J Mol Evol (1996) 43:374–383

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of the Metazoan Protein Kinase C Multigene Family

Michael Kruse,[1] Vera Gamulin,[2] Helena Cetkovic,[2] Zeev Pancer,[1] Isabel M. Müller,[1] Werner E.G. Müller[1]

[1] Abteilung Angewandte Molekularbiologie, Institut für Physiologische Chemie, Universität, Duesbergweg 6, 55099 Mainz, Germany
[2] Department of Molecular Genetics, Institute Ruder Boskovic, 10000 Zagreb, Croatia

Received: 10 January 1996 / Accepted: 12 March 1996

Abstract. Protein kinases C (PKCs) comprise closely related Ser/Thr kinases, ubiquitously present in animal tissues; they respond to second messengers, e.g., Ca^{2+} and/or diacylglycerol, to express their activities. Two PKCs have been sequenced from *Geodia cydonium*, a member of the lowest multicellular animals, the sponges (Porifera). One sponge *G. cydonium* PKC, GCPKC1, belongs to the ''novel'' (Ca^{2+}-independent) PKC (nPKC) subfamily while the second one, GCPKC2, has the hallmarks of the ''conventional'' (Ca^{2+}-dependent) PKC (cPKC) subfamily. The alignment of the Ser/Thr catalytic kinase domains, of the predicted aa sequences for these cDNAs with respective segments from previously reported sequences, revealed highest homology to PKCs from animals but also distant relationships to Ser/Thr kinases from protozoa, plants, and bacteria. However, a comparison of the complete structures of the sponge PKCs, which are—already—identical to those of nPKCs and cPKCs from higher metazoa, with the structures of protozoan, plant, and bacterial Ser/Thr kinases indicates that the metazoan PKCs have to be distinguished from the nonmetazoan enzymes. These data indicate that metazoan PKCs have a universal common ancestor which they share with the nonmetazoan Ser/Thr kinases with respect to the kinase domain, but they differ from them in overall structural composition.

Key words: Sponges — *Geodia cydonium* — Serine/threonine kinases — Phylogeny — Molecular systematics — Molecular evolution

Correspondence to: W.E.G. Müller

Introduction

Metazoan protein kinases C (PKCs) are a family of Ser/Thr kinases involved in the signaling pathway that utilizes second messengers which are generated during the breakdown of membrane phospholipids (survey: Mahoney and Huang 1994). After their discovery by Nishizuka and his colleagues (Takai et al. 1977) PKCs were identified and characterized both biochemically and by molecular cloning (Hardie and Hanks 1995). Initially, the PKCs were identified as Ca^{2+}- and phospholipid-dependent protein kinases (reviewed by Nishizuka 1988; Stabel and Parker 1991). At present, the PKCs are subdivided into three subfamilies: the ''conventional'' (Ca^{2+}-dependent) PKC (cPKC) subfamily, including PKC α/γ, βI, and βII; the ''novel'' (Ca^{2+}-independent) protein kinase C (nPKC) subfamily with PKC δ, ϵ, η, and θ; and the atypical PKC (aPKC) subfamily comprising PKC ζ, and λ (Hardie and Hanks 1995). The physiological function of the latter group is not yet clear.

Experimental evidence has been presented which shows that the PKC-mediated signaling pathway(s) also exists in the lowest multicellular animals, the sponges (Porifera) (Müller et al. 1987; Weissmann et al. 1988; Müller et al. 1990). After incubation of dissociated cells from both *Geodia cydonium* (Müller et al. 1987) and *Microciona prolifera* (Weissmann et al. 1988) with the homologous aggregation factor, the PKC is activated and causes an induction of DNA synthesis via activation of DNA topoisomerase II (Daum et al. 1994).

Based on sequencing data obtained from adhesion molecules and receptors of the sponge *G. cydonium* [a lectin (Pfeifer et al. 1993b), the receptor tyrosine kinase (Schäcke et al. 1994b) comprising extracellularly immu-

noglobulin boxes (Schäcke et al. 1994a) and a homeobox-like gene (Kruse et al. 1994)] it can be deduced that the Porifera branched off from the common metazoan tree over 600 million years ago (Müller 1995). Furthermore, the existence of conserved proteins, structurally and functionally similar to higher metazoan proteins, supports the view that the Porifera and other ''classical'' metazoan animals are of monophyletic origin (Müller 1995).

The phylogenetic relationships of PKCs within the metazoan and nonmetazoan kingdoms have not been presented due to the lack of sequences, especially from species of the lowest animal phylum. In this study we report two nucleotide sequences from *G. cydonium,* one belonging to the cPKC and the second one to the nPKC subfamily. The deduced amino acid (aa) sequences allowed establishment of the evolutionary relationships of the kinase domains among members of this multigene subfamily, not only among PKCs from metazoa but also with Ser/Thr kinases from plants, protozoa, and bacteria. In addition, it is shown that the overall composition of the metazoan PKCs is different from that of nonmetazoan Ser/Thr kinases.

Materials and Methods

Materials. Enzymes for recombinant DNA techniques and vectors were obtained from Stratagene (Heidelberg, Germany).

Sponge. The specimens of *Geodia cydonium* (Porifera, Demospongiae, Geodiidae) were collected near Rovinj (Croatia). The material was immediately frozen in liquid nitrogen and kept in that state until use.

Isolation of Ser/Thr Kinases. cDNA coding for Ser/Thr kinases was isolated from a cDNA library from *G. cydonium* (Pfeifer et al. 1993a).

Screening of the library was performed under low-stringency hybridization conditions of plaque lifts from 3×10^5 pfu on nitrocellulose using the complete PKC-cDNA, isolated from mouse fibroblasts (M25811) (Rose-John et al. 1988), as a probe. Filters were hybridized overnight at 42°C in 35% formamide, 5 × SSC, 0.02% $NaDodSO_4$, 0.1% *N*-laurylsarcosine, and 1% blocking reagent. The dsDNA restriction fragment isolated by microelution (Ausubel et al. 1995) was labeled with digoxygenin-11-dUTP using the random primed labeling kit (Boehringer). Filters were washed twice in 2 × SSC, 0.1% $NaDodSO_4$ at room temperature, followed by additional washing in 0.1 × SSC, 0.1% $NaDodSO_4$ (42°C). Positive clones were detected with an alkaline-phosphatase-conjugated antidigoxygenin antibody using BCIP/NBT as substrate (Blake et al. 1984). Single phage plaques were obtained by three additional screening cycles. Following an *in vivo* excision procedure described by Stratagene, phagemids (pBluescript SK^-) were excised from lambda phages using the filamentous helper phage R408 and the *E. coli* strain XL-1-blue. The positive PKC clones for *G. cydonium* were termed *GCPKC1* and *GCPKC2.*

DNA Sequencing. dsDNA was sequenced by the dideoxy chain termination method (Sanger et al. 1977). After analysis of the 5′ and 3′ ends of the sequence, subclones were constructed either by ligation of restriction fragments or by creating unidirectional deletions and analyzed by end-over-end sequencing.

Sequence Analysis. Prediction of sites and signatures has been performed with programs available in PC/GENE (1995). Homology searches were performed via the E-mail servers at the European Bioinformatics Institute, Hinxton Hall, UK (BLITZ@ebi.ac.uk and FASTA@ebi.ac.uk) and the National Center for Biotechnology Information, National Institutes of Health, MD, USA (BLAST@ncbi.nlm.nih.gov). Sequence comparison as well as the establishment of the dendrogram was achieved with the CLUSTAL program (Higgins and Sharp 1988). This program uses pairwise similarity scores to build the dendrogram.

