Plant phospholipase D mining unravels new conserved residues important for catalytic activity

Yani Arhab, Abdelkarim Abousalham, Alexandre Noiriel

Plant Phospholipase D mining unravels new conserved residues important for catalytic activity

Yani Arhab [‡], *Abdelkarim Abousalham* [‡], *and Alexandre Noiriel* [‡*]

[‡] Univ Lyon, Université Lyon 1, Institut de Chimie et de Biochimie Moléculaires et Supramoléculaires (ICBMS), UMR 5246 CNRS, Métabolisme, Enzymes et Mécanismes Moléculaires (MEM^2), Bât Raulin, 43 Bd du 11 Novembre 1918, F-69622 Villeurbanne cedex, France.

[*] Correspondence should be addressed to : Alexandre Noiriel, Université Lyon 1, Institut de Chimie et de Biochimie Moléculaires et Supramoléculaires (ICBMS), UMR 5246 CNRS, Métabolisme, Enzymes et Mécanismes Moléculaires (MEM^2), Bât Raulin, 43 Bd du 11 Novembre 1918, F-69622 Villeurbanne cedex, France.
Alexandre.Noiriel@univ-lyon1.fr
Tel (+33) 427 465 731

Keywords

Arabidopsis thaliana, C2 domain, HKD motif, Phospholipase D, post-translational modification, transphosphatidylation.

Abstract

Phospholipase D (PLD) is a key enzyme involved in numerous processes in all living organisms. Hydrolysis of phospholipids by PLD allows the release of phosphatidic acid which is a crucial intermediate of multiple pathways and signaling reactions, including tumorigenesis in mammals and defense responses in plants. One common feature found in the plant alpha isoform (PLDα), in some PLD from microbes and in all PLD from eukaryotes, is a duplicated motif named HKD involved in the catalysis. However, other residues are strictly conserved among these organisms and their role remains obscure. To gain further insights into PLD structure and the role of these conserved residues, we first looked for all the plant PLDα sequences available in public databases. With more than 200 sequences retrieved, a generic sequence was constructed showing that 138 residues are strictly conserved among plant PLDα, with some of them identical to residues found in mammalian PLDs. Using site-directed mutagenesis of the PLDα from *Arabidopsis thaliana*, we demonstrated that mutation of some of these residues abolished the PLD activity. Moreover, mutation of the residues around both HKD motifs enabled us to re-define the consensus sequence of these motifs. By sequential deletions of the N-terminal extremity, the minimum length of the domain required for catalytic activity was determined. Overall, this work furthers our understanding of the structure of eukaryotic PLDs and it may lead to the discovery of new regions involved in the catalytic reaction that could be targeted by small molecule modulators of PLDs.

Introduction

The phospholipase D (PLD) superfamily is a diverse group of enzymes found in microbes, fungi, viruses, plants and animals (1). Members of the PLD superfamily include several classes of enzymes that all show activity on the phosphodiester bond; as classical PLDs (EC 3.1.4.4) catalyzing the hydrolysis of the distal phosphoester bond of phospholipids; as cardiolipin synthases and phosphatidylserine synthases, which catalyze a phosphatidyl transfer reaction; and in some bacterial nucleases, such as Nuc, that catalyzes the hydrolysis of DNA phosphodiester bonds (2).

In plants, where PLD activity was first discovered (3), several types of PLDs exist but the alpha isoform (PLDα) is particularly widespread. Different PLD isoforms are implicated in stress physiology and signal transduction *via* hydrolysis of phospholipids. More generally, PLDs are involved in a broad range of cellular and physiological processes (2), notably in numerous pathological disorders in mammals (4–6). As an example, phosphatidic acid (PA), which is a product of the PLD-catalyzed hydrolysis of various phospholipids, is able to recruit and regulate numerous membrane-associated proteins, such as the mammalian target of rapamycin (mTOR) or the proto-oncogenic serine/threonine kinase Raf-1, making PLD an important therapeutic target (5–9).

In addition to their hydrolytic reactions, PLDs are also able to catalyze a transphosphatidylation reaction between a phospholipid and a primary alcohol that permits the exchange of the polar head of phospholipids and, consequently, contributes to the diversity of phospholipids (2).

PLD enzymes can be divided into two distinct groups based on their primary structure: the non-HKD group and the HKD group (2). The non-HKD PLD group is made up of well-studied

enzymes, such as the PLDs from *Streptomyces chromofuscus* (10) or from the spider *Loxosceles* (11) that belong to the superfamily of PLC-like phosphodiesterases (12).The HKD group, which is a member of the PLD superfamily, is characterized by the presence of a conserved His, Lys and Asp in a typical $HXKX_4D$ consensus sequence, also called the HKD motif, where X denotes any kind of residue. This motif is reported to be mostly found in mammals, with the PLD1 and PLD2 isoforms present in *Homo sapiens*; in plants, with the predominant PLDα isoform; in yeast, with SPO14 (13); and in some microbes, with the crystallized PLD from *Streptomyces sp*. strain PMF (14). In plants, mammals and in some bacteria, the HKD motif is duplicated in the primary structure and both motifs are thought to be close together in the tertiary structure, forming a catalytic site at the interface of the bi-lobed enzyme (1).

In plants, as in the model plant *A. thaliana*, multiple PLD genes encoding isoforms are grouped into six classes : α, β, γ, δ, ε and ζ, based on the genic architecture, sequence similarities, domain structures, and biochemical properties (15–17). The conventional PLDα, in contrast to other PLD isoforms, is phosphatidylinositol-4,5-bisphosphate (PIP_2)-independent and requires millimolar levels of Ca^{2+} for *in vitro* activity at a physiological pH (18, 19).

The PLDα isoforms bear a C2 domain of around 150 residues, at the N-terminal extremity, that are involved in the Ca^{2+}-dependent membrane binding and adsorption of the enzyme (15, 20, 21). We have recently demonstrated that the *Arabidopsis* PLDα C2 domain binds phosphatidylglycerol in a Ca^{2+}-independent manner, whilst PA and phosphatidylserine bindings were found to be enhanced in the presence of Ca^{2+} (20). However, because the mature PLDα is cleaved at its N-terminal extremity, it is not yet clear which residues of this C2 domain are necessary for the Ca^{2+}-dependent membrane binding. Indeed, this region undergoes a post-translational modification (PTM), described as a cleavage of thirty to forty residues at the N-

terminal C2 domain extremity. This PTM is found in two situations, either when the PLDα is purified directly from plant tissues, such as *Brassica oleracea* (22), *Ricinus communis* (23), *Oriza sativa* (24), and *Glycine max* (25), or when the protein is purified in a recombinant way. Examples of the latter include *Vigna unguiculata* PLDα, expressed in insect cells (26) or in *P. pastoris* (27), and *A. thaliana* PLDα expressed in *P. pastoris* (20). An exception is the PLD from *Brassica oleracea* (28) expressed in *E. coli*, where the absence of this PTM seems not to impair the catalytic activity. Consequently, the importance and the role of the N-terminal extremity is questionable and requires further investigation. There is no evidence in the literature that PLDα undergoes numerous kinds of PTM in the way that human PLDs do (2), with the exception of phosphorylation that has been observed in a couple of hydroxylated residues (29).

Recombinant plant PLDα is easy to produce in large amounts and requires only Ca^{2+} as a cofactor, compared to other plant isoforms and to mammalian PLDs that require numerous cofactors such as G proteins, PIP_2, fatty acids, etc (2). Moreover, plant PLDα activity can be measured easily, either discontinuously or in a direct and continuous way using an assay based on the chelation-enhanced fluorescence property of 8-hydroxyquinoline following Ca^{2+}-complexation with PLD-generated PA (30). This is in contrast to mammalian PLDs that have to be assayed discontinuously using radiolabeled substrates in end-point kinetics. Hence, plant PLDα is a valuable tool for studying the structural and functional properties of eukaryotic PLDs.

The importance of the invariant charged motif $HXKX_4D$ was first reported and defined in 1995 (31). Eukaryotic PLDs are thought to follow a two-step ping-pong reaction mechanism between the two HKD domains that requires both to be present (32) and intact to result in a biochemical reaction. This means that the duplicated motifs do not function independently, as has been demonstrated for the human PLD1 (33), signifying that a single mutation in these motifs

abolishes the catalytic activity (33). The tridimensional (3D) structure of the endonuclease Nuc from *Salmonella typhimurium*, a prokaryotic PLD family member which bears only one HKD motif, was solved in 1999 (34). The Nuc enzyme crystallizes as a dimer in which the two HKD motifs come together to form a functional catalytic site at the interface of the subunits, and the consensus sequence of the HKD motif has been defined as $HXKX_4DX_6GSXN$ (34). The structural position of the two HKD domains was further confirmed in the crystallized form of the monomeric bacterial PLD from the *Streptomyces sp*. strain PMF (14). Because the HKD motif has only been described in the extended version $HXKX_4DX_6GSXN$, which does not differentiate the first motif from the second one, and also because this version relies on sequence alignment exclusively, we propose a reevaluation of each one of the HKD motifs based on biochemical enzymatic evidence.

The importance of some residues of PLD family members has already been verified, for example in Ymt, the *Yersinia pestis* murine toxin (35), where the H, K, and S residues of both motifs were mutated. In plants, as far as the PLDα isoform is concerned, only a few conserved domains have been identified so far. Mutations of the EFK motif, analogous to the DRY motif, have shown that this motif facilitates the interaction between the Gα subunit of the G-protein and the PLDα isoform (36, 37).

The N-terminus extremity of human PLDs which is not conserved among mammalian PLDs was shown not to be critical. Expression of mutants lacking the first 325 residues of PLD1 (38) or the first 308 residues of PLD2 (39) remained catalytically active. However, the importance of the C-terminal extremity has been proven to be crucial: mutation of the C-terminal threonine of PLD2 by an alanine conferred essentially wild-type PLD activity but the addition of a single residue dramatically altered the catalytic activity (39) and when the carboxyl terminus of PLD1 was

deleted, the enzyme was found to be inactive (38). The addition of a hexahistidine tag at the C-terminal extremity of PLD1 led to a complete attenuation of the activity (40). In cabbage, it is postulated that the C-terminus also contributes to the functional conformation of active PLDα when expressed in *E. coli* (41). In addition, this latter study analyzed the involvement of some conserved residues in the hydrolytic and the transphosphatidylation reactions.

PLDs are now an important target in human health care (42), especially for cancer treatment (43). Since plant and mammalian PLDs share unique and common features, we used PLDα, the major isoform found in plants, in this study to investigate the importance of the conserved residues and to further our understanding of their involvement in PLD catalysis.

Material and Methods

Material

Zeocin was obtained from InvivoGen (Toulouse, France), and Zymolyase and Ampicillin were purchased from Euromedex (Souffelweyersheim, France). *P. pastoris* X-33 strain was from Invitrogen. 1,2-Dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), choline oxidase from *Arthrobacter globiformis*, horseradish peroxidase (type VI), and bovine serum albumin (BSA) were obtained from Sigma-Aldrich Chimie (Saint-Quentin-Fallavier, France). All other chemicals and solvents of highest quality were obtained from local suppliers.

Pichia pastoris transformation

An isolated clone of *P. pastoris* was cultured overnight in 50 mL of Yeast Peptone Dextrose (YPD) medium, at 30°C, and the OD was then measured. Cells were harvested by centrifugation, for 5 min at 1,500 x *g*, and then washed twice with 40 mL of sterile water and twice with 20 mL 1 M sorbitol at room temperature. Cells were finally resuspended in a volume of 1 M sorbitol with an OD equivalent to 75.

Competent cells were transformed with 0.4 – 2.0 µg of AvrII linearized plasmid purified using the Qiaquick kit from Qiagen. Briefly, 80 µL of cells were mixed with DNA in an electroporation cuve and received an electric shock of 2 kV for 6 ms. Cells were then resuspended in 1 mL of 1 M sorbitol and incubated, for 90 min at 30°C, then plated onto solid

YPD with Zeocin at 0.1 mg/mL. After 2 days of incubation at 30°C, clones were then re-streaked onto YPD medium containing Zeocin at 0.2 mg/mL.

Protein expression

Each mutant, and the appropriate controls, was cultured for 48 h at 30°C, under orbital shaking (180 rpm), in 4 mL of YPD medium supplemented with 0.1 mg/mL Zeocin. The cells were then harvested by centrifugation and lysed using either a Mini Beadbeater or a FastPrep-24™ 5G Instrument (Biospec), for 5 times 30 s, with samples conserved on ice in between, using 1 mL of glass beads (425 - 600 µm diameter) in lysis buffer composed of Tris 20 mM pH 8.0, NaCl 150 mM and sucrose 200 mM. The soluble fraction was obtained after 10 min of centrifugation at 9,500 x *g* at 4°C. The protein concentrations were measured using Bradford's method (44) with Bio-Rad Dye Reagent and BSA as the standard.

PLD activity measurement

PLD activity was assayed according to (45). Briefly, the protein extract to be tested was added to the assay mixture (100 µL final volume), composed of 0.4 mM DMPC, 1.24 mM SDS, 1.24 mM Triton X-100, Tris-HCl 50 mM with 20 mM $CaCl_2$ pH 8.0, and incubated for 10 minutes at 30°C. The reaction was stopped by adding 25 µL of 500 mM EDTA pH 8.0 and the PLD activity was revealed by adding 25 µL of chromogenic mixture composed of 280 mM Tris-HCl pH 8.0, 0.5 U choline oxidase, 0.5 U peroxidase, 10 mM 4-aminoantipyrine and 54 mM sodium 2-hydroxy-3,5-dichlorobenzene sulfonate. The PLD-generated choline was quantified with end

point measurements by recording the absorbance at 500 nm, based on a standard curve obtained with pure choline. Control assays were performed simultaneously in the absence of PLD.

Alternatively, PLD activity was assayed with 8-hydroxyquinoline, as described previously (30).

Immunoblotting analysis

Crude protein extracts (10 µg) were separated using SDS-PAGE on 10% gels, as described by Laemmli (46). Proteins were electroblotted onto nitrocellulose membranes for 90 min at 250 mA in a transfer buffer composed of 25 mM Tris, 200 mM glycine, and 20% (v/v) ethanol, using a Trans-Blot apparatus (Bio-Rad). Nonspecific protein binding sites were blocked by incubating membranes for 1 h in a blocking solution containing 5% (w/v) BSA and 0.1% (v/v) Tween 20 in Tris Buffer Saline (TBS). Membranes were then incubated overnight at 4°C with polyclonal antibodies (serum diluted at 1: 4,000), raised in rabbit against a peptide corresponding to the last 117 residues of the *A. thaliana* PLDα C2 domain (20), in TBS containing 0.1% (v/v) Tween 20 and 1% (w/v) BSA. Unbound primary antibodies were removed by four washes (5 min each) with TBS containing 0.1% (v/v) Tween 20 (TBS-T). The membranes were incubated with peroxidase conjugated goat anti-rabbit IgG (Sigma) (diluted 1: 10,000 with TBS-T) for 1 h followed by four washes (5 min each) with TBS-T and then revealed with enhanced chemiluminescence Western blotting solutions (Amersham) according to the manufacturer's instructions.

Because of the numerous clones, the expression level of PLDα cannot be normalized between all samples. The SDS-PAGE technique was used for one out of three independent clones to prove

that the protein was expressed or not and no consideration of relative activity between different mutants can be inferred.

Monoclonal anti-polyhistidine-peroxydase antibody (reference A7058-1VL, Sigma-Aldrich) produced in mouse was incubated overnight at 4°C (dilution 1: 2,000). Unbound antibodies were removed by four washes (5 min each) with TBS-T and revealed with enhanced chemiluminescence Western blotting solutions (Amersham) according to the manufacturer's instructions.

