J Mol Evol (1997) 44:178–188

JOURNAL OF MOLECULAR EVOLUTION

Distinct Annexin Subfamilies in Plants and Protists Diverged Prior to Animal Annexins and from a Common Ancestor

Reginald O. Morgan, M. Pilar Fernandez

Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Oviedo, E-33006 Oviedo, Spain

Received: 30 May 1996 / Accepted: 20 August 1996

Abstract. Annexin homologues in the kingdoms of Planta and Protista were characterized by molecular sequence analysis to determine their phylogenetic and structural relationship with annexins of Animalia. Sequence fragments from 19 plant annexins were identified in sequence databases and composite sequences were also assembled from expressed sequence tags for *Arabidopsis thaliana.* Length differences in protein amino-termini and evidence for unique exon splice sites indicated that plant annexins were distinct from those of animals. A third annexin gene of *Giardia lamblia* (*Anx21*-Gla) was identified as a distant relative to other protist annexins and to those of higher eukaryotes, thus providing a suitable outgroup for evolutionary reconstruction of the family tree. Rooted evolutionary trees portrayed protist, plant, and *Dictyostelium* annexins as early, monophyletic ramifications prior to the appearance of closely related animal annexin XIII. Molecular phylogenetic analyses of DNA and protein sequence alignments revealed at least seven separate plant subfamilies, represented by *Anx18* (alfalfa, previously classified), *Anx22* (thale cress), *Anx23* (thale cress, cotton, rape and cabbage), *Anx24* (bell pepper and tomato p34), *Anx25* (strawberry, horseradish, pea, soybean, and castor bean), *Anx26*-Zma, and *Anx27*-Zma (maize). Other unique subfamilies may exist for rice, tomato p35, apple, and celery annexins. Consensus sequences compiled for each eukaryotic kingdom showed some breakdown of the ''annexin-fold'' motif in repeats 2 and 3 of protist and plant annexins and a conserved codon deletion in repeat 3 of plants. The characterization of distinct annexin genes in plants and protists reflects their comparable diversity among animal species and offers alternative models for the comparative study of structure–function relationships within this important gene family.

Key words: *Giardia lamblia* — Evolutionary divergence — Land plants — Molecular phylogeny — Sequence alignment — Structural features — Subfamily classification

Introduction

The presence of annexins in protists and plants suggests a basic, if not vital, role for these proteins in cell functions common to all eukaryotes. The calcium-dependent phospholipid-binding characteristic of animal annexins is associated with ion channel properties and the inhibition of membrane-associated enzymes such as cytosolic phospholipase A_2 and protein kinase C (Raynal and Pollard 1994). Annexin modulation of plasma membrane events such as exocytosis and of nuclear regulatory processes involving cell growth and development is evident in plants as well as in animals (Battey and Blackbourn 1993; Seals et al. 1994; Wilkinson et al. 1995; Clark and Roux 1995; Proust et al. 1996). To the extent that their actions and species distribution are consistent with a uni-

Abbreviations: aa, amino acid; ANX, annexin protein; *ANX*/*Anx*, human/nonhuman gene locus or DNA; CDS, coding DNA sequence; EST, expressed sequence tag; gDNA, genomic DNA; Gsp, genus/species; Myr, million years; nt, nucleotide; ORF, open reading fame; UTR, untranslated region

Correspondence to: R.O. Morgan; e-mail: mpff@dwarf1.quimica.uniovi.es

versal cellular role, alternative and simpler cell models should then be useful in elucidating their true function(s). One caveat is that the discovery of new annexins from more distant species is making it increasingly difficult to distinguish between orthologous and paralogous genes for valid comparisons. Deciphering the abundant data on annexin structure, function, and regulation requires careful distinction between features that are based on a broad, common homology from those acquired by divergence through gene duplication and selection for novel functions.

Annexins from early diverging eukaryotes such as protists and plants represent key models of the ancient progenitor gene(s) and offer unique insight into the common structure–function requirements that will ultimately define their fundamental physiological role(s). Vertebrate annexins include ten known, unique subfamilies that appear to have limited representation in invertebrates, while other distinct genes from protists, plants, and invertebrates also appear to show greatest similarity with annexins from species limited to the same phylum (Morgan and Fernandez 1995a). We have used rigorous search techniques to cull DNA and protein sequences from the current molecular databases for 43 different eukaryotic species expressing products homologous with annexins. Many of these, including 19 unique fragments from plants, have never been classified, and their uncertain relationship to known annexin subfamilies precludes extrapolation of knowledge about them to other annexins. We have therefore identified and characterized these new protist and plant annexins to clarify their phylogenetic relationships and to compare their structural features to those of animal annexins. This is intended to provide a basis for assessing the relevance of molecular biology studies to the common role and individual functions of specific annexins.

