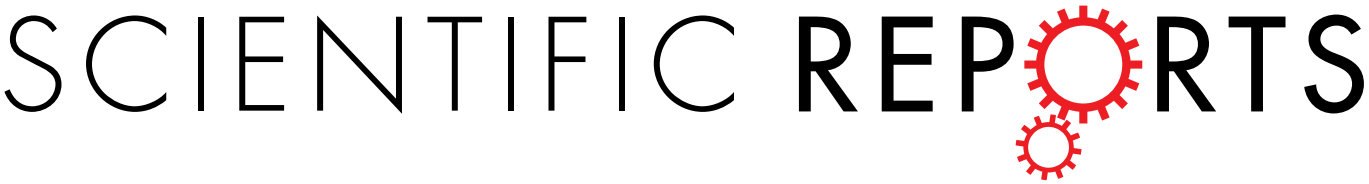

OPEN

Molecular details of secretory phospholipase A_2 from flax (*Linum usitatissimum* L.) provide insight into its structure and function

Received: 27 March 2017
Accepted: 17 August 2017
Published online: 11 September 2017

Payal Gupta[1,2] & Prasanta K. Dash[1]

Secretory phospholipase A_2 (sPLA$_2$) are low molecular weight proteins (12–18 kDa) involved in a suite of plant cellular processes imparting growth and development. With myriad roles in physiological and biochemical processes in plants, detailed analysis of sPLA$_2$ in flax/linseed is meagre. The present work, first in flax, embodies cloning, expression, purification and molecular characterisation of two distinct sPLA$_2$s (I and II) from flax. PLA$_2$ activity of the cloned sPLA$_2$s were biochemically assayed authenticating them as *bona fide* phospholipase A_2. Physiochemical properties of both the sPLA$_2$s revealed they are thermostable proteins requiring di-valent cations for optimum activity.While, structural analysis of both the proteins revealed deviations in the amino acid sequence at C- & N-terminal regions; hydropathic study revealed LusPLA$_2$I as a hydrophobic protein and LusPLA$_2$II as a hydrophilic protein. Structural analysis of flax sPLA$_2$s revealed that secondary structure of both the proteins are dominated by α-helix followed by random coils. Modular superimposition of LusPLA$_2$ isoforms with rice sPLA$_2$ confirmed monomeric structural preservation among plant phospholipase A_2 and provided insight into structure of folded flax sPLA$_2$s.

Utility of flax (*Linum usitatissimum*) to mankind as a valued fibre dates back to more than 8000 years in history[1]. However, as of today, it has become a multipurpose crop that serves as a source of high quality fiber "linen"- a widely used raw material for textile industry. Additionally, the seed oil is an excellent source of omega-3-fatty acid ALA (α-linolenic acid, C:18)[2]. The flax seed is also a repository of active lignan compound "Secoisolariciresinol diglycoside (SDG)"[3] that exhibit health benefits. The oil from flax seed is an excellent industrial solvent being used in paints, and varnishes. Innumerable use of flax increases its value in commercial market making it a high value cash crop. On the other hand, phospholipases have been known to have role in industrial applications amongst which secretory phospholipase A_2 (sPLA$_2$) from animals and microbes have been reported to be used in food industry in emulsification and degumming[4]. However, no industrial application of sPLA$_2$ from plant source is reported as yet[5].

Phospholipase A_2 belongs to a group of hydrolases (EC: 3.1.1.4) that stereo-specifically catalyses the hydrolysis of second acyl ester bond of phospholipids generating free fatty acids (FFAs) and lysophospholipids (LPLs)[6]. Based on the structure, function and evolution, PLA$_2$s are classified into 15 groups (I-XV) belonging to five major types. Amongst them calcium-dependent and independent cytosolic phospholipase A_2, platelet-activating factor acetyl hydrolases, lysosomal phospholipase A_2 and secretory phospholipase A_2 are important[7]. Plant secretory PLA$_2$s belong to group XI, which is further sub-divided into XIA and XIB[6]. In plants, sPLA$_2$s are ubiquitous and are extensively studied enzymes known to be involved in a suite of signal transduction pathways[8,9]. They play myriad roles in biological and metabolic processes leading to growth[10], development[11,12] including plant defence[13] and imparting tolerance against abiotic stress[14,15].

Secretory phospholipase A_2 are low molecular weight proteins (12–18 KDa) characterised by a N-terminal signal peptide, with 12 conserved cysteine residues that form six intra-molecular disulphide bridges[16]. Functional domain of sPLA$_2$s constitutes a signature phospholipase A_2 (PA2c) domain, which is highly conserved from animals to plants[16]. This domain is characterized by presence of a conserved Ca^{2+} binding loop (YGKYCGxxxxGC)

[1]ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi, 110012, India. [2]Department of Biotechnology, Kurukshetra University, Thanesar, 136119, India. Correspondence and requests for materials should be addressed to P.G. (email: payalgupta33@gmail.com) or P.K.D. (email: pdas@nrcpb.org)

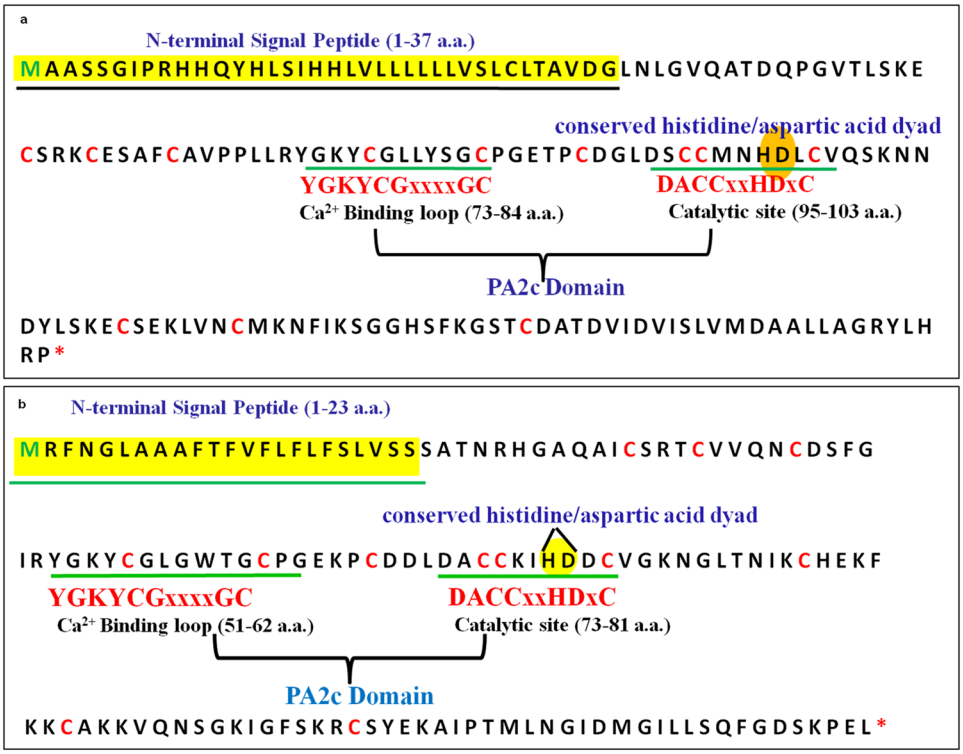

Figure 1. Characteristic structural features of secretory phospholipase A_2 identified in flax sPLA$_2$I and II. (**a**) Characteristic features of secretory phospholipase A_2 identified in LusPLA$_2$I. (**b**) Characteristic features of secretory phospholipase A_2 identified in LusPLA2II. N-terminal signal peptide is highlighted in yellow, conserved calcium binding loop and catalytic site are underlined in green and represented by sequence in red, conserved His/Asp dyad is represented in oval shape, and twelve conserved cysteine residues are shown in red colour.

and a catalytically active motif (DACCxxHDxC) containing conserved His/Asp dyad in the active site. These conserved motifs are necessary for the catalytic activity of sPLA$_2$[16]. His residue at the catalytic site motif is required for nucleophilic attack at the sn-2 acyl bond of phospholipids[17]. These enzymes are heat stable and require micro to milli molar concentration of Ca^{2+} for optimum activity[16].

Although sPLA$_2$s have been identified and purified from a number of plants, detailed functional information about these enzymes in flax is lacking. Among plants, the structure of arabidopsis (*Arabidopsis thaliana*) sPLA$_2$ isoform α and soybean (*Glycine max*) sPLA$_2$ isoform II has been elucidated by homology modelling. Currently, tertiary structure of sPLA$_2$ isoform II from rice (OssPLA$_2$II) is available and is the only testimony (PDB entry 2WG7A)[18] of sPLA$_2$s from plants. However, three-dimensional structures of flax sPLA$_2$s are hitherto not available in any database.

