sPLA2 of the Gram-positive bacterium *S. violaceoruber* [16,31] contains the conserved histidine and aspartic acid residues (Figs 2 and 5). The conserved Ca$^{2+}$-binding domain is absent and the binding of Ca$^{2+}$ is mediated via the aspartic acid and leucine residues upstream of the catalytic site and by the aspartic acid residue adjacent to the catalytic histidine residue (Figs 2 and 5). There are four cysteines that form two putative disulfide bonds, including a bond that connects the catalytic and the Ca$^{2+}$-binding sites (Fig. 5).

In addition to bacteria, sPLA2s containing pfam09056 were also identified in both unicellular and multicellular fungi. The sPLA2 of the fungus *Tuber borchii* [32] is a 23 kDa protein that contains two conserved disulfide bonds analogous to those of the sPLA2 of *S. violaceoruber* (Fig. 5), and the two conserved domains are 42% identical.

There are pfam09056-containing sPLA2s in aquatic invertebrates of the phylum Cnidaria, including the sea anemone *N. vectensis* and the hydrozoan *Hydra magnipapillata*. Cnidarian sPLA2s have the conserved catalytic histidine and aspartic acid residues but contain more numerous cysteine residues and putative disulfide bonds than the corresponding bacterial and fungal sPLA2s (Fig. 5).

Recently, expressed sequence tags (ESTs), coding for potential sPLA2s containing pfam09056, were identified in both unicellular and multicellular fungi. The sPLA2 of the protist *Astrammina rara* (phylum Foraminifera) and the scallop *Mizuhopencten yessoensis* (phylum Mollusca) [33]. Previously, a related sPLA2 was isolated from the venom of another mollusc – the marine snail *Conus magus* – and classified as GIX PLA2 [34] but no conserved domain was identified for this protein in the present study, probably because of the incompleteness of its sequence. However, conserved histidine and aspartic acid active-site motifs are present in the pfam09056-containing sPLA2s of the scallop *M. yessoensis* and the sea anemone *N. vectensis* and in the GIX PLA2 of the *H. carciniopados* and human GV, GIB, GI A and GX PLA2s and otoconin-90. Five disulfide bonds are conserved from the cnidarian to the human. Histidine (H) and aspartic acid (D) are the active-site amino acid residues of the PLA2s, and the corresponding residues are present in otoconin. The basal line represents the mature peptide region.

Fig. 3. Comparison of the disulfide bond positions in the cnidarian *Adamsia carciniopados* and human GV, GIB, GI A and GX PLA2s and otoconin-90. Five disulfide bonds are conserved from the cnidarian to the human. Histidine (H) and aspartic acid (D) are the active-site amino acid residues of the PLA2s, and the corresponding residues are present in otoconin. The basal line represents the mature peptide region.

Fig. 4. Alignment of the conserved domains cd04704 of the *Apis mellifera* GII PL A2 (AmellII_04) and human GII PL A2 (HsapII_04). The Ca$^{2+}$-binding residues tryptophan (W), glycine (G) and aspartic acid (D) are highlighted in blue, the catalytic histidine (H) and aspartic acid (D) are highlighted in red, and the conserved cysteine (C) residues are indicated with ‘#’.
snail C. magus shares 13–33% sequence identity with the bacterial, fungal and cnidarian pfam09056 PLA2s, but lower identity of 9–13% with the cd04706, cd04704 and cd04707 domains – exist and underwent diversification (Fig. 5). Furthermore, the GIX PLA2 of C. magus shares 13–33% sequence identity with the bacterial, fungal and cnidarian pfam09056 PLA2s, but lower identity of 9–13% with the cd04706, cd04704 and cd04707 domains – exist and underwent diversification (Fig. 5).

