THE ROYAL SOCIETY

Evolutionary relationships of Metazoa within the eukaryotes based on molecular data from Porifera

Joachim Schütze[1], Anatoli Krasko[1], Marcio Reis Custodio[1,2], Sofia M. Efremova[3], Isabel M. Müller[1] and Werner E. G. Müller[1*]

[1]*Institut für Physiologische Chemie, Abteilung Angewandte Molekularbiologie, Universität, Duesbergweg 6, 55099 Mainz, Germany*
[2]*Departamento de Histologia e Embriologia, Instituto de Ciêcias Biomédicas, Universidade Federal de Rio de Janeiro, CP 68021, Rio de Janeiro CEP 21941-970, Brazil*
[3]*Biological Institute of Saint-Petersburg University, Oranienbaumskoye sch. 2, Stary Petersburg, Saint-Petersburg 198904, Russia*

Recent molecular data provide strong support for the view that all metazoan phyla, including Porifera, are of monophyletic origin. The relationship of Metazoa, including the Porifera, to Plantae, Fungi and unicellular eukaryotes has only rarely been studied by using cDNAs coding for proteins. Sequence data from rDNA suggested a relationship of Porifera to unicellular eukaryotes (choanoflagellates). However, ultrastructural studies of choanocytes did not support these findings. In the present study, we compared amino acid sequences that are found in a variety of metazoans (including sponges) with those of Plantae, Fungi and unicellular eukaryotes, to obtain an answer to this question. We used the four sequences from 70 kDa heat-shock proteins, the serine–threonine kinase domain found in protein kinases, β-tubulin and calmodulin. The latter two sequences were deduced from cDNAs, isolated from the sponge *Geodia cydonium* for the phylogenetic analyses presented. These revealed that the sponge molecules were grouped into the same branch as the Metazoa, which is statistically (significantly) separated from those branches that comprise the sequences from Fungi, Plantae and unicellular eukaryotes. From our molecular data it seems evident that the unicellular eukaryotes existed at an earlier stage of evolution, and the Plantae and especially the Fungi and the Metazoa only appeared later.

Keywords: calmodulin; heat-shock protein; molecular systematics; Porifera; Ser–Thr kinase domain; β-tubulin

1. INTRODUCTION

Recent analyses of nucleotide (nt) sequences coding for proteins showed that the phylum Porifera (sponges) has a monophyletic origin with the other metazoan phyla (Müller *et al.* 1994; Müller 1995, 1998). This conclusion, drawn from analyses of protein-coding nt sequences, has recently also been supported by rRNA ribosomal data (Cavalier-Smith *et al.* 1996).

Unlike Metazoa, the monophyletic origin of Plantae was established earlier; it can be traced back to Chlorophyta (reviewed in Margulis & Schwartz 1995). The ancestry of Fungi, the third group of multicellular eukaryotes, is less evident; monophyletic and polyphyletic evolution of these spore forming and amastigote organisms are discussed by Margulis & Schwartz (1995).

First efforts to resolve the question of sponge ancestry were undertaken by comparing their cellular biology with that of both unicellular and multicellular animals. Similarities within the ultrastructural level of sponge choanocytes possibly relate them to choanoflagellates (see Taylor 1994). However, detailed inspection of the flagellar appendages did not support this hypothesis (Gallissian & Vacelet 1992); furthermore, the lack of a collar in the monoflagellate cells of amphiblastulae from Calcarea strongly deny a direct phylogenetic relationship between sponges and choanoflagellates (Mehl & Reiswig 1991). Clarification of the hierarchial evolution of multicellular eukaryotes, Metazoa (including sponges), Plantae and Fungi with respect to their links to the unicellular eukaryotes was investigated by using nt data from rDNA sequences. Results were ambiguous; again, the analyses of rRNA sequences pointed towards a common ancestry of Metazoa and Fungi (Wainright *et al.* 1993; Cavalier-Smith *et al.* 1996). However, an extended statistical analysis could not substantiate this grouping (Kumar & Rzhetsky 1996).

The aim of the present study is to evaluate the phylogenetic relationship of the three multicellular eukaryotic subkingdoms to unicellular eukaryotes, with the main emphasis on the positon of sponges. Amino acid (aa) sequences found in all representatives of these subkingdoms, one heat-shock protein (HSP) the serine–threonine (Ser–Thr) kinase domain of protein kinases, β-tubulin and calmodulin, were analysed together with those sequences deduced from sponge cDNAs.

First, polypeptides grouped to the HSPs of the 70 kDa class, HSP70, were investigated. HSPs are highly

*Author for correspondence (wmueller@mail.uni-mainz.de).

Received 21 July 1998 *Accepted* 23 September 1998

conserved throughout living kingdoms. These molecules act as molecular chaperones under conditions of physiological stress (Gething & Sambrook 1992); the chaperones of prokaryotic organisms (reviewed in Georgopoulos *et al.* 1994), which are related to the eukaryotic HSP70 multigene family (reviewed in Günther & Walter 1994), are termed DnaK. They are divided according to their different inducibility into (i) the group of constitutively expressed HSPs, which are also present under non-stressed conditions, and (ii) the group of HSP70 polypeptides, which are induced under temperature shock and other specific stress situations (Subjeck & Shyy 1986; Günther & Walter 1994). The inducible, cytoplasmic HSP70 aa sequences were selected for the analysis described here.

Second, the Ser–Thr kinase domain found in protein kinases from Fungi and Metazoa, but not in Plantae or unicellular eukaryotes (Hardie & Hanks 1995; Kruse *et al.* 1997), was used for the analyses. These enzymes are essential for fungal or metazoan organisms to recognize extracellular signals and to initiate intracellularly appropriate adaptative biological responses.

Third, sequences of β-tubulin, one major element of microtubules (intracellular structures that are employed for a number of functions in eukaryotes, including flagellar motility and chromosome aggregation as well as cell maintenance of cellular morphology) were analysed (see Sullivan 1988). This molecule had previously been chosen for studies in molecular phylogeny (Keeling & Doolittle 1996) but here the sequence from *Geodia cydonium* is included for the first time.

Finally, calmodulin, a protein ubiquitous in eukaryotes, was selected for phylogenetic analysis. It is a Ca^{2+}-binding protein of approximately 150 aa residues that is involved in a wide range of intracellular signalling pathways (see Babu *et al.* 1985). The molecule binds four Ca^{2+} ions in a cooperative fashion, during which it undergoes a conformational change (Ikura *et al.* 1992). The calmodulin sequence from *G. cydonium* was identified for the purpose of the phylogenetic analysis.

2. MATERIALS AND METHODS

(a) *Materials*

Restriction endonucleases and other enzymes for recombinant DNA techniques and vectors were obtained from Stratagene (Heidelberg, Germany), Boehringer Mannheim (Mannheim, Germany), Epicentre Technologies (Madison, WI, USA), and USB (Cleveland, OH, USA).

Specimens of *G. cydonium* (Jameson) (Porifera, Demospongiae, Tetractinomorpha, Astrophorida, Geodiidae) were collected near Rovinj (Croatia).

(b) *Cloning of tubulin from* Geodia cydonium

The complete sponge tubulin cDNA, *GCTUBb*, 1514 bp long, was cloned by polymerase chain reaction (PCR) from a *G. cydonium* cDNA library in γZAP EXPRESS (Pfeifer *et al.* 1993). The degenerate primer 5′-TGRAANCCYTGNARR-CARTC-3′ (where N=A/G/C/T, R=A/G and Y=C/T), designed against the conserved aa segment aa_{127}–aa_{134} (CDCLQGFQ) of β-tubulin (rat; accession number P04691), was used in conjunction with the vector-specific primer T3. The PCR was carried out with a GeneAmp 9600 thermal cycler (Perkin Elmer) with an initial denaturation at 95 °C for 3 min, then 35 amplification cycles of 95 °C for 30 s, 58 °C for 45 s, 72 °C for 2 min and a final extension step at 72 °C for 10 min. The reaction mixture of 50 µl included 20 pmol of the degenerate primer and 10 pmol of the primer T3, 200 µM of each nucleotide, 1 µl of the cDNA library, buffer and 2.5 units of Taq DNA polymerase (Boehringer Mannheim). The PCR product was purified and used as a probe. Screening of the library was performed as previously described (Pfeifer *et al.* 1993). Positive clones were detected with an alkaline-phosphatase-conjugated anti-digoxygenin antibody with BCIP–NBT (5-bromo-4-chloro-3-indolylphosphate–nitroblue tetrazolium) as substrate. The clones obtained were sequenced with an automatic DNA sequenator (Li-Cor 4200).