Data mining

To gather a maximum of PLDα sequences we used as a seed the *A. thaliana* PLDα sequence (U36381) (47) which harbors the following natural variations R26M, A218T, L323S, I498T, and E580G, compared to AT1G52570. We carried out a tBLASTn search on the NCBI database, with default parameters, to acquire a first set of sequences. Then we performed a BLASTp search, using the keyword "PLD", on the Phytozome v12.1 database. With these sequences, we implemented the following criteria (i, ii, and iii) to screen all the remaining plant alpha-type PLD (Table S1) from 31 different databases whether they were properly annotated or not. **(i)** The nucleotide sequences should be complete from the starting methionine to the stop codon, and the length should be within the range of 800 to 840 residues. All partial fragments were excluded, explaining the absence of some organisms such as *Secale cereale*. **(ii)** The sequences should contain all the following markers, from the N-terminus to the C-terminus, that have been found in published PLDα: (a) directly downstream of the starting methionine a consensus sequence HGX_5I; (b) a sequence (I/V)(Y/C)(I/V)XGWS(V/I) between the 220^{th} and 240^{th} residues; (c)

then a conserved (M/V)XT(H/G)D sequence; (d) a first HKD catalytic motif $HXKX_4D$; (e) a conserved motif (F/Y)(I/V)Y(I/V)ENQYF(L/I)GX_4W around the 520^{th} residues with a typical (Y/F)(I/V)XAIRXA motif upstream; (f) then the second HKD motif $HXKX_4D$; (g) and, finally, a consensus sequence PX_2(L/I)T(T/S). **(iii)** It is important that the sequences had been correctly spliced. Twenty-five sequences were manually corrected because some introns were identified and not correctly *in silico* spliced. Typically, for *Aegilops sharonensis* and *A. speltoides,* as for *Triticum durum* and *T. monoccocum, Cannabis sativa, Dianthus caryophyllus, Oryza punctata* and *Pinus taeda,* the same introns were manually spliced both in the N-terminus region and downstream of the second HKD motif. For *Aquilegia coerulea, Citrullus lanatus* and *Ipomoea trifida,* an intron was manually spliced at the N-terminus extremity. Fragments of sequences have been assembled to build the *Arabis alpina, Digitalis purpurea, Dioscorea villosa, Hevea brasiliensis, Picea sitchensis* and *Valeriana officinalis* PLDα sequences. For *Arachis duranensis,* a missing fragment at the N-terminus extremity was added from another coding DNA sequence (CDS). There was an erroneous splicing in the *Pimpinella brachycarpa* sequence, where an intron was found to replace the missing (F/Y)(I/V)Y(I/V)ENQYF(L/I)Gx_4W consensus motif. This was corrected with the appropriate exon. Because there were some X in the sequences of *Chenopodium quinoa*, *Gincko biloba*, *Nelumbo nucifera* and *Nicotiana tabacum*, we replaced those with the corresponding residues found in the corresponding genomic sequence. Also, the sequences of *Oryza glumipatula* and *Quercus rubra* that did not start at the usual methionine were manually corrected by removing the peptide that did not fit the alignment.

In the case where several sequences were identified in a single organism, corresponding to multiple α isoforms, we deliberately selected the closest isoform, in term of sequence identity, to

our PLDα model from *A. thaliana*. The purpose was to retain only one PLD per organism and not to influence the consensus sequence.

It should be stated that a similar PLDα sequence has been found in the bat *Rhinolophus sinicus* recently sequenced (48). First, this sequence bore a striking resemblance to plant PLDα. Secondly, bats, and particularly *Rhinolophus sinicus*, exhibit, in their genomic databases, typical PLDs found in mammals, and, finally, such a plant-like sequence has not been found in complete sequenced genomes from other bats or mammals. Consequently, we presume that this sequence is due to a biological contamination during the sequencing. Nevertheless, because this sequence increased diversity in the sampling of PLDα we retained it, despite its doubtful origin.

Cloning of *PLDα* from *Coccomyxa subellipsoidae C-169.*

This unicellular green alga (49) was obtained from the Van Etten laboratory (University of Nebraska-Lincoln, USA). Cultures were shaken at 60 rpm, for 14 days at 20°C, and grown on Modified Bold's basal Medium (MBBM), containing 10 µg/µL of tetracycline, under a cycle of 10 h of light and 14 h of dark in a New Brunswick Scientific Innova 4340 incubator.

Total mRNAs were extracted using the kit "RNA isolation from plant" from Macherey-Nagel. One µg of RNA was then reverse transcribed using either random hexamers or a polyT primer, and SuperScript™ II Reverse Transcriptase (Invitrogen). PCR reactions were performed using a subset of primers whose designs were based on the genomic sequence (49). Several PCR fragments were obtained, cloned into the pGEM-T Easy vector and sequenced (Eurofins Genomics, Ebersberg, Germany). Three fragments called A, B and C, and covering the whole

sequence, were used. First, fragment A was amplified by PCR using flanking primers bearing the *EcoRI* and *NheI* restriction sites at the 5' and 3' end, respectively. Secondly, fragments B and C were then fused by Splicing by Overlap Extension (SOEing), to obtain the fragment B-C. This was further amplified using flanking primers harboring *NheI* and *NotI* restriction sites, respectively, at the 5' and 3'end. The fragments A and B-C were then digested by EcoRI, NheI and NotI and sequentially ligated together, into the yeast expression vector pGAPZB, to build the full length cDNA of 2460 bp encoding the PLDα of this green alga. The sequence was deposited in Genbank under the reference MG807645.

Cloning of the different constructs of *PLDα* from *Arabidopsis thaliana*.

The 6-His tagged version of AtPLDα was constructed, by PCR, using the forward primer 222 encoding the 6-His tag and the reverse primer 97 (see Table S2 for primer sequences). The PCR product was then digested and cloned into the *SacII* and *NotI* restriction sites of the pGAPZB plasmid. This construction harbors, at the 5' extremity, an *NdeI* restriction site, CATATG, formed by the last codon CAT of the tag and the first ATG of the coding PLD. The double mutant A35H-N36M, of the 6-His tagged version, harbors a second *NdeI* restriction site, CATATG, where the two mutations take place. Therefore, we digested this construction using NdeI to remove the 36-residue long propeptide and ligated back together the remaining extremities to build the Δ36 truncated 6-His tagged version of AtPLDα.

Mutagenesis

AtPLD mutations were generated using the GeneArt® Site-Directed Mutagenesis PLUS system from Invitrogen with cDNA cloned into the pGAPZB as a template (20). Primers listed in Table 1 were used mostly to substitute each targeted residue with an alanyl residue. The *A. thaliana* PLDα cDNA (U36381) was a gift from Dr. X Wang (Donald Danforth Plant Science Center, St Louis, Missouri, USA) and was used as the template to mutate the wild-type version of the protein. The 6 His-tagged version of AtPLDα was used to mutate the N-terminal propeptide.

Random mutations isolation

Several runs of site-directed mutagenesis resulted in random point-mutations, that were further isolated from designed point-mutations. Mutation P208S, initially fused with H661A, was isolated using the two *MfeI* restriction sites located in both the promotor and in the *AtPLD* gene at the nucleotide position 1394. The *MfeI-MfeI* fragment of 1410 bp, bearing the P208S mutation, was then transferred into the wild-type construction, digested by MfeI and then dephosphorylated. The same operation was conducted to isolate the mutation A289V fused with Y659A. The correct orientation of the fragments was verified by PCR, with the appropriate primers, and the entire constructions were then sequenced to check their authenticity of the construct. Single mutations H661A and Y659A have been obtained during other runs of site-directed mutagenesis.

The mutation A85S, fused with H332A, was isolated using the two *HindIII* restriction sites located at positions 434 and 2206 of *AtPLD*. The *HindIII-HindIII* fragment of 1772 bp, bearing

the mutation H332A, was then transferred into the wild-type construct digested by the same enzyme and dephosphorylated. The *HindIII-HindIII* wild-type fragment was also transferred back into the construct containing the random mutation A85S. The same exchange has been conducted to isolate the mutation R65G fused with S348A.

Finally, P415T was separated from the A35L mutation, using the *XhoI-NotI* fragments of 1708 bp to remove P415T, and transferred into the wild-type construct digested under the same conditions. The A35L mutation was isolated, using an *XhoI-XhoI* fragment of 763 bp, and transferred into the 6-His tagged version of AtPLD.

Cloning of the AtPLD minimal domain

Deleted versions of AtPLD, named Δ70 and Δ164, were synthetized using primers that restrict the length to 2,223 or 1,941 bp by omitting at the 5' end 210 or 492 nucleotides, respectively, and these were amplified using the forward primers 244 and 246, respectively, and the reverse primer 97 (the primers are listed in the Table S2), using the Q5® High-Fidelity DNA Polymerase (New England Biolabs). For the constructions NΔ82, NΔ93, NΔ106, NΔ119 and NΔ141, *AtPLDα* was amplified using primers that restrict the length by 246, 279, 318, 357 and 423 nucleotides, respectively, using forward primers 405, 406, 407, 408 and 409 respectively and the reverse primer 97 using the Phusion High Fidelity DNA Polymerase (Thermo Scientific). All constructs harbor a *SacII* restriction site at the 5' end and a *NotI* restriction site at the 3' end. PCR products were purified using the Qiaquick PCR Purification Kit, and then incubated with the Taq DNA polymerase (Euromedex) and dNTP to build an A-tail at both extremities. The products were then ligated into the pGEM®-T Easy vector (Promega), transformed and selected

into JM109 *E. coli* competent cells according to the manufacturer's protocol. To obtain deleted versions of AtPLD named Δ0, Δ35, Δ50, Δ152, a small fragment of the 5' part of the gene was amplified with the Taq DNA polymerase (Euromedex) using forward primers 241, 242, 243, 245, respectively, and the reverse primer 102 for Δ0, Δ35, Δ50 and primer 387 for Δ152. PCR products were ligated into the pGEM®-T Easy vector, as previously described.

For the constructions Δ0, Δ35, Δ50, the plasmid was then transformed into the *dam-3 E. coli* strain GM48 to avoid plasmid methylation of *BclI* sites, and purified. Then, using *SacII* and *BclI* restriction sites, the 5' extremity of the Δ70 minimal domain cloned into the pGEM®-T Easy vector was replaced by the *SacII/BclI* insert of the constructions Δ0, Δ35 and Δ50. For the construction of AtPLD Δ152, using *SacII* and *XhoI* restriction sites, the 5' extremity of the Δ164 minimal domain, cloned into the pGEM®-T Easy vector, was replaced by the *SacII/XhoI* insert of the construction Δ152.

Finally, all constructions were sequenced using the T7 promotor and SP6 promotor primers found in the pGEM®-T Easy vector before being transferred into the pGAPZB *P. pastoris* expression vector, using *SacII* and *NotI* restrictions sites. The final PLD constructions were entirely sequence-verified using primers located on the *pGAP* promotor, the *pGAP* terminator and on the *AtPLD* sequence.

Protein purification

Proteins were extracted in a phosphate buffer containing 50 mM NaH_2PO_4 and 300 mM NaCl pH 8.0. First, the Protino® Ni-TED packed column (Macherey-Nagel) was equilibrated using 2

mL of the phosphate buffer described above. Soluble proteins (2.8 mg) were then loaded onto the column and it was washed with 4 mL of the same buffer. Proteins were eluted with the same buffer supplemented with 250 mM of imidazole and fractions of 250 µL were collected.

Quantitative PCR

Total RNAs were extracted from around 2.5 x 10^7 yeast cells obtained from a 48 h culture using a Monarch® Total RNA Miniprep kit (New England Biolabs), according to the manufacturer's instructions. Briefly, yeast cells were lysed using 1 mL of glass beads (425 - 600 µm), for 5 times 30 seconds, in a bead-beater. Genomic DNA was retained on a Monarch gDNA removal column and then total RNAs were retained on an RNA purification column. Following an on-column DNAse I treatment, for 15 min at room temperature, total RNAs were eluted with water. Each RNA sample (5000 ng) was used for retrotranscription, which was performed under standard conditions with Superscript II reverse transcriptase (Invitrogen) and random hexamer primers. Quantitative PCR was performed on a 5-fold diluted sample using iTaq™ Universal SYBR® Green Supermix (Bio-Rad). A fragment of *AtPLD* (100 bp) was amplified using the primers 5'-GCCTTGGCATGACATTCACT-3' and 5'-GAGCAAGCAGGGTGGTAAAG-3', and the reference gene *ARG4* (50) was also quantified in parallel using 5'-GTGGGCAGAAGGGAAGTTTG-3' and 5'-TTGGTCGTGGAATCTCTGGT-3'. Authenticity of the amplicons was confirmed by sequencing the PCR product. Relative quantification of the copy number was performed using the $2^{-\Delta\Delta Ct}$ method (51) with ΔCt = Ct of target (*AtPLD*) – Ct of reference (*ARG4*).

Statistical analysis

Data are expressed as means ± S.D. Statistical significance was determined by one-way ANOVA with a post-hoc test. Samples were considered to be significantly different for $P < 0.05$.

Results

A new structural model proposed for eukaryotic PLDs

So far, only the crystal structure of the prokaryotic PLD from *Streptomyces sp. PMF* has been solved (14). This structure shows a bilobed enzyme with each lobe bearing an HKD motif. There are no 3D-structures of eukaryotic PLDs, although a SAXS model shows a symmetric enzyme with an ovoid shape and loosely structured tops (52). Based on these observations and on previous results, we decided to build a new model for the structure of the catalytic core of plant PLDα.

PLDs mostly differ in their regulatory domains (C2, PH or PX domains for example) where no sequence homology is found. Four typical motifs called, I, II, III and IV, are well conserved among PLDs and are repeated twice in the primary sequence (1), forming a catalytic core. Consequently, we postulated that the global topology of AtPLDα would be conserved. To understand the topology of PLDs, we used the AtPLDα sequence deleted of its C2 domain. Thus, the catalytic core of the enzyme, starting at residue V151, is 660-residue long and is presumably composed of two lobes, like other HKD-type PLDs (14). Separating the catalytic core in equal halves is a possibility for distinguishing the two lobes. However, an individual lobe secondary structure prediction shows that a common β-strand can align itself between the beginning and the end of the N-terminal lobe. Hence, we decided to build a first lobe shortened by 23 residues and finishing at E457, giving a 307-residue long lobe, and the second lobe, starting at D458, is 353-residue long (Figure S1).

A sequence alignment of the two lobes of AtPLDα shows that the four motifs mostly match together and are quite well aligned, in addition to their similar biochemical properties. The PROMALS3D server, that predicts secondary structures, shows that each motif has the same secondary structure in both lobes. Therefore, we postulate that the two lobes have the same topology (Figure S1).

To describe the common topology found in both lobes we decided to omit the secondary structures of less than four residues and the secondary structures where the sequences do not match together. Thus, we predicted that each lobe would share a topology as follows: β-β-α-β-α-β-α-β-β-β-β-α-α-β (Figure S1).

In parallel, we built a 3D homology model of the AtPLDα catalytic core by using the Phyre[2] server in intensive mode (Figure 1). This server searches for a homologous structure and it used the PLD of *Streptomyces sp*. strain PMF (PDB id: 1F0I), which shares only 18% of sequence identity with AtPLDα. However, both AtPLDα and PLD from *Streptomyces* have the same catalytic function and their four typical motifs are aligned together, so we hypothesized that both proteins would share a similar structure: (i) the two lobes are facing each other i.e. a mirror image, and (ii) the protein is in a horse saddle shape with the catalytic site in the middle. A comparison between the predicted topology (Figure S1) and the topology built on the basis of homology (Figure 1) highlights two major differences. First, there were two supplementary α-helices in the N-terminal lobe, one was modelized at the place of a β-strand at the beginning of the sequence, and the other was non-existent in our prediction. Secondly, the C-terminal lobe was also quite different without the first two β-strands and with an α-helix replacing a β-strand at the end of the sequence. These differences are mostly due to our omission of the shorter secondary structures and those structures that did not match between each lobe.

