A distant evolutionary relationship between GPI-specific phospholipase D and bacterial phosphatidylcholine-preferring phospholipase C

Daniel J. Rigden*

School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK

Received 23 May 2004; accepted 24 May 2004

Available online 15 June 2004

Edited by Robert B. Russell

Abstract In eukaryotes some surface proteins are attached to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. A GPI-specific phospholipase D (GPI-PLD) activity has been characterized and implicated in the regulation of anchoring, thereby influencing the dispersal of anchored proteins or their maintenance on the cell surface, and possibly in cell signalling. Despite its biological and medical importance, little is known of the structure of GPI-PLD. Here, a distant relationship between the catalytic domains of GPI-PLD and some bacterial phospholipases C is demonstrated. A model of the GPI-PLD catalytic site sheds light on catalysis and highlights possibilities for design of improved and more specific GPI-PLD inhibitors. The databases contain hitherto unnoticed close homologues of GPI-PLD from yeast and *Dictyostelium discoideum*.

Keywords: Glycosylphosphatidylinositol anchor; Phospholipase D; Distant homology; Molecular modelling; Trimetal catalytic site

1. Introduction

One way in which eukaryotic cells anchor their proteins on the outer surface of the plasma membrane is the glycosylphosphatidylinositol (GPI) anchor (reviewed in [1]). The structure of the anchor varies between organisms but is typically of the form EthN–P–Man_3GlcN–PI, where EthN is ethanolamine, P is phosphodiester, Man represents mannose, GlcN is glucosamine and PI phosphatidylinositol [2,3]. The structure may be modified by the addition of extra monosaccharide or ethanolamine moieties and it is the carboxy terminus of the anchored protein that forms an amide bond with the ethanolamine residue. The proportion of proteins that are GPI anchored varies between organisms and tissues but the importance of GPI anchoring is reflected in the dramatic effects on growth and development caused by mutations that affect its synthesis [4,5]. Other molecules may also be GPI anchored. Parasitic kinetoplasts like *Trypanosoma cruzi* and Leishmania are particularly rich in non-protein GPI-linked surface molecules that may have roles in virulence and transmission [1].

While synthesis of GPI anchors is now relatively well understood, the reasons, mechanisms and enzymes behind GPI anchor degradation remain far less clear. Endogenous activity specific for GPI anchors is due to both phospholipases C [6] and phospholipases D [7,8]. Interestingly, although GPI-specific phospholipase D (GPI-PLD) is abundant in plasma, it does not release surface proteins from cells: instead, cellular GPI-PLD affects the degree of anchoring of surface proteins, and hence whether the proteins remain bound to the cell or are dispersed, before their arrival at the cell surface [8,9]. Recent data suggest that GPI-PLD can hydrolyze both intermediates along the synthetic pathway of GPI anchors and GPI anchors already attached to proteins [10]. The products of GPI cleavage may also function as second messengers [11] so that GPI-PLD could also have a signalling role [10]. Further stimulating interest in GPI-PLDs are potential roles in important medical conditions. Thus, GPI-PLD is overexpressed in tumour cells and mRNA levels correlate with malignancy [12]. Indeed, GPI-PLD inhibitors retarded growth of certain cancer cells at concentrations similar to those required for enzyme inhibition [13]. Very recently, a correlation between GPI-PLD levels and increased insulin resistance in humans has been reported, suggesting that GPI-PLD could be a chemotherapeutic target for atherosclerosis [14].

No crystal structures are available for GPI-PLDs and relatively little can be inferred from simple sequence comparisons. A β-propeller domain, homologous to that found in integrins, can be confidently detected in the C-terminal portion of the protein [15]. Conservation of its calcium-binding motifs between integrins and GPI-PLD suggests that the C-terminal domain is responsible for the experimentally observed calcium binding [16] and probably functions to colocalize the enzyme to the membrane surface, as for functionally analogous domains in other phospholipases [17]. However, limited proteolysis shows that catalytic activity is associated with approximately 275 N-terminal residues [18]. In other cases, distant homology of various phospholipases with nucleases [19,20] or phosphatases [21,22] has proved useful for the prediction of catalytic residues. Unfortunately, standard database searches reveal no homologues for the GPI-PLD catalytic domain. Nevertheless, it is shown here, by sensitive sequence comparisons and fold recognition, that the catalytic domain of GPI-PLD bears an unexpected distant homology to bacterial phosphatidylcholine-preferring phospholipases C. Structure modelling reveals an unperturbed trimetal catalytic site architecture in GPI-PLDs and enables a possible catalytic

* Fax: +44-151-795-4414.
E-mail address: drigden@liv.ac.uk (D.J. Rigden).

0014-5793/$22.00
doi:10.1016/j.febslet.2004.05.071

mechanism to be proposed. These results should help design tighter binding, more specific inhibitors of GPI-PLD.