**Conserved domain pfam06951 of GXIIA PLA2s**

GXIIA PLA2 was first cloned from human [35]. In the present study, the conserved domain pfam06951 was identified in the GXIIA PLA2s of a large number of vertebrate and invertebrate species, including the basal metazoans N. vectensis (Cnidaria), Trichoplax adhaerens (Placozoa) and Brachionus plicatus (Rotifera), as well as in the nonmetazoan eukaryotes (protists), such as the amoeba Naegleria gruberi (Heterolobosea), Monosiga ovata (Chonnoflagellida), Euglena gracilis (Euglenozoa), Phytophthora infestans (Oomycetes), Thalassiosira pseudonana (Bacillariophyta), Capsaspora owczarzaki (Ichthyosporea), Chromera velia (Alveolata) and Thecamonas trahens (Apusozoa). Furthermore, pfam06951 was also identified in the sPLA2s of the marine algae Micromonas (Viridiplantae) and the prokaryote Planctomyces maris (Bacteria) (Table S1). These observations demonstrate that the conserved domain pfam06951 is widely distributed not only in the sPLA2s of higher animals but also in the sPLA2s of simple eukaryotes and prokaryotes. Sequence alignment of the GXIIA PLA2 conserved domains of organisms ranging from protists to mammals reveals high conservation of the Ca\(^{2+}\)-binding and catalytic sites. The canonical histidine catalytic site C-C-x-x-H-x-x-C motif is highly conserved. In the aspartic acid catalytic motif, the cysteine and aspartic acid residues are also conserved, with the exception of E. gracilis (Euglenozoa), Micromonas pusilla (Viridiplantae) and P. infestans (Oomycetes), in which aspartic acid is replaced by glutamic acid and the first cysteine is absent in all the sequences (Fig. S1). Comparison of the human GXIIA PLA2 (pfam06951) with GIB PLA2 (cd00125) indicated 37% amino acid sequence identity.

Another gene product closely related to GXIIA PLA2 is GXIIB PLA2 [36]. Human GXIIB PLA2 shares 46% sequence identity with GXIIA PLA2. GXIIB PLA2 is a catalytically inactive protein as a result of the substitution of histidine with leucine in the catalytic site (Fig. 2).

**Phylogenetic analysis of sPLA2s**

Phylogenetic analysis of the conserved domains of sPLA2s, carried out with the maximum likelihood (ML) and neighbour-joining (NJ) methods, produced similar tree topologies, suggesting that the members of this protein family have a common and ancient evolutionary origin (Fig. 6). Two major sPLA2 groups – the cd-collection (which includes the cd00125, cd04704, cd04705, cd04706 and cd04707 of unicellular and multicellular organisms) and the pfam-collection (which contains the sPLA2s with the annotated pfam06951 and pfam09056 domains) – exist and underwent distinct trajectories during evolution.
The sPLA₂ forms of bacteria cluster within each PLA₂S subgroup, suggesting that the eukaryote and prokaryote sPLA₂ repertoire share a PLA₂-like common ancestor molecule, which remains to be identified. The cd-collection contains the majority of the sPLA₂ sequences identified, and the clustering observed indicates that discrete relationships between the sPLA₂ protein groups exist (Fig. 6). Members of the cd04704 and cd04705 groups seem to have evolved separately from the other cd-collection members, such as cd00125 and cd04707, and have only been identified in animals.

Discussion
PLA₂ activity was first reported in canine pancreatic juice [37] and sPLA₂S have now been isolated from a large number of snake and other venoms, and also from cells, tissues and body fluids of various unicellular and multicellular organisms [38–40]. Initially, sPLA₂S were classified into two major groups of GI and GII PLA₂S [41], but subsequent sequence homology and conserved disulfide bonds were observed within the other members of the GI/II/V/X PLA₂ col-
lection and the related protein otoconin [42]. This classification has been expanded and refined over the past two decades as information from phylogenetically distinct organisms has accumulated at an accelerating rate [1,2,8,23]. The recent systematic identification of the conserved domains of protein molecules [24,25] allows a novel classification of the sPLA$_2$s (‘cd/pfam classification’) based on the identification of the conserved domains, and a meaningful phylogenetic analysis of the whole range of sPLA$_2$s.

The aim of the present study was to elucidate the origin and evolution of the sPLA$_2$s. The enormous evolutionary span between the organisms expressing sPLA$_2$s (ranging from prokaryotes to metazoan eukaryotes) by necessity introduces a degree of uncertainty in the comparison of protein sequences between such distant phyla. When PLA$_2$ sequences containing conserved domains of different groups are compared, the molecular structure of a particular PLA$_2$ reflects the divergent or convergent evolution of the conserved domain structure and the evolutionary history of the organism in question. In the present study, we investigated the structural variability of the conserved domains of sPLA$_2$s in a wide range of organisms, from bacteria to mammals. A total of 358 conserved domains were identified among 333 sPLA$_2$s sequences. In viruses, sPLA$_2$-like proteins (GVIIA PLA$_2$) have also been reported but their molecular structure, including the conserved domain and the enzymatic activity, differ from those of sPLA$_2$s [43] and thus were not included in the current analysis.

