View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

The first crystal structure of a phospholipase D

Ingar Leiros[1], Francesco Secundo[2], Carlo Zambonelli[2,3], Stefano Servi[4] and Edward Hough[1]*

Background: The phospholipase D (PLD) superfamily includes enzymes that are involved in phospholipid metabolism, nucleases, toxins and virus envelope proteins of unknown function. PLD hydrolyzes the terminal phosphodiester bond of phospholipids to phosphatidic acid and a hydrophilic constituent. Phosphatidic acid is a compound that is heavily involved in signal transduction. PLD also catalyses a transphosphatidylation reaction in the presence of phosphatidylcholine and a short-chained primary or secondary alcohol.

Results: The first crystal structure of a 54 kDa PLD has been determined to 1.9 Å resolution using the multiwavelength anomalous dispersion (MAD) method on a single WO_4 ion and refined to 1.4 Å resolution. PLD from the bacterial source *Streptomyces* sp. strain PMF consists of a single polypeptide chain that is folded into two domains. An active site is located at the interface between these domains. The presented structure supports the proposed superfamily relationship with the published structure of the 16 kDa endonuclease from *Salmonella typhimurium*.

Conclusions: The structure of PLD provides insight into the structure and mode of action of not only bacterial, plant and mammalian PLDs, but also of a variety of enzymes as diverse as cardiolipin synthases, phosphatidylserine synthases, toxins, endonucleases, as well as poxvirus envelope proteins having a so far unknown function. The common features of these enzymes are that they can bind to a phosphodiester moiety, and that most of these enzymes are active as bi-lobed monomers or dimers.

Addresses: [1]Department of Chemistry, Faculty of Science, University of Tromsø, Norway, [2]Instituto di Biocatalisi e Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131 Milano, Italy, [3]Dipartimento di Fisiologica e Biochimica Generale, Universita degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy and [4]CNR, Centro di Studio per le Sostanze Organiche Naturali, Dipartimento di Chimica Politecnico di Milano, Via Mancinelli 7, 20131 Milano, Italy.

*Corresponding author.
E-mail: Edward.Hough@chem.uit.no

Key words: multiwavelength anomalous dispersion (MAD), phosphate-inhibition, phospholipase D (PLD), X-ray crystal structure

Received: **8 December 1999**
Revisions requested: **3 February 2000**
Revisions received: **22 March 2000**
Accepted: **10 April 2000**

Published: **31 May 2000**

Structure 2000, **8**:655–667

0969-2126/00/$ – see front matter
© 2000 Elsevier Science Ltd. All rights reserved.

Introduction

Phospholipase D (PLD) is a ubiquitous enzyme that has been found in a variety of species, for example, plants, mammals and microorganisms. In addition to its hydrolytic activity, PLD catalyzes a transesterification (transphosphatidylation) reaction when alcohol is present as a nucleophilic donor. From this transesterification reaction it is possible to synthesize naturally occurring, low abundance phospholipids such as phosphatidyl-ethanolamine, -serine and -glycerol, from high abundance ones, such as phosphatidylcholine. These and other phospholipids have various applications in the pharmaceutical and food industries, as well as in other fields [1,2]. PLDs from microbial species have a broader specificity for transphosphatidylation activity than PLD from plants (e.g. cabbage [3,4]), which makes the microbial enzymes more interesting for industrial applications of the transphosphatidylation reaction than the plant and mammalian ones. In addition to their contribution to phospholipid metabolism, PLD enzymes are involved in vesicle formation, protein transport, signal transduction and mitosis [5]. One of the products of the hydrolytic action of PLD, phosphatidic acid (PA), is a putative second messenger. Under certain circumstances, the formation of phosphatidic acid is also suggested to cause changes in lipid bilayer properties that would facilitate fusion events and vesicle budding in the course of intracellular membrane traffic [6]. The metabolic fates of PA are the conversion into cytidine diphosphate-diacylglycerol (CDP-DAG) for resynthesis of certain phospholipids; and the formation of two second messengers, diacylglycerol (DAG) and lysophosphatidic acid (LPA), the latter being PA lacking one of the fatty acid chains because of hydrolytic removal [5]. Recently, PLD has been shown to be part of a superfamily [7–9] that includes all the PLDs sequenced so far, cardiolipin synthases, phosphatidyl serine synthases, a *Yersinia* murine toxin, poxvirus envelope proteins and several endonucleases. It is interesting that, with a few exceptions, members of the PLD superfamily share two copies of the conserved sequence motif $HXK(X)_4D$ (in single-letter amino acid code and where X is any amino acid), denoted the HKD motif. The exception being some of the endonucleases, which only have one such domain, and some poxvirus envelope protein homologues, which show only partial conservation of one of the HKD motifs (conserved as $NXK(X)_4D$). It has

been speculated that the lack of the histidine residue in the poxvirus enzymes renders them unable to act as catalysts. Recently, the structure of one of the members of the PLD superfamily, endonuclease (nuc) from *Salmonella typhimurium*, was solved [10]. It is capable of hydrolyzing double- and single-stranded DNA and the artificial subtrate, *bis* (4-nitrophenyl) phosphate [11]. This endonuclease is one of the enzymes in the PLD superfamily that only has one conserved HKD motif and it consists of only 155 residues. It crystallizes as a dimer [10], however, and clearly supports the theory of an ancient gene duplication and splicing event from an ancestral gene for the majority of the members of the PLD superfamily of enzymes [9].

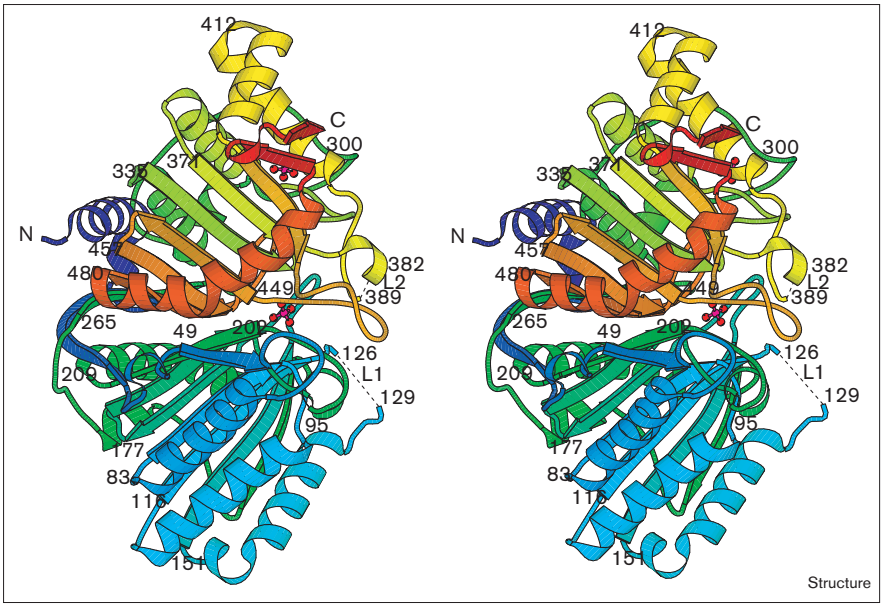
The primary and three-dimensional structures of PLD from *Streptomyces* sp. strain PMF were determined using the multiwavelength anomalous dispersion (MAD) phasing method [12]. Additional solvent-flattening techniques, histogram matching and phase extension of the MAD phases onto the native structure factors extended the working resolution of the maps to 1.4 Å. The present model describes visible regions in electron density and comprises 494 amino acid residues, 2 phosphates and 644 water molecules, with crystallographic R factors of 12.97% and 18.23% for the working and test sets of reflections, respectively.

This paper describes the three-dimensional structure of PLD from *Streptomyces* sp. strain PMF. As sequence information was available for only about 40% of the amino acids, most of the primary sequence was identified from inspection of the 1.4 Å electron-density maps.