2. Materials and methods

Sequences homologous to human GPI-PLD1 were sought in the GenPept database using BLAST and PSI-BLAST [23] and the resulting sequence set was aligned with MUSCLE [24]. Jalview [25] was used for alignment visualization and determination of the three maximally diverse representatives of the GPI-PLD family. ESPRIPT [26] was used for mapping sequence conservation onto structures. The Perl script ps_scan, obtained from website of the PROSITE database of protein sequence motifs and family alignment-derived sequence profiles [27], was used for later motif-based sequence searches. Searches were made in the Smart [28] and Pfam [29] secondary databases, which contain alignments of individual domains or protein sequences. The N-terminal region of human GPI-PLD1 was submitted to the META-server [30], which provides a portal to the leading fold recognition methods, among them FFAS03 [31], 3D-PSSM [32], Genthreader [33] and Bioinbgu [34], and yields consensus predictions of improved reliability [35–37]. These fold recognition methods search for known protein structures compatible with a target sequence and employ sensetive sequence comparisons [31], sometimes combined with consideration of inferred characteristics such as predicted secondary structure and predicted exposure of residues to solvent [32–34]. Secondary structure predictions were made using PSI-PRED [38].

Modelling of the N-terminal portion of human GPI-PLD (Swissprot accession number P80108; [39]), for which a fold recognition alignment was produced, was performed with MODELLER 6 [40] using the structures of *Bacillus cereus* phospholipase C (PDB code 1ah7; [41]) and the α-toxins of *Clostridium absonum* (PDB code 1olp; [42]) and *Clostridium perfringens* (PDB code 1ca1; [17]) as templates. Default regimes of model refinement by energy minimization and simulated annealing were employed. Because of the low sequence similarity between target and template, a rigorous iterative modelling protocol was employed in which 20 models were constructed and analyzed for each alignment variant. These models were analyzed for packing and solvent exposure characteristics with PROSA II [44] and VERIFY_3D [45], and for stereochemical properties with PROCHECK [46]. Possible misalignments were highlighted by positive PROSA II profiles or negative VERIFY_3D profiles which indicate portions of the model that are atypical, compared to experimental protein structures, with respect to solvent exposure of residues (both methods) or residue–residue interactions (PROSA II). Variations in alignment of these regions were tested and resulting profile improvements taken as likely indications of an increasingly accurate target–template alignment. When no further improvements could be made, the model with the highest PROSA II score was taken as the final model. In the reliably modelled region (Fig. 1), the human GPI-PLD plus yeast and *Dictyostelium discoideum* homologues share 11–13% sequence identity with the three templates. Protein structures were superimposed using LSQMAN [47] and visualized using O [48]. Structure figures were made using PyMOL [49] and secondary structures defined with STRIDE [50].

3. Results and discussion

3.1. Sequence analyses and fold recognition

Iterative databases searches with human GPI-PLD1 (one of two closely related human sequences [39]) as a probe revealed a group of mammalian homologues sharing 79–95% sequence identity with the human enzyme. Outside this group only two