Analysis of plant PLDα protein sequences

First, we used various public databases to gather a maximum of α-type plant PLDs, with the PLDα from the model organism *A. thaliana* as bait. All PLDs found needed to have the typical HKD catalytic motif duplicated in their primary structure. We were able to retrieve a complete α-type PLD sequence from >200 different organisms, including mosses, liverworts, Gymnosperms and Angiosperms (Figure 2). At least one PLDα sequence was obtained from most of the organisms we analyzed, and several fragments of PLDα were even found in the fern *Adiantum capillus-veneris* (TST39A01NGRL0007_D22, TST39A01NGRL0026_O22) demonstrating that this lipolytic enzyme is well conserved and distributed among all plant families. Phylogeny analysis shows that PLDα from diverse organisms are often reassembled inside their own families. For example, the PLDα from Brassicaceae are grouped together as are the PLDα from Solenaceae (Figure 2), suggesting that *PLDα* have been maintained in genomes during plant evolution and have evolved from a single ancestor.

Our initial work of collecting PLD sequences from green plants did not allow us to identify sequences in algae using the criteria that we proposed for identifying canonical PLDα. However, an unusual PLD from the microalga *Coccomyxa subellipsoidea* was detected. Strikingly, as described recently (53), this sequence did not harbor one of the two classical HKD motifs, the second one being substituted by an unorthodox HKS motif, and also some internal residues seemed to be missing. It is possible that the catalytic motif is different among Viridiplantae, and that PLD from algae might have a completely different structure compared to PLDα from Embryophytes. However, by examining carefully the recently published genomic sequence of

441 this microalga (49), the edited ORF around the catalytic motifs seems doubtful. We then decided
442 to clone and sequence the cDNA obtained by RT-PCR using primers designed according to the
443 genomic sequence uphill of the presumed ATG and downhill of the presumed stop codon. The
444 cloned sequence obtained showed undoubtedly that this PLD from *C. subellipsoidea* harbors two
445 HKD sequences, and that the previous sequence obtained by *in silico* splicing contained several
446 mistakes. A corrected sequence has been deposited in Genbank under the accession number
447 MG807645. Interestingly, using this microalgal PLD as bait, we were able to identify another
448 PLD of the α type in the diatom *Phaeodactylum tricornutum* (XM_002180282.1). It is of interest
449 because this microalga is a model for lipid engineering, and the PA-releasing PLD is at the
450 crossroads of both phospholipids and triglycerides metabolic pathways.

451 All of the 209 plant PLDα found (Table S1) were then aligned, and a new consensus sequence
452 has been built both to identify the conserved residues and to show its score of conservation
453 (Figure 3). This consensus sequence is 924-residue long, consisting of 807 positions filled with a
454 residue and the remaining 117 positions as gaps in most of the sequences. The PLDs retrieved
455 contain between 800 and 834 residues, with an average of 810 residues. Out of the 924 residues,
456 138 (15%) are strictly conserved among all 209 plant PLDs, 376 (41%) are more than 95%
457 conserved and 479 (52%) are more than 90% conserved. Although some residues, such as those
458 of both HKD motifs were already known to be conserved, this alignment highlights some new
459 regions that are conserved and are probably extremely important for PLD function.

460 Two motifs appear to be particularly well conserved (the following underlined residues are
461 conserved at 100%). First, the consensus P471-H479 (PREPWHDIH) (Figure 3) was identified,
462 as a conserved region between mammalian and yeast PLD (31, 54). Secondly, the consensus
463 G777-D786 (GSANINQRSM), also called the GSRS motif (or conserved region IV, CRIV) (33),

was described as being involved in a critical step of catalysis through modification of seryl residues by transphosphatidylation during the catalytic step of enzyme activation (33). Another consensus sequence F609-G619 (FIYIENQYFLG), that has been used to retrieve PLDα, was also known to be a IYIENQFF motif (or conserved region III, CRIII) (33). It is thought to facilitate the interaction between the choline head of PC and PLD *via* the aromatic residues of this sequence (55). In contrast, the conserved region Ib (CRIb) (33), around residues L314-D331 has undergone more variation during plant evolution, and only L314, G321 and W329 have been strictly conserved (Figure 3).

Moreover, it appears that some of these 138 strictly conserved residues in plant PLDα are also present in human PLDs as well. There are 7 residues only found in PLD2, 1 residue is only found in PLD1, 8 residues are closely similar, but not identical, in both enzymes, and 74 residues are found in both PLD1 and PLD2 (Figure S2). If these residues are conserved in plant and human PLDs, they must be crucial for the catalytic activity and/or the structure of the enzyme, which leads to the hypothesis that these residues have undergone a selective pressure during evolution.

Determining the minimal domain for plant PLDα activity

Although the function of the PLD C-terminal extremity has already been shown to be crucial for catalytic activity, both in plants (41) and mammals (40), the exact function of the N-terminal of the PLDα extremity remains obscure. Using SignalP (56) and the PLDα sequence from *A. thaliana*, no signal peptide was predicted but a cleavage site was identified between residues A35 and N36 with a weak score of 0.168. The overproduction of this protein expressed in *P.*

485 *pastoris*, then purified, led to the identification of N36 as the first residue of the purified protein
486 (20). We subsequently performed serial deletion of the 5' extremity of *AtPLDα* and checked for
487 PLD activity (Figure 4). The PLDαΔ0 was used as the wild-type (wt) representing the same
488 sequence as *PLDα* used in the mutagenesis study but cloned, for convenience, in different
489 restriction sites of pGAPZB. The different constructions were designed by sequential deletions
490 (Figure 4C) of the C2 domain, principally removing each β strand (Figure 4A).

491 With the exception of PLDαΔ70, which will be discussed later, none of the different
492 constructions led to measurable PLD activity, with detected absorbance in the range of the
493 background measured with cells transformed by the empty vector (control) (Figure 4B). PLD
494 activity was measurable only with PLDαΔ0. PLDαΔ35 corresponds to the form of the purified
495 protein found when PLDα is expressed in *P. pastoris* (20), except that a Met is added at its N-
496 terminal extremity. Adding this single residue led to an inactive enzyme. When we checked the
497 expression of the recombinant protein in these different constructs using an antibody raised in
498 rabbits against the 117 residues of the C2 domain, the recombinant protein could be detected
499 from around 90 kDa from the PLDαΔ0 through to the PLDαΔ106 constructs (from a decreasing
500 theoretical size of 91.7 kDa to 80.1 kDa, respectively) (Figure 4D), demonstrating that at least
501 the Δ35, Δ50, Δ82, Δ93, and Δ106 versions are inactive forms of the truncated PLD. Since the
502 antibody was not able to recognize the shorter versions Δ119 to Δ164, we can only speculate that
503 these versions are inactive as well, the C2 domain being drastically shortened or absent in these
504 constructs. The case of Δ70 was, at first, puzzling, with inconsistent results between activity and
505 protein detection depending on the clone tested. Hence, this version of the protein was thought to
506 be particularly labile. We consequently checked both enzyme activity and the presence of the
507 protein on 9 freshly obtained transformants. As shown in Figure 4E, the PLD activity was more

than three times the background of the control for 7 out of 9 transformants. The protein could be detected in clones n°2 to 9, with expression in clone n°1 just visible (Figure 4F), proving that the protein is present. The high variability of PLD activity ranging from 15 to 78% of the wt activity between these nine Δ70 transformants, was observed only for this construction, the usual standard deviation of the mutants of Figure 6 being a maximum of +/- 37% for mutant S348A.

Analysis of the cleavage site of plant PLDα

We have already demonstrated that the integrity of the enzyme is mandatory for catalytic PLD activity since the single deletion of the propeptide led to an inactive enzyme. Surprisingly, the propeptide appears to be cleaved when the PLD is analyzed after purification with hydrophobic/affinity chromatography (20). This cleavage might result from the action of a protease, but, to our knowledge, no such a proteolytic cleavage has been reported for PLDs. There is considerable evidence in the literature that the purified enzyme is truncated by around 35 residues at the N-terminus extremity (20, 22, 24–27, 57). Hypotheses range from the importance of the propeptide in folding of the protein and protecting it from degradation (57), to the idea that it could be a signal peptide (24), potentially addressing the protein to the vacuole (27), or a peptide anchoring the protein to the membrane (58) as in the palmitoylation of human PLD. Suggestions about its function, properties or cleavage are only speculation so far. We should note, however, that evidence of a potential cleavage by a protease is shown by an interaction, described in cardoon, between an aspartic proteinase called cardosin A and the C2 domain of PLDα, leading to cleavage of the C2 domain when the complex was disrupted (59), the binding motifs for the C2 domains were RGD/KGE sequences.

A similar cleavage was observed in the yeast *P. pastoris* when PLDα were recombinantly expressed (20, 27). This yeast contains the proteinase PEP4 (60), a vacuolar aspartic proteinase precursor similar to *S. cerevisiae* PEP4, with 68% of identity and 82% of similarity. Moreover, PEP4 is one of the best matches of Cardosin A homologs amongst yeasts. Cardosin A, PEP4 from *P. pastoris* and PEP4 from *S. cerevisiae* are three aspartic proteinases that bear a matching interacting motif: RGD for cardoon, KGD for *Saccharomyces* and EGK for *Pichia*. The *P. pastoris* strain SMD1168 was consequently used because of its mutation in the aminopeptidase encoded by the gene *pep4*. We used these strains SMD1168 and X33 to produce the *A. thaliana* PLDα fused with a 6xHis-tag at the N-terminus extremity. In both strains, the specific PLD activity in the crude extract was the same whether the protein had its tag or not. Moreover, it was not possible to increase the PLD activity in strain SMD1168 after a purification on a Ni-affinity chromatography (data not shown). Overall, it seems that the 6xHis-tag at the N-terminus extremity was cleaved and did not impair the activity of the protein, and that PEP4 was not responsible for the cleavage.

Consequently, we decided to mutate the residues on both sides of the cleavage site using the 6xHis-tag version of the protein, postulating that modifying the site would prevent its cleavage by a site-specific protease. The activities of single mutants A35L and N36A and the double mutant A35H N36M were then compared in the loading fraction on Ni-TED resin. The expression was checked using an antibody against either the 6xHis-tag or the PLDα-C2 domain (Figure 5). The PLD, fused to its tag, could be detected with each antibody in a crude extract (Figure 5A), indicating that the protein is expressed, and at least part of the protein still has its tag. The presence of the tag was confirmed with the purification of the tagged protein by Ni-affinity chromatography. The PLD was again detected with each antibody in the purified

fractions (Figure 5B), demonstrating that the tag is present in the N-terminal PLD extremity, and that, in these native conditions of purification, the tag is accessible to bind the resin and is not buried inside the protein.

As observed in the SMD1168 strain, in the X33 strain the PLD activity was similar in the wild-type enzyme and in the 6xHis-tag fused enzyme, contrary to a soluble crude extract from a clone transformed with the void plasmid taken as a control (Figure 5C). When the PLD activity was measured in the crude extract, the enzyme activity detected with constructions A35L, N36A and A35H N36M was in the same range as that of the control, indicating that the mutations A35L and N36A around the cleavage site and far from the catalytic region abolished the PLD activity. The result for the double mutant A35H N36M is not definitive as quite low activity (although above the background level) was consistently observed in different sets of experiments.

Unfortunately, the PLD activity in the elution fraction from the Ni-TED column could not be measured as the phosphate buffer used for purification was, strikingly, the only buffer tested that did not allow the wild-type enzyme (*id est* without a His-tag) both to bind to the Ni-TED column in a nonspecific way and to be eluted when imidazole is added. Moreover, the presence of phosphate in the enzymatic test induced a precipitation of the calcium ions added to activate the enzyme.

Finally, the results for the Δ36 truncated 6-His tagged version of the protein are not given as they are equivocal. PLD activity has never been detected in the crude extract of all the transformants tested, but we cannot definitively conclude whether the protein is produced, as multiple attempts to demonstrate the protein expression, by immunoblot, failed to provide reproducible data. This

suggests that the mutated protein is either very poorly expressed or is particularly labile and/or instable.

Mutation of conserved residues among plant PLDα around the HKD motifs

Our initial attempts to work with the PLDα from *A. thaliana* expressed in *E. coli* were unsuccessful. Although the protein was present in high amounts, as revealed by SDS-PAGE analysis in a whole cell extract, no detectable PLD activity could be measured under these conditions in the soluble fraction (data not shown). This result was quite puzzling since the PLDα from *Brassica oleracea* an organism close to *A. thaliana* (both PLDs having 94% identity) was expressed in *E. coli*, purified and found to be active (41). The PLDα from *R. communis* was only expressed as a functional enzyme in *E. coli*, when the 30-residue N-terminal leader sequence was present, the protein being inactive when this sequence was omitted (57). Using 1,2-dipalmitoyl-3-phosphatidyl-[methyl-^{3}H]choline as the substrate, the PLD activity (around 1.0 nmol/min/mg) of the purified wild-type protein has been measured in a different study (36), at eight times lower than the PLD activity detected in the crude extract of the protein expressed in *P. pastoris* (this work). It is possible that only a small amount of the protein folds properly in *E. coli*, and/or that the differences in substrate or in the sensitivity of the technique used to measure the PLD activity leads to this disparity. Moreover, *P. pastoris* seems to be a more interesting expression system because of the N-terminal PTM present in these eukaryotic cells, apparently similar to that found in plants. It is obvious that PLDs expressed in *E. coli* are not maturated: the 6-His tag at the N-terminal extremity permitted purification with nickel-affinity chromatography

596 (36), precluding an autocatalytic mechanism where the PLD cleaves itself. Interestingly, in our
597 experiments the non His-tagged version of AtPLDα was somehow able to bind to a nickel resin
598 in a nonspecific way (data not shown).

599 To verify our initial concept that *P. pastoris* was a suitable expression system for screening
600 mutations of AtPLDα, we first individually mutated the residues of the two HKD motifs (H332,
601 K334 and D339 for the first one; H661, K663, and D668 for the second one). Replacing each of
602 these six residues with an alanyl residue led to a total disappearance of activity, compared to the
603 wild-type version of the protein, the activity being comparable to the background measured in
604 cells transformed by the empty vector (Figure 6A). As has been previously shown for PLD from
605 other organisms, such as *Yersinia pestis* (35), cabbage (41), human (33), yeast SPO14 (33),
606 *Streptoverticillium cinnamoneum* (61), and *Streptomyces septatus* (62), these residues are also
607 involved in the AtPLD catalytic mechanism. Using Western-blot analysis and an antibody
608 directed against the C2 domain of AtPLDα, we showed that the recombinant protein was
609 expressed in a similar fashion as the wild-type protein (Figure 6B).

610 We next searched for other residues that we identified using bioinformatic tools. First, H332
611 thought to be the canonical catalytic residue of the first HKD motif is preceded by another
612 histidyl (H331) that is 100% conserved in all PLDα identified (see Figure 3). The mutation
613 H331A totally abolishes the activity, as does H332A (Figure 6). The mechanism currently
614 accepted for PLD proposes that H332 was responsible for the nucleophilic attack of the substrate
615 which led to a phospho-histidine intermediate (2, 63), but no role has been proposed yet for the
616 downstream residue H331, which also seems to be critical for catalysis. Furthermore, in the
617 second HKD motif, the catalytic aspartyl D668 is followed by another, namely D669. Once
618 again, this residue is strictly conserved among plant PLDα (Figure 3) and in human PLDs

(Figure S2) and its mutation led to an inactive enzyme (Figure 6A). Similarly, all mutations around the second HKD motif (Y659A, I666A, G675A, S676A, A677V, N678A, N680A, R682A, S683A, and M684V) abolished the PLD catalytic activity.

Because until now all designed mutations of conserved residues in this work have abolished the PLD catalytic activity, we decided to check the mutation of a non-conserved residue. The randomly obtained mutation A289V was chosen because this alanyl residue is found in 93% of the sequences identified in this work (Figure 3) and is sometimes replaced by a valyl residue as found in *Rosmarinus officinalis* (Table S1). The A289V mutation did not dramatically alter the activity of the recombinant protein (Figure 6A). Clones with this mutation showed more than 40% of the wild-type activity and a similar level of protein expression, as detected by Western-blot analysis (Figure 6B), demonstrating the potential retention of activity after mutation.