Two main distinct forms of sPLA$_2$s were identified in bacteria: those containing cd04706 (bacteria/plant-specific subfamily of PLA$_2$) and those containing pfam09056 (prokaryotic/fungal PLA$_2$). Notably, no sPLA$_2$s were identified in the Archaea. A marked difference between Bacteria and Archaea is that the predominant lipid constituents of the archaean membranes are prenyl ether lipids, whereas the bacterial membranes contain acyl ester lipids [44–46]. The phospholipid metabolism of bacteria is driven by phospholipases that specifically hydrolyze the acyl ester bonds but are incapable of hydrolyzing the prenyl ether lipids. In light of this observation, it is hypothesized that the sPLA$_2$s of the higher organisms may have their evolutionary origin in Bacteria rather than in Archaea.

Members of the sPLA$_2$s are proposed to share a common ancestry and to have emerged early in evolution, and several theories based upon their sequence similarities and predicted molecular structure have suggested that they evolved rapidly. The GI and GII sPLA$_2$s of Elapidae and Crotalidae snake venoms, respectively, have distinct molecular structures and it has been proposed that they share a common ancestor [22]. Snake venom sPLA$_2$s are suggested to have evolved at an accelerated rate that has resulted in the presence of many variant PLA$_2$ molecules produced by the venom glands [40,47–49].

A recent example of the rapid evolution of sPLA$_2$ genes is the bovine GIID PLA$_2$. In cattle, five duplications have been identified, while a single gene copy is present in the human and rodent genomes. The bovine GIID PLA$_2$s are expressed in the mammary gland and upregulated during the lactating period, and are suggested to participate in the innate immune response [50]. In human and mouse, GIJA, GIIC, GIID, GHE, GIIF and GV PLA$_2$ genes are also the result of gene-duplication events within the same chromosome. In human they are localized in chromosome 1, whereas GIB and GX genes map to chromosomes 12 and 16, respectively. However, comparisons between the gene homologues in human and mouse reveal that species-specific events may occur; for example, the human GIIC PLA$_2$ is a pseudogene, whereas the mouse homologue encodes an enzymatically active protein [51]. Another example of the functional diversity within the members of the GI/II/V/X PLA$_2$ collection are the human and mouse GIJA PLA$_2$s, which are efficient bactericidal enzymes involved in the innate immune response, whereas the closely related digestive enzyme GIB PLA$_2$ is only marginally bactericidal [6,52]. For example, in vitro assays demonstrated that Gram-positive bacteria are killed by sPLA$_2$s and that human GIJA PLA$_2$ is highly potent in this respect, whereas GIB PLA$_2$ is of low efficacy [52].

In Cnidaria, a sister group of the vertebrate clade that diverged more than 500 million years ago [53,54], sPLA$_2$s have been reported that structurally resemble the vertebrate GI and GII PLA$_2$s [19–21,28]. Our current observations indicate that a cnidarian sPLA$_2$ [19] containing cd00125 has disulfide bonds at locations identical to those of the human GV PLA$_2$ and conserved in other members of the GI/II/V/X PLA$_2$ collection. Group I PLA$_2$s of elapid snake venoms have lost the ancestral pancreatic loop present in the cnidarian sPLA$_2$ and also in the mammalian GIB PLA$_2$. Such structural changes have resulted in the appearance of novel functions, including toxicity of the venom sPLA$_2$s [55]. However, the number of cysteine residues and disulfide bonds of sPLA$_2$s vary among closely related animals such as the sea anemones *Adamsia carcinipados*, *Urticina crassicornis*, *Condylactis gigantea* and *N. vectensis* [19–21,28,56]. The variation in the disulfide patterns and other structural features seems to preclude the
exact placement of these cnidarian sPLA_2s in the currently recognized groups of the GI/II/V/X PLA_2 collection [19–21].

