Recognizing the Pleckstrin homology domain fold in mammalian phospholipase D using hidden Markov models

Pamela G. Holbrook[a,*], V. Geetha[b], Michael A. Beaven[a], Peter J. Munson[b]

[a] *Laboratory of Molecular Immunology, National Heart Lung and Blood Institute, National Institutes of Health, Building 10, Room 8N-109, Bethesda, MD 20892, USA*

[b] *Analytical Biostatistics Section, Mathematical and Statistical Computing Laboratory, Center for Information Technology, National Institutes of Health, Building 12A, Room 2041, Bethesda, MD 20892, USA*

Received 6 January 1999; received in revised form 26 February 1999

Abstract Phospholipase D was first described in plant tissue but has recently been shown to occur in mammalian cells where it is activated by cell surface receptors. Its mode of activation by receptors in unclear. Biochemical studies suggest that it may occur downstream of other effector proteins and that small GTP-dependent regulatory proteins may be involved. The sequence in a non-designated region of mammalian phospholipase D1 and 2 shows similarity to a structural domain that is present in signalling proteins that are regulated by protein kinases or heterotrimeric G-proteins. Mammalian phospholipase D has structural similarities with other lipid signalling phospholipases and thus may be regulated by receptors in an analogous fashion.

Key words: Phospholipase D; Pleckstrin homology domain; Hidden Markov models; Phosphatidylinositol-3,4,5-trisphosphate binding protein; β-Adrenergic receptor kinase; Lipid signalling

1. Introduction

Mammalian phospholipase (PL)D is activated in a wide variety of cells by receptor subtypes that also regulate increases in intracellular free calcium [1]. The activated enzyme may generate considerable quantities of a signalling molecule since it specifically hydrolyses phosphatidylcholine (PC) [2], the most abundant phospholipid in cell bilayers. Current models place PLD downstream of the activation of other effector proteins [1,3], a notion supported by the reported absence of recognizable signalling domains from its known sequence structures [4].

The amino acid sequence of two different mammalian PLDs have been deduced [4]. Although originally reported to lack pleckstrin homology (PH) domains [4,5], a recent report indicates the contrary [6]. PLD is stimulated by polyphosphatidylinositols [5,7] which is consistent with the presence of a PH domain [8]. We find that regions of mammalian PLD1 and PLD2 show a substantial sequence similarity to the PH domain of a recently discovered phosphatidylinositol-3,4,5-trisphosphate binding protein (PIP3BP) [9]. Structures of several PH domains have been solved and are similar despite a limited amino acid sequence identity [10,11]. A structure-based hidden Markov model (HMM) built from these solved structures recognizes the PH domain fold in PLD1 and PLD2. Our analysis demonstrates that the predicted secondary structure of the PLD domains includes the seven β-strands and a C-terminal α-helix characteristic of PH domains as well as five conserved amino acids important for interactions with polyphosphatidylinositols. In addition, the PLD PH domains include an unusually large loop between the first and second β-strands that may be a unique site of regulation by divalent cations.

2. Methods

2.1. Assessing homology by pairwise sequence comparison

The recently refined Basic Local Alignment Search Tool (PSI-BLAST) program [12] was used to compare rat PLD2 to sequences in the non-redundant database. PSI-BLAST is often sensitive to weak similarities between proteins that may nonetheless be biologically relevant.

2.2. Assessing homology using family-based models

Family-based HMMs were invoked as an alternative to pairwise sequence comparisons. A HMM is a statistical model similar to a sequence profile but may be estimated from unaligned sequences [13]. HMMs may be based either on the sequence of amino acids [13] or of secondary structure states [14].

2.3. Sequence-based HMM

A broad collection of PH domain sequences was assembled from the 41 sequences in the PH domain 'seed alignment' from the PFAM [15] database together with two additional sequences that were identified by an independent PSI-BLAST search with PIP3BP as the query. A HMM was built using the SAM suite version 1.3.1 [13] with default parameter settings.

2.4. Secondary structure-based HMM

A secondary structure-based HMM was constructed using the method FORESST [14] from the known secondary structure sequences of 10 PH domains designated in the SCOP [16] database with the PDB identifiers 1mai, 1btkA, 1bak, 1btn, 1pms, 1mph, 1dro, 1pls, 1dynA and 1awe. The secondary structures of PLD1 and PLD2 were predicted using the prediction methods PredictProtein (PHD) [17] and Quadratic-Logistic (QL) [18].

