Protein Science (1997), 6:477–480. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Taxonomy and function of C1 protein kinase C homology domains

JAMES H. HURLEY,[1] ALEXANDRA C. NEWTON,[2] PETER J. PARKER,[3] PETER M. BLUMBERG,[4] AND YASUTOMI NISHIZUKA[5]

[1]Laboratory of Molecular Biology, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0580
[2]Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0640
[3]Imperial Cancer Research Fund, Protein Phosphorylation Laboratory, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom
[4]Molecular Mechanisms of Tumor Promotion Section, Laboratory of Cellular Carcinogenesis and Tumor Promotion, National Cancer Institute, Bethesda, Maryland 20892-4255
[5]Biosignal Research Center, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657, Japan

(RECEIVED September 30, 1996; ACCEPTED November 11, 1996)

Abstract: C1 domains are compact α/β structural units of about 50 amino acids which tightly bind two zinc ions. These domains were first discovered as the loci of phorbol ester and diacylglycerol binding to conventional protein kinase C isozymes, which contain two C1 domains (C1A and C1B) in their N-terminal regulatory regions. We present a comprehensive list of 54 C1 domains occurring singly or doubly in 34 different proteins. Many C1 domains and C1 domain-containing proteins bind phorbol esters, but many others do not. By combining analysis of 54 C1 domain sequences with information from previously reported solution and crystal structure determinations and site-directed mutagenesis, profiles are derived and used to classify C1 domains. Twenty-six C1 domains fit the profile for phorbol-ester binding and are termed "typical." Twenty-eight other domains fit the profile for the overall C1 domain fold but do not fit the profile for phorbol ester binding, and are termed "atypical." Proteins containing typical C1 domains are predicted to be regulated by diacylglycerol, whereas those containing only atypical domains are not.

Keywords: diacylglycerol kinase; GTPase activating protein; guanine nucleotide exchange factor; kinase suppressor of Ras; n-chimaerin; phorbol ester; Raf; unc-13; Vav

The protein kinase C family plays a central role in many signal transduction pathways and is an intensively studied example of enzyme regulation by Ca^{2+} and lipids. All protein kinase Cs are characterized by a N-terminal regulatory region and a C-terminal catalytic domain. Protein kinase C regulatory regions have been further subdivided into a pseudosubstrate region, conserved regions C1 and C2, and variable regions (Kikkawa et al., 1989; Bell & Burns, 1991; Dekker & Parker, 1994; Newton, 1995). Dozens of other proteins have sequence homology to one or both of the two conserved regions. Subsequent to its discovery as a conserved region of sequence, C2 has been found to correspond precisely to a distinct structural domain. The term "C2" is widely accepted to denote this domain whether it occurs in a protein kinase C or in one of the dozens of other proteins in which it has been found (Newton, 1995). Here we address the taxonomy of the other conserved half of the protein kinase C regulatory region.

The traditional definition of the protein kinase C C1 region included just over 120 residues, spanning two repeats of a cysteine-rich domain (Kikkawa et al., 1989; Bell & Burns, 1991; Dekker & Parker, 1994; Newton, 1995) in a typical protein kinase C. Most, but not all protein kinase Cs, contain two tandemly repeated copies of this domain. All protein kinase Cs, however, contain at least one cysteine-rich domain. Single copies occur in atypical protein kinase Cs, n-chimaerin, unc-13, the Raf and KSR kinases, and a number of non-kinase oncoproteins, including Vav and Lfc. Some of these single domains bind phorbol esters and diacylglycerol, while others clearly do not (Ono et al., 1989; Ahmed et al., 1991; Burns & Bell, 1991; Kazanietz et al., 1994). Individual domains may have non-equivalent roles within the context of a tandem repeat (Szallasi et al., 1996), but the properties of isolated recombinant first and second domains are not substantially different (Kazanietz et al., 1994). It has become clear over the past five years that the fundamental structural unit is the single, rather than the double, cysteine-rich region.