In the current study, homologues of the vertebrate GXIIX members with highly conserved domain regions were identified for the first time in unicellular organisms, and their function in such organisms remains to be established. In human, GXIIXA PLA_2 is expressed in T lymphocytes and seems to be involved in the regulation of the immune response [57]. Human GXIIXB PLA_2 was recently shown to be involved in the triglyceride metabolism in the liver [58] and is proposed to activate specific receptors that remain to be identified [36].

Life as we know it can be divided into the prokaryotic (cellular organisms that lack a nucleus) Domains of Bacteria (Eubacteria) and Archaea (Archaeabacteria), and the Domain of Eukaryota (organisms consisting of nucleated cells, such as animals, plants and fungi) [44]. Prokaryotes are the oldest cellular life forms on Earth, dating back 3.5–4 billion years and predating eukaryotes by 1 billion years. The current phylogenetic analysis of the conserved domains of sPLA_2s of the representatives of the major prokaryote and eukaryote taxa supports the hypothesis that the sPLA_2s of the eukaryotes, including the Metazoa, Viridiplantae and Fungi, may have shared a common origin with their homologues in bacteria. Two sPLA_2 groups (the cd-collection and the pfam-collection) emerged early in evolution and underwent distinct evolutionary trajectories. Based upon the retrieved data and phylogenetic relatedness, a model for the evolution of the two sPLA_2 group members is proposed centred around gene-duplication and gene-deletion events (Fig. 7A,B). While the two members of the pfam-collection are present in representatives of the Eubacteria and Animalia kingdoms and maintained throughout evolution (Fig. 7B), the cd-collection members seem to have mainly emerged in the Animalia kingdom (Fig. 7A) and their expansion may be associated with the gene duplications that are proposed to have contributed to the increase in organismal complexity during eukaryote evolution [59]. The cd04706 protein members are exclusively found in Eubacteria and Plantae kingdoms and were lost from other life forms. Despite the lack of data from representatives of all clades and the failure to identify homologues it can be hypothesized that in the kingdom Animalia, sPLA_2s with the conserved domains cd00125 and cd04704 were the first members to emerge. Subsequently, several independent gene or genome-duplication events occurred and resulted in the emergence of two novel family members of vertebrate cd04704 and mammalian cd04705. The sequence similarity observed between the vertebrate sPLA_2s containing conserved domains cd04707 and cd00125 and also the mammalian cd04705 with cd04704 suggests that they have a shared common origin and that cd04707 emerged from a gene-duplication event of the cd00125-like ancestral gene at the time of vertebrate emergence and that cd04705 resulted from a later gene-duplication event of the cd04704-like ancestral gene precursor within the mammalian lineage (Fig. 7A). In contrast, only fungi, protists and aquatic invertebrates, such as those of the phyla Cnidaria and Mollusca, have acquired pfam09056-containing sPLA_2s and the homologue gene seems to have been deleted from Protostomes and Deuterostome genomes (Fig. 7B).
The present analysis is based on the highly conserved peptide motifs directly involved in the catalytic function of sPLA₂s, including the Ca²⁺-binding site and the catalytic centre. However, the functional roles of the surrounding domains are at present less well understood (e.g. the functions of the domains flanking the central catalytically active cd04704 of mammalian GIII PLA₂ are unknown) [3] and their study may provide novel insight into the catalytic and noncatalytic functions of sPLA₂s. For instance, binding of sPLA₂ to specific cellular receptors is independent of their catalytic activity, and the protein domains involved in the receptor activation are not yet fully resolved [4]. Other examples are the toxicity of some snake venom PLA₂s, which does not correlate with their catalytic activity [8], and the bactericidal effects of sPLA₂s lacking catalytic activity [60].

It is concluded that the sPLA₂ of eukaryotes share their evolutionary origin with two distinct types of bacterial sPLA₂s. Their evolution and prevalence in genomes seems to be related to the functional constraints of phospholipid metabolism which is a fundamental and conserved process in organisms. Although relatively little is known about the function of prokaryotic sPLA₂s, the large number of distinct sPLA₂ isoforms in metazoans reflects the wide variation of substrate types encountered in the extracellular environment where the enzyme plays many important roles in nutrition, reproduction and immunity.