Interestingly, during the analysis of mutants, PLD activity could not be detected in 3 independent transformants for several mutations of strictly conserved residues: D669A, D689A, P771A, G778A and G798A (Figure 6A). Moreover, the recombinant protein was not detected in these transformants, at least for the single clone that was tested for expression on SDS-PAGE (Figure 6B). Based on our model (Figure 1), we hypothesized that the mutation of these crucial residues somehow destabilizes the 3D structure of the protein, exposing new regions and, possibly, leading to the proteolytic degradation of the protein. Because transformation of *P. pastoris* implies the random integration of the PLD transgene anywhere in the genomic DNA, and to rule out the transgene integration in a poor transcript region, we analyzed the expression of the transgene by qPCR in these particular clones using primers located at the 3' extremity of *AtPLDα* (Figure 7). Compared to the wild-type clone, the expression of the transgene was the same in almost all of these mutants (Figure 7A). However, the recombinant protein in G798A

could not be detected by Western-blot analysis, the protein was barely visible in D689A, the level of expression was quite different among the three clones tested for P771A and G778A, and the expression of D669A was much lower than the wild-type (Figure 7C). These results demonstrate that transcription was not an issue in these mutants, but either translation or folding were impaired, or that PTM or protein degradation occurred.

Mutation of the residues that are the same as human PLD

AtPLDα and both human PLD1 and 2 share the same catalytic motifs, as mentioned previously, but some other residues also share identity with these 3 enzymes (Figure S2). Hence, we investigated the relative importance of some of these conserved residues.

Based on the crystal structure of the PLD from *Streptomyces sp*. strain PMF (14), D202 and D473 were thought to stabilize H470 and H170, respectively, these residues being the histidyls of the two bacterial HKD motifs (64). A basic alignment of AtPLDα and the PLD from *Streptomyces sp*. strain PMF allowed the identification of D407 as the equivalent of D202 and this hypothesis was supported by our models of AtPLDα that clearly placed D407 at a reasonable distance from the catalytic histidine H661 (Figure 1). Identifying the equivalent of D473 is not possible using an alignment, as the C-terminal sequences do not align properly. However, using our models we can identify D689 or E691 as putative analogs to D473 of *Streptomyces sp*. strain PMF. E691 is probably the best candidate considering alignment of the two lobes of AtPLDα (Figure S1). These three residues D407, D689 and E691, are perfectly conserved among plant PLDα (Figure 3) and in human PLD 1 and 2 (Figure S2). When we tested the mutations D407A,

D689A and E691A, the recombinant protein lost its catalytic capacity (Figure 7A). Data for D689A are puzzling as the protein was barely expressed at the protein level compared to the background observed in the control (Figure 7C). Depending on the clone tested and its detected protein expression level, it is possible that this particular mutation destabilized the protein leading to its degradation. As an example, the first clone of mutations P771A and D669A with roughly the same RNA expression RNA level (Figure 7A) and definitely transformed with the transgene (Figure 7B), have quite different protein expression levels (Figure 6C).

Oppositely, mutation of a strictly conserved residue among eukaryotic PLDs as Q522A did not alter the ability of the enzyme to hydrolyze the substrate (Figure 6A). In the PREPWHDIH motif (Figure 3), mutations P401A, R402A, P404A, W405A abolished the PLD activity (Figure 6A). Hence, this sequence between the two HKD motifs, identified as a conserved region between mammalian and yeast PLD (31, 54), is important for plant PLD catalytic activity too. Mutation K261A also suppressed the activity, as well demonstrating that this residue, in common with human PLDs (Figure S2), is important as well.

These residues, found both in plants and in human PLD1 and 2, seem to be crucial for an active form of the enzyme.

Mutation of other conserved residues

At the N-terminal extremity and between the 2 HKD motifs, random mutations A85S, R65G, P208S, and P415T abolished PLD activity (Figure 6A). However, the nature of the replacing

684 residue may explain the loss of activity, for example prolyl residues are known to interrupt
685 secondary structures, and these mutations could extend strands and helixes.

686 In the hydrophobic motif FTVYVVV, mutation F565A abolished PLD activity (Figure 6A).
687 Finally, all the mutations of the PX_2(L/I)T(T/S) motif also suppressed PLD activity as mutations
688 P805A, T809A, T809S and T810A led to an inactive enzyme (Figure 6A). The mutations T809A
689 and T809S indicate that hydroxylation of the residue is not sufficient, and both the length and
690 hydroxylation of the lateral chain of the residue are also essential for a functional enzyme.

691

692

693 # Discussion

694 Interestingly, one of the most diverse regions in the sequence alignment of all plant PLDs is
695 situated at the N-terminal extremity between the 18^{th} and the 64^{th} residues. It is well known that
696 the PLDα purified so far have undergone a PTM described as a cleavage of thirty to forty
697 residues at the N-terminal extremity (20, 22, 24–27, 57). This PTM is found both when PLDs are
698 directly purified from plants and when PLDs are expressed in a recombinant system, such as *P.*
699 *pastoris*, which means that this region is probably recognized both by plant and yeast proteases.
700 It is also possible as well that during the evolution this non-conserved region has progressively
701 evolved in plants with the system responsible for the cleavage, or that this N-terminal region
702 behaves as a fragile protuberance accessible to proteases. Overall, this bioinformatics work
703 highlights new residues that seem to be of prime importance for PLD catalytic activity.

In addition to the fact that adding a single Met before A35, mutating either A35 or N36, or deleting several residues of the C2 domain all abolish the enzyme activity, this study reveals that the N-terminal extremity, its length and its residue composition are crucial for an active form of PLDα. However, the C2 domain does not have to be intact for the enzyme to be active, as β strands 1 to 3 (Figure 4A), unlike other strands of the C2 domain, are not critical for an active Δ70 enzyme. Together, the variable nature of the Δ70 constructs and the collection of inactive constructs suggest that the entirety of the C2 domain is essential.

This work also demonstrates that, (i) at least a fraction of the AtPLDα produced does not undergo a proteolytic cleavage as the tag fused to the protein could be still detected, (ii) adding a 6xHis-tag at the N-terminal extremity does not impair the activity of the AtPLDα as does the same tag but located at the C-terminal extremity (20), (iii) the cleavage site is crucial and has to be preserved for the production of a functional enzyme. It would be quite interesting now to determine the relative proportions of the cleaved and uncleaved forms of PLD, and to check whether the uncleaved form can be purified *via* a Ca^{2+}-mediated hydrophobic interaction chromatography on Octyl-Sepharose as performed for the cleaved form (20). It has been postulated that this interaction is mediated *via* the C2 domain, so a longer C2 domain fused to a 6xHis-tag may modify the PLD interaction with the resin.

It is still unclear which protein is responsible for this PTM. Cleavage of PLDα by itself can be excluded as PLDα expressed in *E. coli* retained its leader sequence (28) and the protein became inactive in this system when the leader sequence was removed by genetic engineering (57). It is puzzling that exactly the same cleavage site could be found when a PLD is expressed in organisms as varied as the yeast *P. pastoris* (27) or the insect *Spodoptera frugiperda* (26). Assuming that an endogenous protease is responsible for this cleavage, yeast would be a

convenient tool for determining which of the different proteases is involved, as several libraries of clones are available where these enzymes are systematically mutated. Yeast genome encodes a total of 121 proteases with a viable knock-out phenotype (65) that could be tested to identify the protease responsible for the cleavage.

It is striking that PLDα are found in all the land plants that we selected, but the origin of this enzyme in plants remains to be elucidated because green algae do not contain such a sequence. The exceptions are *C. subellipsoidea* and *P. tricornutum*, where a horizontal transfer from microbes is suspected. It is possible that other PLD isoforms could replace PLDα as many different isoforms are detected in algae (53). Since algae are now used for biomass and lipid production, and PLD is at the crossroads of triacylglycerol and phospholipids pathways *via* PA production, both identity and implication of the different PLDs need to be determined for effective engineering of these cell factories (66).

Using mutation results, the predicted topology based on primary sequence, the topology built by homology, and by retaining only the conserved sequences, we were able to propose a global structural topology (Figure 8) with two major features: (i) both lobes carry at least four α-helices and six β-strands as the principal secondary structures, (ii) in each lobe, the six β-strands are forming a β-sheet with three α-helices in support on one side (exterior of the protein) and a fourth α-helix on the other side (facing the other lobe). The structural model for AtPLDs could now be tested in order to provide functional insights into structural features and functional aspects of eukaryotic PLDs.

Mutations affecting the activity are mostly found in predicted secondary structures. For plant and mammalian PLD, the C2, PH and PX regulatory domains structure have already been solved

(67–69) but there is still a need to position the regulatory domains relative to the catalytic core. Structural studies, using a homogenously purified plant enzyme (45), are required to identify the position and the dynamics of the C2 domain in plant PLD.

Comparing plant and mammal PLDs is a powerful tool for identifying those residues that have undergone selective pressure during evolution, and, equally, those regions that have freely diverged. The C and N-terminal extremities are perfect examples: among plants the C2 domain is one of the regions with the highest variability and the C-terminal is better conserved. This C2 domain, together with PH and PX regulatory domains, is not restricted to PLDs and is often found in other types of enzymes (2). It has been postulated that the C-terminal extremity of human PLD interacts with the catalytic site (40) and the C-terminal extremity of plant PLD is buried inside the PLD structure (41). Previous attempts to modify this extremity have led to an inactive enzyme (38–41). As shown for mammalian and cabbage PLD (40, 41), and in this work for plant PLDα, the integrity of the C-terminal extremity of PLD is crucial for catalytic activity, suggesting an essential role in the mechanism or the structure, such as a potentially buried terminal peptide. It is possible that the overall structure of the enzyme is not governed by the N-terminal portion that allows the adsorption of the enzyme to the membranes, and, consequently, its substrate, but that the catalytic core may be structured by the C-terminal extremity which has been less affected by evolution. So far, crystal data have only been obtained from bacterial PLDs (14, 35, 70) that do not have regulatory domains, probably because mammalian (2) and plant PLDs (71) that are totally or partially membrane bound, respectively, are more difficult to purify for structural studies. The position of the regulatory domains towards the catalytic core, and their dynamics during membrane adsorption, remain to be elucidated.

771 Out of the 31 strictly conserved residues tested, mutations Q522A and S348A are remarkable
772 because they are the only ones that did not abolish the PLD catalytic activity. These results are
773 quite paradoxical. On the one hand the Q522 residue is always conserved in all PLDα (Figure 3)
774 but, on the other hand, the mutation did not alter the catalytic activity (Figure 6A). It is still
775 possible that this residue is important for other functions *in planta*, such as protein localization or
776 membrane targeting. It would be interesting to test this particular mutation in the plant *pldα1*
777 mutant to check if lipid composition is restored in organs where PLDα plays a key role (72), and
778 if this mutation still permits an adequate response to drought (73). Mutation S348A is also
779 remarkable, not only because this mutant is still active (Figure 6A), but because the S348 residue
780 located just after the first HKD motif is a mirror image of S676 in the second HKD motif defined
781 as $HXKX_4DX_6GSXN$ (15). To our knowledge, it is the first time that this particular residue (or
782 the equivalent in other PLDs) has been mutated and shown to be dispensable for activity.
783 However, mutation S676A resulted in an inactive enzyme (Figure 6A). Consequently, the model
784 proposed for the catalytic site of dimeric Nuc (15), where these two seryl residues hydrogen-
785 bond to the side chain of glutamyl residues has to be modified for the monomeric plant PLD
786 (52), where these two seryl residues are clearly not equivalent. We should mention that none of
787 the hydroxylated residues mutated in this study have ever been shown to be phosphorylated.

788 It is interesting to note that H333D and D340E mutations of PLD from cabbage (corresponding
789 to H332A and D339A in the first HKD motif of PLD from *A. thaliana*), showing no PLD
790 activity *in vitro*, could not be purified by the Ca^{2+}-mediated hydrophobic interaction
791 chromatography on Octyl-Sepharose (41) suggesting that point mutation may provoke a different
792 folding of the protein. It is possible that mutations D669A, D689A, P771A, G778A and G798A
793 dramatically changed the structure and/or the stability of the recombinant protein.

794 The F565 residue belongs to the EKF motif of PLDα1, a motif analogous to the DRY motif
795 present in many GPCRs and through which the enzyme could interact with the heterotrimeric G-
796 protein α-subunit (36). Interestingly, when the same mutant F565A was expressed in *E. coli* (36),
797 any difference in enzyme catalytic activity observed, compared to the wild-type, could probably
798 be explained by the difference in the expression system. *P. pastoris* is known to glycosylate
799 recombinant protein (74), which could inactive the enzyme for this particular mutation. This
800 EKF motif is, nonetheless, mostly conserved among all PLDα (Figure 3). Out of 209 sequences,
801 R564 is found in only 169 sequences having been replaced sometimes by K or P, but E563 and
802 F565 are found in 205 and 208 sequences, respectively, indicating that these two last residues
803 have undergone a severe selective pressure during evolution. F565 is replaced by L565 in the
804 single sequence of *Dichanthelium oligosanthes,* but F565 is found in all other PLDα isoforms as
805 OEL32086.1 or OEL13457.1 in this organism. A sequencing error in this particular sequence
806 cannot be excluded.

807 As far as the mechanisms and catalytic residues are concerned, the two HKD motifs seem not to
808 be equivalent. The first HKD motif has been described as $HXKX_4D$. However, another strictly
809 conserved histidyl residue, H331, is located just before the H332 of the first HKD. This residue
810 is strictly conserved among plant PLDα (Figure 3) and in human PLDs (Figure S2). Mutation
811 H331A led to an inactive enzyme, as did H332A (Figure 6), suggesting that both residues need to
812 be intact to allow a catalytic reaction. The relative role of these two identical residues in close
813 vicinity is not yet clear but, considering the remarkable occurrence of the duplicity of these
814 residues, we propose an extended version of the first HKD motif as $HHXKX_4D$. This duplicate
815 residue is absent in Nuc (15), in Zucchini PLD-like nuclease (75), and in other members of the
816 PLD superfamily such as cardiolipin synthase (2, 76). Interestingly, in bacteria, this duplicate

817 residue is found in alignments of the second motif that could be written as $HHKX_4D$ (14, 76),
818 suggesting that this particular histidyl may be linked to the nature of the PLD substrates.

819 The mutation of either D669A or D668A in the second HKD motif also abolished the PLD
820 activity (Figure 6A). As has been shown in *Y. pestis*, substitution of the conserved aspartic acid
821 in either HKD domain rendered the enzyme Ymt insoluble (35), demonstrating that this residue
822 is probably crucial for the folding or the stability of the protein. Moreover, this residue is not
823 buried inside the enzyme, but instead it is located on the outside of the enzyme in different PLD
824 structures (11, 15). It seems that the mutation of D669 drastically changed the protein as the
825 mutated enzyme either could not be detected in the crude extract (Figure 6B) or it was poorly
826 expressed compared to the wild-type (Figure 7C) depending on the revelation conditions used,
827 and, in particular, the duration of the chemiluminescence detection.

828 Preceding the second HKD motif, mutation Y659A also abolished the PLD activity. Inside this
829 second motif, mutations in I666, as in G675, S676, A677, N678, N680, R682, S683 and M684
830 forming the GSANXNXRSM sequence, abolished the PLD activity when mutated by an alanyl
831 (or by a valyl for A677 and M684). The second HKD motif, described in the literature as
832 $HXKX_4DX_6G(G/S)XN$, can now be proposed as an extended version, for PLDα at least, of this
833 catalytic motif as $YXHXKX_2IXDDX_6GSANXNXRSM$.