Results and discussion

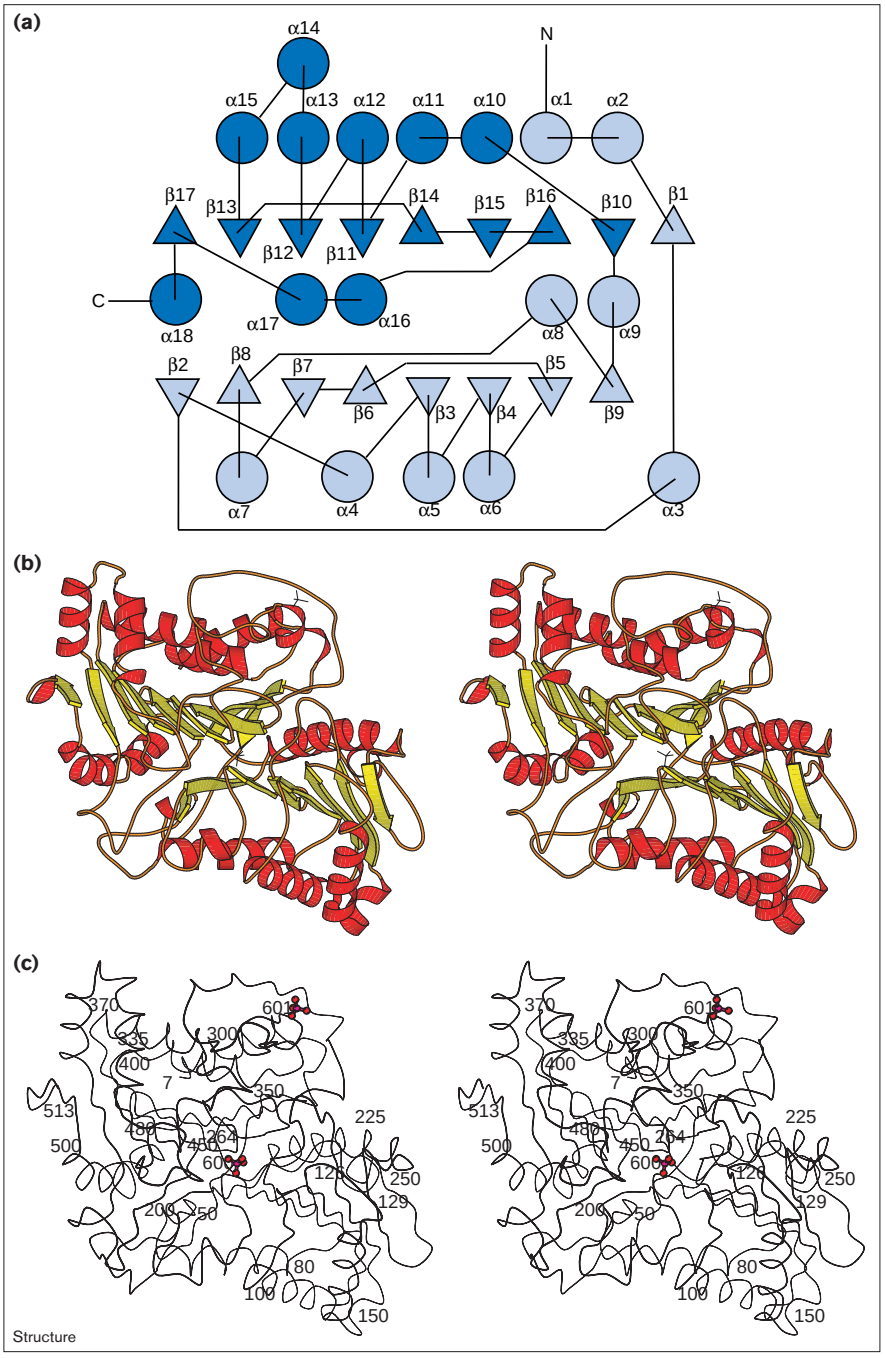
The structure of PLD from *Streptomyces* sp. strain PMF includes 35 secondary structure elements as defined by the Dictionary of Secondary Structures in Proteins (DSSP) [13], forming two β sheets, with nine and eight β strands, respectively. There are 18 α helices that flank the two β sheets; two of these have a somewhat irregular orientation, running diagonally on one side of each of the β sheets. An overall view of the structure of PLD from *Streptomyces* sp. strain PMF is shown in Figure 1 and the topology of the protein is shown in Figure 2a. A stereographic overview of the tertiary architecture of the protein is presented in Figure 2b and a stereographic illustration of the Cα trace is shown in Figure 2c. The two β sheets are sandwiched between surrounding α helices, with the more usual orientation of the helices on the water-exposed side of each β sheet. The α helices α8 and α16–α17 (Figure 2a) lie between the β sheets and are rotated with respect to the general strand direction. The overall structure of PLD from *Streptomyces* sp. strain PMF can be said to be an α-β-α-β-α-sandwich. The overall dimensions of PLD, as seen in Figure 2b, are about 60 Å in diameter with a height of about 48 Å. The depth or length from the presumed membrane-attached side of the enzyme to the opposite end (normal to the paper plane in Figure 2b) is about 42 Å. The structure is composed of two tightly interacting domains with very similar topology, and the overall structure is strikingly

Figure 1



Stereographic illustration of PLD from *Streptomyces* sp. strain PMF colored from dark blue to red according to the amino acid sequence. Two flexible loops (dashed and labeled L1 and L2, accordingly) are seen at the right-hand side, which is the part of the enzyme that docks onto the membrane when hydrolysis takes place. Two phosphate ions (red) are found on the surface of the protein, one located at the interface between two independent crystallographically related molecules, and the other in the active site of the protein, which is located in the center of the figure. This figure was created using BOBSCRIPT [44,45].

Figure 2



Topology and tertiary structure of PLD. **(a)** Overall topology of PLD from *Streptomyces* sp. strain PMF. Modified figure from TOPS [46]. The N-terminal 260 residues are shown in light blue and the remainder of the protein is colored in dark blue, in order to separate the two domains. **(b)** Stereographic overview of the tertiary arrangement of the protein. Helices are shown in red and strands in yellow. The protein is viewed from the outer membrane, and the active-site-bound phosphate can be seen in the center of the protein. This Figure was created using BOBSCRIPT. **(c)** Stereographic presentation of the Cα trace of PLD, made using BOBSCRIPT. The orientation of the protein is as in (b).

similar to the structure of the dimer of the endonuclease (nuc) [10] from *S. typhimurium*.

Although the two rotated α helices in the center of the molecule are distant from the active site, they intersperse the two β sheets in the structure, and it seems probable that they are structurally important. These helices provide the necessary scaffolding or frameworkfor keeping the active site in an open and accessible conformation so that substrate can enter.

The active-site region of the protein is built up of identical residues from each domain, leading to an almost exact twofold axis running through the active-site-bound phosphate present in this structure (see Figure 3).