(c) *Cloning of calmodulin from* G. cydonium

The complete cDNA for calmodulin, *CAMGC*, 643 bp long, was also received by sequential application of PCR, followed by a hybridization procedure with the PCR product previously obtained. The forward primer 5′-AAA/GCTIACIGAT/CGAG/AGAG/AGTIGAC/T-3′ (where I=inosine) was designed against the conserved peptide KLTDEEVD (aa position 119–126 in the human sequence with the accession number P02593). The hybridization temperature was 62 °C.

(d) *Sequence comparisons*

Sequences were analysed by means of the computer program PC/GENE (1995). Similarity searches and sequence retrieval were performed via email servers at the European Bioinformatics Institute, Hinxton Hall, UK (BLITZ and FASTA), and the National Institutes of Health, Bethesda, MD, USA (BLAST). Multiple alignment was performed with CLUSTAL W version 1.6 (Thompson *et al.* 1994); the default options (gap opening, 10.00; gap extension, 0.05; delay divergent sequence, 40%; Blossum series; gap separation distance, 8) were used. The graphic presentation was composed with GeneDoc (Nicholas & Nicholas 1996). The automatic alignment was optimized manually.

Phylogenetic trees were constructed on the basis of aa sequence alignments by neighbour-joining, as implemented in the 'Neighbor' program from the PHYLIP package (Felsenstein 1993). The distance matrix was calculated by using the Dayhoff PAM matrix model as described by Dayhoff *et al.* (1978). The degree of support for internal branches was further assessed by bootstrapping (Felsenstein 1993). Graphical output of the bootstrap figure was produced by the program 'Treeview' (Roderic D. M. Page, University of Glasgow, UK; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

3. RESULTS AND DISCUSSION

(a) *Phylogenetic analysis of 70 kDa heat-shock proteins (HSP70)*

In earlier contributions we reported that the sponge HSP70 aa sequences were grouped together with related sequences from Metazoa into one branch, separated from prokaryotes (Koziol *et al.* 1998); in addition, it was reported that within the phylum Porifera the class of Hexactinellida with its member *Rhabdocalyptus dawsoni* is phylogenetically older than the class of Demospongiae with *G. cydonium*, and the class Calcarea with *S. raphanus* (Koziol *et al.* 1997).

In a more extensive comparison we now present the phylogenetic position of the metazoan HSP70 as related

to sequences from unicellular eukaryotes, Fungi and Plantae (figure 1). The rooted tree, with the bacterial HSP70-homologue DnaK (*Methanosarcina mazei*) as an outgroup, reveals that the unicellular eukaryotes *Leishmania amazonensis* (Euglenozoa) and *Plasmodium cynomolgi* (Alveolata) are positioned at the base of the tree from which Fungi, Plantae and Metazoa branch off later. Bootstrap pseudoreplication analysis reveals that separation of the different taxa of Fungi (Oomycota, Basidiomycota and Ascomycota) and Plantae (Viridiplantae) is not robust.

If we focus on Metazoa, the hexactinellid sponge *R. dawsoni* forms the base of them; the demosponge *G. cydonium* branches off later. The calcarian sponge *S. raphanus* evolved even later and falls together with the cnidarian *Hydra magnipapillata* into one branch. The sequence of HSP70 from *S. raphanus* shares 86% similarity with the sequence from *H. magnipapillata* and a high relationship to that of the platyhelminth *Echinococcus granulosus* (85%) as well as to that of the frog *Xenopus laevis* (87%). However, the bootstrap support for the grouping and the subsequent branching of members of the higher metazoan phyla is weak in some instances (for example, 18% between the sequences of *S. raphanus* and the chicken), a finding that is related to the overall high conservation of HSP70 sequences.

(b) *Phylogenetic analysis of the Ser–Thr kinase domain*

Recently, cDNA sequences encoding calcium-dependent protein kinase C (cPKCs) from the demosponges *G. cydonium* and *Suberites domuncula*, the calcarean *S. raphanus* and the hexactinellid *R. dawsoni* have been isolated and characterized (Kruse *et al.* 1998). It is established that the overall composition of cPKCs with their regulatory domains is restricted to Metazoa (Kruse *et al.* 1996, 1998), whereas the Ser–Thr kinase domain has also been identifed in kinases from Fungi and Metazoa (Hardie & Hanks 1995). Therefore, only this catalytic domain can be used for the present phylogenetic analysis.

The three sponge Ser–Thr kinase domains existing in characteristic cPKCs were compared with related domains of kinases from three Fungi (Ascomycota) and ten other metazoans. Based on the aa aligment of the Ser–Thr kinase domains a phylogenetic tree was constructed and statistically analysed by bootstrapping (figure 2). The rooted phylogenetic tree, using the Tyr kinase domain from the Tyr kinase receptor molecule of *G. cydonium* (Müller & Schäcke 1996) as an outgroup, reveals that the Ser–Thr kinase domain from the hexactinellid *R. dawsoni* branches off first from a common ancestor of the other related domains. The three sequences from Ascomycota appear later. Together with the two sequences from the Demospongiae *G. cydonium* and *S. domuncula*, the sequences known from members of higher metazoan phyla form one separate branch. In this tree the phylogenetic status of diploblasts with Cnidaria, triploblasts with pseudocoelomates (Nematoda), protostomes (Mollusca, Insecta) and deuterostomes (Echinodermata, Vertebrata) is consistent with Hyman's view (1940) of phylogeny. It should be highlighted that the sequence from *R. dawsoni* does not fall into a single branch together with the other sponge sequences; this result might indicate that this hexactinellid sequence is more ancient than the characteristic Ser–Thr kinase domains of Fungi and Metazoa. Owing to the low bootstrap proportion (58%), future analyses might be needed.

(c) *Phylogenetic analysis of β-tubulin*

(i) *Cloning of the β-tubulin gene from* G. cydonium

The nt sequence

The cDNA clone obtained, termed *GCTUBb*, was 1514 bp long (accession number Y09500). The open reading frame (ORF) in *GCTUBb* comprises 1347 bp (the start codon is at nt 70) and predicts a protein 449 aa long, named TBB_GEOCY, with a deduced M_r of 50 352 and pI of 4.43 (PC/GENE (program: Physchem) 1995). A truncated signal polyadenylation site, AATA, (Zarkower *et al.* 1986) is present at nt_{1496}–nt_{1499}.

The deduced aa sequence

Homology searches with the programs BLAST, BLITZ and FASTA (see above) of the sponge deduced aa β-tubulin sequence TBB_GEOCY revealed highest similarity to the sequence from *Caenorhabditis elegans* (96% similar aa). Northern blot analysis with the sponge PCR fragment as a probe displayed one prominent band of approximately 1.6 kb, confirming that a full-length cDNA was isolated (not shown).

The deduced aa sequence displays the characteristic features of β-tubulin. The N-terminus starting with aa MREI (aa_1–aa_4) has consensus with the β-tubulin mRNA autoregulation sequence (PC/GENE; Prosite 1995). The consensus pattern for α,β,γ-tubulin in *G. cydonium* reads GGGTGSG, and is found at aa_{140}–aa_{146} (PC/GENE; Prosite 1995) (figure 3). The four putative GTP-binding sites of β-tubulin (Linse & Mandelkow 1988; Nogales *et al.* 1998) are present in the N-terminal part of the molecule at the following positions; GTP-binding sites I (aa_{140}–aa_{146}), II (aa_{178}–aa_{181}), III (aa_{240}–aa_{244}) and IV (aa_{58}–aa_{67}) (figure 3*a*,*b*).

