Protein Science (1996), *5*:162–166. Cambridge University Press. Printed in the USA.
Copyright © 1996 The Protein Society

FOR THE RECORD

Extending the C2 domain family: C2s in PKCs δ, ϵ, η, θ, phospholipases, GAPs, and perforin

C.P. PONTING[1] AND P.J. PARKER[2]

[1] University of Oxford, Fibrinolysis Research Unit, The Old Observatory, South Parks Road, Oxford, OX1 3RH, United Kingdom
[2] Imperial Cancer Research Fund, Protein Phosphorylation Laboratory, P.O. Box 123, Lincoln's Inn Fields, London, WC2A 3PX, United Kingdom

(RECEIVED September 19, 1995; ACCEPTED October 24, 1995)

Abstract: Various membrane lipid metabolites, generated by phospholipases C and D (PLCs, PLDs), are known to regulate the activities of protein kinases C (PKCs) and GTP-ase activating proteins (GAPs) in a range of cellular processes. Conventional Ca^{2+}-dependent PKCs (α, βI, βII, and γ), PLCs, and various GAPs are all known to contain copies of a phospholipid-binding domain, termed C2 or CalB. Here we recognize that C2 domains are also present in "new" Ca^{2+}-independent PKCs (δ, ϵ, η, and θ), other kinases, a eukaryotic PLD, the *breakpoint cluster region* (*BCR*) gene product, and two further GAPS. Twenty-two previously unrecognized C2 domain sequences are presented, which include a single copy in the mammalian pore-forming proteins, perforin.

Keywords: homology; phospholipid-binding; phosphorylation; signal transduction

Many recent studies have demonstrated the roles of various membrane lipid metabolites in regulating intracellular signaling networks (reviewed in Nishizuka, 1995). Metabolites, such as diacylglycerol (DAG) (generated by phospholipase Cs [PLCs]) and phosphatidic acid (generated by phospholipase D [PLD]), activate protein kinase Cs (PKCs) (Takai et al., 1979) and inhibit Ras GTP-ase activating proteins (GAPs) (Tsai et al., 1989), respectively. Upon activation by Ca^{2+}, phospholipid, and/or DAG, the 12 currently known mammalian PKC isoforms participate in a diverse range of complex cellular processes (reviewed in Dekker & Parker, 1994). PKC specificity and auto-inhibition is provided by a number of noncatalytic regulatory domains, including either one or two DAG-binding domains, and, in the case of Ca^{2+}-dependent PKCs (PKC α, β and γ), a domain, termed C2, which is known to interact with phospholipids in a Ca^{2+}-dependent manner (Bazzi & Nelsestuen, 1990).

Reprint requests to: C.P. Ponting, University of Oxford, Fibrinolysis Research Unit, The Old Observatory, South Parks Road, Oxford, OX1 3RH, United Kingdom; e-mail: chris@biop.ox.ac.uk.

C2 (or "CalB") domains have been identified previously in a number of other protein sequences: synaptotagmins, rabphilin-3A, and Doc2 (two copies each), cytosolic phospholipase A_2 (cPLA2), GAPs, PLCs, RSP5, and Unc-13 (Perin et al., 1990; Sasaki et al., 1990; Maruyama & Brenner, 1991; Bork & Sudol, 1994; Maekawa et al., 1994; Nalefski et al., 1994; Orita et al., 1995). The crystal structure of the first C2 domain of rat synaptotagmin I ($C2_{Syn}$) has been solved recently, revealing an eight-stranded β sandwich containing four aspartic acid residues involved in binding Ca^{2+} (Sutton et al., 1995). Here, we identify 22 novel C2 domain sequences (all pairwise identities <55%) in proteins with lipid-related functions (Fig. 1); several of these sequences possess substitutions within the $C2_{Syn}$ Ca^{2+}-binding site, indicating a potential loss of Ca^{2+}-affinity.