Flax is an economically important field crop. Nearly one-fifth of its transcriptome is unique[19] and that motivated us to characterize the novel secretory PLA$_2$ enzymes of flax and perform detailed structural and functional analysis of LusPLA$_2$s (flax sPLA$_2$I & II). Both sPLA$_2$s in flax were characterised at molecular level by cloning and expressing purified LusPLA$_2$I and LusPLA$_2$II in native form. The PLA$_2$ activity of both flax sPLA$_2$s was bio-assayed using LOX/PLA$_2$ based reaction to authenticate them as *bona fide* phospholipases. In the present study, we cloned, biochemically characterized the flax sPLA$_2$ proteins. Further, we investigated the three-dimensional structure of both the phospholipases of flax based on homology modelling and elucidated the topology of calcium binding loop and catalytic motif site. Our study is the first comprehensive biochemical and molecular analysis providing insight into folded secretory phospholipase A_2 of flax.

Results

Discerning sequence, function, domain and cellular localization of flax sPLA$_2$s. Domains are the functional units of a protein and are known to be evolutionarily conserved[20]. Our exploration of plant specific secretory phospholipase A_2 led to identification of two sPLA$_2$ in flax. A single PA2c domain, the signature domain for sPLA$_2$ (with accession no. cd04706 and PSSM ID: 153095), was identified in LusPLA$_2$I as well as in LusPLA$_2$II proteins that belonged to superfamily cd05417domain (see Supplementary Table S1).

LusPLA$_2$s contained two motifs in the PA2c signature domain, viz. Ca^{2+} binding motif and catalytic active-site motif characterized by the presence of conserved His/Asp dyad (Fig. 1a,b). These motifs are highly conserved in plant sPLA$_2$s[16]. We observed, Ala residue present in the highly conserved catalytic domain (LDACCxxHDxCV) of sPLA$_2$s is replaced by a Ser residue (LDSCCMNHDLCV) in flax sPLA$_2$ isoform II. Twelve conserved Cys residues were identified at positions Cys56, Cys60, Cys65, Cys77, Cys84, Cys90, Cys96, Cys97, Cys103, Cys116, Cys123,

Biological process	Molecular function	Cellular component
Lipid catabolic process	phospholipase A_2 activity	Extracellular region
Growth and development	calcium ion binding	Endoplasmic reticulum
Wounding and pathogen attack	Hydrolase activity	Intracellular organelle
Abiotic and biotic stress	Metal ion binding	Membrane bound
Phospholipid metabolic process	Catalytic activity	Organelle
Primary metabolic process	Protein binding	Cytoplasm
Cellular process	Receptor binding	Endomembrane system
Single-organism metabolic process	Lipase activity	Cell part
Phosphate containing		Cell surface
Anatomical structure		Golgi apparatus
morphogenesis		Intracellular part
Organogenesis		
Response to hormone		
Response to external stimuli		
Biological regulation		
Response to endogenous stimuli		

Table 1. Gene ontology classification of secretory $sPLA_2$ ($LusPLA_2I$ and $LusPLA_2II$) from flax.

Cys140 in $LusPLA_2I$ and Cys35, Cys39, Cys44, Cys55, Cys62, Cys68, Cys74, Cys75, Cys81, Cys92, Cys99, Cys115 in $LusPLA_2II$ (Fig. 1a,b). Other conserved amino acids involved in Ca^{2+} binding and catalysis of the substrate were also present in both $sPLA_2$s in flax[6].

To ascertain (% similarity/dissimilarity) all the plant $sPLA_2$ sequences from flax ($LusPLA_2$), rice ($OssPLA_2$), arabidopsis ($AtsPLA_2$), soybean ($GmsPLA_2$), wheat ($TdsPLA_2$, *Triticum durum*) and snake venom (*Naja kaouthia*) were aligned using Mafft-win and ESPRIPT. All the $sPLA_2$s from plant origin showed a close evolutionary relationship (see Supplementary Fig. S1). The alignment of amino acid sequence among the known plant $sPLA_2$s revealed that $LusPLA_2I$ belonged to group XIB with maximum similarity to soybean (isoform I and II) and arabidopsis ($sPLA_2$ α) while $LusPLA_2II$ belonged to XIA with maximum relatedness to rice isoform I along with arabidopsis $sPLA_2$ β, γ, δ.

We explored occurrence of potential signal peptide in the $LusPLA_2$s by using the SignalP and target organelle by TargetP tools. SignalP results suggested that $LusPLA_2I$ possessed a 37 amino acid (Met1- Gly37) signal peptide at N-terminal end and is cleaved to generate a 129 amino acids long holoenzyme after cleavage. Similarly, $LusPLA_2II$ contained a 23 amino acid (Met1- Ser23) signal peptide at N-terminal end and generates the holoenzyme of 121 amino acids (see Supplementary Fig. S1). TargetP software (http://www.cbs.dtu.dk/services/TargetP/output.php) predicted $LusPLA_2I$ to be secreted into the extracellular space with a reliability class of 4 while iPSORT predicted mitochondria as the target organelle. This prediction in extracellular space and mitochondria can be ascribed to re-localization of $sPLA_2$s upon interaction with other genes or external stimuli. Similarly, TargetP (with a probability of 0.9) predicted $LusPLA_2II$ to be secreted into the extracellular space with a reliability class 2.

Gene ontological classification of flax $sPLA_2$s suggested that these proteins are involved in biological and metabolic processes, including anabolism and catabolism associated with growth and development (Table 1). Apart from growth and development, $LusPLA_2I$ and $LusPLA_2II$ are involved in biological processes such as response to external stimuli and internal stimuli, wounding and pathogen attack, abiotic and biotic stress response. They are also found to be involved in molecular functions like protein/receptor binding, metal ion binding and catalytic activity of substrate (Table 1).

Cloning, expression and purification of *$LusPLA_2$s*. In order to assay the biochemical properties of cloned $sPLA_2$s, both the proteins were expressed and purified as 6xHis-tagged fusion proteins. Both the genes, *$LusPLA_2I$* (502 bp) and *$LusPLA_2II$* (436 bp), were cloned in pENTR/SD/D/TOPO (see Supplementary Figs S2, S3) and mobilized into destination vector pET301/CT-DEST (see Supplementary Figs S4, S5). The successful cloning was confirmed by restriction of pET301/CT-DEST harbouring *$LusPLA_2I$*-6x-His with *Bam*HI and *Nco*I that released a fragment of ~574 bp (see Supplementary Fig. S4) and pET301/CT-DEST harbouring *$LusPLA_2II$*-6x-His with *Bam*HI and *Not*I that resulted in a fragment of 544 bp (see Supplementary Fig. S5). Sequencing of the expression clones confirmed that both the genes (*$LusPLA_2I$* and *$LusPLA_2II$*) were cloned in-frame with downstream 6x-His tag. Both the fusion proteins were obtained by induction (with 0.4 mM IPTG) in BL21-CodonPlus (DE3)-RIPL *E. coli* cells harbouring *$LusPLA_2I$* (Fig. 2a) and *$LusPLA_2II$* (Fig. 2c). Recombinant proteins were purified from the soluble fraction. Presence of an intense band of 22.6 kDa for PLA_2I and 20.6 kDa for PLA_2II was observed in SDS-PAGE. For PLA_2 activity assay, both the proteins were purified from soluble fraction by affinity chromatography (Ni-NTA column). Precise detection of both the PLA_2s were accomplished by western blotting that showed expected molecular weight of $LusPLA_2I$ as 22.6 kDa (Fig. 2b) and for $LusPLA_2II$ as 20.6 kDa (Fig. 2d).

PLA_2 activity assay. Purified PLA_2s obtained through column chromatography were used to assess intrinsic lipase activity. The spectrophotometric assay was used to investigate the catalytic (hydrolysis of phospholipids) ability of the isolated $LusPLA_2I$ and $LusPLA_2II$ proteins. An increase in absorbance at 234 nm upon addition