Data and Methods

The common name annexin (alias lipocortin or calpactin) followed by a Roman numeral is used to designate subfamily or class members of this gene family, while the abbreviations *Anx* for the gene locus and DNA and ANX for protein are, by convention, suffixed with an Arabic numeral to identify the corresponding members. Sequences were identified in the current electronic databases using the BLAST email server of the National Center for Biotechnology Information (Altschul et al. 1990). Consensus amino acid sequences were compiled from a database of aligned annexin sequences for each eukaryotic kingdom by determining the most frequent residues at each position, and these sequences were used to search for additional protein or DNA entries showing high match scores with the consensus. Pairwise alignments with known annexins were evaluated for similarity and homology using the SSEARCH and PRSS programs, respectively, from the FASTA version 2.0 package (Pearson 1990). Annexin nucleotide (nt) and amino acid (aa) sequences were multiply aligned by CLUSTALW (Thompson et al. 1994) and subjected to phylogenetic analysis using computer programs from the Phylogeny Inference Package, PHYLIP version 3.57 (Felsenstein 1989). Options for the latter included repeated, random input of the sequence alignment followed by global rearrangements for maximum likelihood analysis (DNAML), or input of 100 bootstrap alignments for protein distance calculations using the Dayhoff point-accepted mutation matrix (PROTDIST and FITCH). Bootstrap resampling of the input data with random duplication and deletion of individual positions yielded congruent alignments that could be evaluated statistically for the extent to which they left the evolutionary tree (un)altered. The majority-rule consensus trees produced by CONSENSE-PHYLIP bear bootstrap values as measure of certainty about branching topology, while branch lengths in DNAML trees reflect distances between diverging molecular species.

Genus/species (Gsp) abbreviations for previously classified annexins I to XX (Morgan and Fernandez 1995a) include Cel (*Caenorhabditis elegans,* roundworm), Ddi (*Dictyostelium discoideum,* slime mold), Dme (*Drosophila melanogaster,* fruit fly), Gla (*Giardia lamblia*), Hsa (*Homo sapiens,* human), Hvu (*Hydra vulgaris*), and Msa (*Medicago sativa,* alfalfa). The new annexins being considered for phylogenetic classification are named in anticipation of results from the present comparative analyses and are summarized here for future reference. Their proposed subfamily designation with Gsp of origin is followed by the formal taxonomic and familiar names, plus references to the literature and/or electronic sequence databases (db for GenBank/EMBL/DBJ or pir for the NBRF protein database). They include *Anx21*-Gla from *Giardia lamblia* genomic DNA (gDNA) (Townson et al. 1994, db:L27221); *Anx22*-Ath and *Anx23*-Ath of *Arabidopsis thaliana* (thale cress), compiled here as composite cDNA sequences of the expressed sequence tags (ESTs) derived from genome sequencing projects (Newman et al. 1994; CNRS-France, unpublished); *Anx22*-Ath came from db:T04775, T13910, Z18073, and Z25968; *Anx23*-Ath came from db:H36260, H36536, H37398, H76134, H76460, H77008, R29768, R30014, T22046, T22585, T41940, T42209, T43657, T76739, Z17514, Z18518, Z26190, Z32565, Z33916, and Z47714; *Anx23*-Bca from *Brassica campestris* L. ssp. *pekinensis* (Chinese cabbage) 325-bp EST cDNA (Lim et al. 1996, db:L47930) not previously recognized as an annexin; *Anx23*-Bna from *Brassica napus* (rape) 234-bp EST cDNA (Park et al. 1993, db:H74681) not previously recognized as an annexin; *Anx23*-Ghi from *Gossypium hirsutum* (cotton) annexin p34 peptide fragments (Andrawis et al. 1993); *Anx24*-Can from *Capsicum annum* (bell pepper) 1180-bp annexin cDNA (Proust et al. 1996, db:X93308); *Anx24*-Les from *Lycopersicon esculentum* p34 annexin (Smallwood et al. 1990, 1992, db:X63996 and pir:S11461); *Anx25*-Aru from *Armoracia rusticana* (horseradish) gDNA flanking peroxidase genes (Fujiyama et al. 1990, db:D90116) not previously recognized as an annexin; *Anx25*-Fan from *Fragaria* × *ananassa* (strawberry) 1063-bp annexin cDNA (Wilkinson et al. 1995, db:U19941); *Anx25*-Gma from *Glycine max* (soybean) hypocotyl 500-bp annexin EST cDNA (Shi et al. 1995, db:T41436); *Anx25*-Psa from *Pisum sativum* (garden pea) annexin p35 peptide fragments (Clark et al. 1992); *Anx25*-Rco from *Ricinus communis* (castor bean) 511-bp EST cDNA (van de Loo et al. 1995, db:T24154, T24168, T24170), for which overlapping clones yielded a composite sequence and deduced peptide not previously characterized as an annexin; *Anx26*-Zma and *Anx27*-Zma representing *Zea mays* (maize) p33 and p35 annexins, respectively (Battey et al. 1996, db:X98244 and X98245).