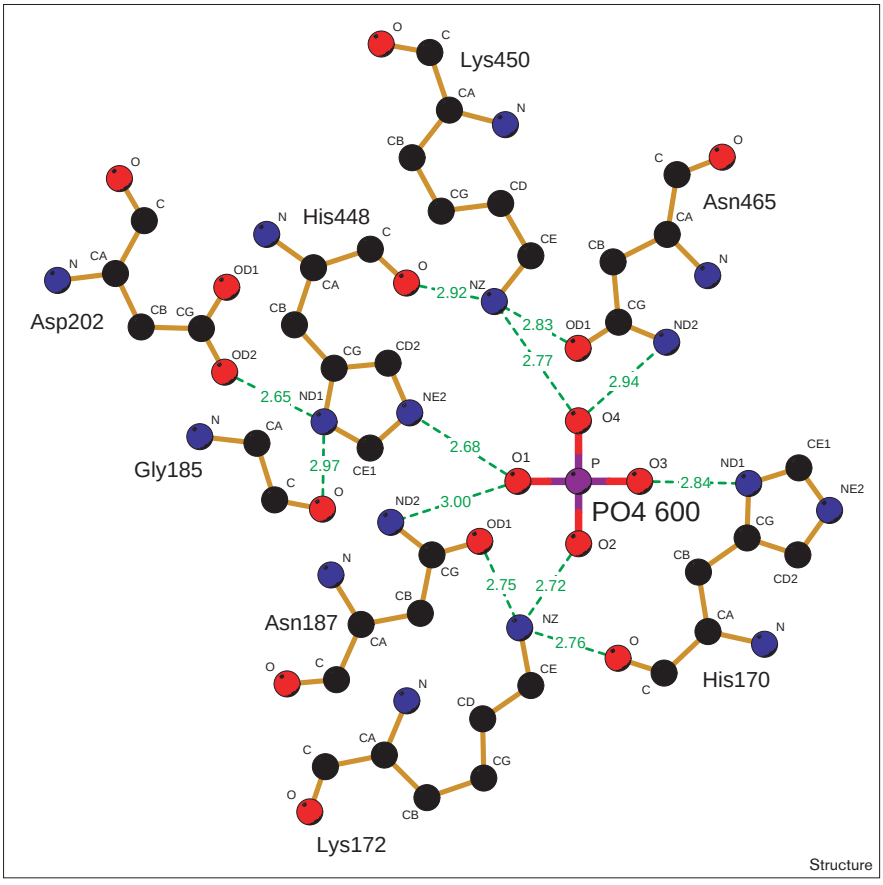
The topology of both β sheets is similar except for the first β strand (β1) in the N-terminal β sheet, which is not present in the C-terminal β sheet. The two β sheets are numbered (0)17652348, where (0) corresponds to β1.

Strands 0, 5, 7 and 8 are antiparallel to the rest. The two β sheets are rotated by about 75° with respect to each other, and the central parts of each domain can be superimposed by a pseudo-twofold rotation axis running almost perfectly through the active site (Figures 2,3 and see later).

Assignment of amino acids

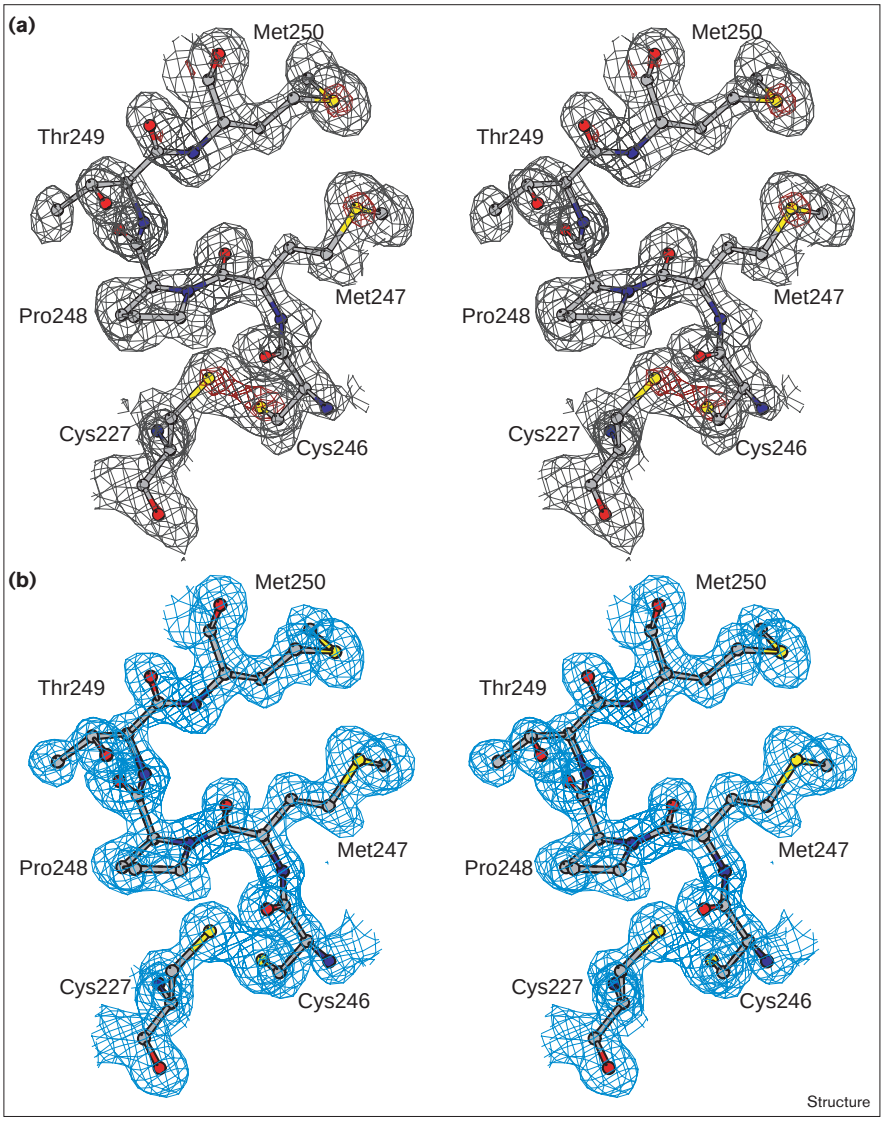
The determination of the tertiary structure of a protein from experimental electron density maps is somewhat hampered by lack of the primary sequence even at 1.4 Å resolution. In this case, however, the shape and quality of the electron density permitted unambiguous assignment of most of the amino acids. The first amino acid residues that were identified and positioned were methionines and cysteines, because their sulfur atoms were visible at a high sigma level in the experimental maps. This is visualized in Figure 4a, where the solvent-flattened map at 1.4 Å resolution is shown with the refined model. The same map at a 5σ level is shown in red. The same view for the final map is shown in Figure 4b. The histidine residues and the aromatic amino acid residues phenylalanine, tyrosine and tryptophan were also readily identified, as were the majority of lysine, arginine, leucine, isoleucine, proline, glycine, alanine and serine sidechains. Valine and threonine residues, having an identical shape, can still be separated because the threonine Oγ1 atom in most cases forms hydrogen bonds to other residues, and the valine sidechain in many cases forms part of a hydrophobic cluster. Aspartic acid/asparagine and glutamic acid/glutamine can, in some cases, be distinguished by their hydrogen-bonding pattern, and in some cases their refined temperature factors. In places where the available sequences had no consensus, however, the choice was made from the fact that asparagine and glutamine are more likely to be buried in the structure than their respective counterparts, aspartic acid and glutamic acid. Comparison of the resulting X-ray sequence with sequences from similar *Streptomyces* and *Streptoverticillium* species showed that the sequence identity is very high. The sequences of several peptide segments from *Streptomyces* sp. strain PMF, which were obtained in our laboratory during the model building and refinement stages, were largely in agreement with the X-ray sequence (Figure 5). The X-ray sequence has enabled the construction of a probe, which at the time of submission has enabled partial gene sequencing in a region including one of the disordered loops (382–389), as well as one of the active-site motifs (448–465). All the conserved regions from the sequences

Figure 3



Two-dimensional presentation of the active site and the bound phosphate molecule. Water molecules in the active site have been omitted for clarity. Atoms are shown in standard colours and bond distances are given in Å. This figure was created using the programs HBPLUS [47] and LIGPLOT [48].

Figure 4



Electron-density maps. **(a)** Experimental solvent-flattened and phase-extended map from DM [41] to 1.4 Å resolution and the refined model. The electron density map is contoured at 1.1 times the rms deviation in gray and at $5\sigma_{rms}$ in red. The two methionine residues and the disulfide bridge are clearly visible in the experimental maps at the high σ level. This figure was created using BOBSCRIPT. **(b)** $2mF_o–DF_c$ SigmaA-weighted [49] electron-density map showing the same region of the protein calculated after the final refinement round, contoured at $1.5\sigma_{rms}$.

match regions in the X-ray structure. Figure 4a shows a portion of the solvent-flattened and phase-extended map from the MAD phases; the final model superimposes well onto the initial maps and allows the inclusion of most of the correct amino acid residues into the experimental maps. There is still some ambiguity, particularly in two loop regions (residues 127–128 and 382–389) where there is no electron density, but overall the final model is of high quality and has stereochemical quality at or even above the average of most of the three-dimensional structures determined at this resolution.