Single domains of typical protein kinase Cs are capable of binding phorbol esters and other protein kinase C activators with high affinity (Ono et al., 1989; Ahmed et al., 1991; Burns & Bell, 1991; Kazanietz et al., 1994). The three-dimensional structures of three different isolated single domains have been determined (Hommel

Reprint requests to: James H. Hurley, Laboratory of Molecular Biology, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0580; e-mail: Hurley@tove.niddk.nih.gov.

et al., 1994; Ichikawa et al., 1995; Zhang et al., 1995; Mott et al., 1996). Single domains of typical protein kinase Cs have been referred to as "cys1" and "cys2" in order of their occurrence within the C1. Single domains, regardless of position, have been called "cysteine-rich domains" (CRD), "cysteine-rich motifs" (CRM), "zinc butterflies," "zinc fingers," or phorbol ester-diacylglycerol binding (pe_dag or dag_pe) domains. All of these terms have proved either too broad or too particular to be completely suitable. Yet other terms have been used for conserved regions of Raf (CR1) and diacylglycerol kinase (C3) isoforms containing these domains. This variation of terminology is awkward for specialists and confusing for non-specialists. It would be more desirable to use a standard terminology that reflects the common structural scaffold of all these domains yet respects their diversity of function and context.

We believe the most straightforward solution is to modify the meaning of a term that is already common currency in the protein kinase C field, and whose previous definition is closely related to the domain at issue. Specifically, we propose that "C1" (protein kinase "C" homology "1") refer to any domain homologous to a single cysteine-rich repeat of protein kinase C, regardless of its function or context. By analogy with the usage for C2 domains in synaptotagmins, we suggest that when two C1 domains occur on a single polypeptide chain, they be designated "C1A" and "C1B." This terminology leaves open the possibility for "C1C" and so on if triple of higher multiples of C1 domains should be discovered in a single protein in the future. The conserved region of typical protein kinase Cs formerly known as "C1" becomes "C1A–C1B" in the new parlance.

Taxonomic classification of C1 domains has important implications for the regulation of C1 domain-containing proteins. We suggest the terms "typical" and "atypical" C1 to refer to subtypes which do or do not bind diacylglycerol and phorbol esters, respectively, following the usage for protein kinase C activation. Some C1 domains have been directly tested for phorbol ester binding (Ono et al., 1989; Ahmed et al., 1991; Burns & Bell, 1991; Kazanietz et al., 1994), and the rest can be tentatively classified as typical or atypical on the basis of sequence. Consensus residues Pro11, Gly23, and Gln27 (Fig. 1) play critical roles in maintaining the precise unzipped structure of the upper β sheet which forms the diacylglycerol binding site. Consensus residues 8, 13, 20, 22, and 24 form the hydrophobic wall around the groove and are believed to insert into the hydrophobic core of the bilayer in typical C1 domains. These second five residues are always Met, Val, Leu, Ile, Phe, Tyr, or Trp in typical C1 domains, but are frequently polar in atypical C1 domains. The sequence and three-dimensional structure (Mott et al., 1996) of the atypical C1 domain of Raf-1 shows an even more dramatic variation from the typical C1 domains: consensus positions 23–27 are entirely absent. Mutation of consensus residues 11 and 27 (Kazanietz et al., 1994) abolishes or greatly diminishes phorbol binding, although other sites have not yet been tested.

The new classification system makes it possible to rapidly identify typical and atypical C1 sequences in newly cloned proteins, and then to predict whether a new C1 domain-containing protein will be regulated by diacylglycerol or not. A comprehensive census of C1 domains was obtained by database searching. Sequences of 25 typical and 10 atypical C1 domains were aligned and used to construct profiles (Gribskov et al., 1990) for typical and atypical C1 domains separately, and for all C1 domains together. These profiles were used to search the SwissProt database. Target sequences were ranked by Z-score. The threshold of significance was set by the highest-scoring false positives, including a multiple C2H2-type zinc finger protein, a LIM domain protein, and a putative multiple disulfide bonded protein. In order to survey C1 domains discovered since the last SwissProt release, searches of GenBank and the science citation index were carried out. Putative C1 domain sequences from these sources were also scored against the overall C1 domain profile. Sequences scoring at or below the false-positive threshold were discarded. The sole exception was diacylglycerol kinase ζ, where the context in which the putative C1 domains occur provided additional support for their identification. The enumeration excludes domains found in sequences of essentially the same protein in different species (orthologues), including cases where different terms are used for the orthologues (e.g., PKCμ/PKD and PKCι/PKCλ). A total of 54 C1 domains in 34 different proteins were identified.