(ii) *Phylogenetic analysis*

In an earlier study it was demonstrated that β-tubulin molecules from Metazoa are more closely related to tubulin sequences from Fungi or Plantae than to those of unicellular eukaryotes (Keeling & Doolittle 1996). In the present phylogenetic analysis the β-tubulin sequence from the sponge *G. cydonium* was included. The unrooted trichotomous tree also shows that, by inclusion of the sponge sequence, the relationships follow the earlier interpretation (Keeling & Doolittle 1996) (figure 4). The robust grouping separates Metazoa from Viridiplantae, Fungi and unicellular eukaryotes. Again owing to the high overall similarity of β-tubulin sequences from the latter taxa (aa identity >90% and similarity >95% between the sponge sequence and β-tubulins from higher metazoans), the resolution of the relationships within Viridiplantae, Fungi and unicellular eukaryotes is low. However, the tree shows that metazoan β-tubulins are more closely related to the fungal sequences (Ascomycota, Basidiomycota and Myxomycota) than to unicellular eukaryotic β-tubulins. In addition, it is evident that the

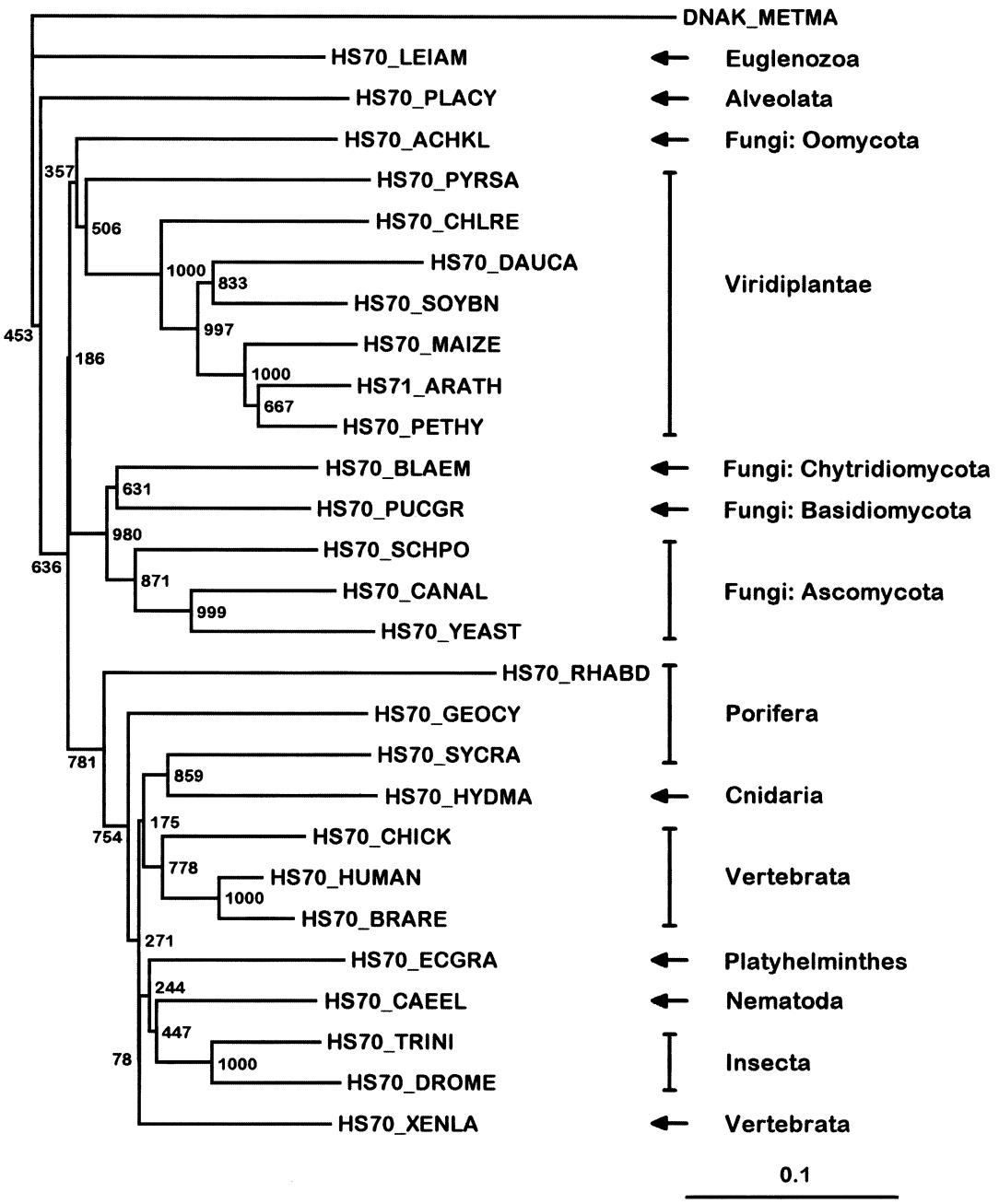

Figure 1. Rooted phylogenetic tree, constructed after alignment of the poriferan HSP70 aa sequences from the Hexactinellida *R. dawsoni* (HS70_RHABD; Y10529), the Demospongia *G. cydonium* (HS70_GEOCY; Y09037) and the Calcarea *S. raphanus* (HS70_SYCRA, Y10530) with the corresponding sequences from I. the unicellular eukaryotes: (i) Euglenozoa: *Leishmania amazonensis* (HS70_LEIAM; Q07437) and (ii) Alveolata: *Plasmodium cynomolgi* (HS70_PLACY; Q05746); II. Fungi: (i) Oomycota: *Achlya klebsiana* (HS70_ACHKL; P41753), (ii) Chytridiomycota: *Blastocladiella emersonii* (HS70_BLAEM; P48720), (iii) Basidiomycota: *Puccinia graminis* (HS70_PUCGR; Q01877) as well as (iv) Ascomycota: *Schizosaccharomyces pombe* (fission yeast) (HS70_SCHPO; Q10265), *Candida albicans* (HS70_CANAL; P41797) and *Saccharomyces cerevisiae* (baker's yeast) (HS70_YEAST; P09435); III. Viridiplantae: *Pyrenomonas salina* (HS70_PYRSA; P37899), *Chlamydomonas reinhardtii* (green algae) (HS70_CHLRE; P25840), *Daucus carota* (carrot) (HS70_DAUCA; P26791), *Glycine max* (soyabean) (HS70_SOYBN; P26413), *Zea mays* (maize) (HS70_MAIZE; P11143), *Arabidopsis thaliana* (mouse-ear cress) (HS70_ARATH; P22953) and *Petunia hybrida* (HS70_PETHY; P09189) and IV. metazoan members: (i); Cnidaria: *Hydra magnipapillata* (HS70_HYDMA; Q05944), (ii) Platyhelminthes: *Echinococcus granulosus* (HS70_ECGRA; U26448), (iii) Nematoda: *Caenorhabditis elegans* (HS70_CAEEL; Z80223), (iv) Insecta: *Trichoplusia ni* (cabbage looper) (HS70_TRINI; U23504) and *Drosophila melanogaster* (HS70_DROME; P11147) and (v) Vertebrata: *Gallus gallus* (chicken) (HS70_CHICK; P08106), *Homo sapiens* (human) (HS70_HUMAN; P11142), *Danio rerio* (zebrafish) (HS70_BRARE; Q90473) and *Xenopus laevis* (African clawed frog) (HS70_XENLA; P02827). The related sequence from the archaebacterium *Methanosarcina mazei* (DNAK_METMA, P27094) was used as outgroup. The analysis was performed by neighbour-joining as described in the text. The numbers at the nodes are an indication of the levels of confidence (percentages) for the branches as determined by bootstrap analysis (1000 bootstrap replicates). The scale bar indicates an evolutionary distance of 0.1 aa substitutions per position in the sequence.