834 Overall, the evolution of an ancestral bacterial dimeric PLD, with an equivalent HKD opposite,
835 could have led to a monomeric PLD in eukaryotes where each HKD motif may have evolved
836 separately.

837

Conclusion

This study identified new key residues, conserved in both plant and human PLDs, which are required for enzyme activity. We were also able to define a new consensus sequence around the two HKD motifs that could be extended beyond these sole residues. The length and residue composition of the PLDα N-terminal extremity were shown to be important, presumably for the folding of the C2 domain of plant PLDs. We also confirmed that the identity of residues at the PLDα C-terminal extremity is crucial, as has been shown previously for mammal PLDs. Overall, these data support the use of plant PLDs as a convenient tool for the study of all HKD-type PLDs, and as a means of elucidating the structure of eukaryotic PLDs.

Author Information

Corresponding Author

*To whom correspondence should be addressed

Tel: +33 4 27 46 57 31

E-mail address: alexandre.noiriel@univ-lyon1.fr

Author Contributions

The manuscript was written from contributions by all the authors. All the authors have given approval to the final version of the manuscript. The authors declare that they have no conflicts of interest with the contents of this article.

Acknowledgments

We would like to thank especially Pavlina Dubois for the cloning of the algal PLD, Pierre S. Garcia for protein alignments, and Mialitiana Solo Nomenjanahary, Karim Bessaa and Houda Abla for their technical help during the mutagenesis steps. We are grateful to Guy Condemine (Université de Lyon) for providing us with the *E. coli* dam- strain, and to James L. Van Etten (University of Nebraska-Lincoln) for providing us with a *Coccomyxa subellipsoidea* C-169 culture. We are indebted to Petar Puijic for his help in disrupting the numerous samples of yeast. Dr Mickael V. Cherrier is gratefully acknowledged for his advice on structural modeling. The English revision by Valerie James is also appreciated. YA was supported by an MENRT grant from the French Ministry of Science and Education. This work was supported by a grant to AN from Ligue contre le cancer, comité de Haute-Savoie.

Abbreviations

PA: phosphatidic acid

PLD: phospholipase D

PTM: post-translational modification

3D: tridimensional

TBS: Tris Buffer Saline

Figures

Figure 1: Modelization of AtPLDα.

The AtPLDα catalytic core model was built by homology using the Phyre[2] v2.0 server in intensive mode and was displayed using Pymol v2.0.4. The α-helices are represented in red and β-strands in yellow, random coils are not displayed. HKD residues are drawn as sticks. N: N-terminus extremity, C: C-terminus extremity.

Figure 2: Phylogenetic relationship and diversity of plant PLDα.

The main plant families are highlighted in color. The PLDα of *Arabidopsis thaliana* is indicated with a black arrow

Figure 3: Residue conservation in PLDα.

The histogram of the consensus sequence indicates the score of conservation of each residue in the alignment of 209 plant PLDα. The Y axis represents the percentage conservation of each residue. Mutated positions in *Arabidopsis thaliana* are indicated with red stars and mutations are shown just below.

Figure 4: Expression of AtPLDα minimal domains.

A) Schematic representation of AtPLDα C2 domain. Residues highlighted in yellow represent Ca^{2+} binding residues, red helices represent α-helices and green arrows are β-strands. Positions of minimal domains are indicated by a black vertical line in the peptide sequence. The blue arrow represents the cleavage position found in recombinant protein after purification. B) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (Δ0), the empty vector (Ct), or the truncated versions of the enzyme. Activities were calculated relative to the Δ0 value. Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to Δ0). The asterisk indicates that the data are represented in part E. C) Validation that the *AtPLDα* transgene was inserted in the genomic DNA of recombinant clones. After DNA extraction, a PCR was run with primers located on the promotor and on the cDNA 790 bp downstream of the end of the C2 domain coding sequence; note the absence of amplicon in the control (Ct). D) Western-blot analysis of the expression of AtPLDα minimal domains. The blotted bands were immmunodetected with a specific AtPLDα C2-domain antibody raised in rabbit, and subsequently visualized using a peroxidase labelled goat anti-rabbit IgG antibody using enhanced chemiluminescence. E) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (Δ0), the empty vector (Ct), or the Δ70 version of the enzyme. Activities were calculated relative to the Δ0 value (100%). Results shown are on n = 3 (Δ0), n = 4 (Ct) or n = 9 (Δ70) independent clones. The horizontal bar represents the mean values. F) Western-blot analysis of the expression of AtPLDα minimal domains. The blotted bands were immmunodetected as mentioned above.

Figure 5: Western-blot analysis of AtPLDα constructs after purification by affinity-chromatography.

The expression was detected in crude protein extracts by immunorevelation of *P. pastoris* clones transformed either with the empty vector (Ct) or expressing the wild-type enzyme (wt), the wild-type enzyme fused with a 6xHis-tag (wt 6xHis) or the different mutants fused with a 6xHis-tag. The blotted bands were immmunodetected with either a specific AtPLDα C2-domain antibody (anti-PLD) or an anti-tag 6xHis peroxidase labelled antibody. A) Crude extract fractions. B) Elution fractions after Ni-TED purification. C) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (wt), the wild-type protein fused with a 6xHis tag (wt 6xHis), the empty vector (Ct), or mutants of the cleavage site. Activities were calculated relative to the wt value (100%). Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to wt 6xHis except A35H N36M 6xHis mutant that is not significant (NS)). The arrow indicates the PLD position at 92.8 kDa.

Figure 6: Relative enzymatic activity and expression of forty-one AtPLDα point mutants.

A) Specific activity was measured with 30 µg of protein from a crude extract of *P. pastoris* expressing either the wt protein (wt) of a molecular mass of 91,7 kDa, the mutated enzymes or transformed with the empty vector (Ct). Activities were then calculated relative to the wt value (100%). Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to wt, except S348A and Q522A mutants).

B) Western-blot analysis of the expression of recombinant AtPLDα mutants. One out of the three clones mentioned above was randomly chosen for SDS-PAGE, the blotted bands were then immmunodetected with a specific AtPLDα C2-domain antibody.

Figure 7: Expression pattern of *AtPLDα* in different mutants.

A) Relative levels of mRNA from three independent clones expressing either the wt protein (wt), the empty vector (Ct), or the mutated enzyme were determined by a real-time polymerase chain reaction (RT-PCR) using the $2^{-\Delta\Delta Ct}$ quantification method and *ARG4* mRNA as an internal reference. Values are the mean ± SD obtained from duplicate samples. B) Validation that the *AtPLDα* transgene was inserted in the genomic DNA of recombinant clones. After DNA extraction, a PCR was run with primers located on the promotor and on the cDNA; note the absence of amplicon in the control (Ct). C) Western-blot analysis of the expression of AtPLDα mutants. The blotted bands were immmunodetected with a specific AtPLDα C2-domain antibody.

Figure 8: Schematic topology of the AtPLDα catalytic core.

Cylinders represent the α-helices and arrows represent the β-strands, N: N-terminus extremity, C: C-terminus extremity. Mutations obtained in this study are displayed as stars. Red and blue stars indicate the mutations that affect or do not affect the catalytic activity of the enzyme respectively. HKD catalytic residues are in underlined bold letters.

Supporting Information

Figure S1: Alignment of lobe 1 (L1) and lobe 2 (L2) of the catalytic core of AtPLDα.

The four motifs I, II, III and IV, adapted from [1], are highlighted respectively in yellow, green, red and blue. Prediction of consensus secondary structures (ss) was made using the PROMALS3D server; α-helices (h) are indicated in purple and β-strands (e) in grey. Secondary structures of less than four residues and secondary structures where the sequences do not match are omitted. HKD motifs are in red bold letters and underlined, mutated residues of this work are indicated in red bold letters.

Figure S2: Alignment of *Homo sapiens* PLD1 and 2 and *Arabidopsis thaliana* PLDα.

Residues highlighted in light grey are identical between HsPLD1, HsPLD2 and AtPLDα. Residues highlighted in dark grey share similar properties between at least three of the four sequences (HsPLD1, HsPLD2 and AtPLDα + plant consensus sequence presented in Figure 2) and are conserved at a minimum of 95 %. Residues highlighted in black share similar properties between at least three of the four sequences (HsPLD1, HsPLD2 and AtPLDα + plant consensus sequence in Figure 2) and are conserved at 100 %. Red stars indicate the mutated residues in AtPLDα.

References

1. Ponting, C. P., and Kerr, I. D. (1996) A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues. *Protein Sci. Publ. Protein Soc.* **5**, 914–922
2. Selvy, P. E., Lavieri, R. R., Lindsley, C. W., and Brown, H. A. (2011) Phospholipase D: enzymology, functionality, and chemical modulation. *Chem. Rev.* **111**, 6064–6119
3. Hanahan, D. J., and Chaikoff, I. L. (1947) A new phospholipide-splitting enzyme specific for the ester linkage between the nitrogenous base and the phosphoric acid grouping. *J. Biol. Chem.* **169**, 699–705
4. Nelson, R. K., and Frohman, M. A. (2015) Physiological and pathophysiological roles for phospholipase D. *J. Lipid Res.* **56**, 2229–2237
5. Bruntz, R. C., Lindsley, C. W., and Brown, H. A. (2014) Phospholipase D signaling pathways and phosphatidic acid as therapeutic targets in cancer. *Pharmacol. Rev.* **66**, 1033–1079
6. Roth, E., and Frohman, M. A. (2018) Proliferative and metastatic roles for Phospholipase D in mouse models of cancer. *Adv. Biol. Regul.* **67**, 134–140
7. Sun, Y., and Chen, J. (2008) mTOR signaling: PLD takes center stage. *Cell Cycle*. **7**, 3118–23
8. Sun, Y., Fang, Y., Yoon, M. S., Zhang, C., Roccio, M., Zwartkruis, F. J., Armstrong, M., Brown, H. A., and Chen, J. (2008) Phospholipase D1 is an effector of Rheb in the mTOR pathway. *Proc Natl Acad Sci U A.* **105**, 8286–91
9. Stieglitz, K. (2017) Structural Insights for Drugs Developed for Phospholipase D Enzymes. *Curr. Drug Discov. Technol.* 10.2174/1570163814666170816112135
10. Imamura, S., and Horiuti, Y. (1979) Purification of Streptomyces chromofuscus phospholipase D by hydrophobic affinity chromatography on palmitoyl cellulose. *J. Biochem. (Tokyo)*. **85**, 79–95
11. Magalhães, G. S., Caporrino, M. C., Della-Casa, M. S., Kimura, L. F., Prezotto-Neto, J. P., Fukuda, D. A., Portes-Junior, J. A., Neves-Ferreira, A. G. C., Santoro, M. L., and Barbaro, K. C. (2013) Cloning, expression and characterization of a phospholipase D from Loxosceles gaucho venom gland. *Biochimie*. **95**, 1773–1783
12. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203
13. Rudge, S. A., and Engebrecht, J. (1999) Regulation and function of PLDs in yeast. *Biochim. Biophys. Acta*. **1439**, 167–174
14. Leiros, I., Secundo, F., Zambonelli, C., Servi, S., and Hough, E. (2000) The first crystal structure of a phospholipase D. *Struct. Lond. Engl. 1993*. **8**, 655–667
15. Li, M., Hong, Y., and Wang, X. (2009) Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochim. Biophys. Acta*. **1791**, 927–935
16. Liu, Q., Zhang, C., Yang, Y., and Hu, X. (2010) Genome-wide and molecular evolution analyses of the phospholipase D gene family in Poplar and Grape. *BMC Plant Biol.* **10**, 117

17. Qin, C., and Wang, X. (2002) The Arabidopsis Phospholipase D Family. Characterization of a Calcium-Independent and Phosphatidylcholine-Selective PLDζ1 with Distinct Regulatory Domains. *Plant Physiol.* **128**, 1057–1068
18. Qin, W., Pappan, K., and Wang, X. (1997) Molecular heterogeneity of phospholipase D (PLD). Cloning of PLDgamma and regulation of plant PLDgamma, -beta, and -alpha by polyphosphoinositides and calcium. *J. Biol. Chem.* **272**, 28267–28273
19. Pappan, K., and Wang, X. (1999) Plant phospholipase Dalpha is an acidic phospholipase active at near-physiological Ca(2+) concentrations. *Arch. Biochem. Biophys.* **368**, 347–353
20. Rahier, R., Noiriel, A., and Abousalham, A. (2016) Functional Characterization of the N-Terminal C2 Domain from Arabidopsis thaliana Phospholipase Dα and Dβ. *BioMed Res. Int.* **2016**, 2721719
21. Tiwari, K., and Paliyath, G. (2011) Cloning, expression and functional characterization of the C2 domain from tomato phospholipase Dα. *Plant Physiol. Biochem. PPB.* **49**, 18–32
22. Abousalham, A., Riviere, M., Teissere, M., and Verger, R. (1993) Improved purification and biochemical characterization of phospholipase D from cabbage. *Biochim Biophys Acta*. **1158**, 1–7
23. Wang, X., Dyer, J. H., and Zheng, L. (1993) Purification and immunological analysis of phospholipase D from castor bean endosperm. *Arch. Biochem. Biophys.* **306**, 486–494
24. Ueki, J., Morioka, S., Komari, T., and Kumashiro, T. (1995) Purification and characterization of phospholipase D (PLD) from rice (Oryza sativa L.) and cloning of cDNA for PLD from rice and maize (Zea mays L.). *Plant Cell Physiol.* **36**, 903–914
25. Abousalham, A., Teissere, M., Gardies, A. M., Verger, R., and Noat, G. (1995) Phospholipase D from soybean (Glycine max L.) suspension-cultured cells: purification, structural and enzymatic properties. *Plant Cell Physiol*. **36**, 989–96
26. El Maarouf, H., Carriere, F., Riviere, M., and Abousalham, A. (2000) Functional expression in insect cells, one-step purification and characterization of a recombinant phospholipase D from cowpea (Vigna unguiculata L. Walp). *Protein Eng*. **13**, 811–7
27. Ben Ali, Y., Carriere, F., and Abousalham, A. (2007) High-level constitutive expression in Pichia pastoris and one-step purification of phospholipase D from cowpea (Vigna unguiculata L. Walp). *Protein Expr Purif*. **51**, 162–9
28. Schäffner, I., Rücknagel, K.-P., Mansfeld, J., and Ulbrich-Hofmann, R. (2002) Genomic structure, cloning and expression of two phospholipase D isoenzymes from white cabbage. *Eur. J. Lipid Sci. Technol.* **104**, 79–87
29. Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* **153**, 1161–1174
30. Rahier, R., Noiriel, A., and Abousalham, A. (2016) Development of a Direct and Continuous Phospholipase D Assay Based on the Chelation-Enhanced Fluorescence Property of 8-Hydroxyquinoline. *Anal Chem*. **88**, 666–74
31. Hammond, S. M., Altshuller, Y. M., Sung, T. C., Rudge, S. A., Rose, K., Engebrecht, J., Morris, A. J., and Frohman, M. A. (1995) Human ADP-ribosylation factor-activated phosphatidylcholine-specific phospholipase D defines a new and highly conserved gene family. *J. Biol. Chem.* **270**, 29640–29643
32. Xie, Z., Ho, W. T., and Exton, J. H. (1998) Association of N- and C-terminal domains of phospholipase D is required for catalytic activity. *J. Biol. Chem.* **273**, 34679–34682