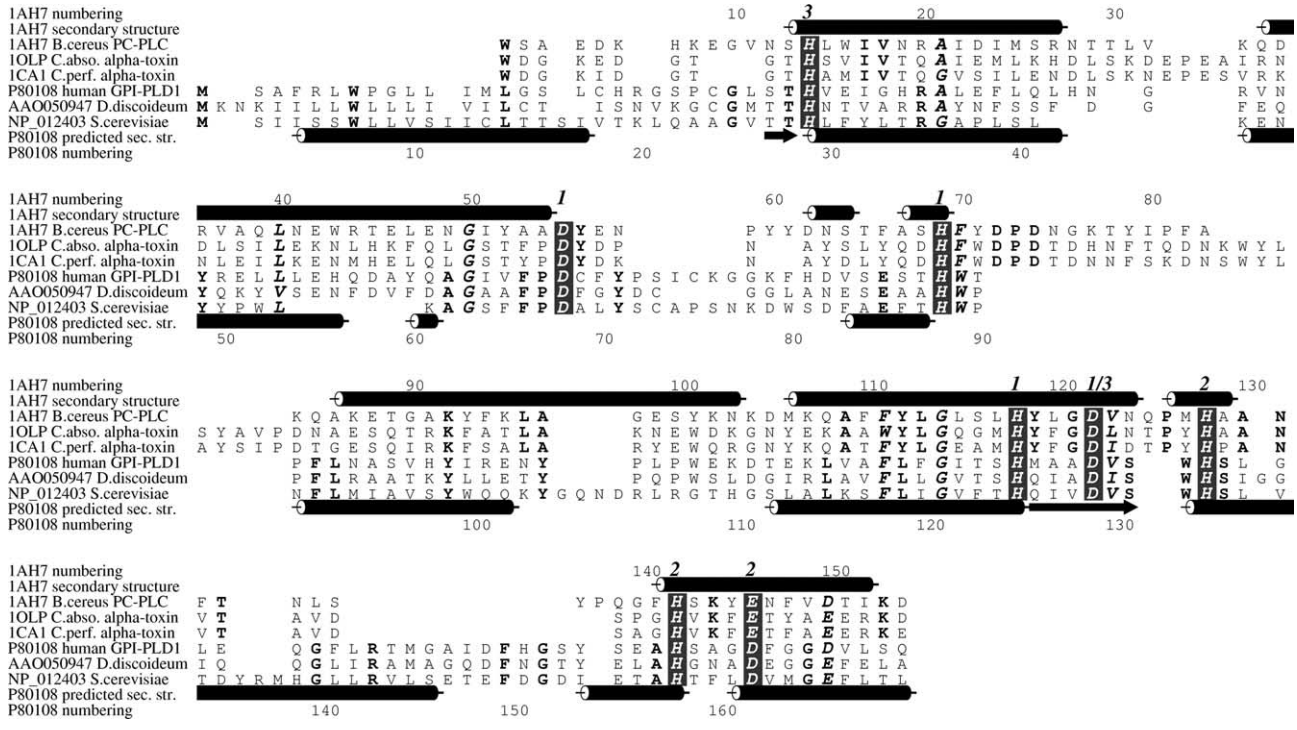

Fig. 1. Sequence alignment of bacterial PLCs and GPI-PLDs. Human GPI-PLD1 (with Swissprot accession number), and its *S. cerevisiae* and *D. discoideum* homologues (each with Genpept accession number), are aligned with the three bacterial PLCs (labelled with PDB code, species and name) used as templates in model construction. The alignment of the PLCs is their structural alignment. The alignment between PLCs and PLDs is the one used to make the final GPI-PLD model. Identities within the sets of PLCs or PLDs are shown in bold face. Residues well conserved between all sequences are additionally italicized. Numbering and actual secondary structure of *B. cereus* PC-PLC are shown above the alignment with numbering and predicted secondary structure of human GPI-PLD1 shown below the alignment. Cylinders represent α- or 3_{10}-helices and arrows are used for β-strands. The presumed zinc ligating residues are shown as white on black and labelled according to which the three bound zincs are predicted to be ligated. The figure was made with ALSCRIPT [66].

other homologues were obtained, from *Saccharomyces cerevisiae* (accession NP_012403) and *D. discoideum* (accession AAO50947), sharing 30% and 20% pairwise sequence identity, respectively, with the human enzyme. In the catalytic region that comprises the first 275 residues or so [18], the pairwise sequence identity with the human enzyme is higher at 36% and 25%, respectively. In each case, the sequences derive from the respective genome projects [51,52] and there is no information regarding function, although the existence of ESTs corresponding to the *D. discoideum* sequence [53] is proof of its expression. This seems to be the first observation of close relatives of GPI-PLDs, likely sharing the same catalytic activity, outside higher eukaryotes. Although not the main focus of this paper, it was curious to note that the calcium-binding motifs within the β-propeller domain shared by the homologous β-propeller domains of integrins and mammalian GPI-PLDs are absent in the *S. cerevisiae* sequence. Further PSI-BLAST searches failed to reveal more distantly related homologues.

More distant homologies, that might offer a route to better understanding the structure–function relationship in GPI-PLDs, were sought through use of the Meta-server, a resource enabling easy analysis of protein sequences by a battery of more sensitive sequence comparisons and fold recognition methods [30]. The catalytic site of GPI-PLDs is known to reside in the N-terminal part of the protein [18]. Searches in the Smart database [28] showed that the integrin-like repeats (smart00191) forming the C-terminal, calcium-binding, non-catalytic β-propeller [15] commenced at residue 378. We therefore submitted the N-terminal 377 residues of human GPI-PLD1 to the Meta-server. Several analyses, including those based on sequence alone [31], produced top scores for the structures of phospholipases C from *B. cereus* and from clostridia species (these also known as the α-toxins). These alignments extended from near the N-terminus to around residue 260, corresponding well with the experimentally determined size for a catalytically active GPI-PLD fragment [18]. In several cases, different putative structural matches scored almost as highly, and other fold recognition methods placed the PLCs in lower positions. Nevertheless, the correctness of the structural correspondence between the GIP-PLDs and the bacterial PLCs was assured by several considerations. First of these was the results of the consensus methods and their comparison to the results of ongoing benchmarking efforts [54]. The latest completed benchmarking results suggest that the Shotgun on 3 consensus method [35] currently distinguishes best between true and false negatives. Human GPI-PLD1 N-terminal region scored 72 by this method, well above the 55 scored by the best-scoring incorrectly assigned fold. Second, there existed an obvious functional similarity between the probe GPI-PLD sequence and the top-scoring PLCs. Furthermore, in the initial fold-recognition alignments six out of nine of the residues ligating the three catalytic site zinc ions of the PLC structures were conserved in the GPI-PLD1 sequence. A good correspondence between the known secondary structure of the PLCs and the predicted secondary structure of human GPI-PLD1 (Fig. 1) was also noted, particularly in the first 150 residues or so. Finally, GPI-PLDs have been experimentally proven to bind catalytically essential zinc [16], although the estimated 10 Zn ions per monomer are well in excess of the three zincs bound per monomer of bacterial PLC.