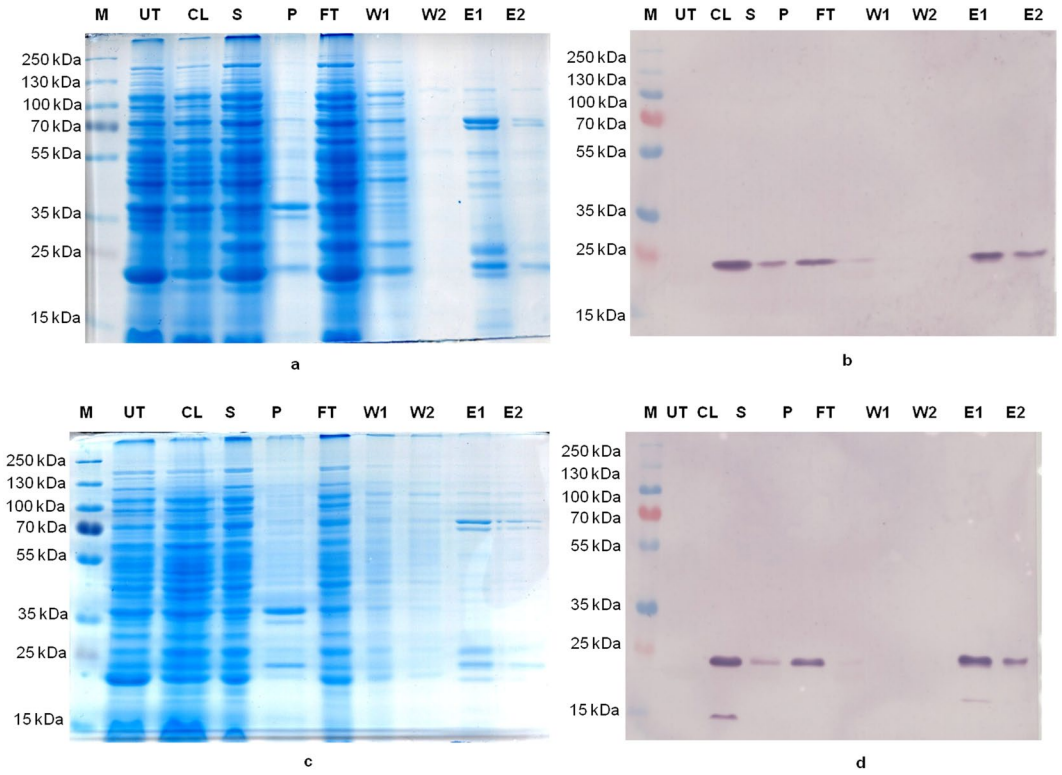

Figure 2. Small scale protein purification of LusPLA$_2$s by Ni-NTA purification column by affinity chromatography. (**a**) SDS-PAGE analysis of the protein fractions of pET301/CT-DEST harbouring LusPLA$_2$I-6xHis fusion protein. (**b**) Detection LusPLA$_2$I-6xHis fusion protein by western blotting using anti-His antibodies. (**c**) SDS-PAGE analysis of the protein fractions of pET301/CT-DEST harbouring LusPLA$_2$II-6xHis fusion protein. (**d**) Detection of LusPLA$_2$II-6xHis fusion protein by western blotting using anti-His antibodies. M - Pageruler plus prestained protein ladder, UT-Untransformed BL21-RIPL, CL-crude lysate of transformed BL21-RIPL, S-supernatent, P-pellet, FT-flow through, W1-first wash, W2- second wash, E1- first elute and E2-second elute.

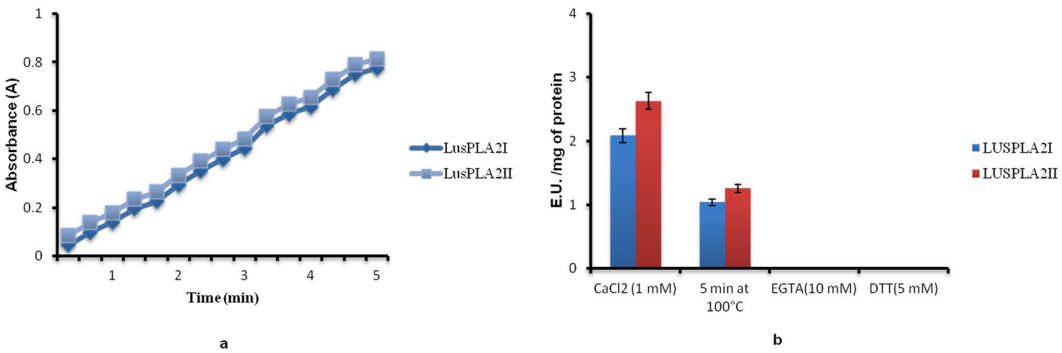

Figure 3. PLA$_2$ activity and biochemical features of recombinant LusPLA$_2$s fusion proteins. (**a**) PLA$_2$ activity measured as increase in absorbance at 234 nm. The absorption was recorded every 20 seconds for 5 minutes. (**b**) Effect of ca^{2+} chelator (EGTA), disulphide bond destabilizer (DTT) and heat inactivation on PLA$_2$ activity of recombinant protein. Data are expressed as mean value ± SD (3 independent experiments).

of phospholipases to substrate confirmed the isolated proteins as *bona fide* phospholipase enzymes (Fig. 3a). Thermostability of proteins was assessed by adding heat attenuated recombinant protein to the assay and was observed that LusPLA$_2$I retained ~50% activity while LusPLA$_2$II retained 48% of its activity (Fig. 3b). Similarly, effect of addition of divalent cations (Ca^{2+}), chelating agents (EGTA) and reducing agents (DTT) on phospholipase activity of flax sPLA$_2$s was assessed by adding $CaCl_2$, EGTA, DTT. While, addition of divalent cation (Ca^{2+}) chelator EGTA (10 mM) to the reaction completely attenuated the enzyme activity (Fig. 3b) of both sPLA$_2$s, addition of disulphide reducing agent DTT (5 mM) to the reaction abolished the enzyme activity (Fig. 3b). However, addition of divalent cation $CaCl_2$ restored the enzymatic properties of both flax sPLA$_2$s (Fig. 3b).

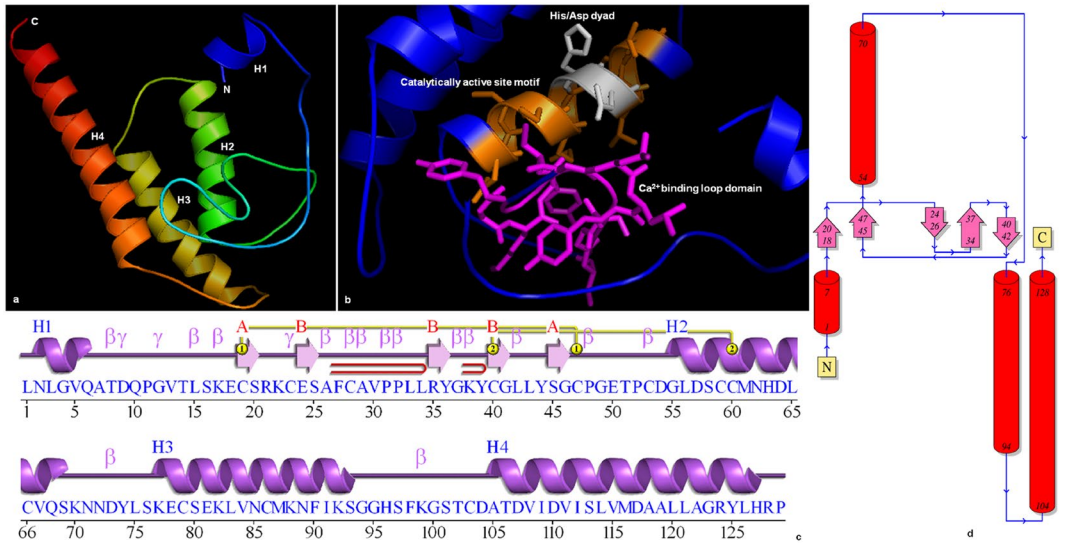

Figure 4. Structure of sPLA$_2$I from flax. (**a**) Stereo ribbon diagram of the LusPLA$_2$I monomer (chain *A*) color-coded from the N-terminus (blue) to the C-terminus (red). Helices (H1–H4) are indicated. (**b**) Ribbon diagram showing the conserved domains of sPLA$_2$. Calcium binding loop is marked in pink, catalytically active site motif is marked in orange and conserved His/Asp dyad in gray. (**c**) Diagram showing the secondary-structure elements of LusPLA$_2$I superimposed on its primary sequence. The labelling of secondary-structure elements is in accordance with *PDBsum* (http://www.ebi.ac.uk/pdbsum): α-helices are labeled H1–H4, five β-strands shown as arrow are labeled as A and B, β-turns and γ-turns are designated by their respective Greek letters (β, γ) and red loops indicate β-hairpins. (**d**) Topology of LusPLA$_2$I protein showing the orientation of α-helices and β-strands.

Physiochemical Properties. The relationship between molecular structure, functional properties and host of operating interaction accounts for the stability and biological activity of a protein. Analysis of physiochemical properties (see Supplementary Table S2) of flax sPLA$_2$I and sPLA$_2$II revealed the calculated molecular mass of LusPLA$_2$I is 17.94 kDa and LusPLA$_2$II is 15.72 kDa. The observed isoelectric point (pI) of LusPLA$_2$I was 6.68 (pI < 7) and LusPLA$_2$II was 8.84 (pI > 7). LusPLA$_2$I was found to abundantly contain leucine followed by serine, cysteine, glycine and valine and deficient in phenylalanine and tryptophan. Likewise, glycine was most abundant amino acid in LusPLA$_2$II followed by lysine, serine, leucine, cysteine and phenylalanine whereas tryptophan was least abundant (see Supplementary Table S3).

Stability of LusPLA$_2$I and LusPLA$_2$II was determined by calculating instability index (II)[21]. The instability index for LusPLA$_2$I was 34.33 and for LusPLA$_2$II was 17.13. Similarly aliphatic index (AI) is another parameter that is positively correlated with the thermal stability of globular proteins[22]. AI of LusPLA$_2$I was found to be 95.72 and was classified as thermally more stable than LusPLA$_2$II having AI of 69.79. Analysis of the Grand Average Hydropathy (GRAVY)[23] revealed LusPLA$_2$I (0.110) to be hydrophobic protein while LusPLA$_2$II with GRAVY score of (-) 0.101 was classified as a hydrophilic protein (see Supplementary Table S2).