Four additional sequences that probably also represent distinct subfamilies were not classified in the absence of a full-length subfamily representative. These included *Anx*-Agr from *Apium graveolens* (celery) 42-kDa annexin peptide fragments (Seals et al. 1994); *Anx*-Les p35 from *Lycopersicon esculentum* (tomato) p35 annexin peptide fragments (Smallwood et al. 1990, pir:S11462, S11463); *Anx*-Mdo from *Malus domesticus* (apple) 287-bp annexin gDNA (Clark and Roux 1995, db:L41393); and *Anx*-Osa from *Oryza sativa* (rice) 361-bp shoot EST cDNA (Sasaki et al. 1994, db:D39787) not previously recognized as an annexin. Three final sequences showing similarity to each other and to the 5′ end of *Anx14*-Ddi (Greenwood and Tsang 1991) included *Anx*-Ath-NT from *Arabidopsis thaliana* (thale cress) 363-bp EST cDNA (Newman et al. 1994, db:T20417), *Anx*-Osa-NT from *Oryza sativa* (rice) 453-bp shoot EST cDNA (Sasaki et al. 1994, db:D48581), and *Anx*-Zma-NT from *Zea mays* (maize) 450-bp endosperm EST cDNA (Shen et al. 1994, db:T18316).

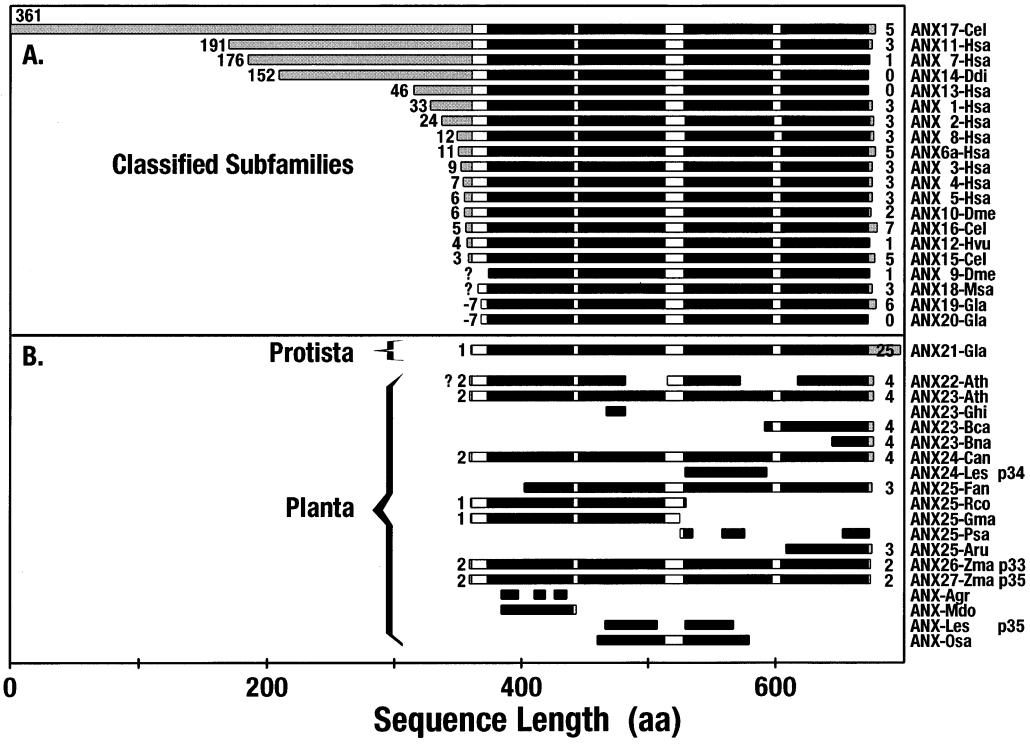

Fig. 1. Protein structures of annexin subfamilies. **A** Each outline depicts the variable amino-terminus (*shaded*), 311-aa tetrad of internal repeats (*solid blocks*) joined by unique linker segments (*open blocks*), and short carboxy-terminus (*shaded*) that are common to representatives of all 20 known annexin subfamilies (Morgan and Fernandez 1995a). They are arranged in order of decreasing amino-terminus size (see *scale*), with the aa lengths of their amino- and carboxy-termini shown on the *left* and *right*, respectively. Only the amino-terminal half of the human annexin VI octad is shown as ANX6a. **B** Regions of the new *Giardia* annexin, ANX21-Gla, and 18 plant sequences under study were aligned with the corresponding regions of known annexin subfamilies and shown with the same *shading patterns* as above. These alignments were used for phylogenetic studies that employed full-length sequences of 311 aa or 933 nt within the tetrad core region homologous for all annexins. Partial sequences were individually compared with the corresponding segments that could be aligned. *Abbreviated names* to the right are those proposed on the basis of the current study and are summarized under Data and Methods.

Results

Annexin Protein Structures