3.2. Model building

The structures of the PLCs from *B. cereus* [41], *C. absonum* [42] and *C. perfringens* [43] all scored comparably well and so were all used as templates in model building of human GIP-PLD1. Although known to resemble bacterial PLCs [19], P1 nuclease was not present in the fold recognition results and was therefore not considered as a template. Successive structural determinations have highlighted the existence of open (active; 3 bound zinc ions) and closed (inactive; 2 bound zinc ions) forms of clostridial PLC [42]. However, the two loops responsible for the largest structural differences between the two forms (residues 63–87 and 135–149 in *C. perfringens* numbering corresponding approximately to stretches 63–83 and 127–142 in *B. cereus* numbering – see Fig. 1) are very different in the PLDs (Fig. 1). The PLDs are substantially shorter in the region of the first loop but conversely have a large insertion in the vicinity of the second. In any case, since we wished to understand catalysis and substrate binding in the PLDs, the open, active forms of the clostridial enzymes (PDB codes 1olp [42] and 1ca1 [43]) were used as templates, along with the *B. cereus* PLC structure (PDB code 1ah7; [41]).

Although only six of the nine zinc-ligating residues of the PLCs were conserved in the initial, fold recognition-derived alignments, examination of the PLD sequence showed that suitable residues, conserved throughout the PLDs, could be aligned with all the ligating residues in order to maintain all three zinc binding sites (Fig. 1). The only difference between the two sets of ligating residues lies at position 146 (*B. cereus* PLC numbering) where the Glu that contributes to coordination of the second zinc ion in the PLCs is conservatively replaced by an Asp in the PLDs. This position is also occupied by Asp in the P1 nucleases that are other known distant relatives of bacterial PLCs [19] that maintain the trimetal site. Since relatively few residues are absolutely conserved between the closely related mammalian GPI-PLDs and the more distant *S. cerevisiae* and *D. discoideum* homologues, these alignment shifts could be made with confidence. They also had favourable effects on the correspondence between predicted PLD secondary structure and actual PLC secondary structure. Absolute conservation between bacterial PLCs and GPI-PLDs was also observed for Gly50 and Gly114. Structural inspection revealed the explanation; these two residues lie at the heart of the interface between their respective helices such that the presence of any side chain would disrupt the helical packing.

Given that alignment errors are one of the major sources of model inaccuracies, the conservation of metal ligands and Gly residues was invaluable, providing reliable markers throughout the sequence alignment. Nevertheless, significant sources of alignment uncertainty remained and were addressed through construction and evaluation of multiple models for alignment variants. One key issue was the alignment at the N-terminus where, in the PLCs, the backbone of the very N-terminal residue, a conserved Trp1, makes two interactions with the third bound zinc. Since PLDs are significantly longer, and not well conserved at their N-terminus (Fig. 1), it seemed unlikely that Met1 of human GPI-PLD1 would take the place of PLC Trp1. Three other options were therefore tested through model construction and analysis with either human GPI-PLD1 His19 (as in the original fold recognition alignments), conserved Leu14 or conserved Trp7 aligned with Trp1 of the PLCs. Profile analysis of the various models with PROSA II [44] and VERIFY_3D [45] strongly favoured the

choice of Leu14, the alternatives each leading to much worse profiles for the N-terminal region. For example, by PROSA II, employing a smoothing window of 11 residues, the mean profile values up to residue 30 for the best overall scoring models were 0.94, 0.13 and 0.35, for the alignments with PLC Trp1 of GPI-PLD1 His19, Leu14 and Trp7, respectively. Circumstantial support for the choice of Leu14 came from the position of conserved Trp132, replacing smaller residues in PLCs. The side chain occupies part of the volume occupied by Trp1 in the templates. The twin replacements of PLC Trp1 with Leu in the PLDs and PLC residue 127 with Trp may therefore be seen as compensatory in terms of volume. Since no template was available for the first 13 residues, this portion was not included in the modelling. Nor could any satisfactory structure be obtained for the 12 residue insertion in PLDs (residues 142–153 in human GPI-PLD1) relative to PLC templates, which was also therefore omitted from the final model.

Although the fold recognition methods aligned GPI-PLD sequence to the complete α-helical PLC domain, a sharp boundary appeared during the modelling process between PLD residues 1–170 (Fig. 1) and residues 171–260. While PROSA II and VERIFY_3D profiles were generally favourable for the former portion, the C-terminal section was characterized by positive PROSA II profiles and negative VERIFY_3D profiles. Furthermore, two β-strands were predicted in the region 1–170 (Fig. 1), known to be essentially all α-helical by reliable fold assignment, within the limits expected by the accuracy of the technique of secondary structure prediction. However, five were predicted for the following smaller section comprising residues 171–260. Allied to a complete lack of any conservation across PLCs and PLDs, these considerations led to the conclusion that the region, GPI-PLD residues 171–260, must differ sufficiently from the PLCs such that the latter no longer constitute useful templates for the region. The final model therefore comprises just the region 1–170 which, nevertheless, contains all catalytic and substrate binding residues.