Mammalian diacylglycerol kinases are the second-largest class of C1 domain containing proteins (Kanoh et al., 1993). It is not yet resolved whether diacylglycerol has a non-substrate role in regulating some or all of the isozymes, although recombinant C1-domain containing fragments of diacylglycerol kinase α do not bind phorbol esters (Ahmed et al., 1991). The C1A domains, but not the C1B domains, of human diacylglycerol kinases β and γ are typical on the basis of fitting the consensus profile. These diacylglycerol kinase isozymes are therefore predicted to bind diacylglycerol at their C1A domain. The diacylglycerol 3-hydroxyl is believed to form three hydrogen bonds with protein and to be entirely excluded from solvent (Zhang et al., 1995). This would completely preclude access by an attacking phosphoryl group. Diacylglycerol binding would therefore presumably have a membrane anchoring or other regulatory role, rather than a catalytic role. Diacylglycerol kinases α, δ, and ϵ possess unusual C1A domains which lack some of the hydrophobic anchor residues (20, 22, 24) but have all of the other typical consensus residues. This suggests that diacylglycerol-promoted membrane insertion is unlikely to occur for the α, δ, and ϵ isozymes, although binding to the diacylglycerol polar backbone cannot be ruled out. The C1B domains of all the diacylglycerol kinase isozymes, and both C1 domains of the ζ isozyme, diverge at almost all typical consensus positions. It is possible that diacylglycerol kinase ζ has diverged so far from other diacylglycerol kinases that the underlying structure of the C1 domain has not been preserved.

Of the atypical C1 domain-containing proteins, protein kinase Cι (known as λ in mouse) and Cζ present two of the least divergent sequences. Protein kinase Cζ deviates from the typical consensus at only two positions, Pro-11 and hydrophobic-20. Protein kinase Cζ is not activated by diacylglycerol, but it is activated *in vitro* by ceramide (Lozano et al., 1994). It is not known with which domain of protein kinase Cζ ceramide interacts. Ceramide, like diacylglycerol, has a minimal polar moiety. It is possible that protein kinase Cζ binds ceramide at its C1 domain by a mechanism analogous to diacylglycerol binding by typical PKCs. Atypical C1 domains of protein kinase Cζ and Cι have also been implicated in protein–protein interactions (Diaz-Meco et al., 1996a, 1996b). Human protein kinase Cμ (known as protein kinase D in mouse) contains a catalytic domain which is more similar to *Dictyostelium* myosin light-chain kinase than to other protein kinase Cs, and was initially classified as an atypical isozyme. Protein kinase D is activated by phorbol ester, however (Valverde et al., 1994). The C1A domain of protein kinase Cμ fits perfectly into the typical C1 consensus pattern. Its C1B domain has only one major violation of