Results and Discussion

Primary Structures of the Sponge cDNAs Coding for Ser/Thr Kinases

The mouse cDNA, coding for PKCα (Rose-John et al. 1988), was used to identify and isolate the corresponding cDNA clones from the marine sponge *G. cydonium.* Two different sequences have been identified; they are termed *GCPKC1* and *GCPKC2.* Five (eight) independent clones all leading to the same sequence *GCPKC1* (*GCPKC2*) have been analyzed.

Clone *GCPKC1* contained a 2,403-bp-long cDNA insert; *GCPKC2* is 2,142 nts long. The nucleotide (nt) sequence as well as the deduced aa sequence of sponge Ser/Thr kinase *GCPKC1* cDNA is shown in Fig. 1, and for *GCPKC2* cDNA in Fig. 2. The open reading frame (ORF) for *GCPKC1* with the ATG-codon for Met is 2,211 bp long (Fig. 1). The typical signal polyadenylation site AATAAA (Zarkower et al. 1986) is present. The ORF for *GCPKC2* has a size of 2,031 nts. The typical polyadenylation site is missing, a fact which was noted already earlier in some cDNAs from *G. cydonium,* like those coding for the S-type lectins (Pfeifer et al. 1993b).

Deduced Amino Acid Sequences of Sponge Ser/Thr Kinases

The aa sequences of sponge Ser/Thr kinases have been deduced from the cDNAs GCPKC1 and GCPKC2 (Figs. 1 and 2). Homology searches both with BLITZ and BLAST programs revealed that the sequences from GCPKC1 and from GCPKC2 must both be classified as belonging to the group of PKC. The two sequences contain the typical Ser/Thr kinase active-site signature (Hanks and Quinn 1991): IIYRDLKLDNVIL (aa 528–540) in GCPKC1 (Fig. 1) and IIYRDLKLDNVLL (467–479) in GCPKC2 (Fig. 2).

GCPKC1

The deduced aa sequence of GCPKC1 (Fig. 1) encodes a 83.2-kDa primary translation product with an estimated pl of 8.3; the instability index is 41.6, indicative for an unstable protein. Northern blot analysis revealed a size for the transcript of *GCPKC1* of 2.8 kb (not

```
GCPKC1 GGTTAAGTTGAGGCCTACTAGCTGACAGCGCAATGGCGTTCGTACGGATCAAGCTGCTGGAGGCGATCGTGGACCACGTCAAGCCCACGG      57
                                        M  A  F  V  R  I  K  L  L  E  A  I  V  D  H  V  K  P  T       19
GCPKC1 ATCCCACTTGCTCTGTCAACATCAAGGAGGCCCTTGCGGGAGAGGACGGGAGAGTCACGTTGGAGCAGCGCAAGAAGACGTTCTTTCCAG     147
       D  P  T  C  S  V  N  I  K  E  A  L  A  G  E  D  G  R  V  T  L  E  Q  R  K  K  T  F  F  P       49
GCPKC1 ACTGGGATCGCTGTTTCGACTCTCACCTAAAGCCGGGTCGACGCATGCAGATCATCGTCAACGACCGCGTGGAATCCTCCCTGCGCCCGC     237
       D  W  D  R  C  F  D  S  H  L  K  P  G  R  R  M  Q  I  I  V  N  D  R  V  E  S  S  L  R  P       79
GCPKC1 TGGCTGAAGTGACGGTGGAAACCGAGGCGTTGGCGACAGAATGTATGGCAGAGGAGGAAGGCAGTGCCGTCAAACTTGCCTTGGACATGA     327
       L  A  E  V  T  V  E  T  E  A  L  A  T  E  C  M  A  E  E  E  G  S  A  V  K  L  A  L  D  M       109
GCPKC1 GACCATCAGGAAAGATGATTCTTCAGGTCAAGCTCTATGGCAGAGAGCACATTGAAGGGCGAGACCTGGCAGTACTGACACCGGAATGGA     417
       R  P  S  G  K  M  I  L  Q  V  K  L  Y  G  R  E  H  I  E  G  R  D  L  A  V  L  T  P  E  W       139
GCPKC1 AGGAATCTCTTCCAAAGAACGCCACAGCTCTGCGAGGGAGACGAGGGGCCATGCACGTGAAGCAGGCACACGTGGAAGACATAAAAGGAC     507
       K  E  S  L  P  K  N  A  T  A  L  R  G  R  R  G  A  M  H  V  K  Q  A  H  V  E  D  I  K  G       169
                                  ††††††††††††††††††††††††††††††††††††††††††††††††††††††††††††††††
GCPKC1 ACCAGTTTGTCAAGAGGTTCTTCCGACGAGCAATTTACTGCTCCCTCTGTCACGAATTCCTCTGGGGATTTACAAAGCAAGGCTACCAGT     597
       H  Q  F  V  K  R  F  F  R  R  A  I  Y  C  S  L  C  H  E  F  L  W  G  F  T  K  Q  G  Y  Q       199
                                              ‡        ‡
       .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
GCPKC1 GCCAAGTGTGCCATTATACAGCCCACAAGAAATGTACTGGCTCAATCCTTGCCAAATGTACTGGAGCTACTGGCAATGAGACTCACTCTA     687
       C  Q  V  C  H  Y  T  A  H  K  K  C  T  G  S  I  L  A  K  C  T  G  A  T  G  N  E  T  H  S       229
       .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
       ‡        ‡                       ‡                    ‡
GCPKC1 AGTTCCTCAGGGAAAGGTTCAAGATTGACATGCCTCATCGGTTCAAAGTCCACAATTTTCTTGGCCCGTCTTTCTGTGACATGTGTGGCC     777
       K  F  L  R  E  R  F  K  I  D  M  P  H  R  F  K  V  H  N  F  L  G  P  S  F  C  D  M  C  G       259
                                                                                 ‡        ‡
GCPKC1 AAATGATGCACGGGATCTTCCGCCAGGGAGCCAAGTGCACAGCGTGTGGTGTGTGTTGCCACATTCGCTGCCAGAAGAACATGCCCCCGC     867
       Q  M  M  H  G  I  F  R  Q  G  A  K  C  T  A  C  G  V  C  C  H  I  R  C  Q  K  N  M  P  P       289
                                        ‡        ‡                    ‡
GCPKC1 TGTGTGGAGTCAATGAGAAGATGTTGGCTGAAGCTCTCAAGAGTGTCGACGAACTCAAGAGAAACAGAAGACTGTCCGCAGGGTCGGATC     957
       L  C  G  V  N  E  K  M  L  A  E  A  L  K  S  V  D  E  L  K  R  N  R  R  L  S  A  G  S  D       319
          ‡
GCPKC1 CGGCGACCACTCCGGGGAGTCCTGGTGCCAAGCCCCTCCCACCTGTCCCCGAGGGGGAGTCAGAGGAGTACATCGAGGTCACTGAAGCCA    1047
       P  A  T  T  P  G  S  P  G  A  K  P  L  P  P  V  P  E  G  E  S  E  E  Y  I  E  V  T  E  A       349
GCPKC1 TGACTCGAGCAGTGTTAGGTCAAACATACAATCTCCCTGGACGAGACGGAGCCCCGCCCATCCCTCCTCGGACGTACAGCCAGCGCGGAG    1137
       M  T  R  A  V  L  G  Q  T  Y  N  L  P  G  R  D  G  A  P  P  I  P  P  R  T  Y  S  Q  R  G       379
GCPKC1 GACACACCTCCAACGGCCACATCACATCGGGCAGCTTCCACGGCTTTGGGCGACCGGCGATGAAGAAGTACAAACTGGAGCAGTTCAAGT    1227
       G  H  T  S  N  G  H  I  T  S  G  S  F  H  G  F  G  R  P  A  M  K  K  Y  K  L  E  Q  F  K       409
GCPKC1 TTCTCAAACTCCTTGGGAAAGGAAGCTTTGGGAAGGTCTTGCTGGCCCAACTGGAGGGGAATGAGCAGTATTTTGCCATCAAGGCCCTCA    1317
       F  L  K  L  L  G  K  G  S  F  G  K  V  L  L  A  Q  L  E  G  N  E  Q  Y  F  A  I  K  A  L       439
       #===== catalytic domain
GCPKC1 AGAAGGACGTAGTACTGGAGGACGATGACGTGGAGGCCACTATGGTGGAGAAGAGACTTCTCGCTCTTGGCTGTAACCATCCTTTCCTCA    1407
       K  K  D  V  V  L  E  D  D  D  V  E  A  T  M  V  E  K  R  L  L  A  L  G  C  N  H  P  F  L       469
GCPKC1 CTCACCTCCACTCCACCTTCCAGACCCCCAGTCACCTGTTCTTTGTGATGGAGTATCTGAATGGCGGCGATCTCATGTATCACATACAGA    1497
       T  H  L  H  S  T  F  Q  T  P  S  H  L  F  F  V  M  E  Y  L  N  G  G  D  L  M  Y  H  I  Q       499
GCPKC1 TTTCTCACAAATTCAAACTCCCCAGAGCAAGGTTCCATGCTGCAGAAATACTCTGTGCTCTTCAGTTCCTCCACAAACAAGGCATTATAT    1587
       I  S  H  K  F  K  L  P  R  A  R  F  H  A  A  E  I  L  C  A  L  Q  F  L  H  K  Q  G  I  I       529
GCPKC1 ACAGAGATCTCAAGTTGGACAACGTGATACTGGACTCTGAGGGTCACTGTAAACTGGCCGACTTTGGCATGTGCAAGGAGAACATCATTG    1677
       Y  R  D  L  K  L  D  N  V  I  L  D  S  E  G  H  C  K  L  A  D  F  G  M  C  K  E  N  I  I       559
GCPKC1 GGTATGCCACTGCAGGCACCTTCTGTGGGACACCAGACTACATATCACCAGAGATCATAAAGGGGAAGAGGTATACATTCTCTGTGGACT    1767
       G  Y  A  T  A  G  T  F  C  G  T  P  D  Y  I  S  P  E  I  I  K  G  K  R  Y  T  F  S  V  D       589
GCPKC1 GGTGGTCTTTTGGAGTCCTCTGCTACGAGATGATTACCGGCCAGTCTCCATTCAGTGGAGAGGATGAGGACGAGTTGTTTGACTCAATCT    1857
       W  W  S  F  G  V  L  C  Y  E  M  I  T  G  Q  S  P  F  S  G  E  D  E  D  E  L  F  D  S  I       619
GCPKC1 GCAACCATCAGGTCTCCTTCTCTCGCTACCTCGACCAAACCACCATCAACTTCCTTGACAAGTTGTTGCAGAGAGATCCAGGGGAGAGGC    1947
       C  N  H  Q  V  S  F  S  R  Y  L  D  Q  T  T  I  N  F  L  D  K  L  L  Q  R  D  P  G  E  R       649
GCPKC1 TGGGCTTGCATCGAAGACAGAGAGAATATCAGAGCACACGCCTTCTTCAGAGAAATTTGACTGTTGTCAAACTGGAGGCCAGGAAACTAA    2037
       L  G  L  H  R  R  Q  R  E  Y  Q  S  T  R  L  L  Q  R  N  L  T  V  V  K  L  E  A  R  K  L       679
                        catalytic domain =====#
GCPKC1 AACCTCCCTTCAAACCCAATGTGAAAGGGGCGTCAGATGCCAGCAACTTTGGCGATGATTTCACCTTCCAGCCAGCCCAGCTCACTCCCA    2127
       K  P  P  F  K  P  N  V  K  G  A  S  D  A  S  N  F  G  D  D  F  T  F  Q  P  A  Q  L  T  P       709
GCPKC1 CTGACACCACACTGGTGATGTCCATCGACCAAACAAACTTCACCGGGTTCTCATTCACTAGTGACCTCTATTCCAAACTCCACTGATATT    2217
       T  D  T  T  L  V  M  S  I  D  Q  T  N  F  T  G  F  S  F  T  S  D  L  Y  S  K  L  H  -         737
GCPKC1 ATCTAACRCTGCTATATATTATAATTGTAATGATTGTGATCTGCCTAAGGCTCAGCTGTAAGTAGAGCGTCTCTCATTTTCTCTGTTTTG    2037
GCPKC1 TAAGTGCTGCTACAGTATGCTGACAATGTGTTTTTAACTGATGTGATCATTTTTTATCCCGAGTGAGTGTATTGTACAATAATAAATGTT    2397
                                                                                  □  □
GCPKC1 GTGTTA(n)                                                                                    2403
```