3.3. Implications for substrate binding and catalysis in GPI-PLDs

Despite the obvious functional similarity between the GPI-PLDs and the bacterial PLCs, they differ in the location of substrate cleavage. PLCs cleave phosphatidylcholine, as an example, to produce phosphorylcholine and diacylglycerol: the products of cleavage of the same substrate by PLD would be choline and diacylglycerol-3-phosphate. How is the PLC vs. PLD difference explained in this case? One possibility would be that the phospho group binds differently so that a conserved catalytic machinery attacks a different side of the phospho group. However, in the model of the catalytic site of human GPI-PLD1, all zinc ligands present in the template PLC structures are fully conserved, with the exception of a single conservative replacement of Glu with Asp (Fig. 1). Structures of bacterial PLCs with bound substrate analogues have revealed that the phospho group is bound by all three zinc ions [55]. From their conservation of the zinc sites, it therefore seems likely that no major differences in substrate phospho group position exist between GPI-PLDs and PLCs. Is it, therefore, that substrate now binds in the opposite orientation, the diacylglycerol portion of PLD substrate binding in the region occupied by the phophocholine moiety in PLCs? While there are sequence differences that could be interpreted as increasing hydrophobicity of the GPI-PLD1 pocket, within the area that binds phosphocholine in PLC, in a way consistent with diacylglycerol binding, such analysis may not be reliable in this case; the homology with templates is very distant and the shape and nature of the binding site could be changed significantly by the missing pieces of the model, the very N-terminus and the 12 residue deletion. Without stronger evidence, the more conservative prediction that the substrate binds similarly overall to both PLCs and PLDs should stand.

Could the explanation for the PLC vs. PLD difference therefore lie in differences in catalytic machinery? Since the visualization of the trimetal centres in PLCs and the distantly homologous P1 nuclease [19], various catalytic mechanisms have been proposed [55–57]. Kinetic data have been obtained that favour activation of a nucleophilic water molecule by a protein side chain rather than one of the zinc ions [58]. Site-directed mutagenesis of conserved acidic residues of the *B. cereus* PLC catalytic site has also been reported [58,59] which strongly implicated Asp55 (*B. cereus* PLC numbering), conserved between PLCs and their distant relative *Penicillium citrinum* P1 nuclease, as being responsible for activation of the water molecule which then nucleophilically attacks the substrate (Fig. 2(a) and (c)). The nucleophilic water molecule has been tentatively identified as one observed crystallographically in complexes of *B. cereus* PLC, both wild-type [55] and Asp55 Asn mutant [57], with substrate analogues (Fig. 2(a)). The water in question is also hydrogen-bonded to Glu4 which is absent in the clostridial PLCs, probably being substituted in these enzymes by conserved Asp2. In contrast, the GPI-PLDs lack acidic side chains in the area, other than Asp68 that corresponds to *B. cereus* PLC Asp55. However, examination of the model reveals a new conserved acidic residue near to Asp68 in the form of Glu85 (Figs. 1 and 2(b)). This is a dramatically non-conservative replacement for Phe or Tyr in the bacterial PLCs (Fig. 1). The conservation of Glu85, reliably placed in the catalytic site through its three residue separation from a zinc ligand (Fig. 1) and by the overlap between predicted GPI-PLD α-helices and a 3_{10} helix in the templates (Fig. 1), is suggestive of an important functional role. This seems unlikely to be direct binding of substrate, since it is predicted to be positioned near the negatively charged phospho group (Fig. 2(b)); the nearest positively charged region of the GPI substrate is the amino group of the glucosamine which cannot bind so close to the site of catalysis. Instead, the position of Glu85 is consistent with its participation in hydrogen bonding, in conjunction with Asp68, to a modelled catalytic water molecule (Fig. 2(b) and (d)). In simple terms, it would therefore functionally substitute Glu4 (*B. cereus*) or Asp2 (clostridia) of the PLCs. Crucially, however, the water molecule would be differently placed with respect to the phospho group of the substrate (Fig. 2) and positioned so that its nucleophilic attack on substrate would lead to PLD-type cleavage (Fig. 2(d)) rather than PLC-type cleavage (Fig. 2(c)). Thus, the prediction is that Asp55/68 conserved between the PLCs and PLDs activates water for subsequent nucleophilic attack. The difference between the two phospholipase classes lies in other conserved acidic residues which contribute to suitably positioning that water molecule. In PLCs, the relevant water molecule may well have been visualized crystallographically [55,57]. In the case of PLDs only a model is available, but it is enough to suggest that if Glu85 contributes to binding of the