An initial comparison of a region ("HR2"), common to PRK1, PRK2, PKC ϵ, and PKC η Ca^{2+}-independent kinases (Palmer et al., 1995), with sequence databases, showed these sequences to be representatives of an homologous family, which also included several other Ca^{2+}-independent PKC sequences. Some similarity of this family to C2 sequences was apparent, yet in isolation was insufficient to propose an evolutionary divergence of HR2 and C2 domains. This similarity was also noted by Sossin and Schwartz (1993); they too were unable to provide statistical evidence for a plausible homology argument. However, comparison of an alignment of known C2 sequences with databases (see below), showed sufficient similarity between the C2 alignment and 6 of the 10 HR2 sequences to positively identify these and the C2s as possessing a common evolutionary ancestor. It is concluded that all 10 HR2 sequences are homologous to C2 domains, and that the HR2 branch of the C2 evolutionary tree contains family members distantly related to, say, that of PKCα.

The strategy used in this analysis was as follows. The similarities between the "HR2" regions of PKC ϵ, PKC η, and human PRK1 and PRK2 kinases were sufficient to indicate a common evolutionary ancestor for these sequences (Z-scores > 8.0σ). Three iterations of a sequence analysis procedure were sufficient to show significant similarities between HR2-like sequences and regions of other protein kinase Cs. This analysis