33. Sung, T. C., Roper, R. L., Zhang, Y., Rudge, S. A., Temel, R., Hammond, S. M., Morris, A. J., Moss, B., Engebrecht, J., and Frohman, M. A. (1997) Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *EMBO J.* **16**, 4519–4530
34. Stuckey, J. A., and Dixon, J. E. (1999) Crystal structure of a phospholipase D family member. *Nat. Struct. Biol.* **6**, 278–284
35. Rudolph, A. E., Stuckey, J. A., Zhao, Y., Matthews, H. R., Patton, W. A., Moss, J., and Dixon, J. E. (1999) Expression, characterization, and mutagenesis of the Yersinia pestis murine toxin, a phospholipase D superfamily member. *J. Biol. Chem.* **274**, 11824–11831
36. Zhao, J., and Wang, X. (2004) Arabidopsis phospholipase Dalpha1 interacts with the heterotrimeric G-protein alpha-subunit through a motif analogous to the DRY motif in G-protein-coupled receptors. *J. Biol. Chem.* **279**, 1794–1800
37. Gookin, T. E., and Assmann, S. M. (2014) Significant reduction of BiFC non-specific assembly facilitates in planta assessment of heterotrimeric G-protein interactors. *Plant J. Cell Mol. Biol.* **80**, 553–567
38. Sung, T. C., Zhang, Y., Morris, A. J., and Frohman, M. A. (1999) Structural analysis of human phospholipase D1. *J. Biol. Chem.* **274**, 3659–3666
39. Sung, T. C., Altshuller, Y. M., Morris, A. J., and Frohman, M. A. (1999) Molecular analysis of mammalian phospholipase D2. *J. Biol. Chem.* **274**, 494–502
40. Liu, M. Y., Gutowski, S., and Sternweis, P. C. (2001) The C terminus of mammalian phospholipase D is required for catalytic activity. *J. Biol. Chem.* **276**, 5556–5562
41. Lerchner, A., Mansfeld, J., Kuppe, K., and Ulbrich-Hofmann, R. (2006) Probing conserved amino acids in phospholipase D (Brassica oleracea var. capitata) for their importance in hydrolysis and transphosphatidylation activity. *Protein Eng. Des. Sel. PEDS*. **19**, 443–452
42. Frohman, M. A. (2015) The phospholipase D superfamily as therapeutic targets. *Trends Pharmacol. Sci.* **36**, 137–144
43. Gomez-Cambronero, J. (2014) Phospholipase D in cell signaling: from a myriad of cell functions to cancer growth and metastasis. *J. Biol. Chem.* **289**, 22557–22566
44. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. **72**, 248–54
45. Arhab, Y., Rahier, R., Noiriel, A., Cherrier, M. V., and Abousalham, A. (2018) Expression and Purification of Recombinant Vigna unguiculata Phospholipase D in Pichia pastoris for Structural Studies. *Methods Mol. Biol. Clifton NJ*. **1835**, 191–201
46. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**, 680–5
47. Dyer JH, Zheng L, and Wang X (1995) Cloning and nucleotide sequence of a cDNA encoding phospholipase D from Arabidopsis (Accession No. U36381) (PGR 95-096). *Plant Physiol.*
48. Dong, D., Lei, M., Hua, P., Pan, Y.-H., Mu, S., Zheng, G., Pang, E., Lin, K., and Zhang, S. (2017) The Genomes of Two Bat Species with Long Constant Frequency Echolocation Calls. *Mol. Biol. Evol.* **34**, 20–34
49. Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S., Pangilinan, J., Pröschold, T., Salamov, A., Schmutz, J., Weeks, D., Yamada, T., Lomsadze, A., Borodovsky, M., Claverie, J.-M.,

Grigoriev, I. V., and Van Etten, J. L. (2012) The genome of the polar eukaryotic microalga Coccomyxa subellipsoidea reveals traits of cold adaptation. *Genome Biol.* **13**, R39
50. Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F. S., and Glieder, A. (2010) Real-time PCR-based determination of gene copy numbers in Pichia pastoris. *Biotechnol. J.* **5**, 413–420
51. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. **25**, 402–408
52. Stumpe, S., König, S., and Ulbrich-Hofmann, R. (2007) Insights into the structure of plant α-type phospholipase D: Structure of plant phospholipase D. *FEBS J.* **274**, 2630–2640
53. Beligni, M. V., Bagnato, C., Prados, M. B., Bondino, H., Laxalt, A. M., Munnik, T., and Ten Have, A. (2015) The diversity of algal phospholipase D homologs revealed by biocomputational analysis. *J. Phycol.* **51**, 943–962
54. Morris, A. J., Engebrecht, J., and Frohman, M. A. (1996) Structure and regulation of phospholipase D. *Trends Pharmacol. Sci.* **17**, 182–185
55. Frohman, M. A., Sung, T. C., and Morris, A. J. (1999) Mammalian phospholipase D structure and regulation. *Biochim. Biophys. Acta*. **1439**, 175–186
56. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*. **8**, 785–786
57. Wang, X., Xu, L., and Zheng, L. (1994) Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from Ricinus communis L. *J. Biol. Chem.* **269**, 20312–20317
58. Sato, H., Watanabe, T., Sagane, Y., Nakazawa, Y., and Takano, K. (2000) Purification and Characterization of Phospholipase D from Cabbage Leaves. *Food Sci. Technol. Res.* **6**, 29–33
59. Simões, I., Mueller, E.-C., Otto, A., Bur, D., Cheung, A. Y., Faro, C., and Pires, E. (2005) Molecular analysis of the interaction between cardosin A and phospholipase D(alpha). Identification of RGD/KGE sequences as binding motifs for C2 domains. *FEBS J.* **272**, 5786–5798
60. Gleeson, M. A., White, C. E., Meininger, D. P., and Komives, E. A. (1998) Generation of protease-deficient strains and their use in heterologous protein expression. *Methods Mol. Biol.* **103**, 81–94
61. Ogino, C., Daido, H., Ohmura, Y., Takada, N., Itou, Y., Kondo, A., Fukuda, H., and Shimizu, N. (2007) Remarkable enhancement in PLD activity from Streptoverticillium cinnamoneum by substituting serine residue into the GG/GS motif. *Biochim. Biophys. Acta*. **1774**, 671–678
62. Uesugi, Y., Mori, K., Arima, J., Iwabuchi, M., and Hatanaka, T. (2005) Recognition of phospholipids in Streptomyces phospholipase D. *J. Biol. Chem.* **280**, 26143–26151
63. Gottlin, E. B., Rudolph, A. E., Zhao, Y., Matthews, H. R., and Dixon, J. E. (1998) Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9202–9207
64. Uesugi, Y., and Hatanaka, T. (2009) Phospholipase D mechanism using Streptomyces PLD. *Biochim. Biophys. Acta*. **1791**, 962–969
65. Spedale, G., Mischerikow, N., Heck, A. J. R., Timmers, H. T. M., and Pijnappel, W. W. M. P. (2010) Identification of Pep4p as the protease responsible for formation of the SAGA-related SLIK protein complex. *J. Biol. Chem.* **285**, 22793–22799

66. Lenka, S. K., Carbonaro, N., Park, R., Miller, S. M., Thorpe, I., and Li, Y. (2016) Current advances in molecular, biochemical, and computational modeling analysis of microalgal triacylglycerol biosynthesis. *Biotechnol. Adv.* **34**, 1046–1063
67. Perisic, O., Fong, S., Lynch, D. E., Bycroft, M., and Williams, R. L. (1998) Crystal structure of a calcium-phospholipid binding domain from cytosolic phospholipase A2. *J. Biol. Chem.* **273**, 1596–1604
68. Zheng, J., Chen, R. H., Corblan-Garcia, S., Cahill, S. M., Bar-Sagi, D., and Cowburn, D. (1997) The solution structure of the pleckstrin homology domain of human SOS1. A possible structural role for the sequential association of diffuse B cell lymphoma and pleckstrin homology domains. *J. Biol. Chem.* **272**, 30340–30344
69. Stahelin, R. V., Karathanassis, D., Murray, D., Williams, R. L., and Cho, W. (2007) Structural and membrane binding analysis of the Phox homology domain of Bem1p: basis of phosphatidylinositol 4-phosphate specificity. *J. Biol. Chem.* **282**, 25737–25747
70. Suzuki, A., Kakuno, K., Iwasaki, Y., Yamane, T., and Yamane, T. (1999) Crystallization and preliminary X-ray diffraction studies of phospholipase D from Streptomyces antibioticus. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 317–319
71. Novák, D., Vadovič, P., Ovečka, M., Šamajová, O., Komis, G., Colcombet, J., and Šamaj, J. (2018) Gene Expression Pattern and Protein Localization of Arabidopsis Phospholipase D Alpha 1 Revealed by Advanced Light-Sheet and Super-Resolution Microscopy. *Front. Plant Sci.* **9**, 371
72. Devaiah, S. P., Roth, M. R., Baughman, E., Li, M., Tamura, P., Jeannotte, R., Welti, R., and Wang, X. (2006) Quantitative profiling of polar glycerolipid species from organs of wild-type Arabidopsis and a phospholipase Dalpha1 knockout mutant. *Phytochemistry*. **67**, 1907–1924
73. Hong, Y., Zheng, S., and Wang, X. (2008) Dual functions of phospholipase Dalpha1 in plant response to drought. *Mol. Plant.* **1**, 262–269
74. Bretthauer, R. K., and Castellino, F. J. (1999) Glycosylation of Pichia pastoris-derived proteins. *Biotechnol. Appl. Biochem.* **30 (Pt 3)**, 193–200
75. Voigt, F., Reuter, M., Kasaruho, A., Schulz, E. C., Pillai, R. S., and Barabas, O. (2012) Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. *RNA N. Y. N.* **18**, 2128–2134
76. Waite, M. (1999) The PLD superfamily: insights into catalysis. *Biochim. Biophys. Acta*. **1439**, 187–197

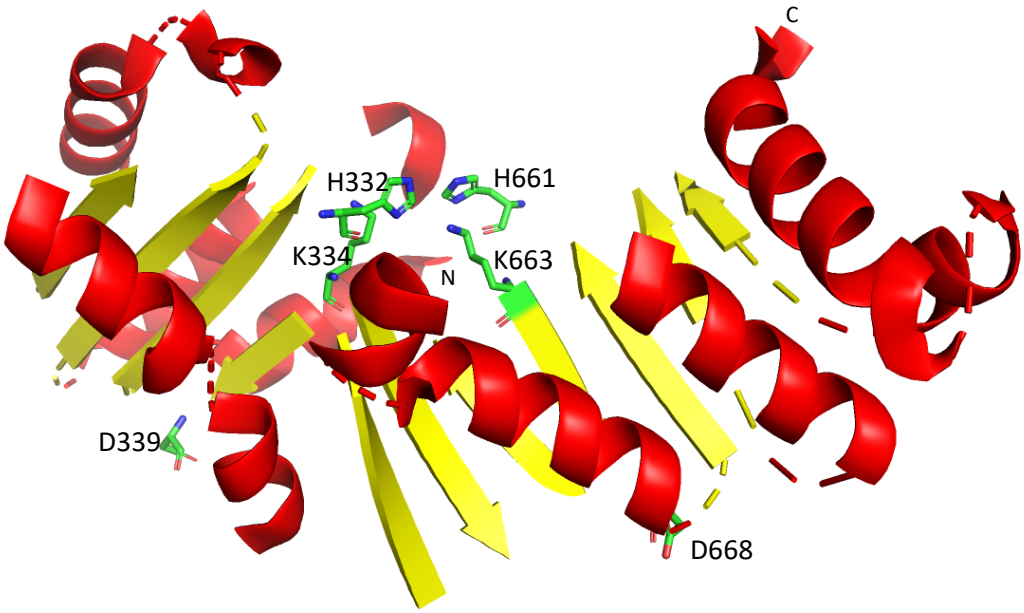

Figure 1 : Modelization of AtPLDα.
The AtPLDα catalytic core model was built by homology using the Phyre[2] v2.0 server in intensive mode and was displayed using Pymol v2.0.4. The α-helices are represented in red and β-strands in yellow, random coils are not displayed. HKD residues are drawn as sticks. N: N-terminus extremity, C: C-terminus extremity.

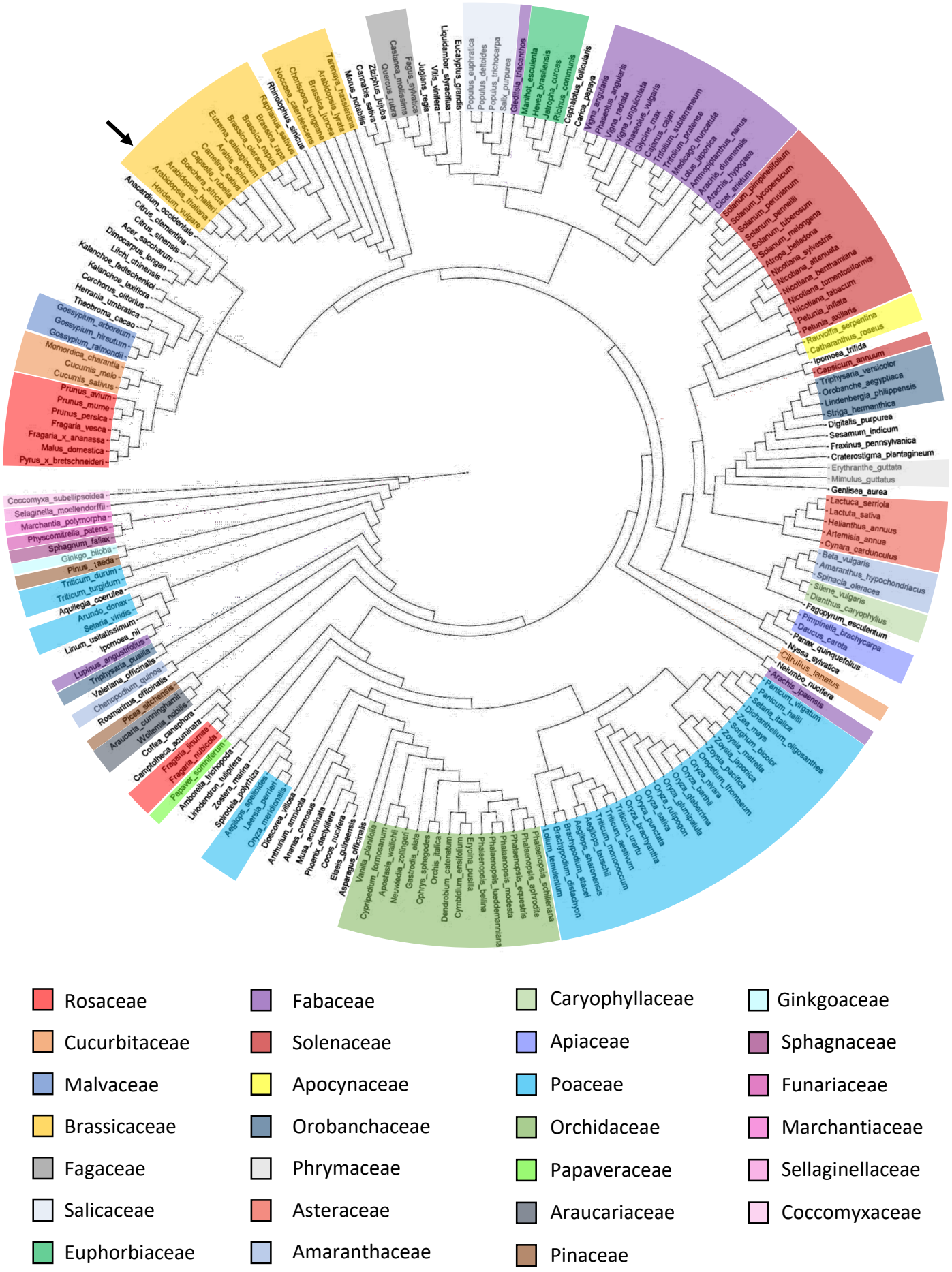

Figure 2 : Phylogenetic relationship and diversity of plant PLDα.
The main plant families are highlighted in color. The PLDα of *Arabidopsis thaliana* is indicated with a black arrow.