Members of the known annexin subfamilies are encoded by paralogous genes with common genomic organization, regulatory properties, and protein structural features. The 5′ coding DNA sequence (CDS) of each annexin exhibits sequence identity and structural homology limited to its respective subfamily. Hence, the sequence length and identity of the corresponding protein amino-termini (Fig. 1) provide a reliable means of inferring subfamily identity except where those amino-termini are short or unknown. The 3′-CDS region contains four (or eight in annexin VI) internally homologous repeats of 68–69 codons linked by repeat-specific segments of four to 16 codons. This latter tetrad core of 311 codons present in all known annexins was used for further structural and evolutionary comparisons.

Candidate sequences homologous to annexins by PRSS-FASTA (including deduced peptides and ambiguously back-translated DNA) were multiply aligned by CLUSTALW and imported into database spreadsheets to compute statistics for nt and aa conservation at each position (see later). A structural outline of the resulting protein alignments (Fig. 1A) obtained for representatives of the 20 classified annexin subfamilies (Morgan and Fernandez 1995a) included two from *Giardia lamblia* (Fiedler and Simons 1995) and a single plant species, ANX18-Msa (Pirck et al. 1994). The relative alignment positions of a third annexin from *Giardia lamblia* and 18 additional peptides from 15 different species of plants (see Data and Methods for names) have been outlined below (Fig. 1B). Annexins summarized in Fig. 1B are proposed members of novel subfamilies on the basis of the current study, including four incomplete structures not specified by a subfamily number. Sizes of both the amino- and carboxy-termini flanking the tetrad core were identical for ANX22-Ath, ANX23-Ath, and ANX24-Can, as for ANX26-Zma and ANX27-Zma, but distinct for ANX25 members. The respective termini are thus diagnostic of structures distinct from animal annexins, but not unique for individual plant subfamilies. Hence, the tetrad core region was the primary focus for phylogenetic analysis and structural comparisons.

Sequence and Structural Notes on Novel Annexins

The expectation of finding vertebrate annexin orthologues in extant plant species was very low, apart from