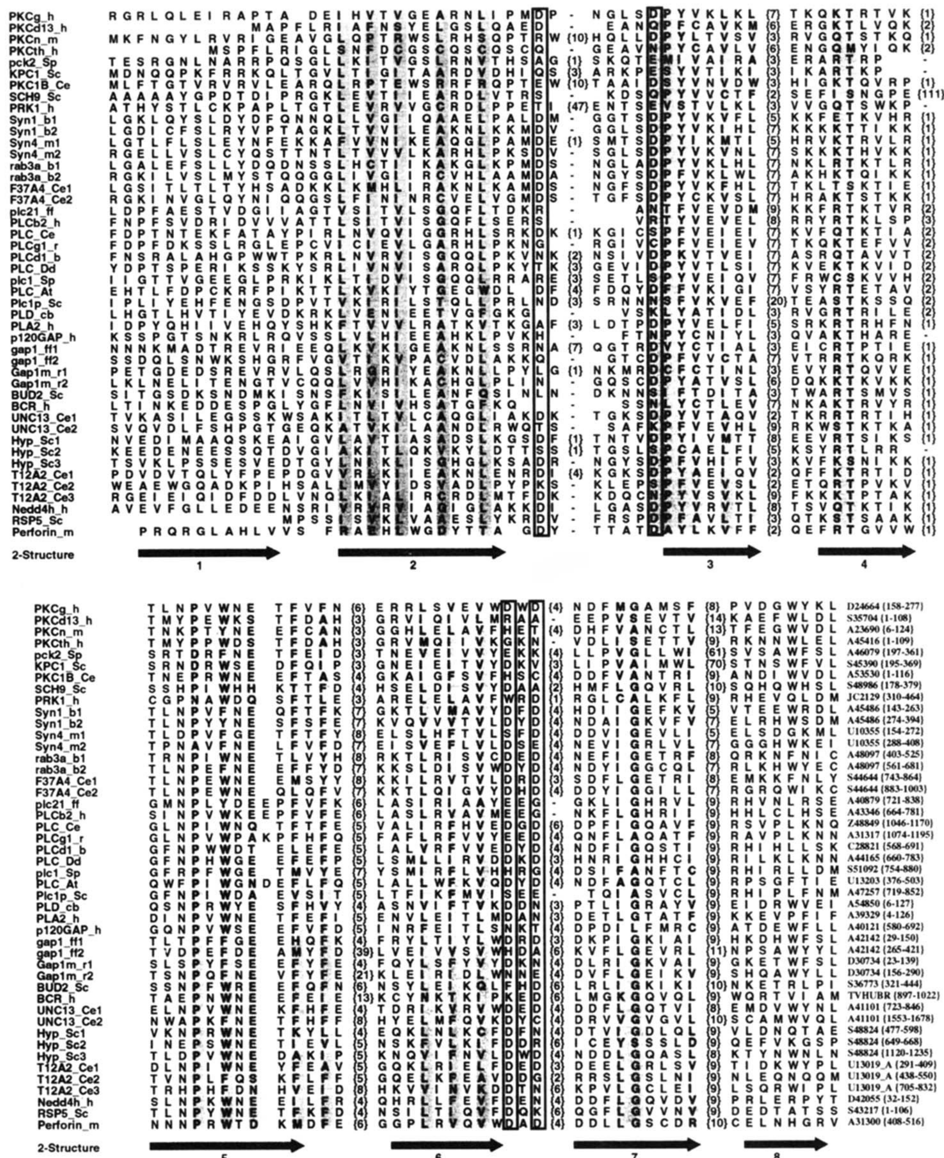

Fig. 1. Multiple sequence alignment of a representative set of C2 domain sequences (all pairwise identities < 55%), displayed using ALSCRIPT (Barton, 1993b) and with relatively conserved positions shaded. This is based on automatically generated alignments and was edited manually to accord with the experimentally determined structure of rat synaptotagmin I ($C2_{Syn}$) (Sutton et al., 1995). Intervening sequences between secondary structures are represented by numbers in parentheses. Accession codes and residue numbers are shown following the alignment. Positions of the four aspartic acid residues of $C2_{Syn}$, which are involved in binding Ca^{2+}, are boxed. Domain limits are uncertain because strands 1 and 8 are poorly conserved; the alignments in these regions are expected to be inaccurate. PKCδ13 and PKCθ do not appear to contain C2 β-strand 1. Given the β-sheet strand order (8-1-2-5) in the $C2_{Syn}$ structure, and the low similarities of putative C2 sequences at their C-termini, it is possible that a subset of these lack strand 1 and, perhaps, strand 8. It is noted that C2 domain constructs consisting only of strands β3-β4-β5 retain partial (Fukuda et al., 1994) and full (Gawler et al., 1995) phospholipid-binding functions in vitro, indicating that the presence of less than eight strands in a subset of C2 homologues is plausible. The C-terminal limit was assigned according to weak similarities between a subset of sequences, and by reference to the third repeat (residues 715–834) of T12A2 (837 residues, total). The putative RSP5 N-terminus, based on the gene sequence, may be inaccurate. A possible fourth C2 in Hyp_Sc between copies 2 and 3 (residues 782–900) scored at the level of the noise in searches and is not included in the figure. A closely similar paralogue of this sequence is contained in GENBANK:SCORFSDNA (*S. cerevisiae* ORF N2250). PKCg, PKCγ; PKCd13, PKCδ13; PKCn, PKCη; PKCth, PKCθ; Syn, synaptotagmin; rab3a, rabphilin-3A; Nedd4h, human homologue of murine Nedd4; h, human; m, murine; Sp, *S. pombe*; Sc, *S. cerevisiae*; b, bovine; Ce, *C. elegans*; ff, fruit fly (*Drosophila*); Dd, *D. discoideum*; At, *A. thaliana*; cb, castor bean; r, rat; other abbreviations in text.

procedure was a two-step process, similar to that used previously (Ponting & Phillips, 1995), which substantially reduced the level of "noise" when compared with single-step methods. Initially, a multiple alignment of previously determined HR2 homologues was compared with current sequence databases (PIR v45 and PATCHX) using a local similarity algorithm (Barton, 1993a). Subsequently, the 300 highest-scoring sequences were compared against a flexible pattern of the most conserved positions using

the Barton and Sternberg (1989) algorithm. Those sequences scoring above the noise, at levels similar to, or higher than, scores of previously determined HR2 sequences, were incorporated within a new multiple alignment, leading to a new iteration. The BLOSUM62 amino acid substitution matrix and a gap penalty of 8 were used throughout, and only single sequences of pairs whose amino acid identities exceeded 55% were retained.

Those sequences found to include HR2s after three iterations were all Ca^{2+}-independent kinases: PKCϵ-isoforms, such as those from *Aplysia californica* AplII, and *Caenorhabditis elegans* PKC1B, and other isoforms: PKCθ, PKCδ13, *Saccharomyces cerevisiae* PKC1, and *Schizosaccharomyces pombe* pck1 and pck2. Surprisingly, all PKCα sequences also scored highly (scores $\approx$ 49), as did inter alia, sequences from a rabphilin-3A-like protein (score 44), an *S. cerevisiae* hypothetical protein (score 64), and castor bean phospholipase D (score 54); HR2 homologues scored between 58 and 99, and the noise plateau was approximately 30. All of the latter sequences correspond to C2 domains, which either were previously known or were later identified as C2 domains. Although it appeared plausible at this stage that Ca^{2+}-independent PKCs contain C2 domains, there was little statistical evidence to support this proposal, and a complementary approach was required.