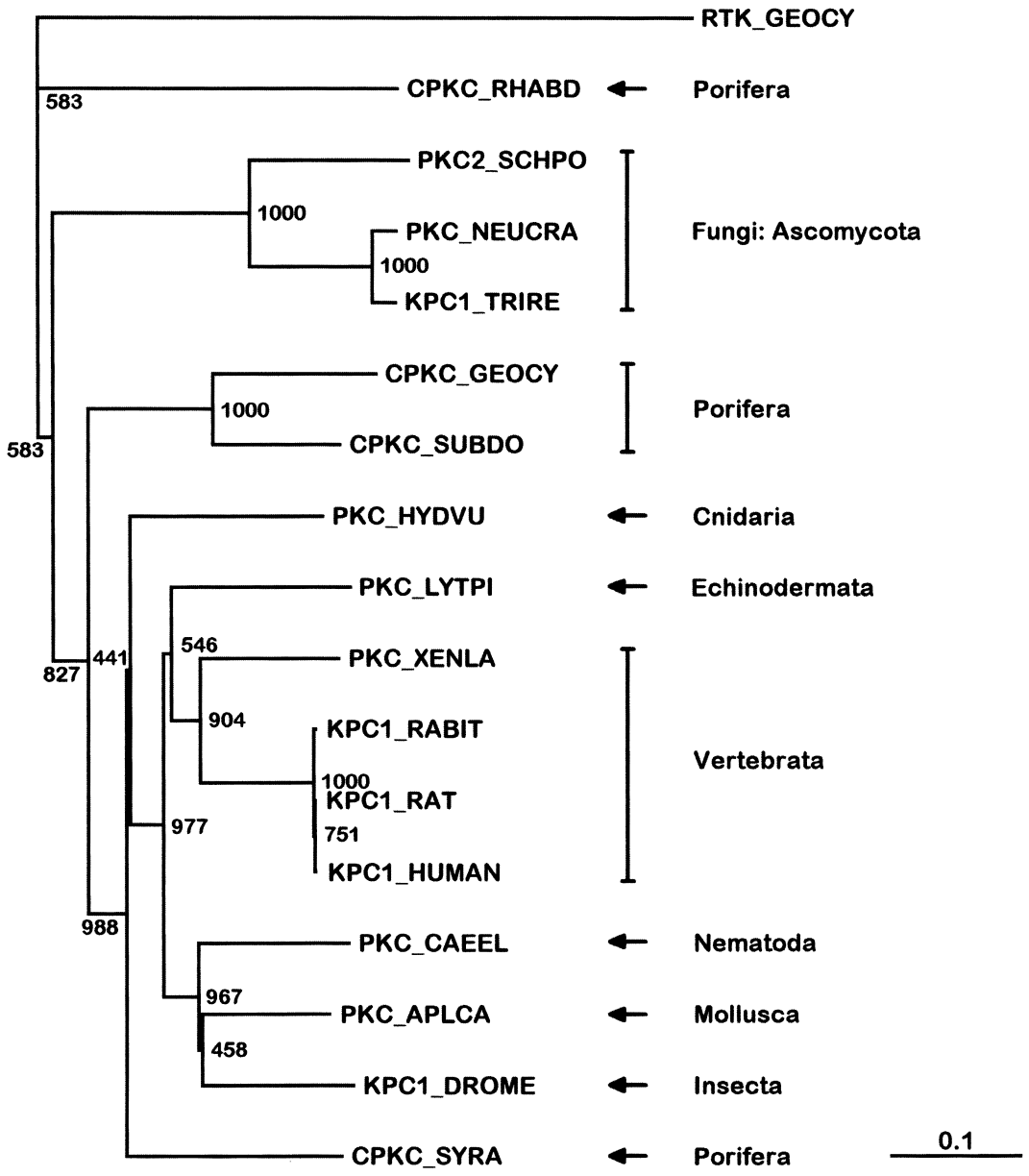

Figure 2. Phylogenetic analysis of Ser–Thr kinase domains present in protein kinases. The deduced sponge sequences from cPKC from *G. cydonium* (Demospongiae) (CPKC_GEOCY; X87683), *Suberites domuncula* (Demospongiae) (CPKC_SUBDO; Y10529), *S. raphanus* (Calcarea) (CPKC_SYRA; Y10530) and *R. dawsoni* (Hexactinellida) (CPKC_RHABD; Y13229) were compared with sequences from I. Fungi: Ascomycota: *Schizosaccharomyces pombe* (PCK2_SCHPO; P36583), *Neurospora crassa* (PKC_NEUCRA; Y12002) and *Hypocrea jecorina* (KPC1_TRIRE; Q99014); II. Metazoa: (i) Cnidaria: *Hydra vulgaris* (PKC_HYDVU; Y12857), (ii) Nematoda: *C. elegans* (PKC_CAEEL; U82935); (iii) Mollusca: *Aplysia californica* (PKC_APLCA; 228058); (iv) Insecta: *D. melanogaster* (KPC1_DROME, P05130), (v) Echinodermata: the sea urchin *Lytechinus pictus* (PKC_LYTPI; U02967), (vi) Vertebrata: frog *Xenopus laevis* (PKC_XENLA; 104167), rabbit *Oryctolagus cuniculus* (KPC1_RABIT; P05772), rat *Rattus norvegicus* (KPC1_RAT; P04410) and human (KPC1_HUMAN; P05771). The tyrosine kinase domain of the receptor tyrosine kinase from *G. cydonium* (RTK_GC; X77528) was chosen as an outgroup for the rooted phylogenetic tree. The numbers at the nodes refer to the levels of confidence as determined by bootstrap analysis (1000 replicates). The scale bar indicates the evolutionary distance.

sequences of the Viridiplantae are nesting within those of unicellular eukaryotes.

(d) *Phylogenetic analysis of calmodulin*

(i) *Cloning of the calmodulin gene from* G. cydonium

The nt sequence

A degenerate oligonucleotide primer, corresponding to the nt sequence of the conserved peptide aa_{119}–aa_{126}, present in human calmodulin (see § 2) was used to detect the corresponding cDNA from *G. cydonium* in the cDNA library. The complete cDNA *CAMGC* (accession number Y16818) obtained is 643 nt long (excluding the poly(A) tail) and comprises an ORF of 447 nt, which encodes 149 aa (figure 5*a*). The putative AUG initiation site displays a strong consensus sequence (A_{-3}/G_{+4}) (Kozak 1991), reading GAAATGG (the putative translation initiation site is underlined).

(*a*)

```
MREIVHLQAGQCGNQIGSKFWEVISDEHGIDATGNYAGDSDLQLERIDVYYNEASGGKYV   60
regulatory                                    GTP binding site: IV
PRGILVDLEPGTMDSVRSGPFGQLFKPDNFVFGQSGAGNNWAKGHYTEGAELVDSVMEKV  120
 IV                  *******
REEAEKCDCLQGFQLCHSLGGGTGSGMGTLLISKIREEFPDRIMNTYSVVPSPKVSDTVV  180
                        I                                   II
EPYNATLSVHQLVENTDESYCIDNEALYDICFRTLKLTTPTYGDLNHLVSLTMSGVTTCL  240

RFPGQLNADLRKLAVNMVPFPRLHFFMPGFAPLTSRGSQSYRALSVPELTQQMFDAKNMM  300
III
AACDPRHGRYLTVAAMFRGRMSMKQVDEQMLNIQNKNSSYFVEWIPNNVKTAVCDIPPRG  360

LKMSATFIGNSTAIQELFKRVSEQFTAMFKRKAFLHWYTGEGMDEMEFTEAESNMNDLVS  420

EYQQYQEATIEEDEEEEYGEEGEPDEEEA                                  449
```

(*b*)

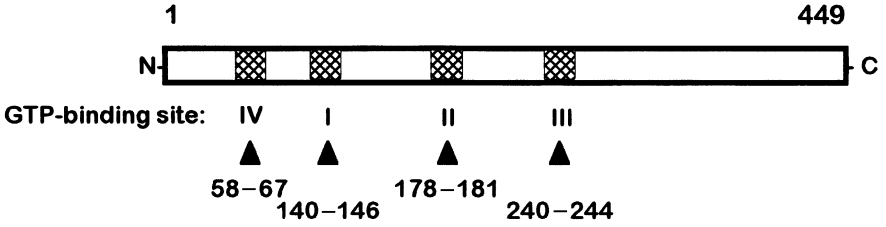

Figure 3. *G. cydonium* β-tubulin. (*a*) The aa sequences of sponge β-tubulin (TBB_GEOCY) deduced from the nt sequence *GCTUBb*. The following signatures in the sequence are marked: the β-tubulin mRNA autoregulation site (regulatory; double underlined), the consensus pattern for α,β,γ-tubulin in (***) and the four putative GTP-binding sites (I–IV; single underlined). (*b*) Schematic representation of the deduced aa sequence of *G. cydonium* β-tubulin.

The deduced aa sequence

According to homology searches with BLAST, BLITZ and FASTA the deduced protein sequence of 149 aa, termed CAM_GEOCY, shows highest similarity to calmodulins. The sponge calmodulin has a putative size (M_r) of 16 776 and an isoelectric point (pI) of 3.99 (PC/GENE 1995). The Northern blot using the sponge PCR fragment of CAM_GEOCY as a probe revealed one band of 0.9 kb, indicating that the full-length cDNA was isolated (not shown).