Fig. 1. Nucleotide sequence (2.4 kb long) and the deduced aa sequence of the *G. cydonium* cDNA *GCPKC1*. The pseudosubstrate segment (†), the zinc fingers (‡), and the phorbol esters/diacylglycerol-binding domain are indicated (···). The boundaries of the catalytic domain at the extreme amino- and carboxy-terminal residues (#) are marked; the protein kinase ATP-binding region signature is *underlined* and the Ser/Thr protein kinase active-site signature is *double underlined. Open squares* mark the putative polyadenylation signal. This sequence has been assigned the EMBL accession number X87684.

shown), indicating that the sequence isolated is of full length.

The sponge Ser/Thr kinase GCPKC1 displays the following boxes. (1) The pseudosubstrate segment (aa 150–169) with its Ala residue surrounded by basic, positively charged aa (Stabel and Parker 1991); (2) the typical phorbol esters/diacyglycerol-binding domain (170–219) with the consensus $Hx_1Fx_{10}Cx_2Cx_3Lx_7YxCx_2Cx_4Hx_2$-$Cx_7C$, (PC/GENE 1995); (3) two zinc fingers of the PKC motif (Stabel and Parker 1991) $Cx_2Cx_{13}Cx_2Cx_7Cx_7C$ (first: 183-186-200-203-211-219; second: 255-258-272-275-283-291); and (4) the Ser/Thr kinase catalytic domain (Hanks and Quinn 1991) (410–662). It is delimited at the amino-terminal boundary Phe (410), which is located seven residues upstream from the aa Gly in the protein kinase ATP-binding region signature which reads