Structural Analysis. Understanding of protein structure is of paramount importance for defining its precise function. Apart from the presence of conserved domains and motifs, the three dimensional structural similarity among sPLA$_2$s from different groups is described solely in groups I-III and group X[17] and are found to contain similar structural motifs. However, the structural data for group XI to which flax sPLA$_2$ belong have not been studied in detail. Thus, we analyzed the secondary structure of both the flax sPLA$_2$ proteins using predict protein software for occurence of alpha helix, extended strands and loops. The results classified LusPLA$_2$I as "All alpha" type since it contained 49.40% alpha helix (>45% H)[24] and LusPLA$_2$II as "Mixed" type that contained 43.75% alpha helix (<45% H)[24] (see Supplementary Fig. S6). Analysis of solvent accessibility composition revealed that both LusPLA$_2$I and LusPLA$_2$II are exposed type proteins with 62.65% and 61.81% residues on exposed surface respectively.

Although three-dimensional structures of plant sPLA$_2$s[18] have been predicted, structural data for flax is unavailable. The homology modelled structures of LusPLA$_2$I (Fig. 4a,b) and LusPLA$_2$II (Fig. 5a,b) were generated using crystal structure of rice sPLA$_2$II[18] (Group-XIB) as template. While, LusPLA$_2$I model had a C-score of 0.75, LusPLA$_2$II had a C-score of 0.83. Additionally, the TM-score of 0.81 ± 0.09 and 0.83 ± 0.08 for LusPLA$_2$I and LusPLA$_2$II respectively confirmed the predicted model with correct topology. Analysis of stereo-chemical quality and accuracy of refined protein model using PROCHECK[25] revealed that dihedral angles of all the residues were located in the most favoured regions (LusPLA$_2$I- 91.0%; LusPLA$_2$II- 90.1%) while 9.0% residues in LusPLA$_2$I and 9.9% in LusPLA$_2$II were present in additionally allowed region of the Ramachandran Plot (see Supplementary Figs S7, S8).

The model is consistent with the secondary structure predictions of PDB sum[26]. Structure of LusPLA$_2$I protein revealed presence of two β- sheets surrounded by four α-helices with one helix on left side and three on right side

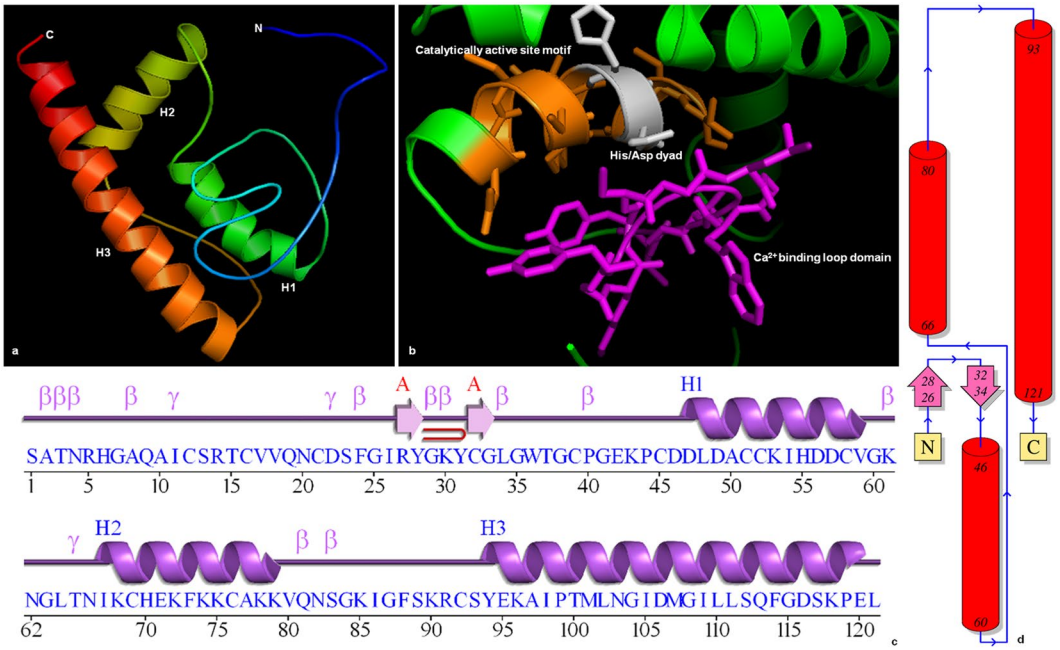

Figure 5. Structure of sPLA$_2$II from flax. (**a**) Stereo ribbon diagram of the LusPLA$_2$II monomer (chain *A*) color-coded from the N-terminus (blue) to the C-terminus (red). Helices (H1–H3) are indicated. (**b**) Ribbon diagram showing the conserved domains of sPLA$_2$. Calcium binding loop is marked in pink, catalytically active site motif is marked in orange and conserved His/Asp dyad in gray. (**c**) Diagram showing the secondary-structure elements of LusPLA$_2$II superimposed on its primary sequence. The labelling of secondary-structure elements is in accordance with *PDBsum* (http://www.ebi.ac.uk/pdbsum): α-helices are labeled H1-H3, two β-strands marked as arrow are labeled as A and B, β-turns and γ-turns are designated by their respective Greek letters (β, γ) and red loops indicate β-hairpins. (**d**) Topology of LusPLA$_2$II protein showing the orientation of α-helices and β-strands.

(Fig. 4c,d). The two β- sheets contained five β-strands out of which β- sheet A contained two parallel β-strands with topology 1X and β- sheet B contained three anti-parallel β-strands with topology 1 1 (Richardson nomenclature)[27]. The five beta strands are arranged in space as β-strands 1 (Cys19-Ser20), β-strands 3 (Arg35-Gly37) and β-strands 5 (Ser45-Gly46) in the same orientation and β-strands 2 (Glu24-Ala26) and β-strands 4 (Cys40-Gly41) in the opposite orientation. The protein structure also comprised of two β hairpins, fifteen β turns[28] (see supplementary Table S4) and three γ turns. Of the four helices, α1 comprising Asn2-Gln6 (5 residues) surrounded the beta sheets on left side along with 15 residues of α2 (Gly55-Ser69), 17 residues of α3 (Lys77-Ser93) and 23 residues of α4 (Ala105-His127) on the right side. The two β-hairpins are incorporated in between β-strand 2–3 and 3–4 belonging to class[29] 8:8 and 3:3 respectively surrounded by α1 (Asn2-Gln6) and α2 (Gly55-Ser69). The γ turns are of inverse type[30,31]. All the twelve Cys residues that form disulfide bonds were harboured between Cys56-Cys84, Cys60- Cys90, Cys65- Cys140, Cys77- Cys97, Cys96- Cys123 and Cys103- Cys116 (Fig. 6a). Disulphide bonds formed between Cys56-Cys84, Cys60- Cys90 and Cys65- Cys140 connect the N- and C- terminal part of the protein, Cys77- Cys97 anchors the Ca^{2+} binding loop to α -helix 2; Cys96- Cys123 and Cys103- Cys116 tether the α-helix 2 to α-helix 3. The four α helices comprised of 60 residue (46.5%) whereas the β sheet comprised of 12 (9.3%) residues in LusPLA$_2$I in flax.

Structure of LusPLA$_2$II revealed presence of single β- sheet and three α-helices (Fig. 5c,d). The single β- sheet contained two anti-parallel β-strands (topology 1) that were arranged in space as β-strand 1 (Arg27-Tyr28) and β-strand 2 (Cys32-Gly33) in the opposite orientation. The structure comprised one β hairpins, thirteen β turns (see supplementary Table S5) and two γ turns. Major part of the protein is composed of three alpha helices such as 13 residues of α1 (Asp47-Val59), 13 residues of α2 (Ile67-Lys79) and 27 residues of α3 (Tyr94-Glu120). These helices start after the beta sheet covering the C-terminal portion of the protein. Only one β-hairpin was present between the strand 1–2 that belongs to class 3:3[29]. All the Cys residues that formed disulfide bonds were harboured between Cys35-Cys62, Cys39- Cys68, Cys44- Cys115, Cys55- Cys75, Cys74- Cys99 and Cys81- Cys92 (Fig. 6b) and stabilized the structure of protein. Disulphide bonds formed between Cys35-Cys62, Cys39- Cys68 and Cys44- Cys115 connect N- terminal to C- terminal portions of the protein; Cys55- Cys75 anchor Ca^{2+} binding loop to α -helix 1 and Cys74- Cys99 and Cys81- Cys92 tether α -helix 1 to α-helix 2. Three α helices comprised 53 residues (43.8%) whereas the β sheet comprised of four (3.3%) residues in LusPLA$_2$ II in flax.