CAM_GEOCY displays four calcium-binding domains (EF-hand), which are characteristic of calmodulins (Cox & Kretsinger 1998). The first domain ranges from aa_{11} to aa_{39}, the second from aa_{45} to aa_{72}, the third from aa_{85} to aa_{114} and the fourth from aa_{122} to aa_{149} (figure 5*a*). The six aa residues involved in binding of Ca^{2+} form the vertices of an octahedron that surrounds the central sharp bend containing the residue Gly. Hydrophobic residues form the inner aspects of the α-helices. The canonical EF-hand starts with Glu (Kretsinger 1991). These four domains were aligned. Domain one of CAM_GEOCY shows the closest relationship to domain three (35% aa identity and 67% similarity) and domain two to domain four (25%, 61%). An unrooted neighbour-joining tree constructed from the four domains shows two clusters, one including domains one and three and the second comprising domains two and four (figure 5*b*).

(ii) *Phylogenetic analysis*

Calmodulin sequences have only been described from eukaryotes, from the unicellular eukaryotes to Metazoa (Hanson & Schulman 1992). In the present study the putative calmodulin from *G. cydonium* was used for phylogenetic analysis including the related sequences from unicellular eukaryotes (grouped to Alveolata), from Viridiplantae, from Fungi (Oomycota and Basidiomycota), and from 11 Metazoa.

The tree was generated and the members of the Alveolata were used as its basis; see figure 6. The statistical bootstrap analysis reveals that from there, the four calmodulin sequences from Viridiplantae branch off first among the multicellular eukaryotes. Between these plant sequences and the Metazoa there are two diverging fungal sequences, one from Oomycota and another from Basidiomycota. Again, the sponge sequence from *G. cydonium* forms the basis of all metazoan calmodulin sequences chosen; it might be suggested that the long branch of the *G. cydonium* sequence contributes to the lack of high resolution of the other metazoan sequences. Although the sponge sequence can doubtless be classified to the calmodulins, the long branch connecting it with the other multicellular eukaryotes indicates a high divergence. Owing to the low statistical significance, a sister-taxon relationship of sequences from members of higher metazoan phyla cannot be established.

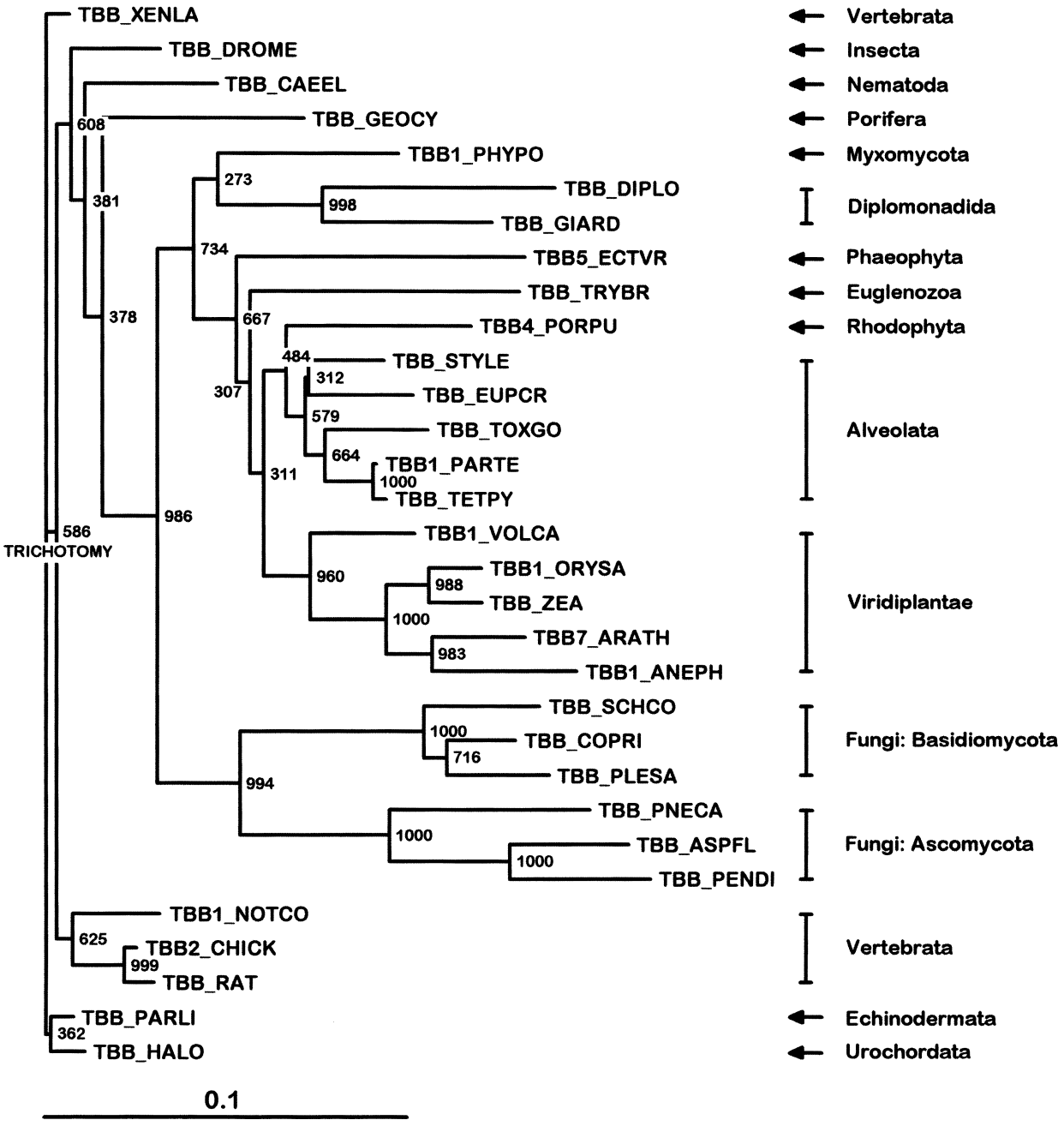

Figure 4. Phylogenetic analysis (unrooted tree construction) of deduced aa sequences of β-tubulin. The deduced aa sequence of the sponge β-tubulin from *G. cydonium* (TBB_GEOCY) was compared with the sequences from I. unicellular eukaryotes: (i) Diplomonadida: Diplomonad ATCC 50330 (TBB_DIPLO; accession number U29441), *Giardia intestinalis* (TBB_GIARD; s00743), (ii) Euglenozoa: *Trypanosoma brucei* (TBB_TRYBR; P04107), (iii) Alveolata: *Stylonychia lemnae* (TBB_STYLE; P11857), *Euplotes crassus* (TBB_EUPCR; P20365), *Toxoplasma gondii* (TBB_TOXGO; P10878), *Paramecium tetraurelia* (TBB_PARTE; P33188), *Tetrahymena pyriformis* (TBB_TETPY; s01769), II. Fungi: (i) Ascomycota: *Pneumocystis carinii* (TBB_PNECA; P24637), *Aspergillus flavus* (TBB_ASPFL; P22012), *Penicillium digitatum* (TBB_PENDI; P53375), (ii) Basidiomycota: *Schizophyllum commune* (TBB_SCHCO; P30668), *Coprinus cinereus* (TBB_COPRI; AB000116), *Pleurotus sajor-caju* (TBB_PLESA; AF008134), III. Myxomycota: *Physarum polycephalum* (TBB_PHYPO; P07436), IV. Plantae: (i) Phaeophyta: *Ectocarpus variabilis* (TBB_ECTVR; P30156), (ii) Rhodophyta: *Porphyra purpurea* (TBB_PORPU; P50262), (iii) Viridiplantae: *Volvox carteri* (TBB_VOLCA; P11482), *Oryza sativa* (TBB_ORYSA; P37832), *Zea mays* (TBB_ZEA; s43328), *Arabidopsis thaliana* (TBB_ARATH; P29515), *Anemia phyllitidis* (TBB_ANEPH; P33630), V. Metazoa: (i) Nematoda: *Caenorhabditis elegans* (TBB_CAEEL; P41937), (ii) Insecta: *Drosophila melanogaster* (TBB_DROME; M20419), (iii) Echinodermata: *Paracentrotus lividus* (TBB_PARLI; P11833), (iv) Urochordata: *Halocynthia roretzi* (TBB_HALO; D89793), (v) *Vertebrata: Xenopus laevis* (TBB_XENLA; P30883), *Gallus gallus* (TBB_CHICK; accession number P32882), *Rattus norvegicus* (TBB_RAT; P04691), *Notothenia coriiceps* (TBB_NOTCO; P36221). The numbers at the nodes refer to the levels of confidence as determined by bootstrap analysis (1000 replicates). The scale bar indicates an evolutionary distance of 0.1 aa substitutions per position in the sequence.