2.5. Z-scores

To evaluate the ability of a HMM to recognize its family members, Z-scores were calculated from log-odds scores [19]. The log-odds scores were obtained from the fit of both PLD domains to the sequence-based and secondary structure-based HMMs. The log-odds score is a measure of how well a given family-based model fits the query sequence as compared to the fit of a 'null' model. These scores were converted to Z-scores which, if the query is not related to the family, should be distributed as a standard normal (gaussian) variable, with zero mean and unit S.D.. Accordingly, Z-scores greater than 3.0 are expected to occur no more than 0.13% of the time. Z-scores are calculated here by subtracting from the query log-odds score the mean log-odds score for a collection of unrelated proteins (after length correction) and dividing the result by the S.D. of those scores. Details of this approach are explored elsewhere [19].

*Corresponding author. Fax: (1) (301) 402 0171.
E-mail: holbroop@fido.nhlbi.nih.gov

PII: S0014-5793(99)00366-X

3. Results and discussion

A pairwise sequence comparison of rat PLD2 with the non-redundant database using PSI-BLAST [12] reveals similarity with a 142 amino acid sequence fragment (216–357) of bovine PIP3BP that includes the sequence (253–355) identified as its C-terminal PH domain [9]. The aligning sequence of PLD2 shows 27% sequence identity (29/108 residues) which is consistent with the notion that the two peptides are homologous [20] although chance alignments with a high identity cannot be ruled out [21]. The regions of PLD2 and PLD1 that align with PIP3BP 253–355 have recognizable features of the first identified PH domains including a glycine (G) as one of the first few amino acids and the nearly invariant tryptophan (W) in the C-terminal α-helix region [11]. The remaining amino acids are characteristically rich in large hydrophobic and positively charged amino acids (Fig. 1).

Because the pairwise sequence alignment between the PIP3BP PH domain and PLD sequences was relatively weak, we turned to family-based models to test whether the aligning regions of PLD1 and 2 are members of the PH domain family. By considering the diverse set forming a protein family, related sequences may be found that are not evident using pairwise sequence comparison techniques alone. The two PH domains of PIP3BP and the putative domains of PLD1 and PLD2 were tested against a sequence-based HMM. The Z-scores indicate that the two PLD sequences (Table 1, column 2) are reliably recognized as members of the PH domain family. Two of the sequences, namely that of β-adrenergic receptor kinase (β-ARK) and pleckstrin gave very high Z-scores as expected since they were included in the training set for the HMM of the PH domain family of proteins.

The secondary structure predictions for the PLD1 and PLD2 domains were compared against a secondary structure-based HMM built using FORESST [14]. Z-scores for the PLD domains were similar in magnitude to those for the predicted sequence of secondary structure states of the PH domains of PIP3BP and pleckstrin (Table 1, column 3). Z-scores of this magnitude or larger would very unlikely occur by chance ($P < 0.002$). Interestingly, the secondary structure prediction for the β-ARK sequence could not be definitively detected as a PH domain. However, when the nuclear magnetic resonance (NMR)-determined secondary structure for β-ARK [22] was tested against the model, the Z-score increased to 4.67 (Table 1). This suggests that the failure to recognize β-ARK as a PH domain was due to a poor quality of prediction, not to a failure of the HMM. Z-scores for the PLD1 and PLD2 random fragments taken from outside the PH domain region were not significant when tested against the sequence-based or structure-based models. The conclusion that the PLD sequences are indeed PH domains is also supported by results from the 3D-1D threading approach of UCLA-DOE which found classic PH domain proteins with top ranking scores [23].