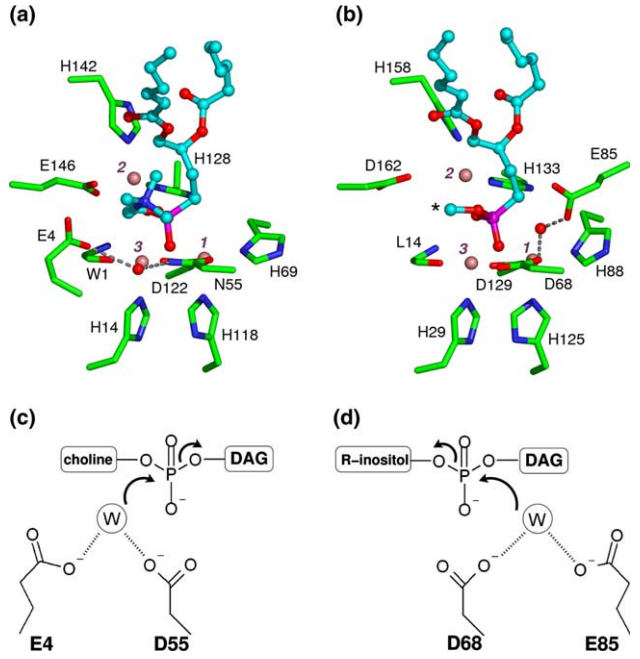

Fig. 2. Comparison of predicted catalytic mechanisms in bacterial PLCs and GPI-PLDs. Comparison of the catalytic sites of (a) *B. cereus* PLC bound to phosphonate substrate analogue (PDB code 1p6d [57]) and (b) human GPI-PLD1 (modelled). In each case, zinc ions (labelled 1–3 according to the numbering of [57]) are shown as pink spheres and the predicted nucleophilic water molecules as red sphere, bound by hydrogen bonds (grey dotted lines). Protein residues binding to either zinc or water are shown as sticks and are labelled. The substrate analogue is shown as ball and stick. In (b) small adjustments have been made to the substrate analogue position in order to ameliorate minor steric clashes with the modelled water molecule. The choline portion of the analogue, not present in the substrates of GPI-PLD, is not shown in (b): an asterisk marks the position of attachment of the inositol moiety. A schematic comparison of bacterial PLCs and GPI-PLDs showing how differently the position of nucleophilic water molecules could lead to different modes of phospholipase activity is shown in (c) *B. cereus* PLC and (d) human GPI-PLD1. DAG stands for diacylglycerol and R for the remainder of the GPI anchor.

nucleophilic water molecule then the water must be positioned so as to lead to PLD-type substrate hydrolysis. Glu85 is thus highlighted as a prime target for study by site-directed mutagenesis. The results of such study would provide a valuable test of the proposed mechanism which, at this stage, is purely hypothetical.

3.4. Prospects for drug design

The trimetal architecture shared by the catalytic sites of bacterial PLCs and GPI-PLDs and relatives is unusual. The current PFAM [29] entry for the bacterial PLC family (PF00882) shows 32 sequences, exclusively bacterial. The S1–P1 nuclease entry (PF02265) shows a distribution that includes 36 eukaryotic sequences, none mammalian, and 7 bacteria. In order to examine further whether the GPI-PLD is the only human representative of this architecture, we formulated a motif, based on the most conserved positions, that would retrieve all known members of the bacterial PLCs, the P1 nucleases and the GPI-PLDs and relatives. Accordingly, a motif search was carried out in the Swissprot database [60] using the following motif [LIFVMW]-x(4)-H-x(3)-D-[LIVMFTCA]-x(2,4)-H-x(8,39)-H-x(3)-[DE], which covers ligands of all three zinc atoms and is represented in *B. cereus* PLC, for example, by residues 117–146. In humans, this motif matches, along with the GPI_PLDs four other proteins, all, from their annotations, unlikely to have the trimetal catalytic sites – a calcium channel subunit (Swissprot entry CCAB_HUMAN), a transcription factor (FXJ1_HUMAN), a Ras activating protein (SNGP_HUMAN) and a zinc finger protein (ZN92_HUMAN). Of course, Swissprot does not include all human proteins, but the motif search results support the idea that the GPI-PLDs are the only human proteins with the PLC-PLD-P1 nuclease class trimetal catalytic site.

The probable uniqueness of the catalytic site architecture in humans has important consequences for the design of GPI-PLD inhibitors for use as treatments for cancer [12,13] or, possibly, atherosclerosis [14]. As the recently designed *B. cereus* PLC inhibitors [61] show, metal ions offer particular opportunities for inhibitor design. Hydroxamic acids derivatives are particularly useful for zinc ligation [62]. Given the similarities described above, it is highly likely that the *B. cereus* PLC inhibitors described [61] will be active towards GPI-PLDs. Other classes of bacterial PLC inhibitors, not specifically those targeted at bound metal, have also been developed and may be useful starting points for GPI-PLD inhibitor design [63]. Indeed, in one case where they have been tested on GPI-PLD, potent inhibition was observed [13]. Knowledge of the catalytic site architecture of GPI-PLDs will certainly help optimize these and the limited number of other known GPI-PLD inhibitors [64]. At present sub-micromolar inhibition of GPI-PLD, an important drug target [12–14], has not been reported. One obvious route might be the combination of the hydroxamic acid derivatives, targeted to the bound metal, with features of product analogue which occupy other parts of the substrate binding site.