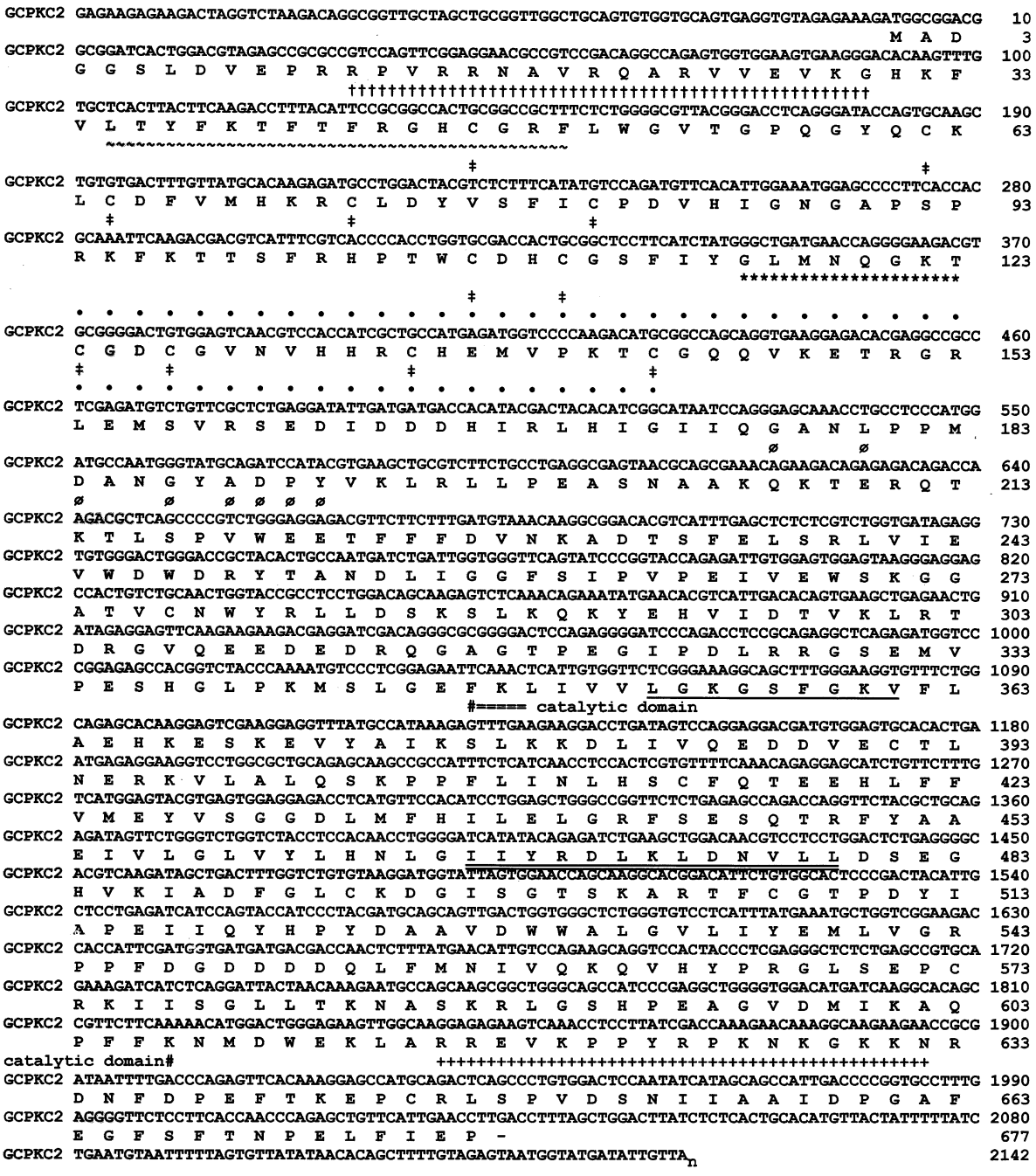

Fig. 2. The 2.1-kb-long nucleotide sequence of the second *G. cydonium* Ser/Thr kinase cDNA, *GCPKC2*, and its deduced aa sequence are shown. The pseudosubstrate segment (†), the ''Greek key'' motif (~~~), the zinc fingers (‡), the phorbol esters/diacylglycerol-binding domain (···), the C2 domain signature (ø), the bipartite nuclear targeting sequence (+), and the ATP/GTP-binding site motif (*) are indicated. The boundary of the catalytic domain (#), the protein kinase ATP-binding region signature (*underlined*), and the Ser/Thr protein kinase active-site signature (*double underlined*) are marked. Accession number X87683.

LGKGSFGKV (414–422) (Hanks and Quinn 1991). The Ser/Thr protein kinase active-site signature is found from aa 528 to 540. The carboxy-terminus of the domain, Arg at position 662, lies six aa downstream of the invariant Arg.

Since the deduced aa of GCPKC1 contains, in addition, (1) the Ser/Thr kinase catalytic domain, including C3, the putative ATP-binding domain (aa 409–462), and C4, the part of the catalytic domain (467–662); (2) the pseudosubstrate segment (150–169); (3) the phorbol esters/diacylglycerol-binding domain (170–219); and (4) two zinc fingers (183–219 and 255–291) (Fig. 3), GCPKC1 has to be classed as a member of the ''novel'' (Ca^{2+}-independent) protein kinase C (nPKC) subfamily (Stabel and Parker 1991; Hardie and Hanks 1995). The nomenclature chosen follows Stabel and Parker (1991).

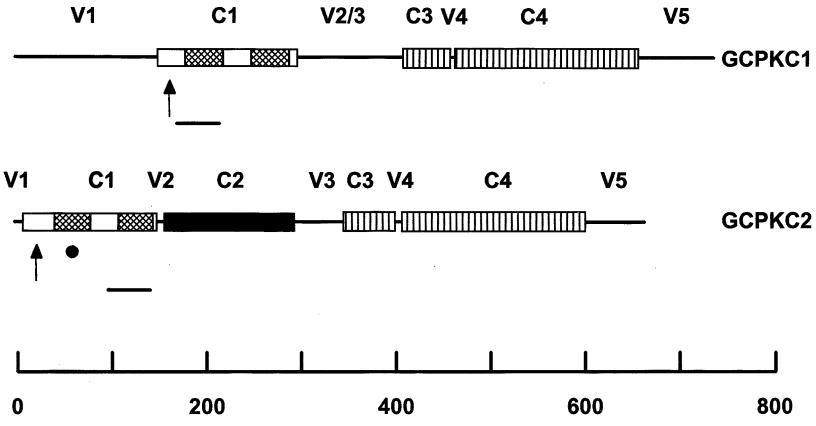

Fig. 3. Domain structure of both deduced aa of GCPKC1 and of GCPKC2. Due to the lack of the C2 domain, but the presence of two zinc fingers within the C1 domain, GCPKC1 belongs to the ''novel'' (Ca^{2+}-independent) protein kinase C subfamily, while GCPKC2, which contains the C2 domain (*filled bar*), has to be grouped to the ''conventional'' (Ca^{2+}-dependent) protein kinase C subfamily. C3 (*striped boxes*) includes the putative ATP-binding domain and C4 (*striped boxes*) is part of the catalytic domain. The conserved (*C*) and variable (*V*) domains are indicated, according to Stabel and Parker (1991). The two *chequered boxes* within the C1 domain show the conserved zinc finger motifs. *Arrow:* pseudosubstrate; *horizontal line:* phorbol esters/diacylglycerol-binding domain; ●: Greek key motif. The *numbers* give the size of the respective sequences in aa.

GCPKC2

The M_r of the deduced aa sequence of *GCPKC2* (Fig. 3) is 76.8 kDa and the pl is 6.5; the instability index is 43.3, characteristic for an unstable protein. Based on Northern blot analysis—size for the transcript: 2.5 kb (not shown)—the *GCPKC2* sequence is full length.

The following units are present in sponge Ser/Thr kinase GCPKC2: (1) the pseudosubstrate segment (13–30) and (2) the ''Greek key'' motif (35–50) with its typical sequence $Lx_2FKx_3Fx_1Gx_4F$. This motif is frequently found in antiparallel β-sheets when four adjacent β-strands are arranged in one to two, two to three, and one to four orientation and might form a hinge (Branden and Tooze 1991). In addition the following typical signatures are present: (3) The typical phorbol esters/diacylglycerol-binding domain (94–143) with $Fx_1Fx_{10}Cx_2Cx_3Ix_2Lx_6Cx_2Cx_4Hx_2Cx_7C$ (PC/GENE 1995), (4) two zinc fingers [first, an incomplete one: (Arg instead of Cys)44-47-62-65-73-81; second: 107-110-124-127-135-143]; (5) an ATP/GTP-binding site motif A (P-loop) Gx_4GKT (116–123) usually not found in PKCs; (6) the C2 domain signature, which is likely involved in Ca^{2+} binding (Stabel and Parker 1991), $Gx_2Lx_3Dx_2GxA$(replaced in GCPKC2 by SDPY (177-180-184-187-189-190-191-192) (PC/GENE 1995); and (7) the Ser/Thr kinase catalytic domain (347–607). The Ser/Thr kinase catalytic domain has the borders Phe (347), located adjacent to the protein kinase ATP-binding region LGKGSFGKV (353–361), and Lys (607). The Ser/Thr protein kinase active-site signature is found at aa 467–479.