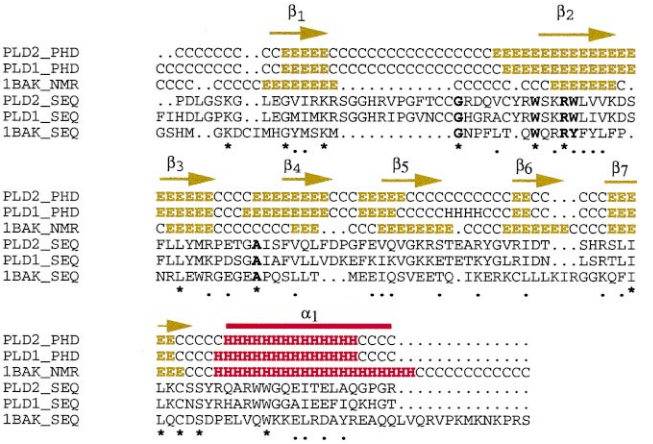

Fig. 1. Alignment of the secondary structure states and amino acid sequence of putative mammalian PLD1, PLD2 and c. elegans PH domains with the NMR structure of β-ARK domain [22]. The three possible secondary structure states are: strand (E), helix (H) and coil (C). The seven β-strands (in green) and 1 α-helix (red) of β-ARK (1bak) are indicated as are the corresponding regions of PLD1 and PLD2. Secondary structure states of the PLD1 and PLD2 domains were predicted using the PHD method [17]. Sequence identities are boxed while sequence similarity together with sequence identity of β-ARK with either of the PLDs is indicated with a dot. The five functionally significant residues that are conserved in β-ARK and PLD are indicated in bold.

The PLC βs and β-ARK both have PH domains and are believed to be activated by receptors via interactions with heterotrimeric G-proteins [24,25]. Structures for the PLC-β PH domains have not been solved although the solution struc-

Table 1
Z-scores for comparison of rPLD1 and rPLD2 to two HMMs of the PH domain family

Protein	Z-scores for sequence-based HMM	Z-scores for structure-based HMM[c]
Experimental		
rPLD1-PH	2.72	2.94
rPLD2-PH	4.01	3.74
PIP3BP-NPH	8.93	3.17
PIP3BP-CPH	12.08	4.52
Positive control[a]		
Pleck-CPH	25.84	3.35 (5.41)[d]
β-ARK1-PH	22.46	0.97 (4.67)[d]
Negative control[b]		
rPLD1-random	−0.14	0.21
rPLD2-random	−0.43	−1.4

[a]Structures for the PH domains of Pleck-C and β-ARK have been determined experimentally and are included as positive controls.
[b]The random PLD sequences were of the same size as the PH domain but chosen from outside the PH domain region of PLD1 and 2 and are included as negative controls.
[c]Secondary structures for the structure-based HMM were predicted from amino acid sequences by the PHD ([17]) method.
[d]Z-scores for the secondary structure-based HMM of Pleck-C and β-ARK are also provided for their experimentally-derived secondary structure states. Z-scores greater than 3.0 are expected to occur no more than 0.13% of the time ($P < 0.0013$).

ture of the PH domain of β-ARK has been determined by NMR [22]. Fig. 1 compares the putative PH domains of PLD1 and PLD2 with that of β-ARK. The predicted sequence of secondary structure states of the PLD1 and PLD2 domains approximates the seven β-strands and C-terminal α-helix of the β-ARK PH domain except that the PHD secondary structure prediction method [17] did not detect a break between the putative β2- and β3-strands of the PLD sequences. A different secondary structure prediction method, the QL method [18], did detect a break in this region. The variable loop connecting the β1- and β2-strands of the putative PLD PH domains is unusually large and contains the motif H(Xn)CC(Xn)C that may bind divalent cations like zinc [26] or calcium.

It is striking that many amino acid residues of the PLD sequences are identical or similar in β-ARK (Fig. 1). The functional significance of five of these residues, G-569, W-576, R-579, Y-580 and A-596 (Fig. 2), is illustrated by the fact that they display 15N and 1H spectral perturbations when the β-ARK PH domain is bound by inositol-1,4,5-trisphosphate (IP3) [22]. This is likely to be due to interactions of positively-charged amino acids in the β1/β2-loop, β2-strand and β3/β4-loop with negatively charged phosphate groups of IP3 analogous to PLCδ [11,27]. The fact that positively-charged amino acids, arginine (R) and lysine (K), are also observed in the putative PLD β1/β2-loop, β2-strand and β3/β4-loop (Fig. 1) provides a structural basis for the observed ability of exogenous PIP2 and PIP3 to stimulate the PLD activity [5,7]. But, since PC is the sole substrate for receptor-activated PLD [2], the possibility exists that the zwitterionic choline-containing headgroup of PC is the physiologically relevant site recognized by PLD PH domains. The PH domains of PLDs 1 and 2 are located in the N-terminal third of the molecule just before the first of four putative catalytic regions [28] (Fig. 3).