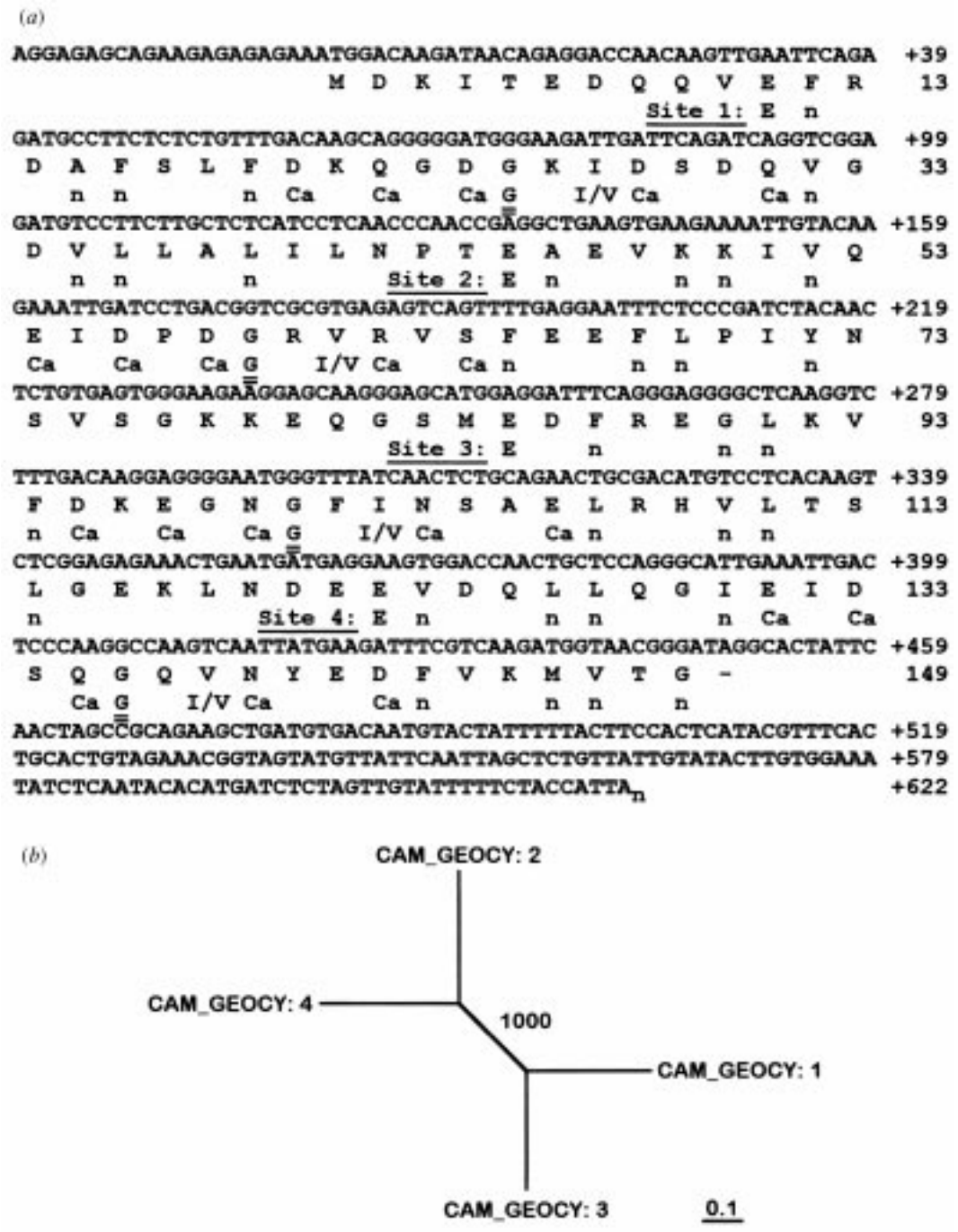

Figure 5. Sponge calmodulin-like protein. (*a*) Nucleotides (*CAMGC*) and deduced aa sequences of the *G. cydonium* calmodulin-like protein (CAM_GEOCY). A_n denotes the poly(A) tract. The positions of the four calcium-binding domains (EF-hand), termed Site 1 to Site 4, are indicated. The characteristic residues of the EF-hand consensus are indicated below the aa sequence; the capital letters denote the respective aa; n refers to hydrophobic residues; Ca indicates the aa involved in Ca^{2+} binding; G marks glycine that occurs at the sharp bend within the helix loop structure (Kretsinger 1991). (*b*) Unrooted phylogenetic tree computed from the four sponge calcium-binding domains CAM_GEOCY 1–4. Two clusters comprising the domains CAM_GEOCY 1 and -3 and CAM_GEOCY 2 and -4 are grouped. The number at the node refers to the level of confidence as determined by bootstrap analysis (1000 bootstrap replicates). The scale bar indicates an evolutionary distance of 0.1 aa substitutions per position in the sequence.

4. CONCLUSIONS

The phylogenetic position of Porifera within metazoan animals has been well established in the past few years based on molecular cloning of the molecules that allow communication of cells among themselves and with their environment (see § 1). Because some of these molecules, especially receptors and their ligands of the adhesion and signal transduction system, are only found in Metazoa, they cannot be used for elucidation of phylogenetic relationships between unicellular eukaryotes and multicellular eukaryotes, Metazoa, Plantae, Fungi and Algae. To solve this problem, data from nt sequences from sponges coding for molecules that also exist in other subkingdoms