Finally, (8) the deduced aa of GCPKC2 shows at the C-terminus a bipartite nuclear targeting sequence, RREVKPPYRPKNKGKKN (616–632); it follows the pattern (R/L)-(R/L)-x_{10}-yyyyy whereby in the y-block at least the residue R or L must occur. Based on findings obtained in higher metazoa that both the inositol lipid cycle (Cocco et al. 1994) and some selected protein kinase C isoforms are found in the nucleus (Schröder et al. 1988; Disatnik et al. 1994), it seems likely that also in sponges related enzymes are present in this organelle.

As GCPKC2 contains, besides the C3 (346–401) and the C4 domain (406–607), a C1 domain (12–153), including the pseudosubstrate segment (13–30) and the two zinc fingers (44–81; 107–143), and in addition a C2 domain (171–295), GCPKC2 belongs to the ''conventional'' (Ca^{2+}-dependent) protein kinase C (cPKC) subfamily (Stabel and Parker 1991; Hardie and Hanks 1995) (Fig. 3).

Ser/Thr Kinase Catalytic Domain

The Ser/Thr kinase catalytic domains of GCPKC1 (GEODIA1) and GCPKC2 (GEODIA2) show all the hallmarks known from sequences of Ser/Thr kinases isolated from higher metazoa (Hanks et al. 1988; Hanks and Quinn 1991) (Fig. 4). The characteristics of the different subdomains, which are all present in the sponge sequences, are the following: Subdomain I contains the ATP-binding signature GxGxxG (the numbers for the respective aa refer to sequence GCPKC1: 415-417-420; Fig. 1); subdomain II contains the invariant Lys (437), required for maximal enzymic activity; subdomain III contains the invariant Glu (456), necessary for ATP binding; subdomain IV contains no invariant; subdomain V contains the residues Met, Tyr, and Leu (486-488-489) involved in the formation of the hyhrophobic pocket surrounding the adenine ring; subdomain VIa has supporting function; subdomain VIb contains the consensus motif $LRDLKx_2N$ (530-531-532-533-534-537) of the catalytic loop; subdomain VII contains the highly conserved triplet DFG (550–552); subdomain VIII, which is the major segment involved in recognition of the substrate, contains the highly conserved S/APE motif (575–577); subdomain IX contains the invariant Asp residue

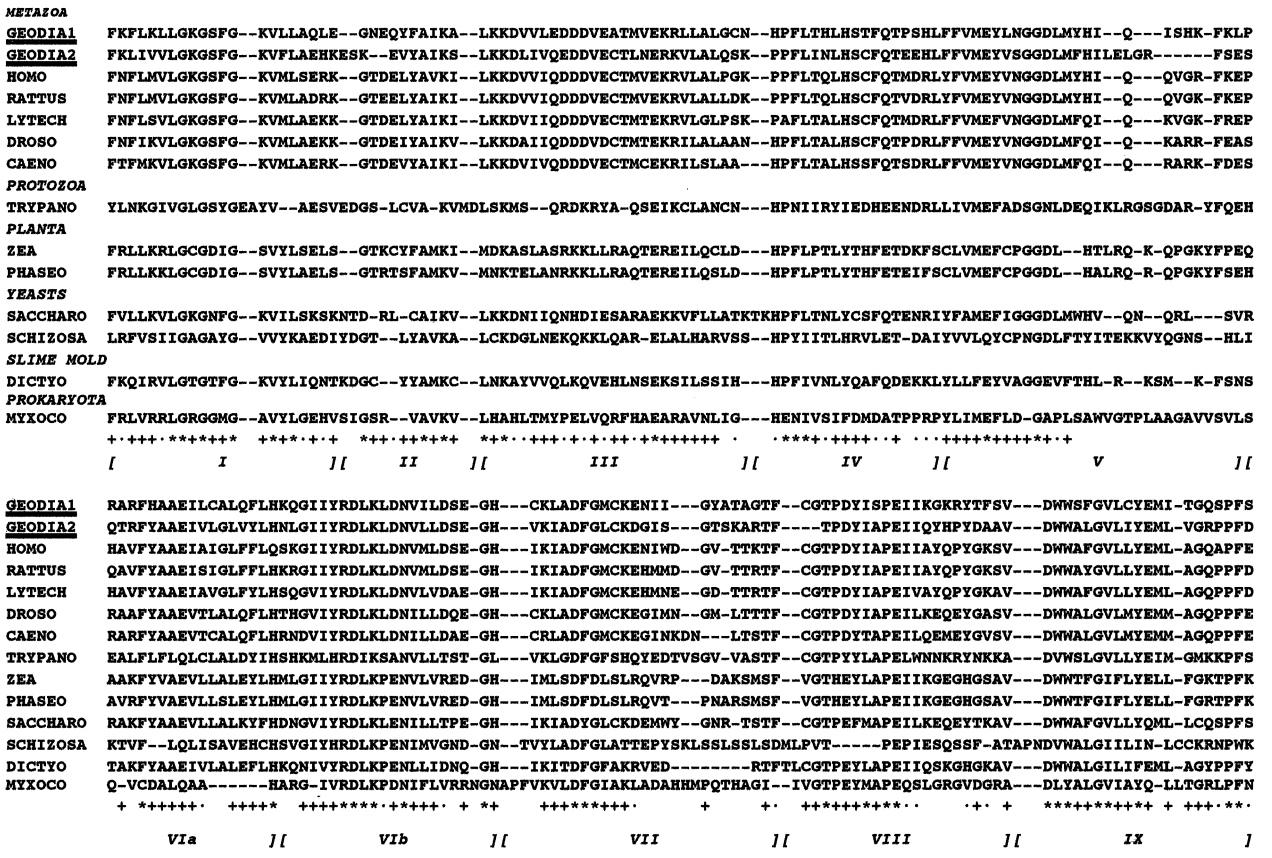

Fig. 4. Alignment of catalytic domains of Ser/Thr kinases from eukaryotic and prokaryotic organisms for construction of the phylogenetic tree. The following sequences have been selected; the last aa position is given. Metazoa: *Homo sapiens* protein kinase C (HOMO; accession number P05127; aa: 542) (Kubo et al. 1987), *Rattus norvegicus* (RATTUS; X07286; aa: 538) (Kikkawa et al. 1987), *Lytechinus pictus* (LYTECH; LP02967; aa: 525) (Rakow and Shen, submitted), *Drosophila melanogaster* (DROSO; P13678; aa: 502) (Schaeffer et al. 1989), *Caenorhabditis elegans* (CAENO; PKC2_CAEEL; aa: 398) (Land et al. 1994), *Geodia cydonium* [GEODIA1 and GEODIA2; GCPKC1 (X87684), GCPKC2 (X87683); aa: 608 and 547]. Protozoa: *Trypanosoma brucei* (TRYPANO; L03778; aa: 225) (Gale and Parsons 1993). Planta: *Zea mays* (ZEA; M62985; incomplete aa sequence) (Biermann et al. 1990), *Phaseolus vulgaris* (PHASEO; J04555; aa: 507) (Lawton et al. 1989). Yeast: *Saccharomyces cerevisiae* (SACCHARO; M32491; aa: 1025) (Levin et al. 1990), *Schizosaccharomyces pombe* (SCHIZOSA; X04728; aa: 240) (McLeod and Beach 1986). Slime mold: *Dictyostelium discoideum* (DICTYO; M38703; aa: 532) (Bürki et al. 1991). Prokaryota: *Myxococcus xanthus* (MYXOCO; M73498; aa: 262) (Munoz-Dorado et al. 1991). Sequence comparison was achieved first using the CLUSTAL program and subsequently by eye inspection to obtain maximum homology. Identities and homologies >90% of related aa (*) and homologies between 50 and 90% (+) and between 40 and 50% (·) are indicated; deletions introduced are denoted by (−). The nomenclature for the subdomains follows that of Hardie and Hanks (1995); they are given in *brackets.*

(589); subdomain X is less conserved; and subdomain XI contains invariant Arg (657) close to the C-terminus of the catalytic domain.