The superimposed structures of OssPLA$_2$ (2WG7A) -LusPLA$_2$I (Fig. 7a) and OssPLA$_2$ (2WG7A) -LusPLA$_2$II (Fig. 7b) exhibited homology in the calcium binding region and alpha helices. Superimposed isoform I and II of flax sPLA$_2$ (Fig. 7c) were also found to be similar in the calcium binding loop and alpha helices.

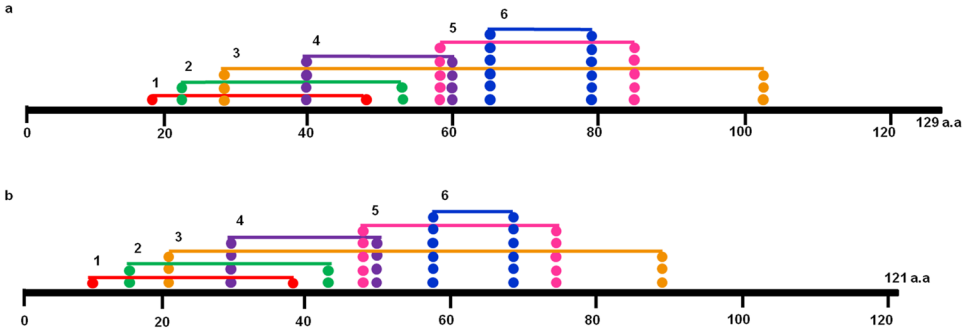

Figure 6. Distribution of six disulphide bridges in $sPLA_2$s in flax. (**a**) In $LusPLA_2I$ all 12 Cys residues are involved in disulphide bond formation. Six potential disulphide bonds are formed between Cys56-Cys84, Cys60- Cys90, Cys65- Cys140, Cys77- Cys97, Cys96- Cys123 and Cys103- Cys116. (**b**) In $LusPLA_2II$ all 12 Cys residues are involved in disulphide bond formation. Six potential disulphide bonds are formed between Cys35-Cys62, Cys39- Cys68, Cys44- Cys115, Cys55- Cys75, Cys74- Cys99 and Cys81- Cys92.

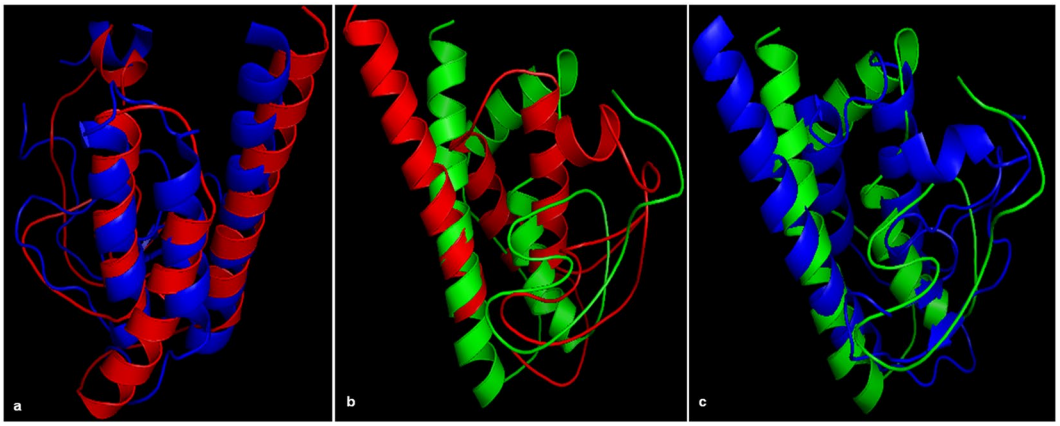

Figure 7. Comparision of tertiary structure of $LusPLA_2$s among themselves and with $OssPLA_2$ (2WG7A). The superimposed structures of $OssPLA_2$ (2WG7A)-$LusPLA_2I$ and $OssPLA_2$ (2WG7A)-$LusPLA_2II$ are similar in the calcium binding loop region and alpha helices. Structure of $OssPLA_2$ shows more identitity to $LusPLA_2I$. Although $LusPLA_2I$ contains 4 α- helices and $LusPLA_2II$ 3 α- helices, the superimposed isoform I and II of flax $sPLA_2$ are also similar in the conserved calcium binding loop and catalytically active site (**a**) Superimposed structure of $OssPLA_2$ (2WG7A) – $LusPLA_2I$. (**b**) Superimposed structure of $OssPLA_2$ (2WG7A) – $LusPLA_2II$. (**c**) Superimposed structure of $LusPLA_2I$- $LusPLA_2II$. Cartoon diagram in red indicates the 3-D structure of $OssPLA_2$ (2WG7A), Cartoon diagram in blue indicates the 3-D model of $LusPLA_2I$ and Cartoon diagram in green indicates the 3-D model of $LusPLA_2II$.

Discussion

Flax entered genomics research lately with decoding of its genome sequence in 2014[32]. Subsequently, genomic information were generated for abiotic stress tolerance in flax[33–36]. Flax is utilized as a multipurpose crop with industrial as well as pharmaceutical use. It yields three economically important products viz. seed oil that is rich in omega-3-fatty acids, bast fiber i.e. linen and nutraceuticals. While, flax seed oil have industrial applications nutraceuticals are used in food industry. Among nutraceuticals, utility of phospholipases from other organisms have been reported in food industry as emulsifiers and degumming agents[4]. However, phospholipase from plant source have not been explored for use in food industry.

In our endeavour, we identified two flax $sPLA_2$s in its genome ($LusPLA_2I$ and $LusPLA_2II$) on the basis of their homology in protein sequence, domain structure and phylogenetic relationship (see Supplementary Fig. S1) with other known plant $sPLA_2$s[37]. A single phospholipase A_2 signature (PA2c) domain with ID cd04706 was identified in both the proteins (see Supplementary Table S1). Secretory PLA_2s containing cd04706 have been identified from many plants including rice[18]. This domain contained a conserved Ca^{2+} binding loop and a catalytic domain with enzymatically active His/Asp dyad. His residue of His/Asp dyad is involved in the deprotonation of ester carbonyl carbon of the substrate and Asp residue interacts with Ca^{2+} cofactor through its β carbonyl group[6]. Conserved residues involved in hydrogen bonding in animals, are found in the Ca^{2+} binding loop of plant phospholipases[38]. Pairwise alignment of flax $sPLA_2$s with plant $sPLA_2$ revealed that all conserved motifs identified in arabidopsis and rice[8] are present in flax. However, minor variations in conserved moieties were observed such as replacement of conserved Asp moiety by His93 in $LusPLA_2II$ and by Ser117 in $LusPLA_2I$. Similar, replacement of conserved

Ala by Ser residue in catalytic domain has been reported in arabidopsis sPLA$_2\alpha$ and citrus[8,39]. Nevertheless, the impact of this replacement on the catalytic activity of enzyme is yet to be elucidated.

In-silico tools predicted LusPLA$_2$II to be secreted into the extracellular space (see Supplementary Table S6). Our result is in accordance with arabidopsis (AtsPLA$_2\beta$ and AtsPLA$_2\gamma$) and rice sPLA$_2$ isoforms (OssPLA$_2\beta$ and OsPLA$_2\gamma$)[14,40] that have been confirmed to be secreted into extracellular space. While, LusPLA$_2$II was localized into extracellular space, LusPLA$_2$I is predicted to be secreted into the mitochondria as well as extracellular space (see Supplementary Table S6). This is commensurate with the detection of sPLA$_2$ activity in mitochondria of durum wheat[41]. Since AtsPLA$_2\alpha$ and OssPLA$_2$I are localized into Golgi bodies and ER[14], we speculate LusPLA$_2$I is secreted into the extracellular space and re-localized to mitochondria or vice-versa on interaction with other proteins or in response to external stimuli. Nuclear re-localization of AtsPLA$_2\alpha$ upon interaction with *AtMYB30*[42] also supports our finding. This is plausible, owing to the plethora of roles played by sPLA$_2$ in various cellular processes.

Ontological classification of flax sPLA$_2$s revealed their involvement in many biological and metabolic processes such as growth and development, lipid catabolic process, auxin response, gravitropism, guard cell movement[43], and root development[44] (Table 1). Several studies have highlighted the involvement of sPLA$_2$ in wounding and pathogen attack[45], cold and salinity tolerance[46]. Recently several isoforms of sPLA$_2$s were reported to be highly up-regulated during water stress in wheat[15], rice[14] and arabidopsis[46]. In wheat, PLA$_2$ activity concomitantly increases during drought stress[15] while sPLA$_2$s have been implicated in hyperosmotic stress in *Chlamydomonas*[47].