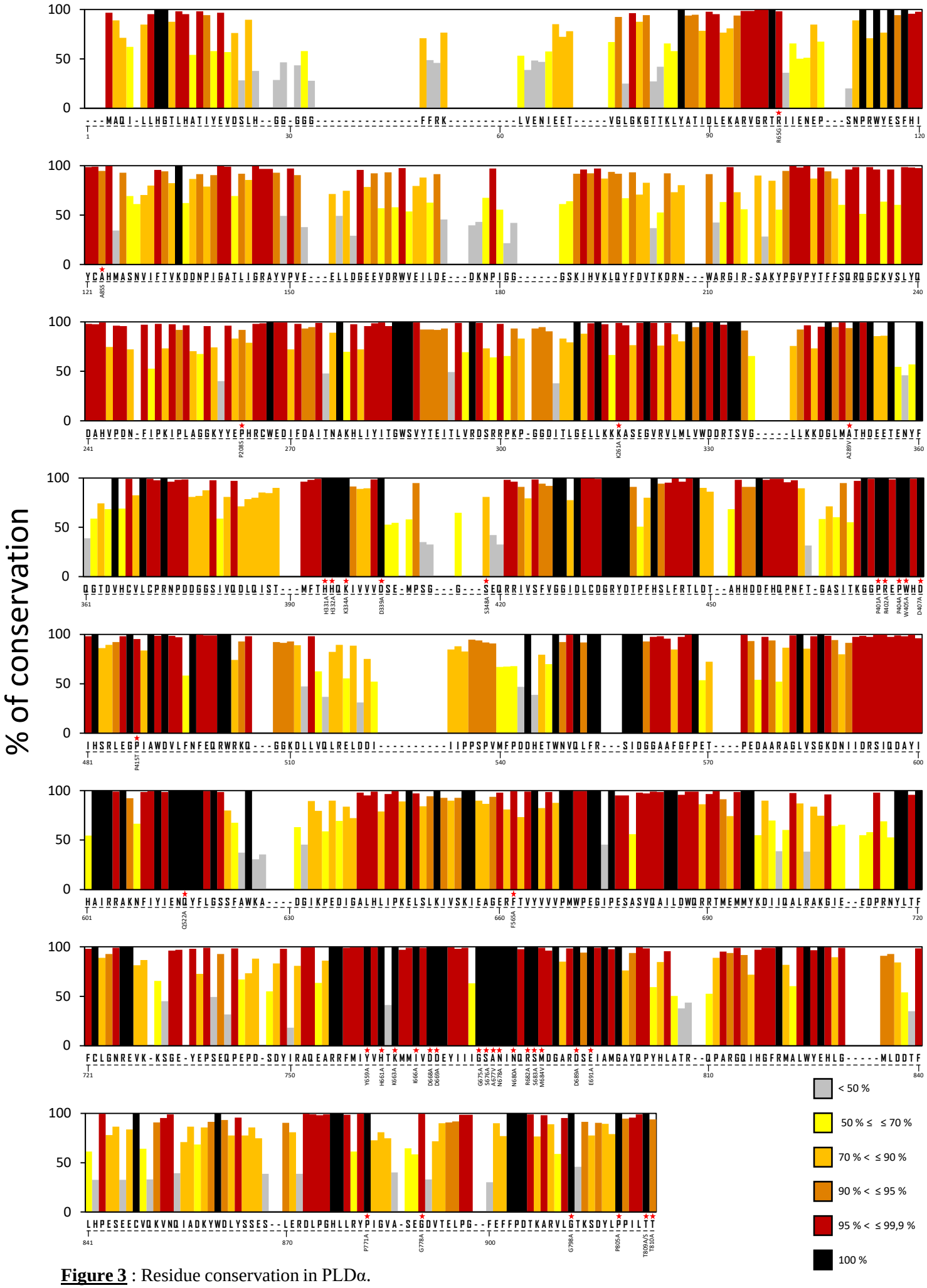

Figure 3 : Residue conservation in PLDα.
The histogram of the consensus sequence indicates the score of conservation of each residue in the alignment of 209 plant PLDα. The Y axis represents the percentage conservation of each residue. Mutated positions in *Arabidopsis thaliana* are indicated with red stars and mutations are shown just below.

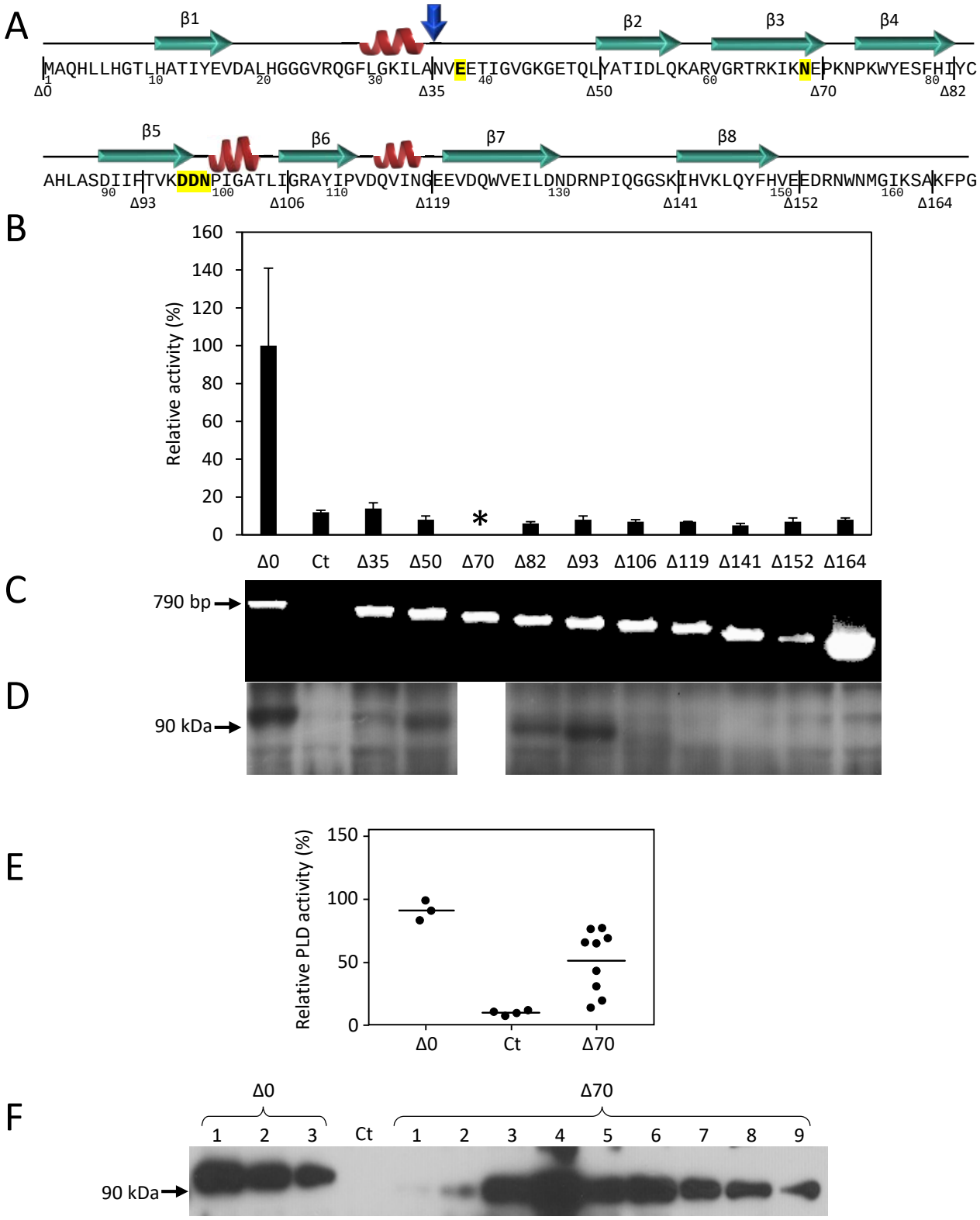

Figure 4 : Expression of AtPLDα minimal domains.

A) Schematic representation of AtPLDα C2 domain. Residues highlighted in yellow represent Ca2+ binding residues, red helices represent α-helices and green arrows are β-strands. Positions of minimal domains are indicated by a black vertical line in the peptide sequence. The blue arrow represents the cleavage position found in recombinant protein after purification. B) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (Δ0), the empty vector (Ct), or the truncated versions of the enzyme. Activities were calculated relative to the Δ0 value. Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to Δ0). The asterisk indicates that the data are represented in part E. C) Validation that the *AtPLDα* transgene was inserted in the genomic DNA of recombinant clones. After DNA extraction, a PCR was run with primers located on the promotor and on the cDNA 790 bp downstream of the end of the C2 domain coding sequence; note the absence of amplicon in the control (Ct). D) Western-blot analysis of the expression of AtPLDα minimal domains. The blotted bands were immmunodetected with a specific AtPLDα C2-domain antibody raised in rabbit, and subsequently visualized using a peroxidase labelled goat anti-rabbit IgG antibody using enhanced chemiluminescence. E) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (Δ0), the empty vector (Ct), or the Δ70 version of the enzyme. Activities were calculated relative to the Δ0 value (100%). Results shown are on $n = 3$ (Δ0), $n = 4$ (Ct) or $n = 9$ (Δ70) independent clones. The horizontal bar represents the mean values. F) Western-blot analysis of the expression of AtPLDα minimal domains. The blotted bands were immmunodetected as mentioned above.

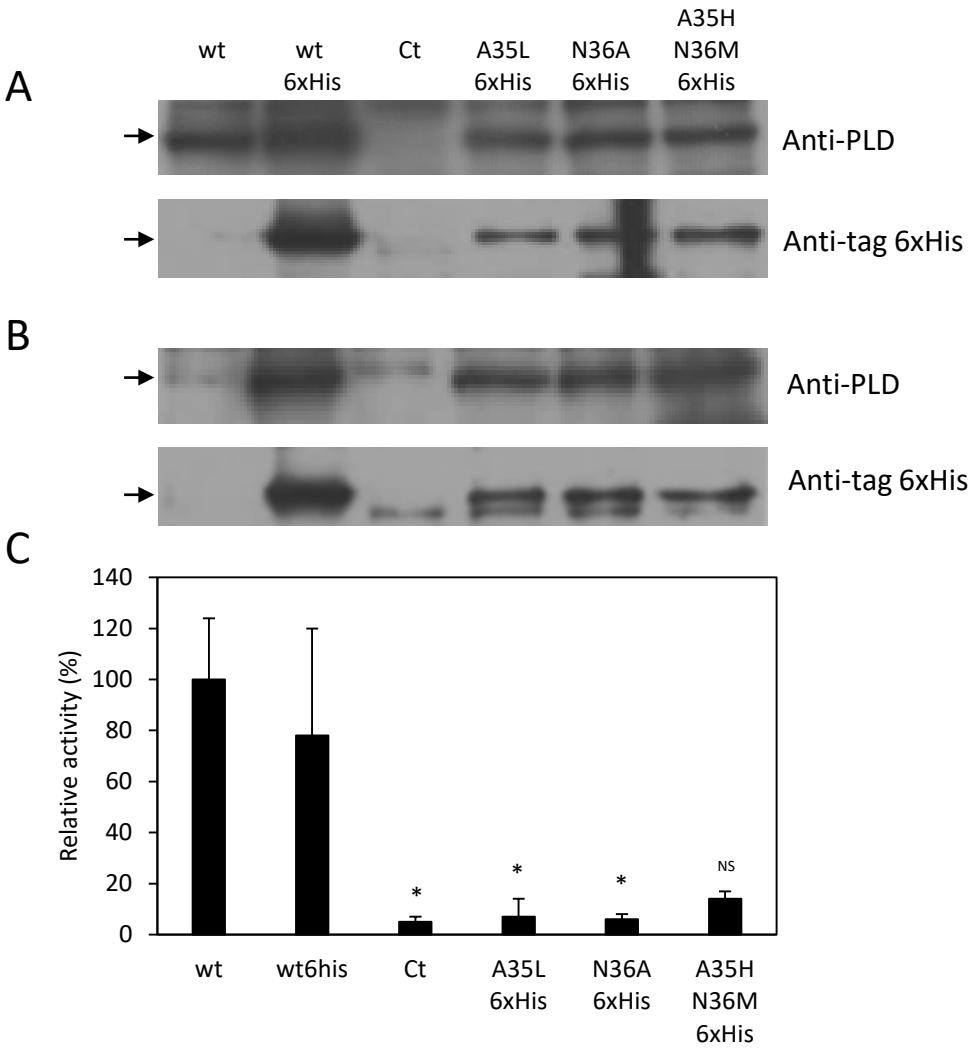

Figure 5 : Western-blot analysis of AtPLDα constructs after purification by affinity-chromatography.

The expression was detected in crude protein extracts by immunorevelation of *P. pastoris* clones, transformed either with the empty vector (Ct) or expressing the wild-type enzyme (wt), the wild-type enzyme fused with a 6xHis-tag (wt 6xHis) or the different mutants fused with a 6xHis-tag. The blotted bands were immmunodetected with either a specific AtPLDα C2-domain antibody (anti-PLD) or an anti-tag 6xHis peroxidase labelled antibody. A) Crude extract fractions. B) Elution fractions after Ni-TED purification. C) The PLD activity was measured in crude extracts of *Pichia pastoris* expressing either the wt protein (wt), the wild-type protein fused with a 6xHis tag (wt 6xHis), the empty vector (Ct), or mutants of the cleavage site. Activities were calculated relative to the wt value (100%). Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to wt 6xHis, except A35H N36M 6xHis mutant that is not significant (NS)). The arrow indicates the PLD position at 92.8 kDa.

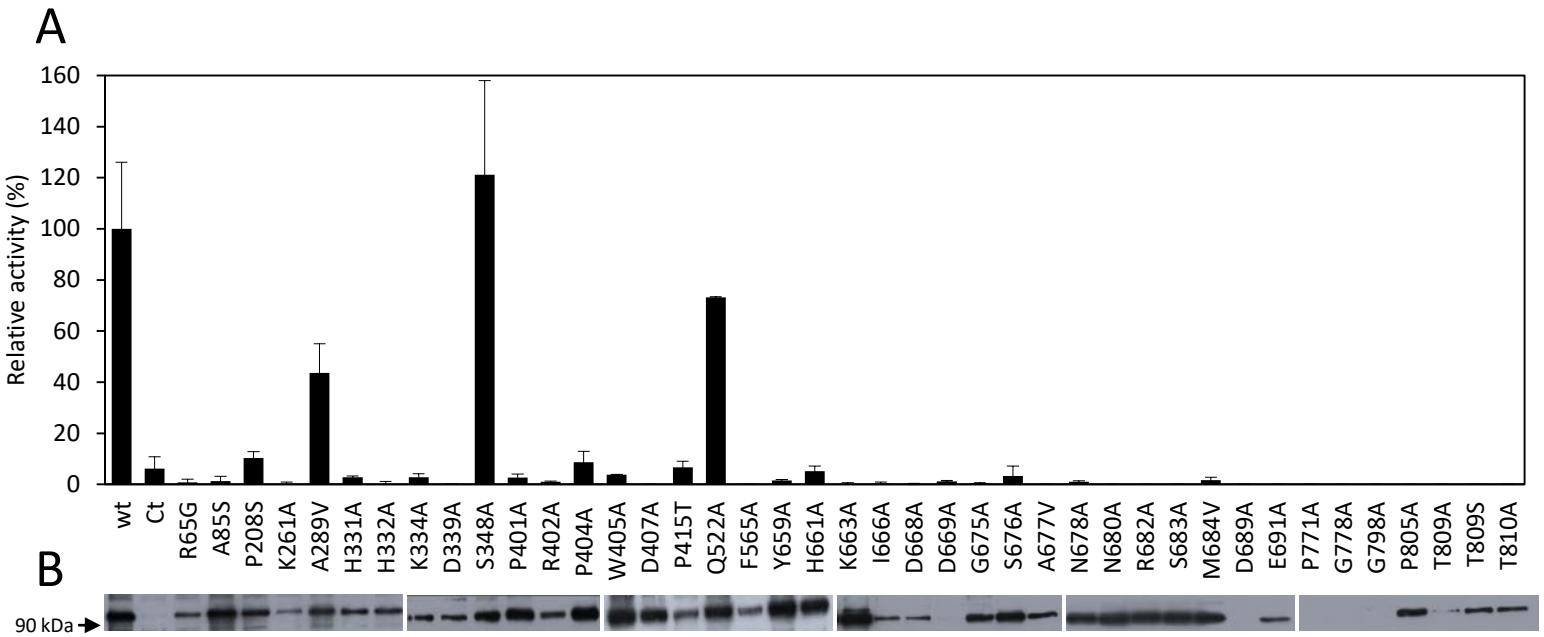

Figure 6 : Relative enzymatic activity and expression of forty-one AtPLDα point mutants.