Phylogenetic Analysis of Ser/Thr Kinase Domain

Twelve out of 135 Ser/Thr kinase sequences listed in the EMBL Sequence Database (1994) as well as in the summary given by Hardie and Hanks (1995) have been selected for analysis. Only those sequences from different phyla have been included which show the Ser/Thr kinase signatures. Two mammalian Ser/Thr kinases (rat and human), PKCs, have been selected. In addition, one further kinase from a deuterostomian species has been included from *Lytechinus pictus.* From invertebrates, sequences to be grouped into the cPKC or the nPKC subfamily are known from *Drosophila melanogaster*, *Aplysia californica*, and *Caenorhabditis elegans*. From yeasts, sequences that are thought to belong to the same groups have been cloned from *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Hardie and Hanks 1995). The other kinase sequences used for the comparison are from protozoa, plants, slime mold, and bacteria; they belong to other subfamilies of Ser/Thr kinases. The representatives chosen from them display the highest similarities to the sequences from *G. cydonium.*

The following sequences, spanning the Ser/Thr kinase domain, have been used for the comparison: Eukaryota: Metazoa: Deuterostomia: Chordata: *Homo sapiens* [Ca^{2+}- and phospholipid-dependent PKC βII, according to the classification of Hardie and Hanks (1995)], *Rattus norvegicus* [Ca^{2+}- and phospholipid-dependent PKC ζ].

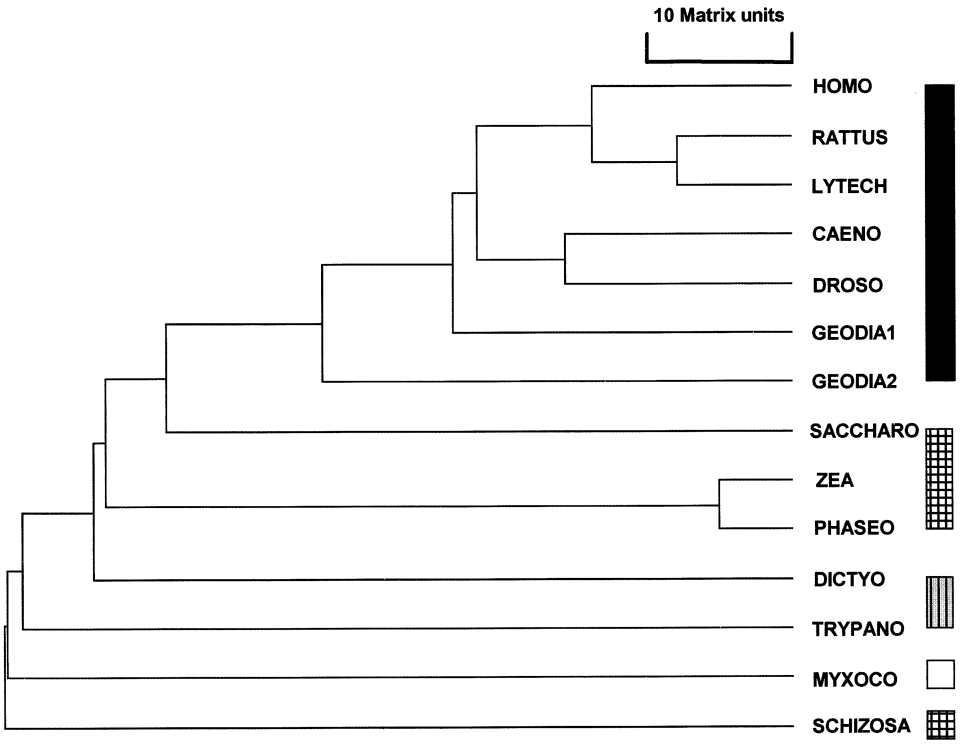

Fig. 5. Phylogenetic tree based on similarity-matrix analysis of the Ser/Thr kinase domain present in sequences of the following 14 organisms. Metazoa (*black bar*): HOMO (*Homo sapiens*), RATTUS (*Rattus norvegicus*), the sea urchin LYTECH (*Lytechinus pictus*), CAENO (*Caenorhabditis elegans*), DROSO (*Drosophila melanogaster*), and the sponge GEODIA1 and GEODIA2 (*Geodia cydonium*); planta/fungi (*crosshatched*) SACCHARO (*Saccharomyces cerevisiae*), PHASEO (*Phaseolus vulgaris*), ZEA (*Zea mays*), and SCHIZOSA (*Schizosaccharomyces pombe*); protozoa (*vertically*) DICTYO (*Dictyostelium discoideum*), TRYPANO (*Trypanosoma brucei*), prokaryota/bacteria (*open bar*) MYXOCO (*Myxococcus xanthus*). For sequence comparison the CLUSTAL program (PC/GENE 1995) was applied; the relationships between the species, based on the number of identical amino acids, are given in matrix units.

Echinodermata: *L. pictus* (unclassified); Protostomia: Arthropoda: *D. melanogaster* (cPKC), Nematoda: *C. elegans* (nPKC); Protozoa: Flagellata: *Trypanosoma brucei* (''other protein kinase family''); Planta: Dicotyledoneae: Leguminosae: *Phaseolus vulgaris* (belonging to the AGC group of Ser/Thr kinases), Monocotyledoneae: Glumiflorae: *Zea mays* (belonging to the AGC group of Ser/Thr kinases); Mycophyta/yeast: Eumycophyta: Eumycetes: *S. cerevisiae* (nPKC ?), *S. pombe* (nPKC ?), Myxomycetes/protozoa: *Dictyostelium discoideum* (cyclic nucleotide-regulated protein kinase family) and from Prokaryota/bacteria: Gracilicutes: Scotobacteria: Myxobacterales: *Myxococcus xanthus* (''other protein kinase family''). They are compared with the sequences from the poriferan, metazoan organism, *G. cydonium* [GCPKC1 (GEODIA1) nPKC; GCPKC2 (GEODIA2) cPKC]. Due to the low sequence identity with these sequences, the one from *A. californica* was not included (<15% aa similarity to *G. cydonium* sequences).

Amino acid sequences were truncated prior to analysis in order to exclude regions that are less conserved. Therefore, we restricted the sequences to the Ser/Thr kinase catalytic domain. Analysis was performed first with the CLUSTAL program. At this stage it was apparent that the subdomains X and XI are less suitable for inclusion in the comparison; thus, only subdomains I–IX were chosen. The plant sequences *P. vulgaris* and *Z. mays* contain, between subdomains VII and VIII, an interspersed sequence of 83 and 81 nts, respectively; these were excluded also. The maximum number of matches was established by eye inspection, orienting at the positions of the highly conserved aa (Fig. 4). The values for the overall identity/homology were identity or >90% homology (with respect to aa belonging to the same group), 17.5%; homologies between 50 and 90%, 40.4%; and 40–50% homology, 13.1%.