Biochemical characterization of both LusPLA$_2$ proteins revealed them to be *bona fide* phospholipase A$_2$. Detailed molecular analysis by SDS-PAGE and western blot analysis revealed that observed molecular weight of recombinant proteins were ~22.6 kDa and ~20.6 kDa for LusPLA$_2$I (Fig. 2a,b) and LusPLA$_2$II (Fig. 2c,d) respectively. The observed molecular weight, after accounting for 6x-his tag (~4.7 kDa), of LusPLA$_2$I (~17.9 kDa) and LusPLA$_2$II (~15.7 kDa) respectively is in congruence with their predicted molecular weight. PLA$_2$/LOX- coupled spectrophotometric assay of both the proteins revealed that cloned sPLA$_2$s of flax are *bona fide* lipases as the increase in absorbance results from the release of free linoleate from PC$_{LIN}$ by the activity of LusPLA$_2$I and LusPLA$_2$II (Fig. 3a). Detailed, enzymatic assay of both proteins revealed requirement of micro to mili molar concentration of calcium for their optimum activity. Complete abolition of enzymatic activity by addition of 10 mM EGTA, a potent Ca^{2+} chelator, confirmed our observation (Fig. 3b). The inhibition of enzyme activity due to unavailability of Ca^{2+} can be ascribed to destabilization of transition-state intermediate[48]. Similarly, addition of 5 mM DTT, a disulphide bond reducing agent, completely prevented the protein activity (Fig. 3b). This can be attributed to the destabilization of intra-molecular disulphide bridges[17] prevalent in LusPLA$_2$I and LusPLA$_2$II protein. Similarly, thermostability assay of both the recombinant proteins viz. LusPLA$_2$I and LusPLA$_2$II revealed 50% of their enzymatic activity is retained after boiling at 100°C for 5 minutes (Fig. 3b). It has been reported that wheat sPLA$_2$ isoform III retains ~45% of enzyme activity after boiling at 100°C for 5 minutes[49]. We belive structural stability provided by six disulphide bridges account for the thermo-stability of the proteins.

Analysis of physiochemical properties of both the proteins revealed they have different pI values for LusPLA$_2$I (6.68) and LusPLA$_2$II (8.84) (see Supplementary Table S2). The difference in pI makes LusPLA$_2$I to be acidic while LusPLA$_2$II to be basic. Acidic sPLA$_2$s are also reported from rice (isoform I and III) and soybean-XIA-1 where as slightly alkaline sPLA$_2$s are reported from arabidopsis and soybean-XIA-2[16]. Neutral sPLA$_2$s are reported from carnation (*Dianthus caryophyllus*) and tomato (*Solanum lycopersicum*)[16]. The amino acid composition of both proteins revealed that LusPLA$_2$I is rich in hydrophobic amino acids while LusPLA$_2$II is abundant in hydrophilic amino acid followed by charged amino acid lysine (see Supplementary Table S3). Another parameter, protein instability index of both proteins revealed them to be highly stable. Usually, proteins with instability index values greater than 40 are considered to be unstable proteins[21]. Both LusPLA$_2$s were found to be highly stable as indicated by instability index value below 40 (LusPLA$_2$I- 34.33 and LusPLA$_2$II- 17.13) and higher aliphatic index (LusPLA$_2$I- 95.72 and LusPLA$_2$II- 60.79) (see Supplementary Table S2). Among the two, LusPLA$_2$I is more stable than LusPLA$_2$II. The lower stability of LusPLA$_2$II is indicative of structural flexibility. Additionally, structural flexibility of LusPLA$_2$II can also be ascribed to abundance of glycine moieties that are generally found at the surface of the proteins, within the loops or coils and provide flexibility. Grand average hydropathy results revealed that LusPLA$_2$I with GRAVY score of 0.110 is hydrophobic and LusPLA$_2$II with GRAVY score of (-) 0.101 is hydrophilic (see Supplementary Table S2). A negative GRAVY score indicates soluble nature of the protein. Abundance of glycine with negative GRAVY score further explains the hydrophilic nature of LusPLA$_2$II.

Insight into the secondary structure of flax sPLA$_2$s indicated that they are dominated by alpha helices followed by random coils (see Supplementary Fig. S6). This is corroborated by the three-dimensional model that revealed the presence of 4 α-helices in LusPLA$_2$I (Fig. 4) and 3 α-helices in LusPLA$_2$II (Fig. 5). The precise function of a protein depends on interaction of its exposed surface with solvent. It was found that both the proteins contain more than 60% residues exposed on the surface. This property might account for the involvement of sPLA$_2$s in host of biological processes. Our result suggests, the three-dimensional model of both sPLA$_2$s of flax generated by I-TASSER[50,51] was of correct topology due to higher C-score value of the predicted model. The C-score is a measure of quality of the predicted model and its value ranges from (-) 5 to 2. A higher C-score indicates a high quality model and C-score > (-) 1.5 indicates the correct folding[52]. The C-score values of 0.75 and 0.83 obtained for LusPLA$_2$I and LusPLA$_2$II respectively indicate that both the protein models are of good quality and correct folding.

Similarly, the TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means random similarity. The desired TM score of 0.81 ± 0.09 and 0.83 ± 0.08 for LusPLA$_2$I and LusPLA$_2$II respectively indicate the correct topology of modelled protein (Figs 4a and 5a). These results were further corroborated by the stereochemical stability assessment by Ramachandran plot analysis (see Supplementary Figs S6, S7). Occurrence of 90% or more residues in the most favoured region of Ramachandran plot, classified the refined models of LusPLA$_2$

proteins to be of good quality (see Supplementary Figs S6, S7). The arrangement of 6 disulphide bonds formed by 12 conserved Cys residues is also in accordance with known plant $sPLA_2$s (Fig. 6a,b)[16].

Both the models were based on principle template crystal structure of *Oryza sativa* $sPLA_2II$ (2WG7A)[18]. Despite the deviations in primary and secondary structures, the tertiary structure of flax $sPLA_2$s is well conserved in the regions essential for precise function of the enzyme viz, Ca^{2+} binding loop and catalytic active site containing conserved His/Asp dyad (Figs 4b,c,d and 5b,c,d). This was commensurate with the superimposed structures of $OssPLA_2$-$LusPLA_2I$, $OssPLA_2$-$LusPLA_2II$ and $LusPLA_2I$-$LusPLA_2II$ (Fig. 7). Superimposed structure of $OssPLA_2$-$LusPLA_2I$ revealed that they were similar and both $OssPLA_2$ and $LusPLA_2I$ contained four α-helices. While Ca^{2+} binding loop did not cover α-helix, catalytically active site motif covered most of the region of α-helix 2 in both the proteins. $LusPLA_2II$ contained only three α-helices. The α-helix covering the N-terminal end of $OssPLA_2$ was missing in $LusPLA_2II$. In $LusPLA_2II$ also, Ca^{2+} binding loop did not cover α-helix and catalytically active site motif covered most of the region of α-helix 2. However, the connecting structures varied in their length and morphology among the two $LusPLA_2$s and the template (Fig. 7).

In summary, our work provides first insight into the structure and catalytic mechanism of two $sPLA_2$s in flax. In this endeavour, we expressed and purified active $LusPLA_2I$ and $LusPLA_2II$ from soluble fraction. We carried out the first biochemical characterization of $sPLA_2$s from flax and provides insight into the mechanism of $sPLA_2$ enzyme. The holomeric structure of both the proteins revealed that they are of high quality and topology. Such biochemical and structural analysis providing insight into the structure and function of an important class of protein is required for fine tuning of their *in planta* expression for improving performance of flax during stress.

Methods

Sequence Retrieval and analysis of $sPLA_2$. Protein sequence of secretory phospholipase A_2 of arabidopsis (*Arabidopsis thaliana*), rice (*Oryza sativa*), soybean (*Glycin max*), citrus (*Citrus sinensis*), wheat (*Triticum durum*), flax (*Linum usitatissimum*) and snake venom (*Naja kaouthia*) were retrieved from the sequence repository of the NCBI database (www.ncbi.nlm.nih.gov/) (see Supplementary Table S6). Snake venom $sPLA_2$ sequence was included as an out-group in the study. The conserved domains specific to secretory phospholipase A_2 were identified using the NCBI Conserved Domain Database[53] CDD v3.14–47363 PSSMs (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) annotations. The motifs were predicted using MEME suite[54] (http://meme.nbcr.net/meme/cgi-bin/meme.cgi). The parameter of motif size was set at six amino acids as minimum and 50 amino acids as maximum. The data were analysed using SignalP for post-translational modifications[55]. Putative localization of flax $sPLA_2$s were predicted by TargetP[56] and iPSORT[57]. The sequences of proteins were aligned using Mafft-win[58] and viewed using ESPRIPT[59] and phylogenetic tree was constructed by maximum likelyhood method using MEGA6[60]. UniProt database (http://www.uniprot.org) was used to assign gene ontology classifications[61].

Cloning, expression and purification of $sPLA_2$ protein. Total RNA isolated from flax leaves was used for cDNA synthesis using SuperScript™ III First-Strand Synthesis supermix following the manufacturer's instructions (Life technologies Corporation). Sequence of flax $sPLA_2$ isoform *$LusPLA_2I$* (Genbank accession: KU361324) and *$LusPLA_2II$* (Genbank accession: KU361324) were amplified using primer pairs listed in Supplementary Table S7. The amplified gene products were cloned in Gateway entry vector pENTR/SD/D/TOPO (Life technologies Corporation) as per manufacturer's instructions and authenticity was confirmed by sequencing. For protein expression, the *$LusPLA_2I$* and *$LusPLA_2II$* genes were mobilized from entry vector into the destination vector, pET301/CT-DEST (Invitrogen Corporation) to generate the expression clones pET301/CT-DEST harbouring *$LusPLA_2I$*-6xHis and pET301/CT-DEST harbouring *$LusPLA_2II$*-6xHis as per manufacturer's instructions.