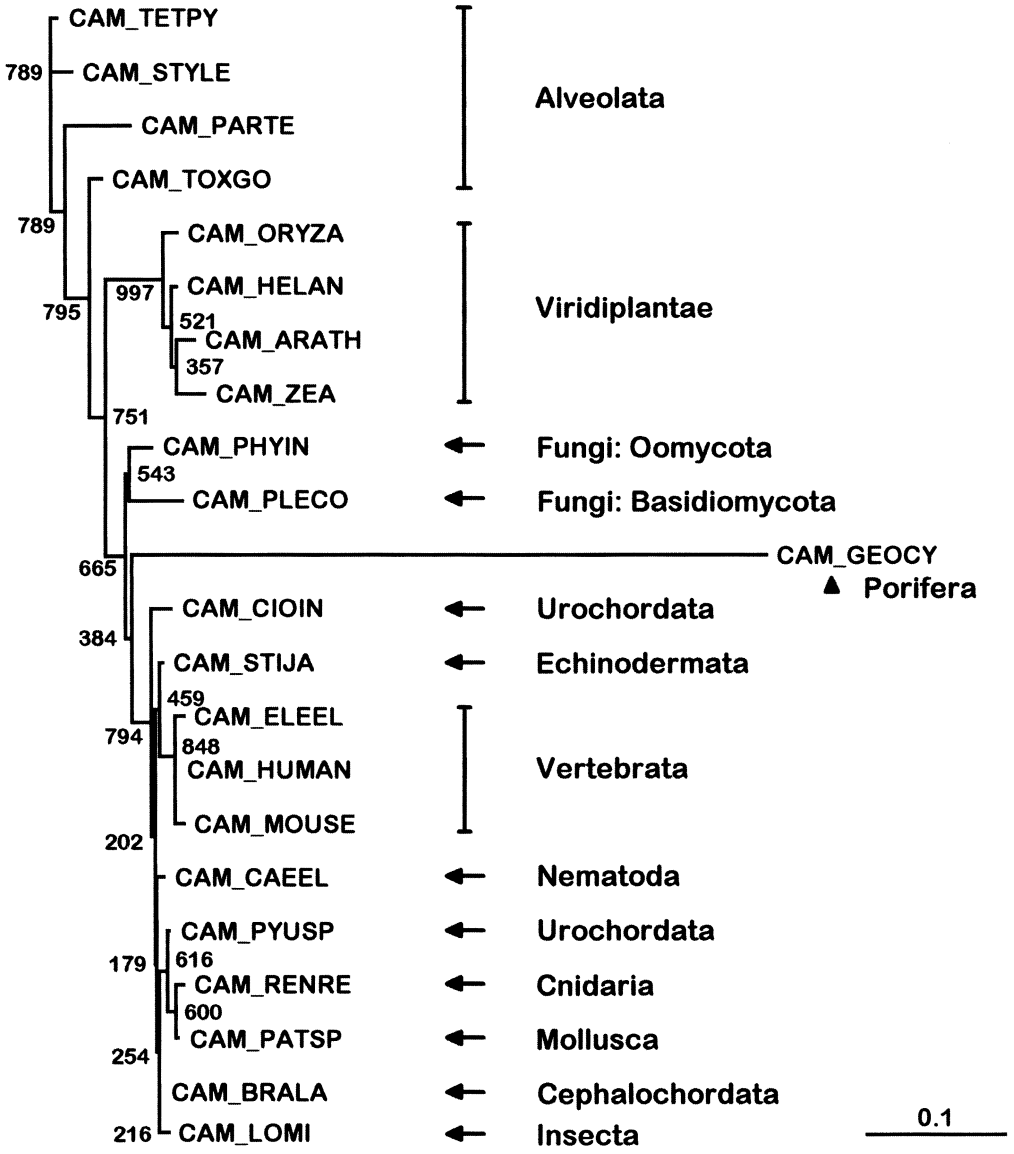

Figure 6. Phylogenetic analysis of calmodulin (unrooted). After alignment of deduced aa sequences of the sponge calmodulin from *G. cydonium* (CAM_GEOCY), this sequence was compared with calmodulin-related sequences from I. the unicellular eukaryotes: (i) Alveolata: *Tetrahymena pyriformis* (Oligohymenophora) (CAM_TETPY; g2144831), *Stylonychia lemnae* (Spirotrichea) (CAM_STYLE; g115528), *Paramecium tetraurelia* (CAM_PARTE; accession number 809298), *Toxoplasma gondii* (Coccidia) (CAM_TOXGO; Y08373); II. Viridiplantae: *Oryza sativa* (Magnoliophyta) (CAM_ORYZA; AF042839), *Helianthus annuus* (Magnoliophyta) (CAM_HELAN; U79736), *Arabidopsis thaliana* (Magnoliophyta) (CAM_ARATH; P25069), *Zea mays* (Magnoliophyta) (CAM_ZEA; 1076792); III. Fungi: (i) Oomycota: *Phytophthora infestans* (CAM_PHYIN; P27165), (ii) Basidiomycota: *Pleurotus cornucopiae* (CAM_PLECO; P11120) and IV. Metazoa: (i) Cnidaria: *Renilla reniformis* (Cnidaria) (CAM_RENRE; P02596), (ii) Nematoda: *Caenorhabditis elegans* (Nematoda) (CAM_CAEEL; AF016429), (iii) Insecta: *Locusta migratoria* (CAM_LOMI; g71667), (iv) Mollusca: *Patinopecten* sp. (CAM_PATSP; P02595), (v) Urochordata: *Ciona intestinalis* (Urochordata) (CAM_CIOIN; Y13578) and *Pyuridae* gen. sp. (Urochordata) (CAM_PYUSP; P11121), (vi) Cephalocordata: *Branchiostoma lanceolatum* (Cephalochordata) (CAM_BRALA; g115508), (vii) Echinodermata: *Stichopus japonicus* (CAM_STIJA; PS00018), and (viii) Vertebrata: human (CAM_HUMAN; P02593), mouse (CAM_MOUSE; g2119362), chicken (CAM_CHICK; g2392137), *Electrophorus electricus* (Teleostei) (CAM_ELEEL; P02594). The numbers at the nodes refer to the level of confidence as determined by bootstrap analysis (1000 replicates). The scale bar indicates an evolutionary distance.

have been used. Whereas the HSP70 chaperone proteins of eukaryotes are distantly related to the prokaryotic DnaK molecules (Koziol *et al.* 1998) and the Ser–Thr kinase domain is found in enzymes both from eukaryotes and from prokaryotes (Kruse *et al.* 1996), the families of calmodulin and β-tubulins are present only in eukaryotes.

The phylogenetic framework for all four selected molecules from sponges, HSP70s, Ser–Thr kinase domains, calmodulin and β-tubulin, shows that Metazoa branch off as the last subkingdom together with the Fungi from the Plantae and the unicellular eukaryotic ancestor(s). Hence, it can be assumed that the postulated close relationship of choanoflagellates to sponges, as concluded from rRNA sequence data (Wainright *et al.* 1993), does not reflect the real process of evolution, as has already been proposed (Kumar & Rzhetsky 1996). To support this conclusion, sequences encoding proteins from choanoflagellates have been analysed.

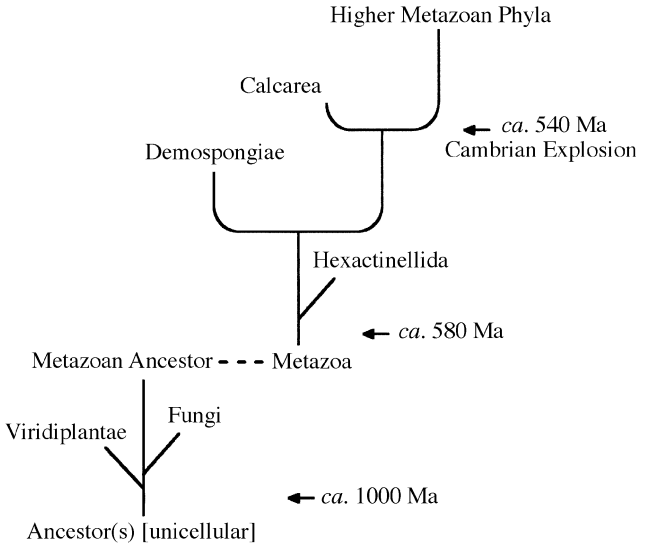

Figure 7. Proposed branching order of the three major subkingdoms, Viridiplantae, Fungi and Metazoa, evolving from ancestral unicellular eukaryotes. Porifera form the basis of the Metazoa. With the classes Hexactinellida, Demospongiae and Calcarea, Porifera have a common ancestor with higher metazoan phyla. The approximate dates of divergence are indicated.

Based on sequence comparisons using primarily α- and β-tubulin, elongation factor 1α and actin, Baldauf & Palmer (1993) previously proposed the close phylogenetic relationship between multicellular eukaryotes and Metazoa while the base of a phylogenetic tree is represented by sequences from unicellular eukaryotes. Therefore, the data included in our analyses unequivocally show that Porifera have to be grouped to the branch of Metazoa and do not display a close, direct relationship to unicellular eukaryotes.

Based not only on nt sequence data but also on fossil records, the transition to multicellularity has been calculated to have taken place about 1000 million years ago (Ma) in red algae (Knoll 1992; Kumar & Rzhetsky 1996). Later in evolution the green algae evolved, 700 Ma (Knoll 1992; Kumar & Rzhetsky 1996); sponges, the first metazoan fossils, have been dated back to 580 Ma (Li *et al.* 1998). These sponges have siliceous spicules and belong to the class of Demospongiae (Li *et al.* 1998). Hence, sponges lived 40–50 million years before the Cambrian explosion (Valentine *et al.* 1996), the time of main divergence of metazoan phyla (Valentine 1994). Based on the extent of aa substitutions of two galectins from *G. cydonium*, it had been calculated that these molecules diverged from the galectin isolated from the nematode *Caenorhabditis elegans* 800 Ma (Hirabayashi & Kasai 1993; Pfeifer *et al.* 1993); this result supported the conclusion that sponges existed before the Cambrian explosion. The branching order originating from ancestral unicellular eukaryotes via Viridiplantae–Fungi to Porifera, the simplest metazoans, follows both the published fossil data (see above) and the sequence data given here (figure 7). In addition, the reported data support recent findings that indicate that, among the three classes of Porifera, Hexactinellida, Demospongiae, and Calcarea, Hexactinellida is the phylogenetically oldest taxon, and Calcarea is the class most closely related to higher metazoan phyla (Koziol *et al.* 1997, 1998) (figure 7).

In conclusion, phylogenetic analyses of the four deduced aa sequences from sponges (HSP70, Ser–Thr kinase domain, β-tubulin and calmodulin) (i) support earlier findings indicating that all metazoans including the phylum Porifera are of monophyletic origin (Müller 1995, 1998) and (ii) extend the view that the Metazoa are more closely related to other multicellular eukaryotes (Fungi) than to unicellular eukaryotes.