It was decided to repeat the iterative scanning procedure, not using HR2 sequences as a basis set, but using previously established C2 sequences, taken from Sutton et al. (1995). After four iterations, a representative set (<55% pairwise identities) of 43 C2 domain sequences was established, containing both previously known and newly identified members (scores 54–86, top noise 54, level of noise plateau, 42). These sequences show similarities at several positions, and are proposed to form an homologous family. This set included the HR2 sequences of PKC isoforms δ, ϵ, and η, pck1, AplII and PKC1B, but not PRK1, PKC θ and pck2, which fell at or below the noise threshold. However, having established that all of these HR2 sequence share a common ancestor, and that a large fraction of these are homologous to C2 domains, it is logical to conclude that all such HR2 sequences are C2 homologues. This argument uses the determination of evolutionary relationships in different, yet overlapping, regions of the C2 phylogenetic tree, to establish their common ancestry. The remaining "HR2-like" C2 sequences were appended to the representative set, making a total of 46 sequences.

Identification of C2 domains in Ca^{2+}-independent "new" nPKCs δ, ϵ, η, and θ, as well as in Ca^{2+}-dependent "conventional" cPKCs α, βI, βII, and γ, shows that of the mammalian PKCs, only the "atypical" aPKCs ζ, λ, and μ do not appear to possess this domain (Fig. 2). Yeast PKCs and SCH9 kinase (Toda et al., 1988) and vertebrate PRKs contain one C2. A separate sequence motif (termed "HR1" by Palmer et al., 1995), present in triplicate in PRK1 and PRK2, appears in duplicate in the yeast PKCs (*S. pombe* pck1 and pck2, *S. cerevisiae* PKC1 and orthologues; Fig. 2).

Phospholipase C isoforms from mammals, *Drosophila*, yeast, slime mold (*Dictyostelium discoideum*) and *Arabidopsis thaliana*, each possess a C2 domain, within a region known as the extended Y box. Phospholipases A_2, from mammals, chicken, and fish, have been shown elsewhere (Nalefski et al., 1994), to contain C2 domains at their N-terminus. The only currently available phosphatidylcholine-hydrolyzing phospholipase D sequence, from castor bean (Wang et al., 1994), has been shown here to contain a C2 domain at its N-terminus. A single C2 domain in human p120GAP, and two each in rat and *Drosophila* Gap1m, had previously been noted (Maekawa et al., 1994). However, single C2 domains in three other GAPs had not: BUD2 from *S. cerevisiae*, and human *breakpoint cluster region* (*BCR*) and *active BCR-related* (*ABR*) gene products.

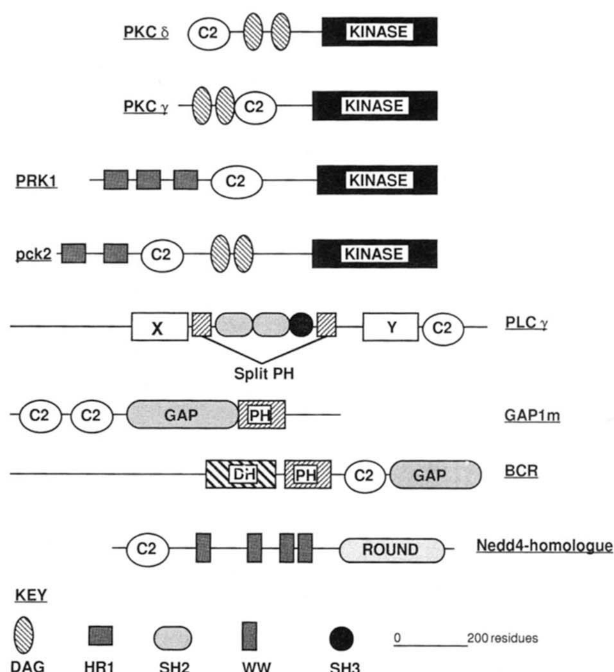

Fig. 2. Domain organization of several C2-containing proteins. Domains are shown approximately to scale, with respect to numbers of amino acids. DAG, a cysteine-rich DAG/phorbol ester-binding domain (C1); DH, Dbl-homologous domain; PH, pleckstrin homology domain; SH2, SH3, Src-homology-2 and -3 domains; other abbreviations in text.