**Material and methods**

**Database mining and data collection**

Secreted PLA₂ sequences were retrieved from the publicly available protein databases of NCBI (http://www.ncbi.nlm.nih.gov) and Swiss-Prot (http://www.uniprot.org) using the Basic Local Alignment Search Tool (BLASTp) algorithm [61] and default settings. Database searches were performed using the peptide sequences of the human GIB PLA₂ (P00630), honeybee venom GIII PLA₂ (P00630), human GXIIA PLA₂ (Q9BZM1), O. sativa GXIB PLA₂ (Q9XG81) and S. violaceoruber PLA₂ (Q6UV28). In addition, PLA₂ and PLA₂-like protein sequences were identified in the Microbial and Eukaryotic Genome database (http://www.ncbi.nlm.nih.gov/genome) following a similar strategy. Searches were also performed on nucleotide database data and covered all completed genomes in the NCBI genome database (1359 bacterial, 79 archaeal and 231 eukaryotic genomes; October 2010 release), and also available EST data using the tBLASTn and sequence matches with an e-value of < 10 were retrieved and their sequence analysed. The deduced protein sequences were obtained using the BCM Search Launcher (http://searchlauncher.bcm.tmc.edu/seq-util/Options/sixframe.html) and compared with available homologue data.

**In silico sequence annotations**

The conserved domains of sPLA₂ were identified using the NCBI Conserved Domains Database CDD-27036 PSSMs (http://www.ncbi.nlm.nih.gov/cdd) annotation by using the sequences identified in this study as queries. The result includes an alignment between the query and the search model consensus sequence, the expected-value for the alignment, the identity (name) of the conserved domain and the location of the conserved domain in the query sequence [25]. Histidine active-site and aspartic acid active-site protein motifs were identified based on the PROSITE database (http://au.expasy.org/prosite) pattern annotation. The localization of disulfide bonds was retrieved from the Swiss-Prot database and from published data.

**Sequence comparisons and phylogenetic analysis**

Pairwise and multiple protein sequence alignments were carried out using the ClustalW2 program [62] available from EBI (http://www.ebi.ac.uk/Tools/msa/clustalw2) and the default parameters. The percentage of sequence similarities (based on the observed substitutions of one amino acid for another in homologous proteins) and identities between the sPLA₂ were calculated based upon protein alignments using the GeneDoc interface (http://www.psc.edu/biomed/genedoc). Phylogenetic analysis of the conserved domain sequences of bacterial, fungal and metazoan PLA₂ was performed using 53 taxa representatives. The protein alignment produced was submitted to PROTEST analysis (http://darwin.uvigo.es/software/prottest.html) to select the best model of protein evolution that fits the data set [63]. Phylogenetic analyses were conducted using the NJ [64] and ML methods, and reliability for internal branching was assessed using the bootstrap method [65]. NJ analysis was performed using MEGA4 programme [66] and the p-distance amino acid model with 1000 bootstrap replicates. All positions containing alignment gaps and missing data were eliminated (pairwise deletion option) and a total of 215 positions were analyzed in the final data set. The ML tree (PhyML, v3.0 aLRT) [67] was constructed in Phylogeny.fr web interface (http://phylogeny.lirmm.fr/phylo.cgi/index.cgi). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.167) and four gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (γ = 1.804) and analysis was performed using 100 bootstrap replicates. Graphical representation and edition of the phylogenetic tree were performed with TREEDYN (v198.3, http://www.treedyn.org).
Both methods produced similar tree topologies and the NJ bootstrap consensus tree was selected and taxa clades with support values < 50% collapsed.

Acknowledgements

The authors thank Deborah Power for critical reading of the manuscript and anonymous referees for valuable suggestions. Supported by the Research Fund of Turku University Hospital and the Portuguese National Science Foundation (FCT)/CCMAR pluriannual grant.

References

14 Renetseder R, Brunie S, Dijkstra BW, Drenth J & Sigler PB (1985) A comparison of the crystal structures of phospholipases A₂ from bovine pancreas and Crota-
Conserved domains of sPLA2s

T. J. Nevalainen et al.


**Supporting information**

The following supplementary material is available:

**Fig. S1.** Alignment of the conserved domain pfam06951 sequences of 20 GXIIA PLAs.

**Table S1.** Conserved domains of 333 prokaryotic and eukaryotic secreted phospholipases A2.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.