Previous studies of PLD in permeabilized cells and cell-free systems have shown that PLD is stimulated by the small GTP-dependent regulatory protein, ADP-ribosylation factor (ARF) [5,7,29]. ARF may interact with PLD1 in the region between its first and second catalytic regions [30] (Fig. 3). The above findings, the reported absence of signalling domains from PLD and the ability of phorbol esters to stimulate the

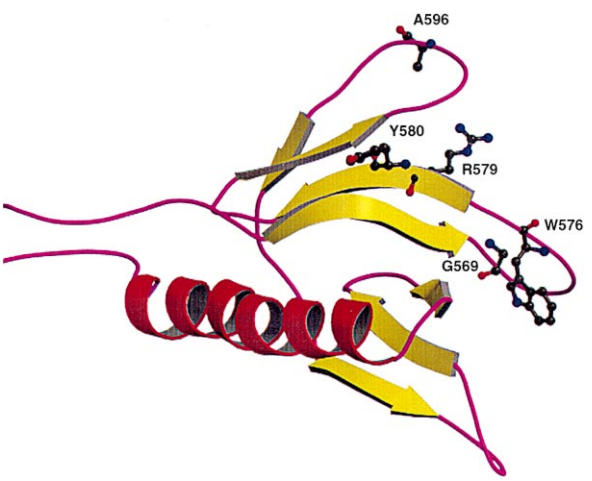

Fig. 2. The solution structure of β-ARK [22] indicating five residues that are perturbed by IP3 binding and which are also conserved in PLD1 and PLD2. This figure was prepared in MOLSCRIPT ([32]) and rendered in Raster 3D [33].

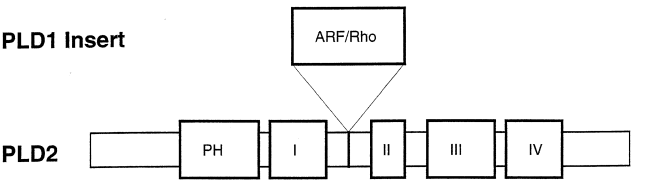

Fig. 3. The location of the PH domains within the PLD1 and PLD2 sequence relative to other regions (designated I–IV) of possible 'catalytic' function that are highly conserved in the PLD family [28]. PLD1 has an insert between regions I and II that may be a site for interaction with small G-proteins like ARF [30].

enzyme have contributed to the notion that PLD is activated by receptors in a non-conventional manner that may be downstream of PLC activation [1,3] and involve small G-proteins like ARF [31] rather than the heterotrimeric G-proteins or other signalling complexes. The presence of a PH domain in PLD1 and PLD2 raises the possibility that PLD may also be regulated by receptors in a more direct manner analogous to the PLCs involving heterotrimeric G-proteins or kinase activation [24]. In keeping with regulation by G-proteins, it is now recognized that a sizable portion of the total genome of c. elegans is dedicated to signaling through G-protein coupled receptors [34] and its apparent PLD2 homologic has a similarly placed domain homologous to the PH domain of mammalian PLD (Fig. 1). Several targeting mechanisms have been proposed for PH domains [11] most notably interactions with (1) the anionic phospholipids PIP2 and PIP3, (2) the βγ-subunit of heterotrimeric G-proteins, (3) protein kinase C and (4) phosphoserine and phosphothreonine residues of proteins. Any or all to these mechanisms may target mammalian PLD to membranes when a ligand binds to its receptor.