The *E.coli* (CodonPlus (DE3)-RIPL cells (Agilent Technologies) cells harbouring the *$LusPLA_2I$*- 6xHis and *$LusPLA_2II$*-6xHis recombinant plasmid were grown at 37°C in LB medium containing 100 µg/ml carbenicillin until $OD_{600} = 0.6$ and induced with 0.4 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 4 h. The cells were harvested by centrifugation at 10,000 rpm, 4°C for 10 min and lysed by re-suspending in native lysis buffer containing lysozyme and sonicated at 24–25% amplitude 30 sec ON/OFF cycle on ice for 12–15 minutes ON condition. The recombinant proteins were purified by using Ni-NTA affinity chromatography by using QIAexpress® Ni-NTA Fast Start kit by following manufacturer's instructions. The purity and yield of recombinant protein were analysed by SDS-PAGE and protein content was determined by Bradford's method[62].

SDS-PAGE and western blotting. The recombinant fusion protein $LusPLA_2I$-6xHis and $LusPLA_2II$-6xHis were resolved on a 12% (w/v) SDS-PAGE as described earlier[63] and visualized after coomassie brilliant blue staining. For western blot analysis, both the proteins were resolved on 12% acrylamide gel and transferred to PVDF (polyvinylidene diflouride) membrane using electro-blot system (Major science, USA) at constant voltage of 100 V, 15°C for 3 h in 1X Towbin buffer (25 mM Tris, 192 mM glycine, 20% methanol and 0.05% SDS). The western blotting was performed using WesternBreeze® Chemiluminescent Kit (Life technologies Corporation), as per manufacturer's instructions.

PLA_2 activity determination. PLA_2 activity was evaluated using spectrophotometric method based on PLA_2/lipoxygenase (LOX) coupled reactions[15]. The PLA_2 activity was assayed by addition of the recombinant protein to 4 E.U. LOX, 0.5 mM 1,2-diacyl-sn-glycero-3-phosphocholine (PC_{LIN}) and 1 mM $CaCl_2$ in 2 ml 50 mM sodium borate buffer (pH 9.0). Increase in absorbance at 234 nm was monitored. The effect of 10 mM ethylene glycol bis (-2 amino ethyl ether) – N, N, N, N – tetra acetic acid (EGTA), a Ca^{2+} chelator and 5 mM dithiothreitol (DTT), a disulphide bond reducing agent on enzyme activity was evaluated. Effect of heat (100°C for 5 min)[15] on protein activity was studied to ascertain thermostability of proteins.

Physiochemical characterisation and secondary structure analysis. To gather the structural information about flax $sPLA_2$s, physiochemical properties and secondary structure was predicted for both the

proteins. Physio-chemical characterization was performed using the Expasy's ProtParam server (http://us.expasy.org/tools/protparam.html)[64]. Isoelectric point (pI), molecular weight, instability index[21], aliphatic index[22] and grand average hydropathy (GRAVY)[23] were predicted to estimate the stability of the protein. Predict Protein server (https://www.predictprotein.org/)[65] was employed for secondary structure prediction of both LusPLA$_2$ isoforms. Presence of alpha helices, strands and loops were analyzed and exposed surface for solvent accessibility of protein was calculated by Predict Protein server.

Homology modelling and evaluation. Multiple online tools are available for predicting the three dimensional model of protein such as Chunk-TASSER[66], Meta-TASSER[67], Pro-sp3-TASSER[68], TASSER-VMT[69] and I-TASSER[70]. Initial model of flax sPLA$_2$I and sPLA$_2$II was constructed using fully automated I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) that combined threading and *ab-initio* modelling for the prediction which automatically selected rice sPLA$_2$II as the template for modelling. The initial model generated using I-TASSER was refined further by ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner/). The refined models were checked for stereo-chemical properties by Ramachandran Plot analysis using RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php)[71]. The residues falling in the disallowed region of the Ramachandran Plot were further refined by Modloop, which automatically modelled the loops of protein structures (https://modbase.compbio.ucsf.edu/modloop/)[72]. The protein model was refined until all the amino acid residues fell in the favoured region of the Ramachandran plot (RAMPAGE). The validation of the modelled structure was performed to determine the accuracy of secondary structure predictions and modelling by PDBSum (https://www.ebi.ac.uk/pdbsum/)[26]. PROCHECK was used to confirm if all the residues were falling in the most favoured regions of the Ramachandran Plot[25]. Structure visualization was performed with pymol (http://www.pymol.org). The predicted models of proteins were submitted to Protein Model Database (PMDB)[73] and have been assigned identifier PM0080416 (LusPLA$_2$I) and PM0080415 (LusPLA$_2$II).

Data availability. All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

1. Karg, S. New research on the cultural history of the useful plant *Linum usitatissimum* L.(flax), a resource for food and textiles for 8,000 years. *Veg. Hist. Archaeobot.* **20**, 507–508 (2011).
2. Rodriguez-Leyva, D., Bassett, C. M., McCullough, R. & Pierce, G. N. The cardiovascular effects of flaxseed and its omega-3 fatty acid, alpha-linolenic acid. *Can. J. Cardiol.* **26**, 489–496 (2010).
3. Touré, A. & Xueming, X. Flaxseed lignans: source, biosynthesis, metabolism, antioxidant activity, bio-active components, and health benefits. *Compr. Rev. Food. Sci. Food. Saf.* **9**, 261–269 (2010).
4. De Maria, L., Vind, J., Oxenbøll, K. M., Svendsen, A. & Patkar, S. Phospholipases and their industrial applications. *App. Microbiol. Biotechnol.* **74**, 290–300 (2007).
5. Mansfeld, J. Plant phospholipases A$_2$: perspectives on biotechnological applications. *Biotechnol. Lett.* **31**, 1373 (2009).
6. Gupta, P., Saini, R. & Dash, P. K. Origin and evolution of group XI secretory phospholipase A$_2$ from flax (*Linum usitatissimum*) based on phylogenetic analysis of conserved domains. 3 Biotech 10.1007/s13205-017-0790-x (2017).
7. Six, D. A. & Dennis, E. A. The expanding superfamily of phospholipase A$_2$ enzymes: classification and characterization. *BBA-Mol. Cell. Biol. L.* **1488**, 1–9 (2000).
8. Lee, H. Y. *et al.* Secretory low molecular weight phospholipase A$_2$ plays important roles in cell elongation and shoot gravitropism in *Arabidopsis*. *Plant Cell* **15**, 990–2002 (2003).
9. Seo, J. *et al.* Phospholipase A$_2\beta$ mediates light-induced stomatal opening in *Arabidopsis*. *J. Exp. Bot.* **59**, 3587–3594 (2008).
10. Kim, J. Y. *et al.* Characterization of the full-length sequences of phospholipase A$_2$ induced during flower development. *Biochim. Biophys. Acta.* **1489**, 389–392 (1999).
11. Kim, H. J. *et al.* Endoplasmic reticulum and golgi localized phospholipase A$_2$ plays critical roles in *Arabidopsis* pollen development and germination. *Plant Cell* **23**, 94–110 (2011).
12. May, C., Preisig-Muller, R., Hohne, M., Gnau, P. & Kindl, H. A phospholipase A$_2$ is transiently synthesized during seed germination and localized to lipid bodies. *Biochim. Biophys. Acta.* **1393**, 267–276 (1998).
13. Pohnert, G. Phospholipase A$_2$ activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. *Plant Physiol.* **129**, 103–111 (2002).
14. Singh, A. *et al.* Rice phospholipase A superfamily: organization, phylogenetic and expression analysis during abiotic stresses and development. *PLoS One* **7**, e30947 (2012).
15. Verlotta, A., Liberatore, M. T., Cattivelli, L. & Trono, D. Secretory phospholipases A$_2$ in durum wheat (*Triticum durum* Desf.): gene expression, enzymatic activity, and relation to drought stress adaptation. *Int. J. Mol. Sci.* **14**, 5146–5169 (2013).
16. Lee, H. Y., Bahn, S. C., Shin, J. S., Hwang, I. & Back, K. Multiple forms of secretory phospholipase A$_2$ in plants. *Prog. Lipid Res.* **44**, 52–67 (2005).
17. Mansfeld, J., Gebauer, S., Dathe, K. & Ulbrich-Hofmann, R. Secretory phospholipase A$_2$ from *Arabidopsis thaliana*: insights into the three-dimensional structure and the amino acids involved in catalysis. *Biochemistry* **45**, 5687–5694 (2006).
18. Guy, J., Stahl, U. & Lindqvist, Y. Crystal structure of a class XIB phospholipase A$_2$: rice (*Oryza sativa*) isoform-2 PLA$_2$ and an octanoate complex. *J. Biol. Chem.* **284**, 19371–19379 (2009).
19. Venglat, P. *et al.* Gene expression analysis of flax seed development. *BMC. Plant Biol.* **11**, 74 (2011).
20. Fong, J. H. & Marchler-Bauer, A. Protein subfamily assignment using the Conserved Domain Database. *BMC Res. Notes* **1**, 114 (2008).
21. Guruprasad, K., Reddy, B. V. & Pandit, M. W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* **4**, 155–161 (1990).
22. Ikai, A. J. Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898 (1980).
23. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
24. Rost, B. PHD: Predicting one-dimensional protein structure by profile-based neural networks. *Meth. in Enzym.* **266**, 525–539 (1996).
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
26. Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **33**, 266–268 (2005).
27. Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 167–339 (1981).