Phosphate geometry and environment, identification of an active site

As determined initially, the structure contained a tungstate ion, which was used to solve the phase problem (see the Materials and methods section). The crystallization, however, was carried out using a citrate-phosphate buffer. Refinement of the model that was built from the MAD data against the data collected in the presence of inorganic phosphate clearly revealed one phosphate molecule occupying the same position as the tungstate molecule. Phosphate and tungstate were later shown to be competitive inhibitors for the enzyme (SS, unpublished observations). In addition, a second phosphate molecule was identified on the surface of the enzyme. The two phosphate molecules located in the native structure were refined to the expected tetrahedral geometry; however, their environments are very different. The best-defined phosphate molecule is bound in the active site and has a well-defined overall shape and hydrogen-bonding pattern. This phosphate molecule appeared both in the

Figure 5

Multiple sequence alignment of PLD encoded from *Streptomyces* (St.ac, *Streptomyces acidomyceticus*; St.an, *Streptomyces antibioticus*; St.sp, *Streptomyces* sp.) and *Streptoverticillium* (Stv.cinn, *Streptoverticillium cinnamoneum*) species, and portions of the sequence of PLD from *Streptomyces* sp. strain PMF (pept.) obtained in our laboratory, or at a later stage from the gene sequence (residues 371–484), and on top the experimentally determined sequence of PLD from *Streptomyces* sp. strain PMF obtained using X-ray crystallography (X-ray). Secondary structure elements are represented by red cylinders (α helices) or yellow arrows (β strands) and are as defined by DSSP [13]. Light grey regions denote residues conserved in at least four sequences describing bacterial PLDs. Boxes represent the duplicated and conserved motifs 1–4 [9]. In these regions, the sequences for nuc (aligned twice and denoted nuc-A and nuc-B in the alignment) from *Salmonella typhimurium* and human PLD1 have been added. Asterisks in the catalytically critical regions (motifs 3 and 4) denote residues conserved in most PLD superfamily members. Brackets (e.g., <19>) denote the number of residues intervening the aligned residues. Dashes represent deletions or regions in the X-ray sequence (residues 127–128 and 382–389) with undefined electron density. The figure was made using ALSCRIPT [50].

$2mF_o–DF_c$ and in the $mF_o–DF_c$ σA-weighted maps when the native structure factors were refined against the model. The second phosphate molecule, however, is on the interface between three independent molecules and could only be refined when the occupancy was lowered to 0.5. The hydrogen-bonding pattern of the two molecules are also different with the active-site-bound phosphate forming six protein interactions as can be seen in Figure 3, whereas the other phosphate has three protein interactions only (one from each molecule). The active-site-bound phosphate molecule might well adopt the same orientation as the phosphate head group of a phospholipid substrate during initial binding to the enzyme and subsequent hydrolysis. Thus, the structure of PLD with phosphate in the active site might provide insight into the binding of phospholipid, and identify which residues are responsible for substrate binding and catalysis. It can be seen in Figure 3 that the residues in close contact with the phosphate head group of the substrate are histidine, lysine and asparagine, with one of each from each of the two domains in the protein. The most interesting residues, from a catalytic point of view, are the two histidines. They have relatively tight interactions with the phosphate group and are the most potent ones to perform the hydrolysis of the bound phospholipid in the substrate-binding pocket of the enzyme.

Multiple sequence alignment [7,9,14] has shown that the endonuclease nuc from *S. typhimurium*, a small 155 residue-protein, shows a surprising sequence similarity to mammalian, plant and bacterial PLDs. This is particularly true for the residues in the vicinity of the active site. As mentioned already, multiple sequence alignment has identified a PLD superfamily, in which the common denominator is that the proteins have been shown to be, or are assumed to be, phosphate-binding proteins. In contrast to the majority of the members of the PLD superfamily, nuc is a small protein, and the primary structure has only one repeat of the predicted active-site residues. In the three-dimensional structure, however, nuc forms a dimer with a crystallographic twofold axis directly through the active-site-bound tungstate ion [10]. Therefore, the active sites and their surroundings in the structures of PLD from *Streptomyces* sp. strain PMF and nuc from *S. typhimurium* are very similar. One important difference, however, is the coordination of the two histidine residues in the active sites of the two structures. In the three-dimensional structure of nuc, they are restricted by the crystallographic symmetry and have identical surroundings and charge. Both histidine residues have hydrogen bonds from their imidazole ring nitrogens (Nδ1 and Nε2) to the tungstate molecule and to one glutamic acid. This leads to a bridging of the imidazole rings between the glutamic acid and the tungstate molecule. In PLD, only one of the histidine residues (His448) is bridged by hydrogen bonds in this fashion, and instead of a glutamic acid, an aspartic acid (Asp202) forms a hydrogen bond to one of the imidazole nitrogens (Nδ1). The other histidine residue (His170) is only hydrogen bonded to the phosphate molecule, and in this case, Nδ1 of the imidazole ring participates in this bond. Oδ1 of a second aspartic acid (Asp473) is 5.3 Å away from the Nδ1 atom of His170. It is strange that these two residues are so far apart and this might be caused by the binding of the phosphate molecule, which might be able to penetrate deeper into the active site and replace the protein–protein intramolecular interactions by the relatively strong hydrogen bond between His170 and one of the phosphate oxygens. This is indicated by the tungstate-inhibited structure of PLD from *Streptomyces* sp. strain PMF in which the sidechain of His170 is oriented with the Nε2 atom pointing directly towards the tungsten atom. The distance between the tungsten atom and the Nε2 atom is 2.33 Å. In this orientation, His170 Nδ1 forms an interaction with Asp473 Oδ1. This could indicate that His170 is the nucleophile in the catalytic reaction, as the other probable protein nucleophile, His448, is positioned considerably further away from the tungsten atom, with a distance of about 3.5 Å between the tungsten atom and the His448 Nε2 atom (data not shown). The role of the second histidine (His448) is then most likely to activate a water molecule being the second nucleophile in the reaction mechanism. It should be noted that this observation is in disagreement with what has been found for PLD from *Streptomyces antibioticus*, in which the catalytic nucleophile has been located to the C-terminal domain (His450 in PLD from *Streptomyces antibioticus*) [15].

Comparison with other lipolytic enzymes

The three-dimensional structures of many lipolytic enzymes have been determined. This includes the secretory and cytosolic phospholipases A_2 ($sPLA_2$ [16] and $cPLA_2$ [17]) the phosphatidyl-inositol (PI), phosphatidyl-choline (PC) specific phospholipase C and α toxin (PI-PLC from both bacterial [18] and mammalian sources [19], PC-PLC from *Bacillus cereus* [20], and the α toxin from *Clostridium perfringens* [21]), as well as many triacylglycerol lipases, including cutinase [22], and fungal lipases [23]. All members of the triacylglycerol lipase family share the common α/β-hydrolase fold, having one mainly parallel β sheet (except for strand 2), surrounded on both sides by α helices. Some of the family members have an extra antiparallel strand (strand 9). In order to search for structural similarities among the above mentioned structures of lipolytic enzymes, the starting point is to look for conservations within their overall fold. No such conservation can be seen that includes all the members of the lipolytic enzymes but most of the members have an α/β-hydrolase fold, which is common to many esterases and other hydrolytic enzymes [24,25], or they show similarity to this fold. One example is $cPLA_2$, which shows conservation in most of the strand directions and in that it contains a nucleophilic elbow. The reaction mechanism of $cPLA_2$ is distinct from other acyl hydrolases, however, as it contains a catalytic diad instead of the usual triad [17].