The degree of identity, given as the number of identical aa per 100, has been calculated. The results revealed that the metazoan PKCs obtained from *H. sapiens, R. norvegicus, L. pictus, D. melanogaster, C. elegans*, and *G. cydonium* are closely related among themselves and show only lower homology to the Ser/Thr kinases of (1) the protozoan sequence from *T. brucei*, (2) the plant sequences *P. vulgaris* and *Z. mays*, (3) the yeast sequences from *S. cerevisiae* and *S. pombe*, and (4) the slime mold/protozoa *D. discoideum*. (5) The Ser/Thr kinase from the bacterial organism *M. xanthus* still displays a 27% identity with the corresponding human sequence (Fig. 5).

After calculating the pairwise similarity scores (PC/

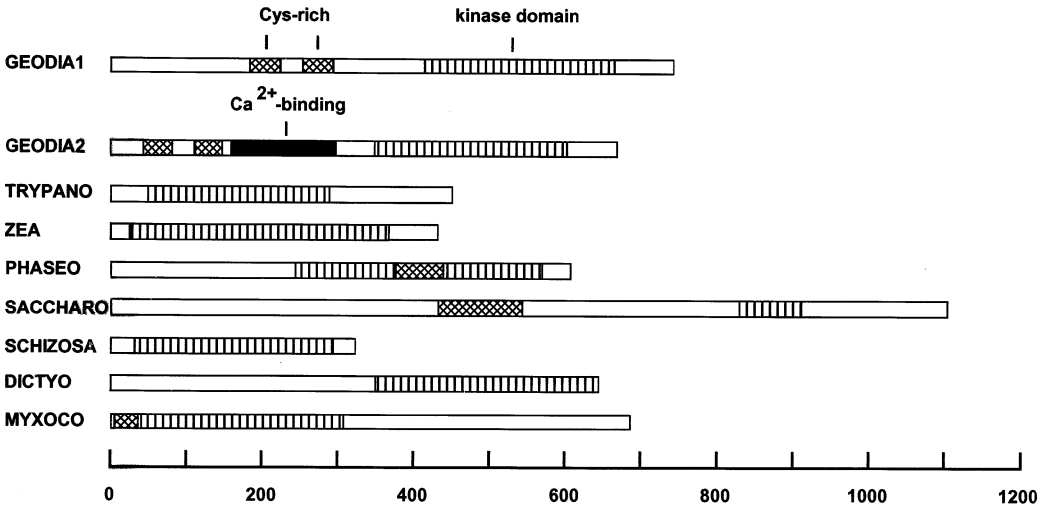

Fig. 6. Comparison of the domain structures of the metazoan PKC, here with the examples of the sponge GCPKC1 and GCPKC2, with the nonmetazoan Ser/Thr kinases. The origins of the sequences are shown in the legend to Fig. 5. The presentation, according to Hardie and Hanks (1995), in Cys-rich repeats, C2 [Ca^{2+}-binding site], and kinase domain is applied. The length of the segments/total deduced protein is given in a *scale* of aa numbers.

GENE 1995) the relationships among the PKCs are depicted in a dendrogram (Fig. 5). It shows that all selected metazoan sequences are closely related and fall into one branch. Separated from them are the sequences derived from the plants *S. cerevisiae, Z. mays,* and *P. vulgaris.* As expected, only distantly related to the metazoan sequences are those from Protozoa (*D. discoideum* and *T. brucei*). The lowest percent identity with the selected Ser/Thr kinase is the one obtained from the bacterium *M. xanthus* and the fission yeast *S. pombe.* The data also show that the sequences from *S. cerevisiae* and *S. pombe* are only distantly related to each other, suggesting a very early divergence of these two fungi from each other.

Phylogenetic Analysis of the Structure of the PKCs

As summarized (Stabel and Parker 1991; Hardie and Hanks 1995), the isoforms of the mammalian PKCs are separated into the cPKC, nPKC, and aPKC. The two sequences identified in *G. cydonium* fall exactly into the groups of metazoan PKCs—GCPKC1 into the subfamily nPKC and GCPKC2 into subfamily cPKC.

A comparison of these two sequences with the nonmetazoan Ser/Thr kinases revealed distinct differences (Fig. 6). Unlike in GCPKC2, none of the other kinases shows a typical C2 domain, and unlike in GCPKC1 and GCPKC2, there were no conventional conserved zinc finger motifs. *T. brucei* and *Z. mays* kinase consist only of a kinase domain (20–285 and 29–370, respectively); *P. vulgaris* kinase shows within its kinase domain (235–568) interspersed Cys-rich repeats (379–421); *S. cerevisiae* contains only a distantly related C2 domain (672–767), which lacks the consensus $[AG]x_2Lx_3Dx_2GxSDPY$, Cys-rich repeats (427–531) and the kinase region (830–949); *S. pombe* and *D. discoideum* contains only the kinase segment (17–300 and 342–648), resp.); and *M. xanthus* again, besides the kinase region (59–322), contains Cys-rich repeats (10–56).

Consequently, the metazoan PKCs, shown here for the cPKC and the nPKC from *G. cydonium,* have to be distinguished from the nonmetazoan (protozoan, plant, yeast, and bacterial) PKCs on the basis of their domain composition.

Conclusion

In the present study two cDNAs, *GCPKC1* and *GCPKC2,* coding for PKCs have been cloned from *G. cydonium,* a species belonging to the lowest animal phylum, the sponges. GCPKC1 can be classified as belonging to the ''novel'' (Ca^{2+}-independent) PKC subfamily and GCPKC2 to the ''conventional'' (Ca^{2+}-dependent) PKC subfamily. The alignment and the subsequent construction of a phylogenetic tree (based on the Ser/Thr kinase domain of the PKCs) of the sponge sequences using the hitherto-known invertebrate Ser/Thr kinase sequences as well as two representatives of vertebrate PKCs revealed a common origin for PKCs from animals. Furthermore, the animal kinase domains also display homologies with those from plants, protozoa, and bacteria. This suggests that the Ser/Thr kinase domain has a universal common ancestor, as is known for other enzymes, e.g., for RNA polymerases (Pühler et al. 1989). However, the overall structures of the Ser/Thr kinases with respect to domains, other than the catalytic domain, indicate that composition of the PKCs from metazoans is identical, but differs from those of nonmetazoans.

This finding of a common ancestor for the Ser/Thr kinase domain also sheds new light on the nature of

signal molecules to which organisms from bacteria to metazoa react. Previously it was suggested that only unicellular and multicellular eukaryotes, but not prokaryotes, react to authentic signal molecules (Stabel and Parker 1991). This conclusion was impressively supported by Csaba and his colleagues (1994), who were able to demonstrate that the unicellular *Tetrahymena* reacts to ''vertebrate'' hormones, e.g., insulin, epinephrine, and histamine. Signal molecules that resemble insulin, calmodulin, and chorionic gonadotropin have been described as being produced in bacteria (Gorby et al. 1988). According to the homology comparison summarized here, these organisms contain Ser/Thr kinases. To the best of our knowledge, sequences for Ser/Thr kinases of archaebacteria have not been cloned yet.

The finding that the overall structure of the metazoan PKCs is different from those found in the nonmetazoans also suggests different functions of Ser/Thr kinases in these organisms: (1) The nonmetazoan sequences do not contain the potential Ca^{2+}-binding site and hence these enzymes might not be crucially regulated by this cation; (2) they lack the typical pseudosubstrate structure, suggesting a different regulation mechanism of the catalytic site; and (3) they show Cys-rich sequences that are not in accordance with the typical zinc fingers found in metazoan PKCs, suggesting a differing mode of interaction between the Ser/Thr kinases and nucleic acid or other proteins, respectively.

Based on these data it is apparent that the Ser/Thr kinase domain of metazoan PKCs is a block that shares homology with other eukaryotes and prokaryotes. This finding also suggests an identical or similar enzymatic mechanism. However, the presence of new domains in the metazoan PKCs indicates different manners of regulation of the kinase activity.