28. Hutchinson, E. G. & Thornton, J. M. A revised set of potentials for beta turn formation in proteins. *Protein Sci.* **3**, 2207–2216 (1994).
29. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759–777 (1989).
30. Milner-White, E. J., Ross, B. M., Ismail, R., Belhadj-Mastefa, K. & Poet, R. One type of gamma turn, rather than the other, gives rise to chain reversal in proteins. *J. Mol. Biol.* **204**, 777–782 (1988).
31. Rose, G. D., Gierasch, L. M. & Smith, J. A. Turns in peptides and proteins. *Adv. Prot. Chem.* **37**, 1–109 (1985).
32. Wang, Z. *et al*. The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J.* **72**, 461–473 (2012).
33. Dash, P. K. *et al*. Genome-wide analysis of drought induced gene expression changes in flax (*Linum usitatissimum*). *GM crops* & *food* **5**, 106–119 (2014).
34. Dash, P., Gupta, P. & Rai, R. Hydroponic method of halophobic response elicitation in flax (*Linum usitatissimum*) for precise down-stream gene expression studies. *Int. J. Trop. Agri.* **33**, 1079–1085 (2015).
35. Gupta, P. & Dash, P. K. Precise method of *in situ* drought stress induction in flax (*Linum usitatissimum*) for RNA isolation towards down-stream analysis. *Ann. Agri. Res.* **36**, 10–17 (2015).
36. Shivaraj, S. M. *et al*. Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (*Linum usitatissimum*). *Sci. Rep.* **7**, 46137 (2017).
37. Mariani, M. E. *et al*. In silico and *in vitro* characterization of phospholipase A_2 isoforms from soybean (*Glycine max*). *Biochimie* **94**, 2608–2619 (2012).
38. Murakami, M. *et al*. Recent progress in phospholipase A_2 research: from cells to animals to humans. *Prog. Lipid Res.* **50**, 152–192 (2011).
39. Liao, H. L. & Burns, J. K. Light controls phospholipase $A_2\alpha$ and β gene expression in *Citrus sinensis*. *J. Exp. Bot.* **61**, 2469–2478 (2010).
40. Ståhl, U. *et al*. Plant low-molecular-weight phospholipase A_2s (PLA_2s) are structurally related to the animal secretory PLA_2s and are present as a family of isoforms in rice (*Oryza sativa*). *Plant Mol. Biol.* **41**, 481–490 (1999).
41. Trono, D., Soccio, M., Laus, M. N. & Pastore, D. The existence of phospholipase A_2 activity in plant mitochondria and its activation by hyperosmotic stress in durum wheat (*Triticum durum* Desf.). *Plant Sci.* **199**, 91–102 (2013).
42. Froidure, S. *et al*. AtsPLA$_2$-α nuclear relocalization by the *Arabidopsis* transcription factor AtMYB30 leads to repression of the plant defense response. *Proc. Natl. Acad. Sci. USA* **107**, 15281–15286 (2010).
43. Ryu, S. B. Phospholipid-derived signaling mediated by phospholipae A in plants. *Trends Plant Sci.* **9**, 229–235 (2004).
44. Lee, O. R. *et al*. Phospholipase A_2 is required for PIN-FORMED protein trafficking to the plasma membrane in the *Arabidopsis* root. *Plant Cell* **22**, 1812–1825 (2010).
45. Laxalt, A. M. & Munnik, T. Phospholipid signalling in plant defence. *Curr. Opin. Plant Biol.* **5**, 332–338 (2002).
46. Narusaka, Y. *et al*. Expression profiles of *Arabidopsis* phospholipase A IIA gene in response to biotic and abiotic stresses. *Plant Cell Physiol.* **44**, 1246–1252 (2003).
47. Munnik, T. & Meijer, H. J. Osmotic stress activates distinct lipid and MAPK signalling pathways in plants. *FEBS Lett.* **498**, 172–178 (2001).
48. Scott, D. L. *et al*. Interfacial catalysis: the mechanism of phospholipase A_2. *Science* **250**, 1541–1546 (1990).
49. Verlotta, A. & Daniela, T. "Expression, purification and refolding of active durum wheat (*Triticum durum* Desf.) secretory phospholipase A_2 from inclusion bodies of *Escherichia coli*". *Protein Expr. Purifi.* **101**, 28–36 (2014).
50. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
51. Yang, J. *et al*. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
52. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Struct., Funct., Bioinf.* **68**, 1020 (2007).
53. Marchler-Bauer, A. *et al*. CDD: NCBI's conserved domain database. *Nucleic Acids Res. gku1221* (2014).
54. Bailey, T. L. *et al*. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res. gkp335* (2009).
55. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
56. Emanuelsson, O., Nielsen, H., Brunak, S. & Von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
57. Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. & Miyano, S. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298–305 (2002).
58. Katoh, K., Kuma, K. I., Miyata, T. & Toh, H. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome inform.* **16**, 22–33 (2005).
59. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).
60. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
61. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res. gku989* (2014).
62. Bradford, M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
63. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
64. Gasteiger, E. *et al*. Protein identification and analysis tools on the ExPASy server. *Humana Press* 571–607 (2005).
65. Rost, B., Yachdav, G. & Liu, J. The predictprotein server. *Nucleic Acids Res.* **32**, 321–326 (2004).
66. Zhou, H. & Skolnick, J. Ab initio protein structure prediction using chunk-TASSER. *Biophys. J.* **93**, 1510–1518 (2007).
67. Zhou, H., Pandit, S. B. & Skolnick, J. Performance of the Pro-sp0-TASSER server in CASP8. *Proteins: Struct. Funct. Bioinf.* **77**, 123–127 (2009).
68. Zhou, H. & Skolnick, J. Protein structure prediction by pro-Sp3-TASSER. *Biophys. J.* **96**, 2119–2127 (2009).
69. Zhou, H. & Skolnick, J. Template-based protein structure modeling using TASSERVMT. Proteins: Struct. *Funct. Bioinf.* **80**, 352–361 (2012).
70. Zhang, Y. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins: Struct. Funct. Bioinf.* **77**, 100–113 (2009).
71. Lovell, S. C. *et al*. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Struct. Funct. Bioinf.* **50**, 437–50 (2003).
72. Fiser, A. & Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500–2501 (2003).
73. Castrignanò, T., De Meo, P. D. O., Cozzetto, D., Talamo, I. G. & Tramontano, A. The PMDB protein model database. *Nucleic Acids Res.* **34**, 306–309 (2006).

Acknowledgements

The work was carried out under in-house research (RPP- Sanction no 7–5/2010-IC-IV) and ICAR-NPTC to P.K.D. by Indian Council of Agricultural Research (ICAR), New Delhi, India. Financial support in the form of research scholarship to P.G. is acknowledged.

Author Contributions

P.G. and P.K.D. designed and carried out the experiments. P.G. analyzed bioinformatic data and prepared the manuscript with contribution from P.K.D.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10969-9

Competing Interests: The authors declare that they have no competing interests.

Accession codes: Following accession codes were used in the study. AGD95019 (*Triticum durum* $sPLA_2$ isoform I), AGD95020 (*T. Durum* $sPLA_2$ isoform II), AGD95021 (*T. Durum* $sPLA_2$ isoform III), AGD95022 (*T. Durum* $sPLA_2$ isoform IV), BAS81761 (*Oryza sativa* $sPLA_2$ isoform I), BAS83356 (*O. sativa* $sPLA_2$ isoform II), BAS85989 (*O. sativa* $sPLA_2$ isoform III), BAT14386 (*O. sativa* $sPLA_2$ isoform IV), AEC06025 (*Arabidopsis thaliana* $sPLA_2$ isoform I), AEC06913 (*A. thaliana* $sPLA_2$ isoform II), AEE85635 (*A. thaliana* $sPLA_2$ isoform III), AEE85634 (*A. thaliana* $sPLA_2$ isoform IV), KU361324 (*Linum ussitatisimum* $sPLA_2$ isoform I), KU361325 (*L. ussitatisimum* $sPLA_2$ isoform II), NP_001240142 (*Glycine max* $sPLA_2$ isoform I), NP_001241376 (*G. max* $sPLA_2$ isoform II), BAA36404 (*Naja kaouthia*).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.