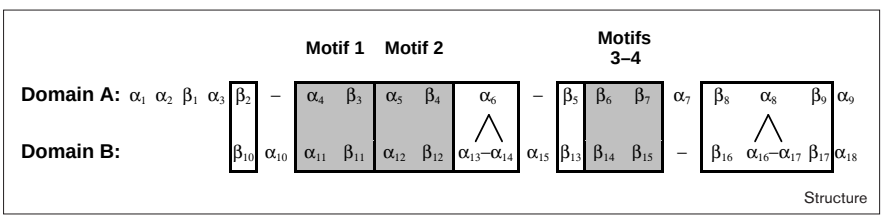
A DALI (distance matrix alignment) [26] database search for similar three-dimensional structures using the refined model of PLD as the target, did not produce any deposited structures that are strikingly similar to PLD, except for the already mentioned structure of nuc [10]. The best fit in the database search was obtained for cutinase from *Fusarium solani* (PDB code 1cex [22]), a member of the lipase family and one of the smallest of the α/β-hydrolases. Cutinase has only 197 residues, a Z score of 3.3, as defined by the DALI software (structure pairs with Z scores below 2 are structurally dissimilar), and a root mean square (rms) deviation of 4.48 Å for the 86 aligned Cα atoms. Optimization of the structural alignment using the available tools in the graphical program O [27] gave an rms deviation of 1.87 Å for 63 aligned atoms. Superpositioning of these atoms in the two structures showed that a region including the strands β3, β4 and β5 in PLD superimposes well onto strands β3, β5 and β6 in cutinase (following the α/β-hydrolase nomenclature). Otherwise, the structures are completely different. Rat PI-PLC-δ1 (isozyme δ-1, PDB code 1djx [28]) had a Z score of 3.0 and an rms value of 10.9 Å for 117 Cα atoms, which could be optimized to 2.11 Å for 56 Cα atoms in O

[27]. It was striking that the same region in PLD superimposed onto both cutinase and PI-PLC-δ1, especially considering that cutinase is a mixed α-β-α-structure with a parallel β sheet, whereas the aligned part of PI-PLC-δ1 is an α/β (TIM) barrel. Another structural alignment database search was performed using the program DEJAVU [29], again looking for proteins having a similar overall fold as PLD. The best score, as decided from graphical inspection of two superpositioned models, was obtained against the eukaryotic DNA polymerase processivity factor PCNA (PDB code 1plq [30]) with a score value of 1.74 and an rms deviation of 1.88 Å for six fitted secondary structure elements, all being β strands. The β strands fitted were β1, β10, β11, β14, β15 and β16 in PLD. In addition to these six β strands, another β strand aligns well but has different strand direction in the two proteins (β12). From Figure 2a, it can be seen that all these β strands are close to each other and are essential parts of one of the β sheets in the structure.

Sequence and structural comparison to other members of the PLD superfamily

Sequence alignment of PLD from several *Streptomyces* species shows that a number of regions are conserved among all the species. Calculating the sequence identity between the experimental X-ray sequence of PLD from *Streptomyces* sp. strain PMF and PLDs from other *Streptomyces* sequences, showed that the sequences from *Streptomyces acidomyceticus* [31], *Streptomyces antibioticus* [32] and *Streptomyces* sp. [33] have 74.3%, 72.1% and 69.9% sequence identity, respectively, when compared with the sequence obtained from the refined X-ray structure of PLD from *Streptomyces* sp. strain PMF. A recently deposited sequence encoding PLD from *Streptoverticillium cinnamoneum* [34] has a surprisingly high sequence identity of 78.3% when compared with the X-ray sequence of PLD from *Streptomyces* sp. strain PMF. An overview of the aligned sequences of PLD from *Streptomyces* and *Streptoverticillium* species, as well as the experimental X-ray sequence of PLD from *Streptomyces* sp. strain PMF is shown in Figure 5. This figure also includes the conserved regions of the sequences of nuc [35] from *S. typhimurium* aligned twice, and human PLD1 [14]. A sequence alignment of one of the domains of the X-ray sequence of PLD from *Streptomyces* sp. strain PMF against the other domain did not yield any striking results apart from regions already found to be duplicated in the primary structures of PLD from both bacterial, plant, yeast and mammalian sources [7,9,14]. These involve the active-site regions including the conserved residues responsible for substrate binding and hydrolysis, as well as regions of more general character. From a structural point of view, the topology diagram shown in Figure 2a, clearly indicates that the two domains have a common origin, and the structure clearly consists of two domains related by a pseudo-twofold rotation axis running through the active-site-bound phosphate. The region most clearly duplicated in the two domains, however, is the active-site region containing residues 170–187 and 448–465. These regions mainly share the same primary and secondary structural elements. The conserved aspartic acid in the $HXK(X)_4D$ motif that exists for many members of the PLD superfamily (Asp177 and Asp455 in PLD from *Streptomyces* sp. strain PMF), was initially believed to be involved directly in the catalytic process. Similarly to what is already shown in the crystal structure of nuc [10], these these conserved residues are indeed positioned a long way (about 30 Å) from the active site of PLD from *Streptomyces* sp. strain PMF. As in nuc, these residues are involved in several intramolecluar interactions. Except for the already mentioned region, there are no striking sequence similarities from one domain to the other. The suggestion of an ancient gene duplication and fusion event seems to be valid. When residues from the active-site regions in both domains were chosen and structurally aligned using the tools available in O [27], several secondary structural elements could be superpositioned onto each other. If the enzyme was picked twice and one molecule was superpositioned onto the other, a total of 251 Cα atoms from each molecule could be superpositioned with an rms deviation of only 2.05 Å. Regarding the topology of the two domains, there seems to be a conservation of the directionality and position of the β strands in each of the two β sheets (the one exception being the very first β strand, β1, which is not duplicated), and most of the α helices are positioned roughly in the same fashion in the two domains, protecting the mainly hydrophobic β strands from the hydrophilic surroundings. Results from the

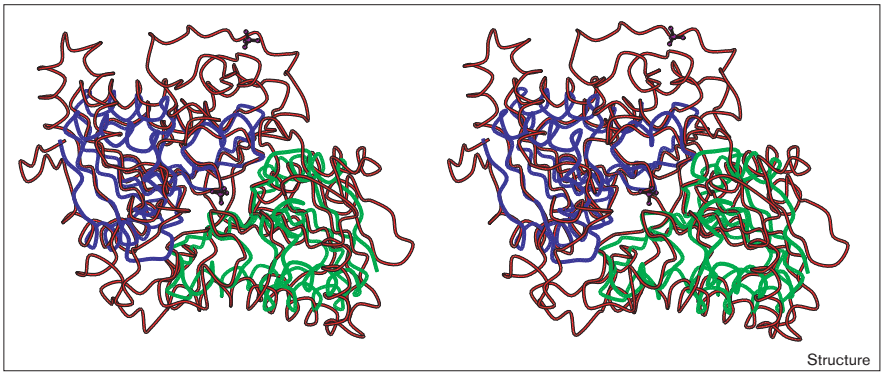
Figure 6



Simplified internal sequence and structural alignment of PLD from *Streptomyces* sp. strain PMF. PLD is divided into two domains and one domain is aligned onto the other. Filled boxes represent predicted conserved motifs from multiple sequence alignment; bordered boxes represent structurally aligned motifs. The regions denoted motifs 1–4, are the motifs already found to be conserved from multiple sequence alignment of PLD and its homologues [9].

Figure 7

Structural alignment of phosphate-inhibited PLD from *Streptomyces* sp. strain PMF (red) onto a dimer of tungstate-inhibited nuc (PDB code 1bys) from *Salmonella typhimurium* (blue and green).



multiple sequence alignment and structural alignment are summarized in Figure 6.