A) Specific activity was measured with 30 µg of protein from a crude extract of *P. pastoris* expressing either the wt protein (wt) of a molecular mass of 91,7 kDa, the mutated enzymes or transformed with the empty vector (Ct). Activities were then calculated relative to the wt value (100%). Values are the mean ± SD obtained from three independent clones ($p<0.05$ for all constructs compared to wt, except S348A and Q522A mutants).

B) Western-blot analysis of the expression of recombinant AtPLDα mutants. One out of the three clones mentioned above was randomly chosen for SDS-PAGE, the blotted bands were then immmunodetected with a specific AtPLDα C2-domain antibody.

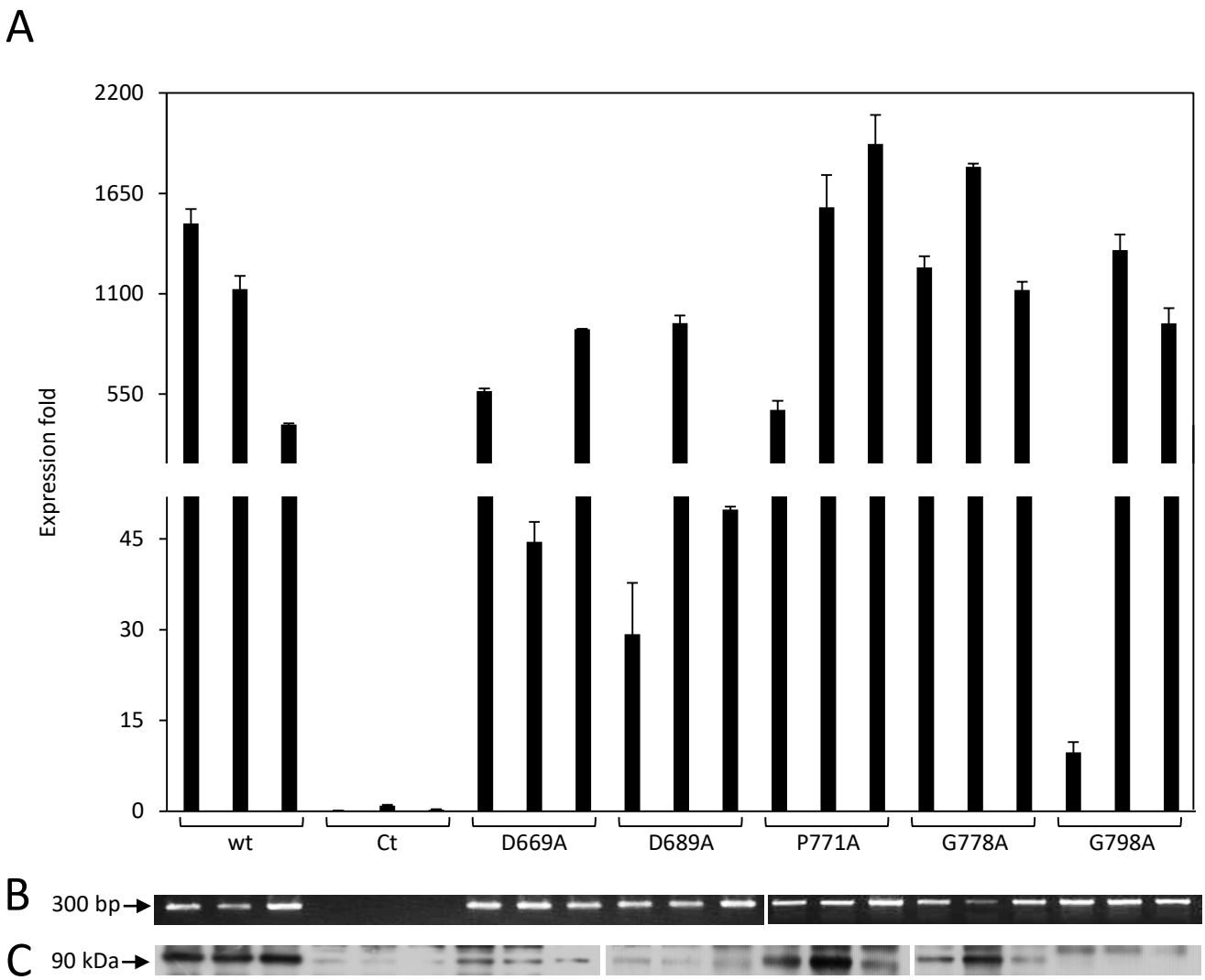

Figure 7 : Expression pattern of *AtPLDα* in different mutants.
A) Relative levels of mRNA from three independent clones expressing either the wt protein (wt), the empty vector (Ct), or the mutated enzyme were determined by a real-time polymerase chain reaction (RT-PCR) using the $2^{-\Delta\Delta Ct}$ quantification method and *ARG4* mRNA as an internal reference. Values are the mean ± SD obtained from duplicate samples. B) Validation that the *AtPLDα* transgene was inserted in the genomic DNA of recombinant clones. After DNA extraction, a PCR was run with primers located on the promotor and on the cDNA; note the absence of amplicon in the control (Ct). C) Western-blot analysis of the expression of AtPLDα mutants. The blotted bands were immmunodetected with a specific AtPLDα C2-domain antibody.

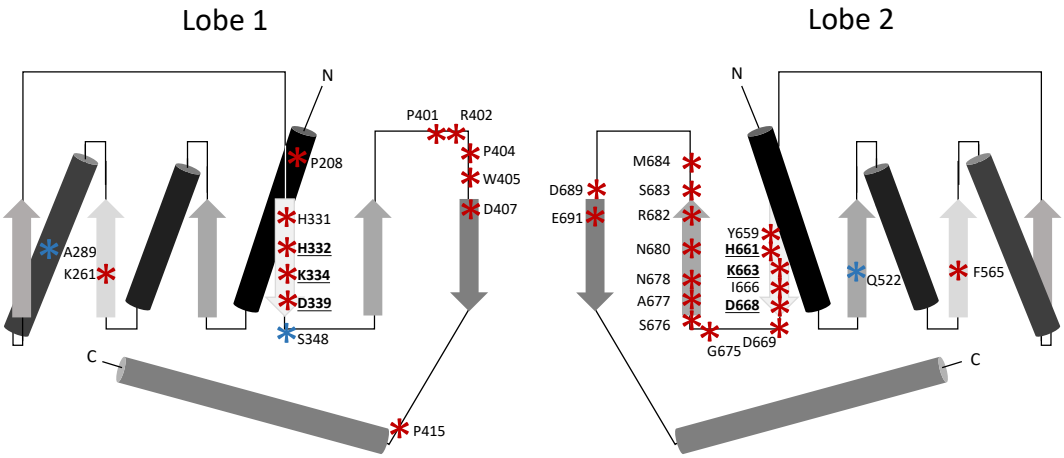

Figure 8 : Schematic topology of the AtPLDα catalytic core.
Cylinders represent the α-helices and arrows represent the β-strands, N: N-terminus extremity, C: C-terminus extremity. Mutations obtained in this study are displayed as stars. Red and blue stars indicate the mutations that affect or do not affect the catalytic activity of the enzyme respectively. HKD catalytic residues are in underlined bold letters.

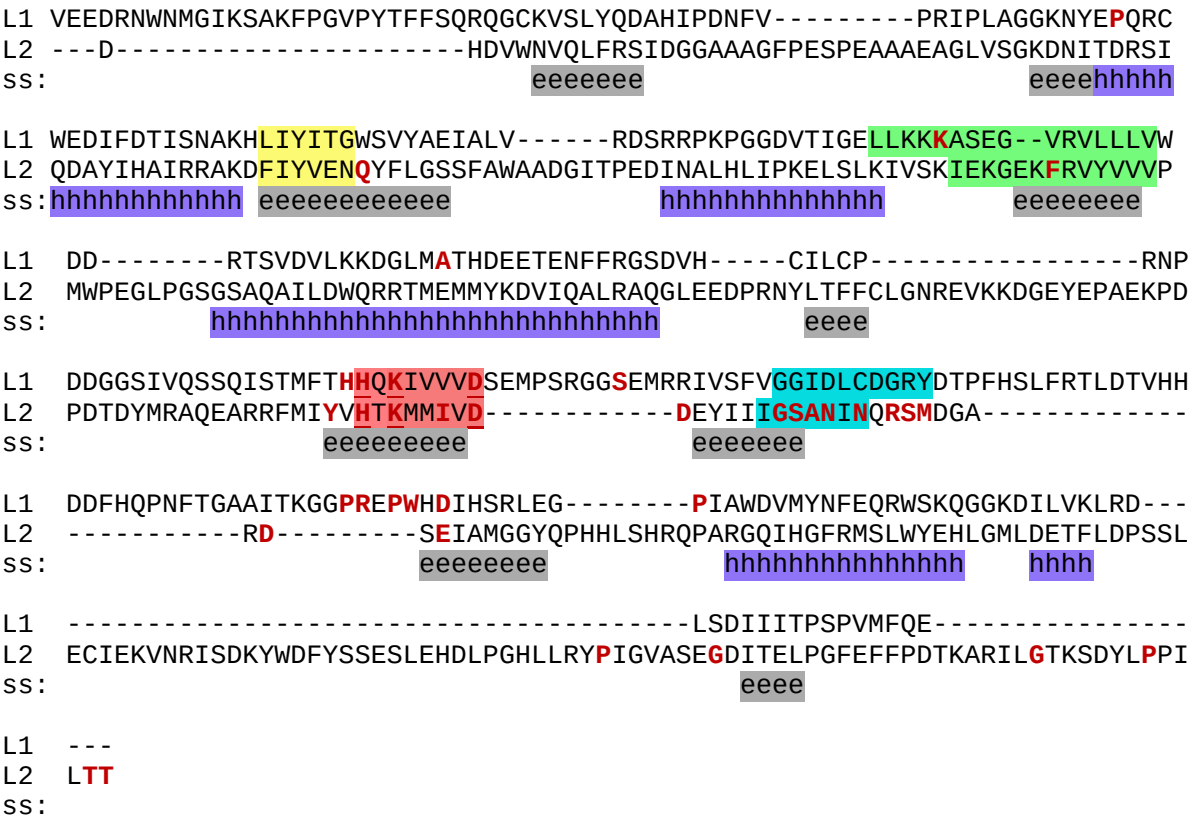

Figure S1 : Alignment of lobe 1 (L1) and lobe 2 (L2) of the catalytic core of AtPLDα.
The four motifs I, II, III and IV, adapted from [1], are highlighted, respectively, in yellow, green, red and blue. Prediction of consensus secondary structures (ss) was made using the PROMALS3D server; α-helices (h) are indicated in purple and β-strands (e) in grey. Secondary structures of less than four residues and secondary structures where the sequences do not match are omitted. HKD motifs are in red bold letters and underlined, mutated residues of this work are indicated in red bold letters.

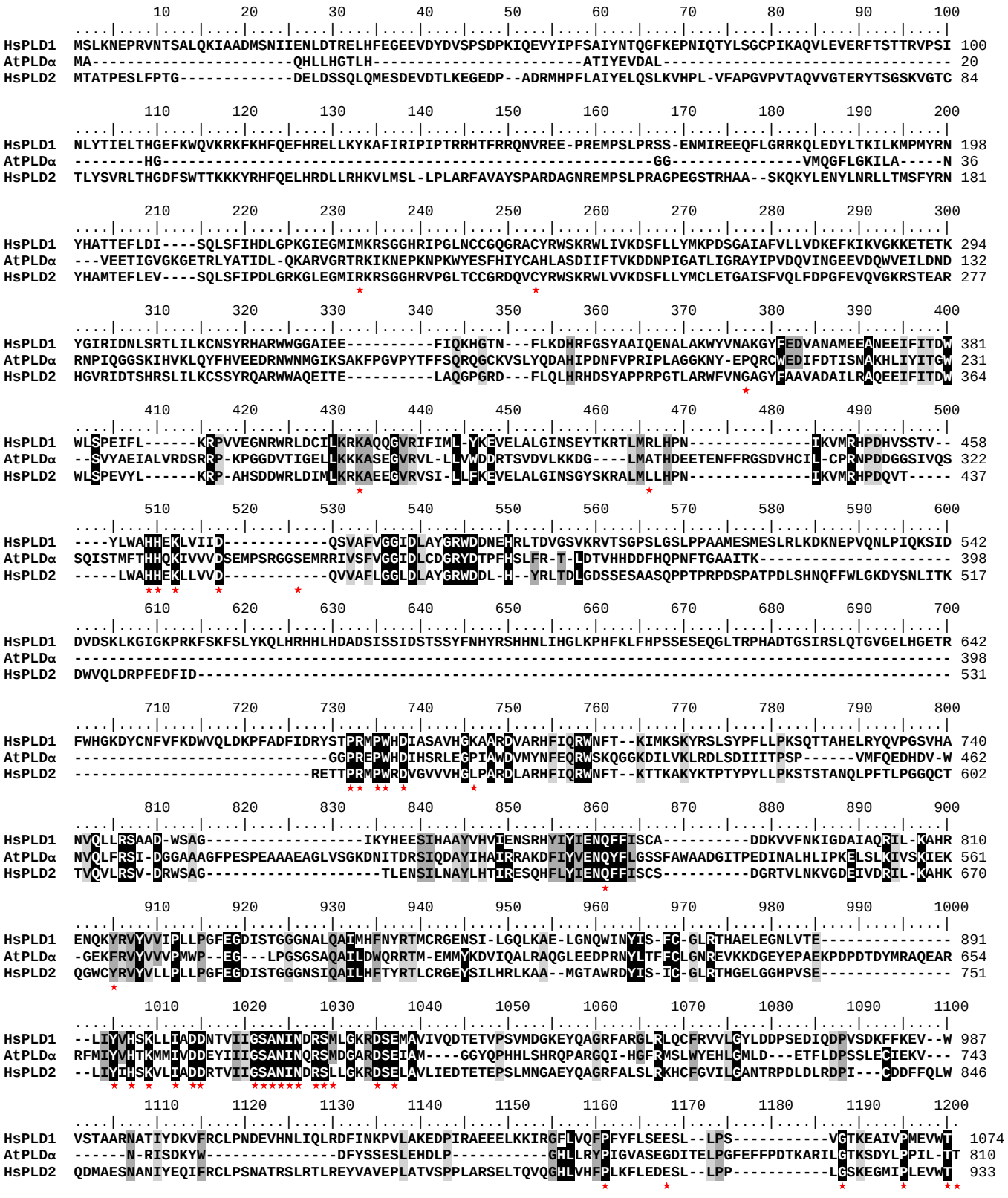

Figure S2 : Alignment of *Homo sapiens* PLD1 and 2 and *Arabidopsis thaliana* PLDα.
Residues highlighted in light grey are identical between HsPLD1, HsPLD2 and AtPLDα. Residues highlighted in dark grey share similar properties between at least three of the four sequences (HsPLD1, HsPLD2 and AtPLDα + plant consensus sequence presented in Figure 2) and are conserved at a minimum of 95 %. Residues highlighted in black share similar properties between at least three of the four sequences (HsPLD1, HsPLD2 and AtPLDα + plant consensus sequence in Figure 2) and are conserved at 100 %. Red stars indicate the mutated residues in AtPLDα.