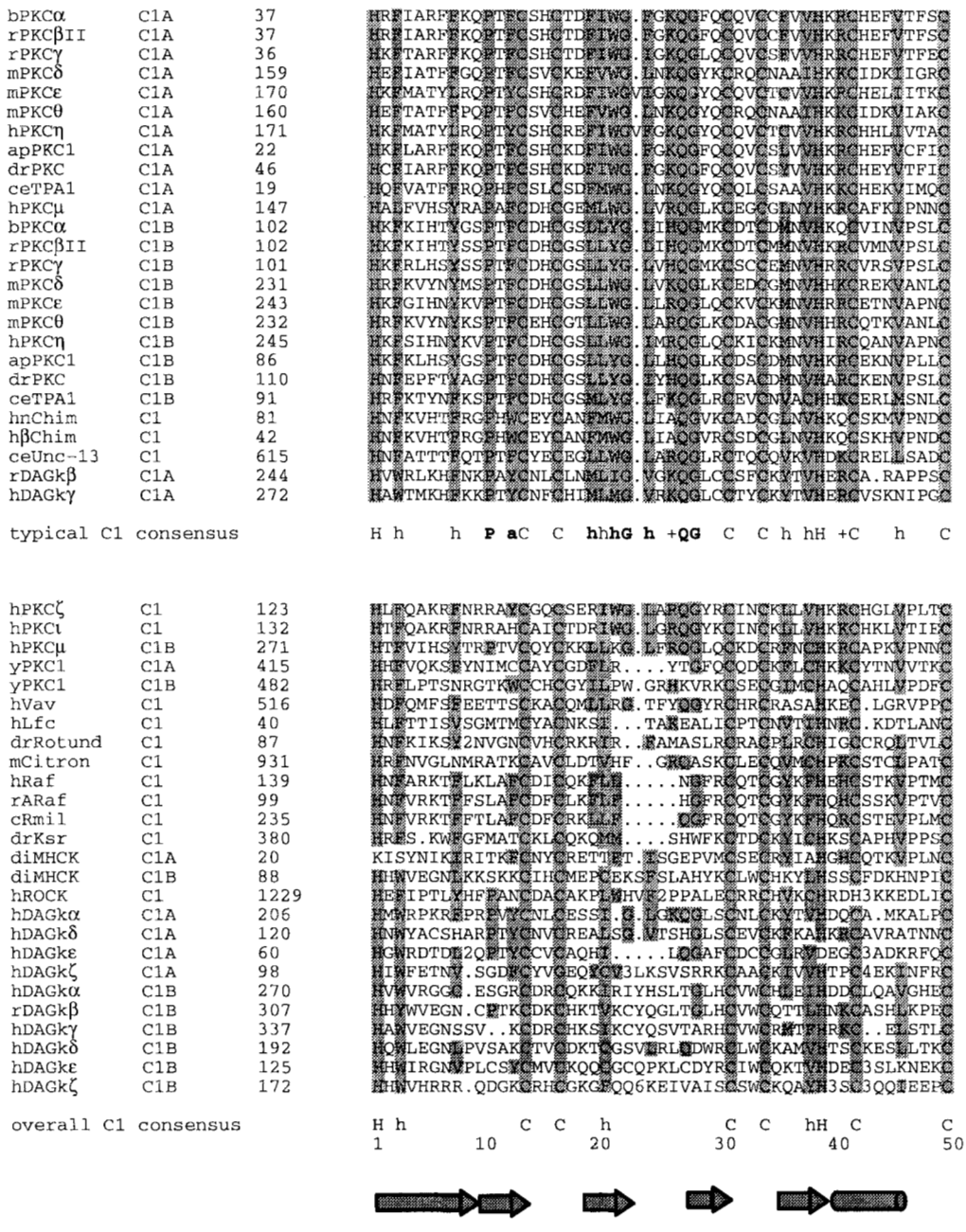

Fig. 1. Alignment of representative C1 domains. Numerals n within sequences indicate insertion of n amino acid residues, and periods indicate gaps. The consensus profiles shown below the alignment include shaded residues which strictly fit the profile as well as those in which there are not more than three violations. Uppercase letters in the profile are amino acid residues, lowercase letters and symbols are: h, large hydrophobic (Trp, Tyr, Phe, Leu, Ile, Val, Met); a, aromatic; +, basic. Secondary structures are shown for the mouse protein kinase Cδ C1B domain (Zhang et al., 1995). Accession codes are from Swissprot (P or Q) or GenBank. Source organisms are: b, bovine; r, rat; m, murine; h, human; ap, *Aplysia californica*; dr, *Drosophila melanogaster*; ce, *Caenorhabditis elegans*; y, *S. cerevisiae*; c, chicken; di, *D. discoideum*. A copy of this alignment is available at http://www-mslmb.niddk.nih.gov/hurleygroup.html, where we plan to update it periodically to reflect new developments.

the profile, a Lys at consensus position 22 which could potentially "snorkel" out of the lipid core of the bilayer. The good fit to the typical C1 domain profile is in accord with experiments, and supports the classification of protein kinase Cμ/protein kinase D as "typical" despite their low overall sequence similarity to other protein kinase Cs.

The best-characterized atypical C1 domain is that of the Raf protein kinase, which has been shown to bind Ras and phosphatidylserine (Ghosh et al., 1996), although Ras binding to Raf also requires the N-terminal RBD domain of Raf. In addition to mammalian atypical protein kinase Cs and Raf, protein kinases with atypical C1 domains include the yeast protein kinase C homologue PKC1, the kinase-suppressor of Ras KSR, the R-mil oncogene, *Dictyostelium discoideum* myosin heavy chain kinase, and the Rho-activated coiled-coil containing kinase p160-ROCK. *Saccharomyces cerevisiae* PKC1 is activated by the *S. cerevisiae* Rho homologue

RHO1 (Kamada et al., 1996), which binds to the region containing the C1 domains (Nonaka et al., 1995). It is unknown whether the primary Rho binding site of p160-ROCK is within the C1 domain, since this sequence contains another putative Rho binding sequence outside the C1 domain (Ishizaki et al., 1996). The common thread linking these atypical C1-domain containing kinases is their interaction with small G-proteins.