When the topology of PLD is compared with the reported topology and 'domain organization' between two crystallographically related molecules in the crystal structure of nuc [10], it becomes apparent that the topology, that is the strand direction and the orientation of the α helices, to some extent is conserved from one structure to the other. It is possible to align the primary structure of nuc twice against the sequence of PLD as determined using X-ray crystallography; that is, each domain of PLD can be aligned against the primary structure of nuc. It is also possible to align the structure of a dimer of nuc from *S. typhimurium* onto the structure of PLD from *Streptomyces* sp. strain PMF. Doing this gives an overall rms deviation of roughly 4 Å for all Cα atoms of nuc onto the most similar residues in PLD. This alignment is shown in Figure 7, and shows that the two active-site regions are indeed very similar, although more distant regions differ more. It seems that the structure of nuc from *S. typhimurium* provides a minimal structural scaffold for hydrolysis of the phosphodiester backbone common within the majority of the members of the PLD superfamily of proteins. Recombinant nuc is a small protein consisting of only 155 amino acid residues. The secondary structure elements seem to be densely packed, having very short interconnecting loops. As nuc is known to recognize both single- and double-stranded DNA as well as the artificial phosphodiester substrate *bis* (4-nitrophenyl) phosphate [11], it might be less specific than most other members of the PLD superfamily, and therefore have a more solvent-exposed active site than that, for example in the structure of PLD from *Streptomyces* sp. strain PMF, in which the active site is deeply buried within the protein. In order to determine which residues are responsible for the substrate specificity of PLD, the regions between the active site and the surface need to be examined in detail. The active-site entrance (entrance of the substrate-binding pocket) is cone-shaped with a width across the top of about 30 Å. The active-site-bound phosphate is approximately 13 Å from residues on the lip of the entrance. In the recent structure of the human cytosolic PLA_2 [17]), it is speculated that a flexible lid is the reason for the interfacial activation [36] of that enzyme. In PLD, two such flexible regions (loops) are found, both being positioned relatively close to the active site, in the region of the enzyme believed to be in contact with the membrane when phospholipid catabolism takes place. When one of the domains in the protein is superpositioned onto the other as described previously, the two loops can be superpositioned as well; that is, they occupy roughly the same position in space. One of the loops is of particular interest (residues 382–389) both in that it is relatively long and that it is absolutely conserved within all *Streptomyces* and the one *Streptoverticillium* sequence identified so far. The function of the loops might be to act as a lid in order to shield the hydrophobic parts of the active site from the surroundings, and/or to embed the enzyme into the membrane in order for catalysis to happen. In view of the long distance from the active site to the surface of the protein, it is likely that the protein is in some way brought in closer contact with the lipid bilayer membrane and thereby the substrate upon catalysis.

Biological implications

The phospholipase D (PLD) superfamily of enzymes consists of members as diverse as PLDs, a *Yersinia* murine toxin, endonucleases, phosphatidyl-serine synthases, cardiolipin synthases and poxvirus envelope proteins of unknown function. The common denominator of these proteins is their ability to bind a phosphodiester moiety in their active site. The active site region of the PLD superfamily members consists of a conserved motif, the $HXK(X)_4D$ motif (or the HKD motif). Most of these proteins contain two such motifs, a few exceptions are reported, for example, the 16 kDa endonuclease nuc [35] which contains one such motif only. The crystal structure of nuc [10] reveals that it is crystallized as a dimer with a crystallographic diad running through the active site to form a single active site from a dimer.

Unlike nuc, the crystal structure of the 54 kDa PLD from *Streptomyces* sp. strain PMF reported here contains two domains and makes up one active site from a monomer. Except for the differences in size between the two proteins, the overall impression is that they are exceptionally similar considering the fact that the two enzymes are evolutionarily very divergent. The proposed reaction mechanism for nuc [10] is very likely to be valid both for PLDs and for other members of the PLD superfamily being able to perform catalysis on phosphodiester bonds.

The fact that the crystal structures of PLD from *Streptomyces* sp. strain PMF and the published crystal structure of nuc have overall similarity will be helpful in pointing out the respective regions in the two structures that are crucial for the different substrate specificities of nuc and PLD.

Materials and methods

Crystallization conditions and heavy-atom derivatives

PLD from *Streptomyces* sp. strain PMF was purified as described previously [37]. Crystals were obtained using the sitting-drop vapor diffusion method and two sets of conditions: 0.2 M ammonium acetate, 0.1 M citrate/phosphate buffered at pH 5.4 and 27.5% polyethylene glycol (PEG) 4K; or 0.2 M ammonium sulfate, 0.1 M sodium acetate buffered at pH 5.4 and 27.5% PEG 4K. The first set of conditions gave crystals that were larger and diffracted to higher resolution than the latter. Details about crystallization will be presented elsewhere [38]. Crystals from both conditions belong to the monoclinic spacegroup $P2_1$ with cell parameters $a = 57.28$ Å, $b = 57.42$ Å, $c = 68.7$ Å, $\beta = 93.17°$. The incorporation of sodium tungstate was carried out by backsoaking the crystals in a phosphate- and sulfate-free stabilizing solution prior to the addition of 10 mM of sodium tungstate. Both the crystallizing and the stabilizing solutions also serve as cryoprotecting solutions, because of the relatively high concentration of PEG 4K. All data for the MAD experiment were measured from the same crystal at 100K.

Data collection and processing

Data for the native data sets with and without bound phosphate were collected at the Swiss-Norwegian Beamline (SNBL), ESRF (Grenoble, France) using a MAR Research 300 imaging plate detector system, and data for the MAD experiment were collected at station BM14, ESRF (Grenoble, France), using a MAR Research 345 imaging plate detector system in the 345 mm scanning mode with 150 μm pixel size. A XANES (X-ray absorption near edge structure) scan of the W L_{III}- absorption edge clearly showed the incorporation of an anomalous scatterer into the soaked crystal (Figure 8a).

Data sets were collected at the whiteline (pk; λ1 in Figure 8b), having the maximum anomalous signal (maximum f″), the inflection point (pi; λ2 in Figure 8b), for the rising edge where the dispersive signal is at its minimum (highest absolute value of f′), the second inflection point (pi2; λ3 in Figure 8b), on the falling edge where the dispersive signal is at its maximum (lowest absolute value of f′), and finally a remote wavelength (re; λ4 in Figure 8b), in this case chosen at 0.873 Å, where the dispersive signal is known to have a small value, and the anomalous signal still is of considerable size. Data collection statistics for the MAD data collection as well as the same statistics regarding the data collected on the native enzyme with and without bound PO_4 is presented in Table 1.

Data collected for MAD purposes were indexed with the program DENZO [39], using fixed unit cell parameters and crystal orientation matrices for all four wavelengths in order to provide the same scaling on all data sets. Scaling was carried out using SCALEPACK [39], retaining the anomalous differences and the original indices of all reflections collected. Each of the resulting reflection files were subsequently processed with the CCP4 [40] programs AGROVATA, truncated using TRUNCATE, and the reflection files for all wavelengths were merged together with SCALEIT, with the remote wavelength chosen as the reference. Statistics from the data reduction are given in Table 2.

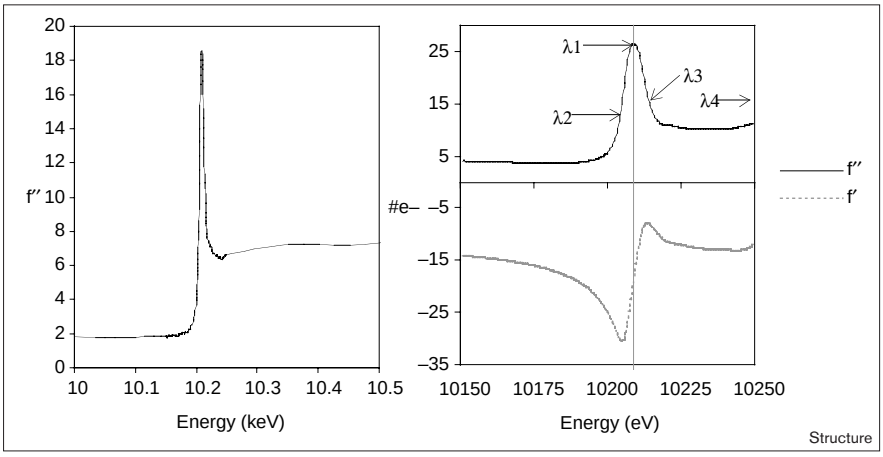
Structure determination and model building

The single-heavy-atom position was readily determined from the anomalous Patterson maps, and was refined in MLPHARE [40]. Refinement statistics are given in Table 3.