The remaining C2s identified here occur in a *C. elegans* rabphilin-3A-like putative protein (F37A4_Ce) (two copies); three or four copies each in *C. elegans* (T12A2) and *S. cerevisiae* (Hyp_Sc) orthologues; a second C2 in Unc-13 (residues 1563–1680); one C2 in a human Nedd4 homologue (Nedd4h_h), which also contains four WW domains; and, surprisingly, a single C2 in mammalian perforins, which are pore-forming proteins. Initially, it was considered that perforin may have been identified as a "false positive." Of all C2-containing proteins, only perforin contains an N-terminal leader pre-peptide, typical of secreted or cytoplasmic granule-stored proteins. Moreover, it contains an EGF-like domain (which, possessing disulphide bonds, is unlikely to function intracellularly) immediately N-terminal to the putative C2 domain sequence. However, perforin has been shown to bind to membranes in a Ca^{2+}-dependent, EDTA-reversible manner at 4 °C (Müller & Tschopp, 1994); if subsequently incubated at 37 °C, perforin "inserts," i.e., becomes chelator nonextractable. It is noted that all four aspartic acid residues, known to be involved in Ca^{2+} ligation in the $C2_{Syn}$ crystal structure, are conserved in the perforin putative C2 sequence (Fig. 1). It would seem likely that the initial, reversible, Ca^{2+}-dependent membrane-binding event is conferred on perforin by a domain, homologous to C2s.

Several sequences, previously proposed to be C2 homologues, showed levels of similarity to known C2 sequences at, or below,

the "noise" level, and, therefore, were unable to be identified positively as encoding C2 domains. These include a region encompassing residues 672–766 of *S. cerevisiae* PKC1 (Levin et al., 1990); PKC1, however, does contain a C2 domain between residues 207–363. Similarly, proposed C2s in yeast Tor2 and in the neurofibromatosis gene product (NF-1) (Stephens et al., 1993) could not be proven. Other proposals that yeast VPS34 and mammalian and plant phosphatidylinositol 3-kinases (PI3Ks) contain C2s (Stephens et al., 1993; Welters et al., 1994) also could not be proven, although a common evolutionary origin of these sequences was established. In this case, although these sequences scored among the highest "noise" scores, an attempt to find "missing link" sequences, with sufficient similarities to both established C2 sequences and PI3K sequences, was unsuccessful.

One of the features of many of these C2 domain-containing proteins is that they contain multiple "membrane-interacting" domains. This is most obviously the case for proteins that have two or more C2 domains (synaptotagmins, rabphilin-3A, etc.), but is nevertheless no less true for PKCs, which encode phospholipid-dependent DAG-binding domains (C1) in addition to C2 domains and also phospholipases (PLC, PLD) that bind their substrates (and products) through domains independent of the C2 domains aligned here. The case for the phospholipases is defined explicitly for cPLA2, where the C2-deleted protein retains its ability to interact with and hydrolyze monomeric phospholipid substrate, whereas the C2 domain alone displays Ca^{2+}-dependent phospholipid-binding properties (Nalefski et al., 1994). For synaptotagmin, the presence of two C2 domains has been shown to evoke a cooperative interaction with membranes (Damer & Creutz, 1994). It is likely that the combination of domains in these other proteins similarly confers cooperative interactions with membranes. At this facile level, the identification of the V0 domains of the nPKC subclass as C2-related domains, provides a satisfying symmetry with the better understood "Ca^{2+}-dependent" C2 domains of the cPKC subclass. While Ca^{2+} evidently regulates membrane occupancy of the cPKC proteins (see, for example, Kiley et al., 1990) it can be surmised that the V0/C2 domains of the nPKCs do so in a manner independent of Ca^{2+}. As typified by the behavior of activated, i.e., DAG-bound, PKC, which becomes stably associated with the membrane (as evidenced by the translocation assay for activation; see Kraft & Anderson, 1983), perhaps these combinations of domains confer a stability to the membrane-associated state that is not adequately supported by a single contact.