Acknowledgments. Supported by grants from the Deutsche Forschungsgemeinschaft (Mü 348/12-1) and the International Human Frontier Science Program (RG-333/96-M).

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (1995) Current protocols in molecular biology. John Wiley and Sons, New York

Biermann B, Johnson EM, Feldman LJ (1990) Characterization and distribution of a maize cDNA encoding a peptide similar to the catalytic region of second messenger dependent protein kinases. Plant Physiol 94:1609–1615

Blake MS, Johnston KH, Russell-Jones GJ, Gotschlich EC (1984) A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-antibody on Western blots. Anal Biochem 136: 175–179

Branden C, Tooze J (1991) Introduction to protein structure. Gerland, New York

Bürki E, Anjard C, Scholder JC, Reymond CD (1991) Isolation of two genes encoding putative protein kinases regulated during *Dictyostelium discoideum* development. Gene 102:57–65

Cocco L, Martelli AM, Gilmour RS (1994) Inositol lipid cycle in the nucleus. Cell Signal 6:481–485

Csaba G (1994) Phylogeny and ontogeny of chemical signalling: origin and development of hormone receptors. Int Rev Cytol 155:1–48

Daum G, Eisenmann-Tappe I, Fries HW, Troppmair J, Rapp UF (1994) The ins and outs of Raf kinases. Trends Biochem Sci 19:474–480

Disatnik MH, Winnier AR, Mochly-Rosen D, Arteaga CL (1994) Distinct responses of protein kinase C isoenzymes to c-erbB-2 activation in SKBR-3 human breast carcinoma cells. Cell Growth Differ 5:873–880

EMBL Sequence Database, Release 40 (September 1994)

Gale MJ, Parsons M (1993) A *Trypanosoma brucei* gene family encoding protein kinases with catalytic domains structurally related to Nek1 and NIMA. Mol Biochem Parasitol 59:111–121

Gorby GL, Robinson EN, Barley LR, Clemens CM, McGee ZA (1988) Microbial invasion: a covert activity? Can J Microbiol 34:507–512

Hanks SK, Quinn AM (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 200A:3–37

Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42–52

Hardie G, Hanks S (1995) The protein kinase factsbook: protein-serine kinases. Academic Press, London

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Kikkawa U, Ogita K, Ono Y, Asaoka Y, Shearman MS, Fujii T, Ase K, Sekiguchi K, Igarashi K, Nishizuka Y (1987) The common structure and activities of four subspecies of rat brain protein kinase C family. FEBS Lett 223:212–216

Kruse M, Mikoc A, Cetkovic H, Gamulin V, Rinkevich B, Müller IM, Müller WG (1994) Molecular evidence for the presence of a developmental gene in the lowest animals: identification of a homeobox-like gene in the marine sponge *Geodia cydonium.* Mech Ageing Dev 77:43–54

Kubo K, Ohno S, Suzuki K (1987) Primary structures of human protein kinase C beta-I and beta-II differ only in their C-terminal sequences. FEBS Lett 223:138–142

Land M, Islas-Trejo A, Freedman JH, Rubin CS (1994) Structure and expression of a novel, neuronal protein kinase C (PKC1B) from *Caenorhabditis elegans.* PKC1B is expressed selectively in neurons that receive, transmit, and process environmental signals. J Biol Chem 269:9234–9244

Lawton MA, Yamamoto RT, Hanks SK, Lamb CJ (1989) Molecular cloning of plant transcripts encoding protein kinase homologs. Proc Natl Acad Sci USA 86:3140–3144

Levin DE, Fields O, 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

Mahoney CW, Huang KP (1994) Molecular and catalytic properties of protein kinase C. In: Kuo JF (ed) Protein kinase C. Oxford University Press, New York, pp 16–63

McLeod M, Beach D (1986) Homology between the ran 1+ gene fission yeast and protein kinases. EMBO J 5:3665–3671

Müller WEG (1995) Molecular phylogeny of metazoa [animals]: monophyletic origin. Naturwissenschaften 82:321–329

Müller WEG, Rottmann M, Diehl-Seifert B, Kurelec B, Uhlenbruck G, Schröder HC (1987) Role of the aggregation factor in the regulation of phosphoinositide metabolism in sponges. Possible consequences on calcium efflux and on mitogenesis. J Biol Chem 262:9850–9858

Müller WEG, Ugarkovic D, Gamulin V, Weiler E, Schröder HC (1990) Intracellular signal transduction pathways in sponges. Electron Microsc Rev 3:97–114

Munoz-Dorado J, Inouye S, Inouye M (1991) A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. Cell 67:995–1006

Nishizuka Y (1988) The molecular heterogeneity of protein kinase-C and its implications for cellular regulation. Nature 334:661–665.

PC/GENE (1995) Data Banks CD-ROM; Release 14.0. IntelliGenetics, Inc. Mountain View, CA

Pfeifer K, Frank W, Schröder HC, Gamulin V, Rinkevich B, Müller IM, Müller WEG (1993a) cDNA cloning of the polyubiquitin gene from the marine sponge *Geodia cydonium* which is preferentially expressed during reaggregation of cells. J Cell Science 106:545–554

Pfeifer K, Haasemann M, Gamulin V, Bretting H, Fahrenholz F, Müller WEG (1993b) S-type lectins occur also in invertebrates: unusual subunit composition and high conservation of the carbohydrate recognition domain in the lectin genes from the marine sponge *Geodia cydonium.* Glycobiology 3:179–184

Pühler G, Leffers H, Gropp F, Palm P, Klenk HP, Lottspeich F, Garrett RA, Zillig W (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci USA 86:4569–4573

Rakow TL, Shen SS (submitted) Molecular cloning and characterization of protein kinase C from the sea urchin *Lytechinus pictus.* (cited from: EMBL Sequence Database, Release 40)

Rose-John S, Dietrich A, Marks F (1988) Molecular cloning of mouse protein kinase C (PKC) cDNA from Swiss 3T3 fibroblasts. Gene 74:465–471

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schäcke H, Müller WEG, Gamulin V, Rinkevich B (1994a) The Ig superfamily includes members from the lowest invertebrates to the highest invertebrates. Immunol Today 15:497–498

Schäcke H, Schröder HC, Gamulin V, Rinkevich B, Müller IM, Müller WEG (1994b) Molecular cloning of a receptor tyrosine kinase from the marine sponge *Geodia cydonium:* a new member of the receptor tyrosine kinase class II family in invertebrates. Mol Membr Biol 11:101–107

Schaeffer E, Smith D, Mardon G, Quinn W, Zuker C (1989) Isolation and characterization of two new *Drosophila* protein kinase C genes, including one specifically expressed in photoreceptor cells. Cell 57:403–412

Schröder HC, Rottmann M, Wenger R, Bachmann M, Dorn A, Müller WEG (1988) Studies on protein kinases involved in regulation of nucleocytoplasmic mRNA transport. Biochem J 252:777–790

Stabel S, Parker PJ (1991) Protein kinase C. Pharmacol Ther 51:71–95

Takai Y, Kishimoto A, Inoue M, Nishizuka Y (1977) Studies on a cyclic nucleotide independent protein kinase and its proenzyme in mammalian tissues. J Biol Chem 252:7603–7609

Weissmann G, Riesen W, Davidson S, Waite M (1988) Stimulus coupling in marine sponge cell aggregation: lipid metabolism and the function of exogenously added arachidonic and docosahexaenoic acids. Biochim Biophys Acta 960:351–364

Zarkower D, Stephenson P, Sheets M, Wickens M (1986) The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA *in vitro.* Mol Cell Biol 6:2317–2323