The relative occupancy for λ4 in Table 3 can be seen to deviate from the expected values, because the data were collected on the falling edge of the fluorescence scan where it is expected to have the lowest absolute value of f′. In fact it refines to a value that is very close to what is obtained along the rising edge. Nonetheless, λ4 contributes significantly to improve the quality of the MAD phases, and for that reason was included.

The figure of merit (FOM) for the refined position of the heavy atom was 0.52 for all reflections up to 1.9 Å resolution. The resulting phases were then subjected to solvent flattening and phase extensioning using

Figure 8



Two representations of a fluorescence scan of the tungsten L_{III} edge for the crystal used in the MAD data collection experiment. **(a)** The raw data of the energy scan. **(b)** A normalized scan from CHOOCH [51], estimating the dispersive contribution (f′). The wavelengths used in the MAD data collection are indicated in (b).

Table 1

Data collection statistics for the native and MAD data sets.

Crystal	Space group	Resolution (Å)	a (Å)	b (Å)	c (Å)	β (°)	Source
MAD, WO_4	$P2_1$	1.9 (1.7)*	57.61	56.90	68.95	93.67	ESRF, BM14
Native, PO_4	$P2_1$	1.4	57.28	57.42	68.70	93.17	ESRF, BM01

*The number in the parentheses is the resolution for the remote wavelength of the MAD data collection.

DM [40,41] to 1.7 Å resolution giving a FOM of 0.66 for all reflections in the remote data set and an electron-density map of excellent quality. As native data had already been collected to 1.4 Å resolution, however, these data were solvent-flattened and phase-extended in DM using the MAD phases from MLPHARE [40]. The resulting FOM was 0.59, and visual examination of the electron-density map on a graphical workstation, comparing it with the solvent-flattened MAD map to 1.7 Å resolution, clearly showed that the higher resolution map was superior, revealing more structural detail than the 'lower' resolution one, for instance the directionality of the protein backbone. Both maps were used in the model building process, however, because the overall connectivity of the 1.7 Å map was somewhat better.

The initial atomic model was built from the experimental maps using the graphical program O [27]. The quality of the maps allowed the placement of 465 residues into the electron density. This accounts for about 95% of the present model comprising 494 amino acid residues. As our information was only for peptide fragments, sequences encoding PLD from similar species were used in the process of tracing the amino acid backbone. These sequences were from *Streptomyces acidomyceticus* [31], *Streptomyces antibioticus* [32] and a sequence from *Streptomyces* sp. [33], which is somewhat different from the one reported here. Sequenced peptide fragments from the *Streptomyces* sp. PLD herein reported comprised 107 amino acids and gave useful information in the initial model building steps.

Refinement

This initial model was subjected to refinement in X-PLOR [42] using all reflections in the range 15–1.9 Å. The R values after the refinement were 33.4% and 29.3% for the test and working sets of reflections, respectively. The subsequent refinement and model building rounds were performed using all data in the range 15–1.4 Å. After convergence in X-PLOR, refinement was continued using SHELXL [43]. Statistics from the final refinement round can be seen in Table 4. The

Table 2

Statistics from data reduction.

Crystal	λ (Å)	Resolution (Å)	No. measured*	Redundancy	Total	Unique reflections completeness†	Friedel pairs completeness‡ (%)	R_{merge}§ (%)	I/σI# (%)
MAD, pk	λ1 = 1.21427	1.9 (2.0–1.9)	129,032	3.7 (3.5)	35,237	99.8 (99.7)	99.2 (96.1)	5.2 (20.2)	9.4 (3.5)
MAD, pi	λ2 = 1.21475	1.9 (2.0–1.9)	129,022	3.7 (3.5)	35,229	99.8 (99.5)	99.2 (96.2)	4.8 (22.1)	9.7 (3.2)
MAD, pi2	λ3 = 1.21381	1.9 (2.0–1.9)	135,807	3.7 (3.5)	35,241	99.9 (99.6)	99.2 (94.7)	4.8 (23.4)	10.8 .8 (3.1)
MAD, re	λ4 = 0.88553	1.7 (1.79–1.7)	185,333	3.8 (3.7)	48,958	99.6 (99.8)	99.3 (99.4)	4.3 (25.0)	11.8 (2.9)
Native, PO_4	0.873	1.4 (1.48–1.4)	261,275	3.0 (3.0)	86,900	99.2 (99.2)	–	5.8 (20.4)	7.4 (3.4)

Numbers in parentheses represent values for the highest resolution bin. *Number of observations prior to scaling. †Completeness of the unique data sets. ‡Completeness of the anomalous data sets. §$R_{merge} = (\Sigma_h\Sigma_i|I_i(h)-<I(h)>|/(\Sigma_h\Sigma iI(h)$, where $I_i(h)$ is the ith measurement of reflection h and <I(h)> is the weighted mean of all measurements of h. #Intensity signal-to-noise ratio. pk, pi, pi2 and re are defined in the text and Figure 8.

Table 3

Statistics after refinement of the heavy-atom position in MLPHARE.

Crystal	λ (Å)	Resolution (Å)	f′*	f″†	R_{cullis}‡			Phasing power§	
					(acentric)	(centric)	(anom)	(acentric)	(centric)
WO_4, re	λ3 = 0.88553	1.7	0.000	6.854	–	–	0.77	–	–
WO_4, pi	λ2 = 1.21475	1.9	−18.822	10.836	0.70	0.57	0.77	1.63	1.36
WO_4, pk	λ1 = 1.21427	1.9	−10.789	17.515	0.81	0.73	0.68	1.06	0.92
WO_4, pi_2	λ4 = 1.21381	1.9	−18.408	8.938	0.72	0.58	0.80	1.53	1.39

*Refined dispersive occupancy relative to the remote wavelength as given by MLPHARE. †Refined anomalous occupancy from MLPHARE. ‡$R_{cullis} = \Sigma h|\,|F_{deriv}(h)|-|F_{nat}(h)|\,|-|F_H|\,|/\Sigma h|\,|F_{deriv}(h)|-|F_{nat}(h)|\,|$, where F_H is the calculated heavy-atom structure factor. §The phasing power is defined as (rms F_H/rms lack-of-closure) summed over the reflections used in the heavy-atom refinement.

Table 4

Refinement summary.

Resolution range (Å)	10.0–1.40
Number of unique reflections	86,730
Number of non-hydrogen protein atoms	3731
Number of double conformations	17
Number of phosphates	2
Number of waters	644
R value (%)*	12.97
Free R value (%)†	18.23
Deviation from ideal geometry‡	
Bond lengths (Å)	0.014
Bond angles (°)	1.959
B values (Å^2)	
Mainchain atoms	11.37
Sidechain atoms	13.89
Phosphate molecules	13.52
Water molecules	28.69
All atoms	14.94
Ramachandran plot (%)	
In most favored regions	87.2
In additional allowed regions	12.8

*R value $= \sum||F_o|-|F_c||/\sum|F_c|$. †A 5% subset of the reflections was excluded from all steps of the refinement and used for R_{free} [52] calculation purposes. No intensity cutoff was applied during refinement. ‡Refinement was carried out in SHELXL, the coordinates were thereafter brought into X-PLOR for calculation of deviation from the Engh and Huber [53] dictionary.

present model is connected into one polypeptide chain except for two loop regions comprising residues 127–128 and 382–389, both being close to the active site, however the latter being the closest and thought to be important in either substrate recognition, or acting as a lid to cover the hydrophobic parts of the active site from the surroundings. It is an interesting fact that this loop is conserved for all the sequences of bacterial origin.