The idea that combined/duplicated domains may be an important factor in stabilizing the membrane association of these proteins might also be applied in a predictive capacity. Thus, PRK1 is known to be a lipid-activated protein kinase (Palmer & Parker, 1995 and references therein). If, as shown here, the second conserved region within the PRK1 regulatory domain (HR2) is indeed a C2 domain that binds lipids, then one might surmise that the first conserved region (HR1) also confers some membrane-interacting property.

Are these membrane-binding properties of C2 domains, whether regulated by Ca^{2+} or not, the totality of C2 domain function? There is evidence that C2 domains also interact with non-phospholipid targets. For example, in several synaptotagmins, the first C2 domain binds syntaxin, a plasma membrane protein, and the second C2 domain binds clathrin AP-2 (Li et al., 1995). Similarly, synaptotagmin II binds inositol 1,3,4,5-tetrakisphosphate (Fukuda et al., 1994). These observations suggest that C2 domains, in addition to binding phospholipids, may also interact with other (membrane-bound) targets. Whether such additional targets control the membrane-binding or are the raison d'être of the C2 domains, or whether these domains simply act to recruit other functional domains from within the homologous polypeptide chain, remains an open issue.

Note added in proof: A recent study (Brose N, Hofmann K, Hata Y, Südhof TC. 1995. Mammalian homologues of *Caenorhabditis elegans unc-13* gene define novel family of C_2-domain proteins. *J Biol Chem 270*:25273–25280.) has identified independently several of the C2 domain sequences discussed here.

Acknowledgments: C.P.P. is a Medical Research Council (UK) Training Fellow and is a member of the Oxford Centre for Molecular Sciences. C.P.P. thanks Dr. C.M. Dobson for providing a stimulating working environment. We thank an anonymous reviewer whose comments led to a considerable improvement of Figure 1.

References

Barton GJ. 1993a. An efficient algorithm to locate all locally optimal alignments. *Comput Applic Biosci 9*:729–734.

Barton GJ. 1993b. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng 6*:37–40.

Barton GJ, Sternberg MJE. 1989. Flexible pattern sequence patterns. A sensitive method to detect weak structural similarities. *J Mol Biol 212*:389–402.

Bazzi MD, Nelsestuen GL. 1990. Protein kinase C interaction with calcium: A phospholipid-dependent process. *Biochemistry 29*:7624–7630.

Bork P, Sudol M. 1994. The WW domain: A signalling site in dystrophin? *Trends Biochem Sci 19*:531–533.

Damer CK, Creutz CE. 1994. Synergistic membrane interactions of the two C2 domains of synaptotagmin. *J Biol Chem 269*:31115–31123.

Dekker LV, Parker PJ. 1994. Protein kinase C – A question of specificity. *Trends Biochem Sci 19*:73–77.

Fukuda M, Aruga J, Niinobe M, Aimoto S, Mikoshiba K. 1994. Inositol-1,3,4,5-tetrakisphosphate binding to C2B domain of IP4BP/synaptotagmin II. *J Biol Chem 269*:29206–29211.

Gawler DJ, Zhang LJW, Moran MF. 1995. Mutation-deletion analysis of a Ca^{2+}-dependent phospholipid binding (CalB) domain within p120 GAP, a GTPase-activating protein for p21 *ras*. *Biochem J 307*:487–491.

Kiley S, Schaap D, Parker PJ, Hsieh LL, Jaken S. 1990. Protein kinase C heterogeneity on GH_4C_1 rat pituitary cells. *J Biol Chem 265*:15704–15712.

Kraft AS, Anderson WB. 1983. Phorbol esters increase the amount of Ca^{2+}, phospholipid-dependent protein kinase associated with plasma membrane. *Nature 301*:621–623.

Levin DE, Fields FO, Kunisawa R, Bishop JM, Thorner J. 1990. A candidate protein kinase C gene, *PKC1*, is required for the *S. cerevisiae* cell cycle. *Cell 62*:213–224.

Li C, Ullrich B, Zhang JZ, Anderson RGW, Südhof TC. 1995. Ca^{2+}-dependent and -independent activities of neural and non-neural synaptotagmins. *Nature 375*:594–599.

Maekawa M, Li S, Iwamatsu A, Morishita T, Yokota K, Imai Y, Kohsaka S, Nakamura S, Hattori S. 1994. A novel mammalian Ras GTP-ase activating protein which has phospholipid-binding and Btk homology regions. *Mol Cell Biol 14*:6879–6885.