4. Conclusions

The biological and medical importance of GPI-PLD is becoming ever more apparent so that the lack of structural information for the enzyme will increasingly be felt. Until a crystal structure is available the distant homology with bacterial PLCs reported here, and the catalytic site model that could thereby be constructed, can be used to guide experimental work. Specifically, the prediction that GPI-PLD contains a zinc trimetal centre, once confirmed experimentally, should facilitate inhibitor design, while the catalytic mechanism tentatively proposed here can be readily tested by site-directed mutagenesis. The GPI-PLD homologues in *S. cerevisiae* and *D. discoideum* are interesting subjects for future study given the different role of GPI-proteins in the former [65] and the ease of genetic manipulation of the latter.

References

[1] Ferguson, M.A. (1999) J. Cell. Sci. 112, 2799–2809.
[2] Homans, S.W., Ferguson, M.A., Dwek, R.A., Rademacher, T.W., Anand, R. and Williams, A.F. (1988) Nature 333, 269–272.
[3] Ferguson, M.A., Homans, S.W., Dwek, R.A. and Rademacher, T.W. (1988) Science 239, 753–759.
[4] Benghezal, M., Benachour, A., Rusconi, S., Aebi, M. and Conzelmann, A. (1996) EMBO J. 15, 6575–6583.

[5] Nozaki, M., Ohishi, K., Yamada, N., Kinoshita, T., Nagy, A. and Takeda, J. (1999) Lab. Invest. 79, 293–299.
[6] Hereld, D., Hart, G.W. and Englund, P.T. (1988) Proc. Natl. Acad. Sci. USA 85, 8914–8918.
[7] Davitz, M.A., Hereld, D., Shak, S., Krakow, J., Englund, P.T. and Nussenzweig, V. (1987) Science 238, 81–84.
[8] Scallon, B.J., Fung, W.J., Tsang, T.C., Li, S., Kado-Fong, H., Huang, K.S. and Kochan, J.P. (1991) Science 252, 446–448.
[9] Metz, C.N., Brunner, G., Choi-Muira, N.H., Nguyen, H., Gabrilove, J., Caras, I.W., Altszuler, N., Rifkin, D.B., Wilson, E.L. and Davitz, M.A. (1994) EMBO J. 13, 1741–1751.
[10] Mann, K.J., Hepworth, M.R., Raikwar, N.S., Deeg, M.A. and Sevlever, D. (2004) Biochem. J. 378, 641–648.
[11] Tsujioka, H., Misumi, Y., Takami, N., Ikehara, Y. and Tujioka, H. (1998) Biochem. Biophys. Res. Commun. 251, 737–743.
[12] Xiaotong, H., Hannocks, M.J., Hampson, I. and Brunner, G. (2002) Clin. Exp. Metastasis 19, 199–291.
[13] Brunner, G., Zalkow, L., Burgess, E., Rifkin, D.B., Wilson, E.L., Gruszecka-Kowalik, E. and Powis, G. (1996) Anticancer Res. 16, 2513–2516.
[14] Kurtz, T.A., Fineberg, N.S., Considine, R.V. and Deeg, M.A. (2004) Metabolism 53, 138–139.
[15] Springer, T.A. (1997) Proc. Natl. Acad. Sci. USA 94, 65–72.
[16] Li, J.Y., Hollfelder, K., Huang, K.S. and Low, M.G. (1994) J. Biol. Chem. 269, 28963–28971.
[17] Naylor, C.E., Eaton, J.T., Howells, A., Justin, N., Moss, D.S., Titball, R.W. and Basak, A.K. (1998) Nat. Struct. Biol. 5, 738–746.
[18] Heller, M., Butikofer, P. and Brodbeck, U. (1994) Eur. J. Biochem. 224, 823–833.
[19] Volbeda, A., Lahm, A., Sakiyama, F. and Suck, D. (1991) EMBO J. 10, 1607–1618.
[20] Stuckey, J.A. and Dixon, J.E. (1999) Nat. Struct. Biol. 6, 278–284.
[21] Stonehouse, M.J., Cota-Gomez, A., Parker, S.K., Martin, W.E., Hankin, J.A., Murphy, R.C., Chen, W., Lim, K.B., Hackett, M., Vasil, A.I., and Vasil, M.L. (2002) Mol. Microbiol. 46, 661–676.
[22] Zambonelli, C. and Roberts, M.F. (2003) J. Biol. Chem. 278, 13706–13711.
[23] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[24] Edgar, R.C. (2004) Nucleic Acids Res. 32, 1792–1797.
[25] Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004) Bioinformatics 20, 426–427.
[26] Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) Bioinformatics 15, 305–308.
[27] Hulo, N., Sigrist, C.J., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. and Bairoch, A. (2004) Nucleic Acids Res. 32, D134–D137.
[28] Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P. and Bork, P. (2004) Nucleic Acids Res. 32, D142–D144.