Regulators and effectors of the small G-proteins Rho, Rac, and Cdc42 form another major class of C1 domain containing proteins. In addition to yeast PKC1 and p160-ROCK, these include the typical C1-containing chimaerins, which are Rho-specific GTPase activating proteins (GAPs), and the atypical C1-containing Vav, Lfc, rotund, and citron proteins. The C1 domain of the Vav proto-oncogene product is required for its transforming activity (Coppola et al., 1991), while that of Lfc is not (Whitehead et al., 1995). The oncogene proteins Vav and Lfc both contain dbl homology (DH) domains which are characteristic of RhoA and Cdc42-specific guanine nucleotide exchange factors (Hart et al., 1994), but the biochemical activity of the Vav and Lfc DH domains are unknown. Rotund is a Rac-specific GAP (Agnel et al., 1992) and citron is a putative Rac1 and RhoA effector (Madaule et al., 1995). As for the protein kinase Cs, different types of C1 domains appear to connect quite disparate signal inputs and outputs.

In conclusion, the reclassification of C1 domains within PKC as C1A and C1B acknowledges the potential for the independent action of these domains. Furthermore, the functional classification of these C1 domains as typical and atypical accounts for their observed properties either as diacylglycerol targets or not. This clearly has a predictive value that will guide future experimentation.

Note added in proof: Sakane et al. recently reported diacylglycerol kinases α, β, and γ do not bind phorbol ester (Sakane F, Kai M, Wada I, Imai S, Kanoh H. 1996. *Biochem J 318*:583–590).

References

Agnel M, Roder L, Vola C, Griffin-Shea R. 1992. A *Drosophila* rotund transcript expressed during spermatogenesis and imaginal disc morphogenesis encodes a protein which is similar to human Rac GTPase-activating (Rac-GAP) proteins. *Mol Cell Biol 12*:5111–5122.

Ahmed S, Kozma R, Lee J, Monfries C, Harden N, Lim L. 1991. The cysteine-rich domain of human proteins, neuronal chimaerin, protein kinase C and diacylglycerol kinase binds zinc. *Biochem J 280*:233–241.

Bell RM, Burns DJ. 1991. Lipid activation of protein kinase C. *J Biol Chem 266*:4661–4664.

Burns DJ, Bell RM. 1991. Protein kinase C contains two phorbol ester binding domains. *J Biol Chem 266*:18330–18338.

Coppola J, Bryant S, Koda T, Conway D, Barbacid M. 1991. Mechanism of activation of the vav protooncogene. *Cell Growth Differ 2*:95–105.

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

Diaz-Meco MT, Municio MM, Sanchez P, Lozano J, Moscat J. 1996a. Lambda-interacting protein, a novel protein that specifically interacts with the zinc finger domain of the atypical protein kinase C isotype lambda/iota and stimulates its kinase activity in vitro and in vivo. *Mol Cell Biol 16*:105–114.

Diaz-Meco MT, Municio MM, Frutos S, Sanchez P, Lozano J, Sanz L, Moscat J. 1996b. The product of par-4, a gene induced during apoptosis, interacts selectively with the atypical isoforms of protein kinase C. *Cell 86*:777–786.

Ghosh S, Strum JC, Sciorra VA, Daniel L, Bell RM. 1996. Raf-1 kinase possesses distinct binding domains for phosphatidylserine and phosphatidic acid. *J Biol Chem 271*:8472–8480.

Gribskov M, Luthy R, Eisenberg DE. 1990. Profile analysis. *Methods Enzymol 183*:146–159.

Hart MJ, Eva A, Zangrill D, Aaronson SA, Evans T, Cerione RA, Zhang Y. 1994. Cellular transformation and guanine nucleotide exchange activity are catalyzed by a common domain on the dbl oncogene product. *J Biol Chem 269*:62–65.