References

[1] Exton, J.H. (1997) Physiol. Rev. 77, 303–320.
[2] Holbrook, P.G., Pannell, L.K., Murata, Y. and Daly, J.W. (1992) J. Biol. Chem. 267, 16834–16840.
[3] Hodgkin, M.N., Pettitt, T.R., Martin, A., Michell, R.H., Pemberton, A.J. and Wakelam, M.J. (1998) Trends Biochem. Sci. 23, 200–204.
[4] Martin, A., Saqib, K.M., Hodgkin, M.N., Brown, F.D., Pettit, T.R., Armstrong, S. and Wakelam, M.J. (1997) Biochem. Soc. Trans. 25, 1157–1160.
[5] Hammond, S.M., Jenco, J.M., Nakashima, S., Cadwallader, K., Gu, Q., Cook, S., Nozawa, Y., Prestwich, G.D., Frohman, M.A. and Morris, A.J. (1997) J. Biol. Chem. 272, 3860–3868.
[6] Steed, P.M., Clark, K.L., Boyar, W.C. and Lasala, D.J. (1998) FASEB J. 12, 1309–1317.
[7] Brown, H.A., Gutowski, S., Moomaw, C.R., Slaughter, C. and Sternweis, P.C. (1993) Cell 75, 1137–1144.
[8] Harlan, J.E., Hajduk, P.J., Yoon, H.S. and Fesik, S.W. (1994) Nature 371, 168–170.
[9] Tanaka, K., Imajoh-Ohmi, S., Sawada, T., Shirai, R., Hashimoto, Y., Iwasaki, S., Kaibuchi, K., Kanaho, Y., Shirai, T., Terada, Y., Kimura, K., Nagata, S. and Fukui, Y. (1997) Eur. J. Biochem. 245, 512–519.
[10] Lemmon, M.A., Ferguson, K.M. and Schlessinger, J. (1996) Cell 85, 621–624.
[11] Rebecchi, M.J. and Scarlata, S. (1998) Annu. Rev. Biophys. Biomol. Struct. 27, 503–528.
[12] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acid Res. 25, 3389–3402.
[13] Hughey, R. and Krogh, A. (1996) Comput. Appl. Biosci. 12, 95–107.
[14] Di Francesco, V., Garnier, J. and Munson, P.J. (1997) J. Mol. Biol. 267, 446–463.

[15] Sonnhammer, E.L., Eddy, S.R. and Durbin, R. (1997) Proteins 28, 405–420.
[16] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[17] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[18] Munson, P.J., Di Francesco, V. and Porrelli, R.N. (1994) Proc. 6th Int. Conf. Int. Syst. Sci. V, 375–384.
[19] Geetha, V., Di Francesco, V., Garnier, J. and Munson, P.J. (1998) Protein Sci. 7, 81.
[20] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[21] Brenner, S.E., Chothia, C. and Hubbard, T.J. (1998) Proc. Natl. Acad. Sci. USA 95, 6073–6078.
[22] Fushman, D., Najmabadi-Haske, T., Cahill, S., Zheng, J., LeVine, H. and Cowburn, D. (1998) J. Biol. Chem. 273, 2835–2843.
[23] Fischer, D. and Eisenberg, D. (1996) Protein Sci. 5, 947–955.
[24] Rhee, S.G. and Bae, Y.S. (1997) J. Biol. Chem. 272, 15045–15048.
[25] Touhara, K., Inglese, J., Pitcher, J.A., Shaw, G. and Lefkowitz, R.J. (1994) J. Biol. Chem. 269, 10217–10220.
[26] Hyvonen, M. and Saraste, M. (1997) EMBO J. 16, 3396–3404.
[27] Ferguson, K.M., Lemmon, M.A., Schlessinger, J. and Sigler, P.B. (1995) Cell 83, 1037–1046.
[28] Morris, A.J., Engebrecht, J. and Frohman, M.A. (1996) Trends Pharmacol. Sci. 17, 182–185.
[29] Cockcroft, S., Thomas, G.M., Fensome, A., Geny, B., Cunningham, E., Gout, I., Hiles, I., Totty, N.F., Truong, O. and Hsuan, J.J. (1994) Science 263, 523–526.
[30] Katayama, K., Kodaki, T., Nagamachi, Y. and Yamashita, S. (1998) Biochem. J. 329, 647–652.
[31] Mitchell, R., McCulloch, D., Lutz, E., Johnson, M., MacKenzie, C., Fennell, M., Fink, G., Zhou, W. and Sealfon, S.C. (1998) Nature 392, 411–414.
[32] Kraulis, P.J. (1991) J. Appl Crystallogr. 24, 946–950.
[33] Merritt, E.A. and Bacon, D.J. (1997) Methods Enzymol. 277, 505–524.
[34] Bargmann, C.I. (1998) Science 282, 2028–2033.