Accession numbers

The atomic coordinates have been deposited in the Protein Data Bank with accession code 1F0I and will be on hold for one year.

Acknowledgements

The Norwegian Research Council and CNR, PF Biotecnologie (CNR Target Project in Biotechnology) is acknowledged for financial support. ESRF is acknowledged for financial support for the BM14 data collection. The staff at the ESRF beamline BM14 and at the Swiss-Norwegian CRG Beamline (SNBL) are all acknowledged for expert technical assistance.

References

1. D'Arrigo, P. & Servi, S. (1997). Using phospholipases for phospholipid modification. *Trends Biotechnol.* **15**, 90-96.
2. Servi, S. (1999). Phospholipases as synthetic catalysts. *Topics Curr. Chem.* **200**, 127-157.
3. Shimbo, K., Yano, H. & Miyamoto, Y. (1989). Two *Streptomyces* strains that produce phospholipase D with high transphosphatidylation activity. *Agric. Biol. Chem.* **53**, 3083-3085.
4. Shimbo, K., Yano, H. & Miyamoto, Y. (1990). Purification and properties of phospholipase D from *Streptomyces lydicus*. *Agric. Biol. Chem.* **54**, 1189-1193.
5. Exton, J.H. (1994). Phosphatidylcholine breakdown and signal transduction. *Biochim. Biophys. Acta* **1212**, 26-42.
6. Liscovitch, M. (1996). Phospholipase D: role in signal transduction and membrane traffic. *J. Lip. Med. Cell Signal.* **14**, 215-221.
7. Koonin, E.V. (1996). A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem. Sci.* **21**, 242-243.
8. Morris, A.J., Engebrecht, J. & Frohman, M.A. (1996). Structure and regulation of phospholipase D. *Trends Pharmacol. Sci.* **17**, 182-185.
9. Ponting, C.P. & 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.* **5**, 914-922.
10. Stuckey, J.A. & Dixon, J.E. (1999). Crystal structure of a phospholipase D family member. *Nat. Struct. Biol.* **6**, 278-284.
11. Zhao, Y., Stuckey, J.A., Lohse, D.L. & Dixon, J.E. (1997). Expression, characterization, and crystallization of a member of the novel phospholipase D family of phosphodiesterases. *Protein Sci.* **6**, 2655-2658.
12. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
13. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
14. Hammond, S.M., *et al*., & 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.
15. Iwasaki, Y., Horiike, S., Matsushima, K. & Yamane, T. (1999). Location of the catalytic nucleophile of phospholipase D of *Streptomyces* antibioticus in the C-terminal half domain. *Eur. J. Biochem.* **264**, 577-581.
16. Dijkstra, B.W., Drenth, J., Kalk, K.H. & Vandermaelen, P.J. (1978). Three-dimensional structure and disulfide bond connections in bovine pancreatic phospholipase A2. *J. Mol. Biol.* **124**, 53-60.
17. Dessen, A., *et al*., & Somers, W. S. (1999). Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. *Cell* **97**, 349-360.
18. Heinz, D.W., Ryan, M., Bullock, T.L. & Griffith, O.H. (1995). Crystal structure of the phosphatidylinositol-specific phospholipase C from *Bacillus cereus* in complex with myo-inositol. *EMBO J.* **14**, 3855-3863.
19. Essen, L.O., Perisic, O., Cheung, R., Katan, M. & Williams, R.L. (1996). Crystal structure of a mammalian phosphoinositide-specific phospholipase C delta. *Nature* **380**, 595-602.
20. Hough, E., *et al*., & Derewenda, Z. (1989). High-resolution (1.5 Å) crystal structure of phospholipase C from *Bacillus cereus*. *Nature* **338**, 357-360.
21. Naylor, C.E., *et al*., & Basak, A. K. (1998). Structure of the key toxin in gas gangrene. *Nat. Struct. Biol.* **5**, 738-746.
22. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997). Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis. *J. Mol. Biol.* **268**, 779-799.
23. Derewenda, Z.S., Derewenda, U. & Dodson, G.G. (1992). The crystal and molecular structure of the *Rhizomucor miehei* triacylglyceride lipase at 1.9 Å resolution. *J. Mol. Biol.* **227**, 818-839.
24. Derewenda, Z.S. & Derewenda, U. (1991). Relationships among serine hydrolases: evidence for a common structural motif in triacylglyceride lipases and esterases. *Biochem. Cell. Biol.* **69**, 842-851.
25. Schrag, J.D. & Cygler, M. (1997). Lipases and α/β hydrolase fold. *Methods Enzymol.* **284**, 85-107.
26. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
27. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
28. Essen, L.O., Perisic, O., Katan, M., Wu, Y., Roberts, M.F. & Williams, R.L. (1997). Structural mapping of the catalytic mechanism for a mammalian phosphoinositide-specific phospholipase C. *Biochemistry* **36**, 1704-1718.
29. Kleywegt, G.J. & Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525-545.
30. Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. & Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233-1243.
31. Hasegawa, M., Ota, N. & Aisaka, K. (1992). DNA sequence encoding *Streptomyces acidomyceticus* phospholipase D-K. *Japanese patent JP 1992088981-A/1*
32. Iwasaki, Y., Nakano, H. & Yamane, T. (1994). Phospholipase D from *Streptomyces antibioticus*: cloning, sequencing, expression, and relationship to other phospholipases. *Appl. Microbiol. Biotechnol.* **42**, 290-299.
33. Takahara, M., Houriyou, K. & Imamura, S. (1993). DNA having genetic information of phospholipase D-P and its use. *Japanese patent JP 1993252935-A/1*

34. Ogino, C., *et al*., & Fukuda, H. (1999). Purification, characterization, and sequence determination of phospholipase D secreted by *Streptoverticillium cinnamoneum*. *J. Biochem. (Tokyo)* **125**, 263-269.
35. Pohlman, R.F., Liu, F., Wang, L., More, M.I. & Winans, S.C. (1993). Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101. *Nucleic Acids Res.* **21**, 4867-4872.
36. Nalefski, E.A., *et al*., & Clark, J. D. (1994). Delineation of two functionally distinct domains of cytosolic phospholipase A2, a regulatory Ca^{2+}-dependent lipid-binding domain and a Ca^{2+}-independent catalytic domain. *J. Biol. Chem.* **269**, 18239-18249.
37. Carrea, G., D'Arrigo, P., Piergianni, V., Roncaglio, S., Secundo, F. & Servi, S. (1995). Purification and properties of two phospholipases D from *Streptomyces* sp. *Biochim. Biophys. Acta* **1255**, 273-279.
38. Leiros, I., Hough, E., D'Arrigo, P., Carrea, G., Pedrocchi-Fantoni, G., Secundo, F. & Servi, S. (2000). Crystallization and preliminary X-ray diffraction studies of Phospholipase D from *Streptomyces* sp. *Acta Crysallogr. D* **56**, 466-468.
39. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
40. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
41. Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.
42. Brünger, A.T. (1992). *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, CT.
43. Sheldrick, G.M. & Schneider, T.R. (1997). SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319-343.
44. Kraulis, P. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
45. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132-134, 112-133.
46. Westhead, D.R., Slidel, T.W., Flores, T.P. & Thornton, J.M. (1999). Protein structural topology: automated analysis and diagrammatic representation. *Protein Sci.* **8**, 897-904.
47. McDonald, I.K. & Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.
48. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.
49. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140-149.
50. Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.
51. Ewans, G. (1998). http://lagrange.mrc-lmb.cam.ac.uk/doc/Gwyndaf/Chooch.html.
52. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.
53. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392-400.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.