Hommel U, Zurini M, Luyten M. 1994. Solution structure of a cysteine-rich domain of rat protein kinase C. *Nature Struct Biol 1*:383–387.

Ichikawa S, Hatanaka H, Takeuchi Y, Ohno S, Inagaki F. 1995. Solution structure of cysteine-rich domain of protein kinase Cα. *J Biochem 117*:566–574.

Ishizaki T, Maekawa M, Fujisawa K, Okawa K, Iwamatsu A, Fujita A, Watanabe N, Saito Y, Kakizuka A, Morii N., Narumiya S. 1996. The small GTP-binding protein Rho binds to and activates a 160 kDa Ser/Thr protein kinase homologous to myotonic dystrophy kinase. *EMBO J 15*:1885–1893.

Kamada Y, Qadota H, Python CP, Anraku Y, Ohya Y, Levin DE. 1996. Activation of yeast protein kinase C by Rho1 GTPase. *J Biol Chem 271*:9193–9196.

Kanoh H, Sakane F, Imai S, Wada I. 1993. Diacylglycerol kinase and phosphatidic acid phosphatase-enzymes metabolizing lipid second messengers. *Cell Signal 5*:495–503.

Kazanietz MG, Bustelo XR, Barbacid M, Kolch W, Mischak H, Wong G, Pettit GR, Bruns JD, Blumberg PM. 1994. Zinc finger domains and phorbol ester pharmacophore: Analysis of binding to mutated form of protein kinase Cζ and the *vav* and *c-raf* proto-oncogene products. *J Biol Chem 269*:11590–11594.

Kikkawa U, Kishimoto A, and Nishizuka Y. 1989. The protein kinase C family: Heterogeneity and its implications. *Annu Rev Biochem 58*:31–44.

Lozano J, Berra E, Municio MM, Diaz-Meco MT, Dominguez I, Sanz L, Moscat J. 1994. Protein kinase Cζ isoform is critical for κB-dependent promoter activation by sphingomyelinase. *J Biol Chem 269*:19200–19202.

Madaule P, Furuyashiki T, Reid T, Ishizaki T, Watanabe G, Morii N, Narumiya S. 1995. A novel partner for the GTP-bound forms of *rho* and *rac*. *FEBS Lett 377*:243–248.

Mott HR, Carpenter JW, Zhong S, Ghosh S, Bell RM, Campbell SL. 1996. The solution structure of the Raf-1 cysteine-rich domain: A novel Ras and phospholipid binding site. *Proc Natl Acad Sci USA 93*:8312–8317.

Newton AC. 1995. Seeing two domains: C1 and C2. *Current Biol 5*:973–976.

Nonaka H, Tanaka K, Hirano H, Fujiwara T, Kohno H, Umikawa M, Mino A, Takai Y. 1995. A downstream target of *RHO1* small GTP-binding protein is *PKC1*, a homolog of protein kinase C, which leads to activation of the MAP kinase cascade in *Saccharomyces cerevisiae*. *EMBO J 14*:5931–5938.

Ono Y, Fujii T, Igarashi K, Takayoshi R, Tanaka C, Ushio K, Nishizuka Y. 1989. Phorbol ester binding to protein kinase C requires a cysteine-rich zinc-finger-like sequence. *Proc Natl Acad Sci USA 86*:4868–4871.

Szallasi Z, Bogi K, Gohari S, Biro T, Acs P, Blumberg PM. 1996. Non-equivalent roles for the first and second zinc fingers of protein kinase Cδ. *J Biol Chem 271*:18299–18301.

Valverde AM, Sinnett-Smith J, Van-Lint J, Rozengurt E. 1994. Molecular cloning and characterization of protein kinase D: A target for diacylglycerol and phorbol esters with a distinctive catalytic domain. *Proc Natl Acad Sci USA 91*:8572–8576.

Whitehead I, Kirk H, Tognon C, Trigo-Gonzalez G, Kay R. 1995. Expression cloning of *lfc*, a novel oncogene with structural similarity to guanine nucleotide exchange factors and to the regulatory region of protein kinase C. *J Biol Chem 31*:18388–18395.

Zhang G, Kazanietz MG, Blumberg PM, Hurley JH. 1995. Crystal structure of the Cys2 activator-binding domain of protein kinase Cδ in complex with phorbol ester. *Cell 81*:917–924.