The sequences reported here are deposited in the EMBL database: *G. cydonium* β-tubulin (*TUBGCb*) Y17002 and *G. cydonium* calmodulin (*CAMGC*) Y16818.

Supported by a grant from the Deutsche Forschungsgemeinschaft (Mü 348/12-1), the International Human Frontier Science Program (RG-333/96-M) and from CNPq and PRONEX of the Brazilian Government.

REFERENCES

Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. 1985 Three-dimensional structure of calmodulin. *Nature* **315**, 37–40.

Baldauf, S. L. & Palmer, J. D. 1993 Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natn. Acad. Sci. USA* **90**, 11 558–11 562.

Cavalier-Smith, T., Allsopp, M. T. E. P., Chao, E. E., Boury-Esnault, N. & Vacelet, J. 1996 Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Can. J. Zool.* **74**, 2031–2045.

Cox, J. A. & Kretsinger, R. H. 1998 EF-hand calcium-binding domain. The Prosite 'textbook'. Available from A. Bairoch, Department of Medical Biochemistry, University of Geneva, Switzerland. Oxford: Intelli Genetics Inc.

Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. 1978 A model of evolutionary change in protein. In *Atlas of protein sequence and structure* (ed. M. O. Dayhoff), vol. 5, suppl. 3, pp. 345–352. Washington, DC: National Biomedical Research Foundation.

Felsenstein, J. 1993 *PHYLIP, version 3.5.* Seattle: University of Washington.

Gallissian, M. F. & Vacelet, J. 1992 Ultrastructure of the oocyte and embryo of the calcified sponge, *Petrobiona massiliana* (Porifera, Calcarea). *Zoomorphology* **112**, 133–141.

Georgopoulos, C., Liberek, K., Zylicz, M. & Ang, D. 1994 Properties of the heat shock proteins of *Escherichia coli* and the autoregulation of the heat shock response. In *The biology of heat shock proteins and molecular chaperones* (ed. R. I. Morimoto, A. Tisserères & C. Georgopoulos), pp. 209–249. New York: Cold Spring Harbor Laboratory Press.

Gething, M. J. & Sambrook, J. 1992 Protein folding in the cell. *Nature* **355**, 33–45.

Günther, E. & Walter, L. 1994 Genetic aspects of the hsp70 multigene family in vertebrates. *Experientia* **50**, 987–995.

Hanson, P. I. & Schulman, H. 1992 Neuronal Ca^{2+}/calmodulin-dependent protein kinases. *A. Rev. Biochem.* **61**, 559–601.

Hardie, G. & Hanks, S. 1995 *The protein kinase factsbook: protein-tyrosine kinases*. London: Academic Press.

Hirabayashi, J. & Kasai, K. 1993 The family of metazoan metal-independent β-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* **3**, 297–304.

Hyman, L. H. 1940 *The invertebrates*, vol. 1 *(Protozoa through Ctenophora)*. New York: McGraw-Hill.

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B. & Bax, A. 1992 Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638.

Keeling, P. J. & Doolittle, W. F. 1996 Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Molec. Biol. Evol.* **13**, 1297–1305.

Knoll, A. H. 1992 The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627.

Kozak, M. 1991 An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887–903.

Koziol, C., Leys, S. P., Müller, I.M. & Müller, W. E. G. 1997 Cloning of Hsp70 genes from the marine sponges *Sycon raphanus* (Calcarea) and *Rhabdocalyptus dawsoni* (Hexactinellida). An approach to solve the phylogeny of sponges. *Biol. J. Linn. Soc.* **62**, 581–592.

Koziol, C., Kobayashi, N., Müller, I. M. & Müller, W. E. G. 1998 Cloning of sponge heat shock proteins: Evolutionary relationships between the major kingdoms. *J. Zool. Syst. Evol. Res.* **36**, 101–109.

Kretsinger, R. H. 1991 The EF-hand, homologs and analogs. In *Novel calcium-binding proteins* (ed. C. W. Heizmann), pp. 17–37. Berlin: Springer-Verlag.

Kruse, M., Gamulin, V., Cetkovic, H., Pancer, Z., Müller, I. M. & Müller, W. E. G. 1996 Molecular evolution of the metazoan protein kinase C multigene family. *J. Molec. Evol.* **43**, 374–383.

Kruse, M., Müller, I. M. & Müller, W. E. G. 1997 Early evolution of Metazoan serine/threonine- and tyrosine kinases: identification of selected kinases in marine sponges. *Molec. Biol. Evol.* **14**, 1326–1334.

Kruse, M., Leys, S. P., Müller, I. M. & Müller, W. E. G. 1998 Phylogenetic position of Hexactinellida within the phylum Porifera based on amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. *J. Molec. Evol.* **46**, 721–728.

Kumar, S. & Rzhetsky, A. 1996 Evolutionary relationships of eukaryotic kingdoms. *J. Molec. Evol.* **42**, 183–193.

Li, C. W., Chen, J. Y. & Hua, T. E. 1998 Precambrian sponges with cellular structures. *Science* **279**, 879–882.

Linse, K. & Mandelkow, E. M. 1988 The GTP-binding peptide of ß-tubulin. *J. Biol. Chem.* **263**, 15 205–15 210.

Margulis, L. & Schwartz, K. V. 1995 *Five kingdoms*. New York: Freeman and Company.

Mehl, D. & Reiswig, H. M. 1991 The presence of flagellar vanes in choanomeres of Porifera and their possible phylogenetic implications. *Z. Zool. Syst. Evol. Forsch.* **29**, 312–319.

Müller, W. E. G. 1995 Molecular phylogeny of metazoa (animals): monophyletic origin. *Naturwissenschaften* **82**, 321–329.

Müller, W. E. G. 1998 Molecular phylogeny of Eumetazoa: experimental evidence for monophyly of animals based on genes in sponges (Porifera). *Prog. Molec. Subcell. Biol.* **19**, 89–132.

Müller, W. E. G. & Schäcke, H. 1996 Characterization of the receptor protein-tyrosine kinase gene from the marine sponge *Geodia cydonium*. *Prog. Molec. Subcell. Biol.* **17**, 183–208.

Müller, W. E. G., Müller, I. M. & Gamulin, V. 1994 On the monophyletic evolution of the Metazoa. *Brazil. J. Med. Biol. Res.* **27**, 2083–2096.

Nicholas, K. B. & Nicholas, H. B. Jr 1996 *GeneDoc version 1.1.004L*; www.cris.com/~ketchup/genedoc.shtml.

Nogales, E., Wolf, S. G. & Downing, K. H. 1998 Structure of the αβ-tubulin dimer by electron crystallography. *Nature* **391**, 199–203.

PC/GENE 1995. *Data Banks CD-ROM*; release 14.0. Mountain View, CA: IntelliGenetics.

Pfeifer, K., Haasemann, M., Gamulin, V., Bretting, H., Fahrenholz, F. & Müller, W. E. G. 1993 S-type lectins occur also in invertebrates: high conservation of the carbohydrate recognition domain in the lectin genes from the marine sponge *Geodia cydonium*. *Glycobiology* **3**, 179–184.

Subjeck, J. R. & Shyy, T. T. 1986 Stress protein systems of mammalian cells. *Am. J. Physiol.* **17**, C1–C17.

Sullivan, K. F. 1988 Structure and utilization of tubulin isotypes. *A. Rev. Cell Biol.* **4**, 687–716.

Taylor, F. J. R. 1994 The role of phenotypic comparisons in the determination of protist phylogeny. In *Early life on earth* (ed. S. Bengtson), pp. 312–326. New York: Columbia University Press.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22,** 4673–4680.

Valentine, J. W. 1994 The Cambrian explosion. In *Early life on earth* (ed. S. Bengtson), pp. 401–412. New York: Columbia University Press.

Valentine, J. W., Erwin, D. H. & Jablonski, D. 1996 Developmental evolution of metazoan bodyplan: the fossil evidence. *Devl Biol.* **173**, 373–381.

Wainright, P. O., Hinkle, G., Sogin, M. L. & Stickel, S. K. 1993 Monophyletic origins of the Metazoa: an evolutionary link to fungi. *Science* **260**, 340–342.

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*. *Molec. Cell Biol.* **6**, 2317–2323.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.