Maruyama IN, Brenner S. 1991. A phorbol ester/diacylglycerol-binding protein encoded by the *unc-13* gene of *Caenorhabditis elegans*. *Proc Natl Acad Sci USA 88*:5729–5733.

Müller C, Tschopp J. 1994. Resistance of CTL to perforin-mediated lysis – Evidence for a lymphocyte membrane protein interacting with perforin. *J Immunol 153*:2470–2478.

Nalefski EA, Sultzman LA, Martin DM, Kriz RW, Towler PS, Knopf JL, Clark JD. 1994. Delineation of two functionally distinct domains of cytosolic phospholipase A_2, a regulatory Ca^{2+}-dependent lipid-binding domain and a Ca^{2+}-independent catalytic domain. *J Biol Chem 269*:18239–18249.

Nishizuka Y. 1995. Protein kinase C and lipid signaling for sustained cellular responses. *FASEB J 9*:484–496.

Orita S, Sasaki T, Naito A, Komuro R, Ohtsuka T, Maeda M, Suzuki H,

Igarashi H, Takai Y. 1995. Doc2: A novel brain protein having two repeated C2-like domains. *Biochem Biophys Res Commun 206*:439–448.

Palmer RH, Parker PJ. 1995. Expression, purification and characterisation of the ubiquitously expressed protein kinase C-related kinase 1 – PRK1. *Biochem J 309*:315–320.

Palmer RH, Ridden J, Parker PJ. 1995. Cloning and expression patterns of two members of a novel protein-kinase-C-related kinase family. *Eur J Biochem 227*:344–351.

Perin MS, Fried VA, Mignery GA, Jahn R, Südhof TC. 1990. Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. *Nature 345*:260–261.

Ponting CP, Phillips C. 1995. DHR domains in syntrophins, neuronal NO synthases and other intracellular proteins. *Trends Biochem Sci 20*:102–103.

Sasaki T, Kikuchi A, Araki S, Hata Y, Isomura M, Kuroda S, Takai Y. 1990. Purification and characterization from bovine brain cytosol of a protein that inhibits the dissociation of GDP from and the subsequent binding of GTP to *smg* p25A, a *ras* p21-like GTP-binding protein. *J Biol Chem 265*:2333–2337.

Sossin WS, Schwartz JH. 1993. Ca^{2+}-independent protein kinase Cs contain an amino terminal domain similar to the C2 consensus sequence. *Trends Biochem Sci 18*:207–208.

Stephens LR, Jackson TR, Hawkins PT. 1993. Agonist-stimulated synthesis of phosphatidylinositol(3,4,5)-triphosphate: A new intracellular signalling system? *Biochim Biophys Acta 1179*:27–75.

Sutton RB, Davletov BA, Berghuis AM, Südhof TC, Sprang SR. 1995. Structure of the first C_2 domain of synaptotagmin I: A novel Ca^{2+}/phospholipid-binding fold. *Cell 80*:929–938.

Takai Y, Kishimoto A, Iwasa Y, Kawahara Y, Mori T, Nishizuka Y. 1979. Calcium-dependent activation of a multifunctional protein kinase by membrane phospholipids. *J Biol Chem 254*:3692–3695.

Toda T, Cameron S, Sass P, Wigler M. 1988. *SCH9*, a gene of *Saccharomyces cerevisiae* that encodes a protein distinct from, but functionally and structurally related to, cAMP dependent protein kinase catalytic subunits. *Genes & Dev 2*:517–527.

Tsai MH, Yu CL, Wei FS, Stacey DW. 1989. The effect of GTPase activating protein upon Ras is inhibited by mitogenically responsive lipids. *Science 243*:522–525.

Wang X, Xu L, Zheng L. 1994. Cloning and expression of phosphatidylhydrolyzing phospholipase D from *Ricinus communis* L. *J Biol Chem 269*:20312–20317.

Welters P, Takegawa K, Emr SD, Chrispeels MJ. 1994. AtVPS34, a phosphatidylinositol 3-kinase of *Arabidopsis thaliana*, is an essential protein with homology to a calcium-dependent lipid binding domain. *Proc Natl Acad Sci USA 91*:11398–11402.