[29] Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004) Nucleic Acids Res. 32, D138–D141.
[30] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[31] Rychlewski, L., Jaroszewski, L., Li, W. and Godzik, A. (2000) Protein Sci. 9, 232–241.
[32] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J. (2000) J. Mol. Biol. 299, 499–520.
[33] Jones, D.T. (1999) J. Mol. Biol. 287, 797–815.
[34] Fischer, D. (2000) in: Pacific Symposium on Biocomputing (Altman, R.B., Dunker, A.K., Hunter, Lauderdale, K. and Klein, T.E., Eds.), pp. 119–130, World Scientific, Singapore.
[35] Fischer, D. (2003) Proteins 51, 434–441.
[36] Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) Bioinformatics 19, 1015–1018.
[37] Lundstrom, J., Rychlewski, L., Bujnicki, J. and Elofsson, A. (2001) Protein Sci. 10, 2354–2362.
[38] Jones, D.T. (1999) J. Mol. Biol. 292, 195–202.
[39] Tsang, T.C., Fung, W.-J.C., Levine, J., Metz, C.N., Davitz, M.A., Burns, D.K., Huang, K.-S. and Kochan, J.P. (1992) FASEB J. 6, A1922–A1922.
[40] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[41] Hough, E., Hansen, L.K., Birknes, B., Jynge, K., Hansen, S., Hordvik, A., Little, C., Dodson, E. and Derewenda, Z. (1989) Nature 338, 357–360.
[42] Clark, G.C., Briggs, D.C., Karasawa, T., Wang, X., Cole, A.R., Maegawa, T., Jayasekera, P.N., Naylor, C.E., Miller, J., Moss, D.S., Nakamura, S., Basak, A.K. and Titball, R.W. (2003) J. Mol. Biol. 333, 759–769.
[43] Naylor, C.E., Eaton, J.T., Howells, A., Justin, N., Moss, D.S., Titball, R.W. and Basak, A.K. (1998) Nat. Struct. Biol. 5, 738–746.
[44] Sippl, M.J. (1993) Proteins 17, 355–362.
[45] Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Nature 356, 83–85.
[46] Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993) J. Appl. Crystallogr. 26, 283–290.
[47] Kleywegt, G.J. (1996) Acta Cryst. D 52, 842–857.
[48] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Cryst. A 47, 110–119.
[49] DeLano, W.L., 2004. Available from http://www.pymol.org.
[50] Frishman, D. and Argos, P. (1995) Proteins 23, 566–579.
[51] Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Science 274, 563–567.
[52] Glockner, G., Eichinger, L., Szafranski, K., Pachebat, J.A., Bankier, A.T, Dear, P.H., Lehmann, R., Baumgart, C., Parra, G., Abril, J.F., Guigo, R., Kumpf, K., Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A., and Noegel, A.A. (2002) Dictyostelium genome sequencing consortium. Nature 418, 79–85.
[53] Urushihara, H., Morio, T., Saito, T., Kohara, Y., Koriki, E., Ochiai, H., Maeda, M., Williams, J.G., Takeuchi, I. and Tanaka, Y. (2004) Nucleic Acids Res. 32, 1647–1653.
[54] Rychlewski, L., Fischer, D. and Elofsson, A. (2003) Proteins 53, 542–547.
[55] Hansen, S., Hough, E., Svensson, L.A., Wong, Y.L. and Martin, S.F. (1993) J. Mol. Biol. 234, 179–187.
[56] Sundell, S., Hansen, S. and Hough, E. (1994) Protein Eng. 7, 571–577.
[57] Antikainen, N.M., Monzingo, A.F., Franklin, C.L., Robertus, J.D. and Martin, S.F. (2003) Arch. Biochem. Biophys. 417, 81–86.
[58] Martin, S.F. and Hergenrother, P.J. (1998) Biochemistry 37, 5755–5760.
[59] Martin, S.F., Spaller, M.R. and Hergenrother, P.J. (1996) Biochemistry 35, 12970–12977.
[60] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) Nucleic Acids Res. 31, 365–370.
[61] Martin, S.F., Follows, B.C., Hergenrother, P.J. and Franklin, C.L. (2000) J. Org. Chem. 65, 4509–4514.
[62] Babine, R.E. and Bender, S.L. (1997) Chem. Rev. 97, 1359–1472.
[63] Roberts, M.F., Wu, Y., Zhou, C., Geng, D. and Tan, C. (1996) Adv. Enzyme Regul. 36, 57–71.
[64] Low, M.G. and Stutz, P. (1999) Arch. Biochem. Biophys. 1371, 332–339.
[65] de Nobel, H. and Lipke, P.N. (1994) Trends Cell. Biol. 4, 42–45.
[66] Barton, G.J. (1993) Protein Eng. 6, 37–40.