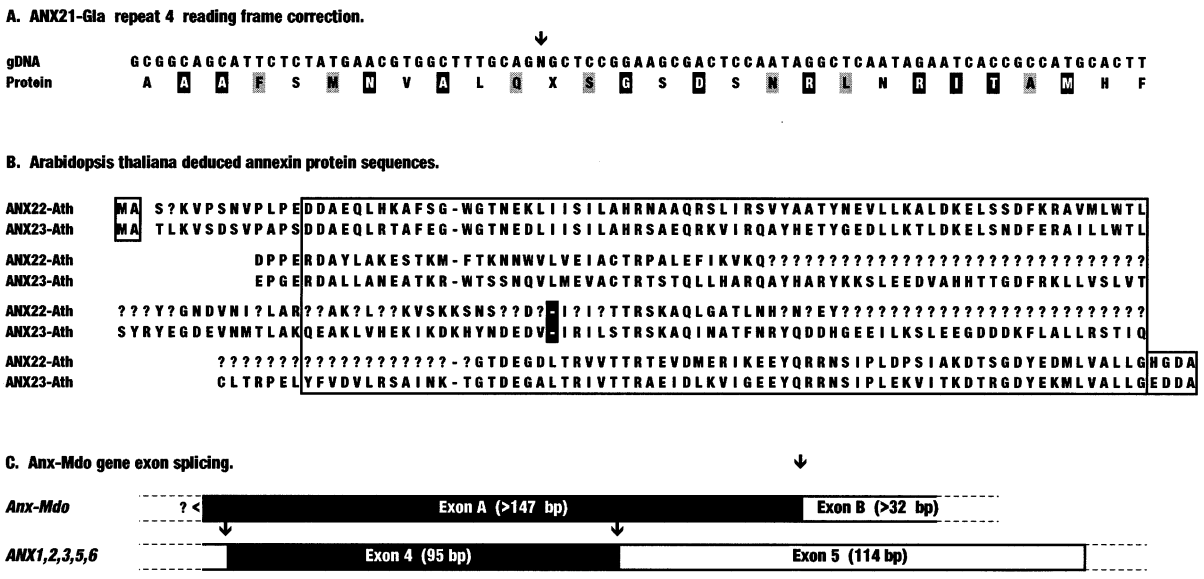

Fig. 2. Structure and alignment of selected annexin sequences. **A** The open reading frame (ORF) of repeat 4 in *Anx21*-Gla (Townson et al. 1994) was corrected by insertion of an ambiguous nt *N* (below *arrow*) corresponding to position 9846 in the complementary sequence of db:L27221. The resulting translation shift from ORF-1 to 3 gave a deduced aa sequence over the remainder of the repeat that retained significant identity (*reverse highlight*) and conservative aa replacement (*shaded residues*) with respect to ANX19-Gla. **B** Putative annexin proteins ANX22-Ath and ANX23-Ath (317 aa) were deduced from connected DNA sequences compiled from two separate alignments of four and 20 *Arabidopsis thaliana* EST sequences, respectively (Newman et al. 1994), that showed significant homology with other annexins. They are shown from their amino-termini (2 aa), through each of the four vertically aligned, homologous repeats (*large boxed area*), to their carboxy-termini (4 aa). *Question marks* signify unknown residues in ANX22-Ath and a *reverse-highlighted gap* in the middle of repeat 3 marks the deletion site common to other plant annexins. **C** Lack of correspondence between the exon 4 splice sites common to annexins I, II, III, V, and VI and that of aligned exon A (>147 bp, incomplete at 5′ end) in gDNA for *Anx*-Mdo (Clark and Roux 1995).

the structural considerations above. For example, the ten known human annexins have a mean aa identity between them of 49.8 ± 4.1% (range 44.0–58.5%) and similar estimated mutation rates of 1% aa replacements every 9.0 ± 3.3 Myr (Morgan and Fernandez 1995a). This implies pairwise divergence times of 250–700 Myr, depending on aa differences and individual protein mutation rates. The upper limit was given by the 55.8% aa difference between two slowly mutating annexins, ANX2-Hsa with 1% aa replacements every 16.7 Myr, and ANX7-Hsa with 1% aa replacements every 8.5 Myr. Conversely, ANX4-Hsa and ANX5-Hsa had only a 41.5% aa difference and more typical mutation rates of 1% aa replacements every 9.6 and 7.9 Myr, respectively, indicating that they may have duplicated about 200–300 Myr ago. These estimates assume a constant protein clock and their accuracy is dependent on knowledge of species divergence times. Yet they are consistent with the known existence of ANX2, ANX5, and ANX7 in amphibians, estimated to have diverged from mammals 300–350 Myr ago (Gould 1993). The limited representation of vertebrate annexins in invertebrates further suggests that animal annexins are unlikely to exist in plants, as the latter are believed to have diverged from eukaryotic lineage about 1,000 Myr ago (Doolittle et al. 1996).

A new annexin homologue from *Giardia lamblia,* originally sequenced as genomic DNA flanking the pyruvate-flavodoxin oxidoreductase gene (Townson et al. 1994), has not been previously characterized. It showed 52% nt identity to *Anx19*-Gla over 939 nt and 27.2% aa identity to ANX3-Hsa over 309 aa of its tetrad core region. Significance testing by PRSS-FASTA indicated that the optimal alignment score of 283 for the latter comparison had a probability of random occurrence equal to 3.93×10^{-15}. This index of homology with an animal annexin and the fact that the new annexin was equally distant (dissimilar) from all previously identified annexins suggested that the three *Giardia* annexins could collectively serve as an ideal outgroup to root evolutionary trees of annexins from higher eukaryotes. Following the convention of numbering annexin subfamilies in order of their identification, we shall call this gene *Giardia lamblia* annexin XXI, abbreviated *Anx21*-Gla. Proper alignment of the gDNA sequence of *Anx21*-Gla with other annexins required insertion of a single, ambiguous base early in the fourth internal repeat (Fig. 2A). The resulting shift and correction of the open reading frame retained significant homology for the deduced protein and preserved key aa known to be strongly conserved in other annexins (see consensus sequences later).

In view of the structural uniqueness of individual annexin subfamilies (Fig. 1), we sought to identify full-length cDNA and proteins among plant annexins that could serve as prototype representatives for new sub-

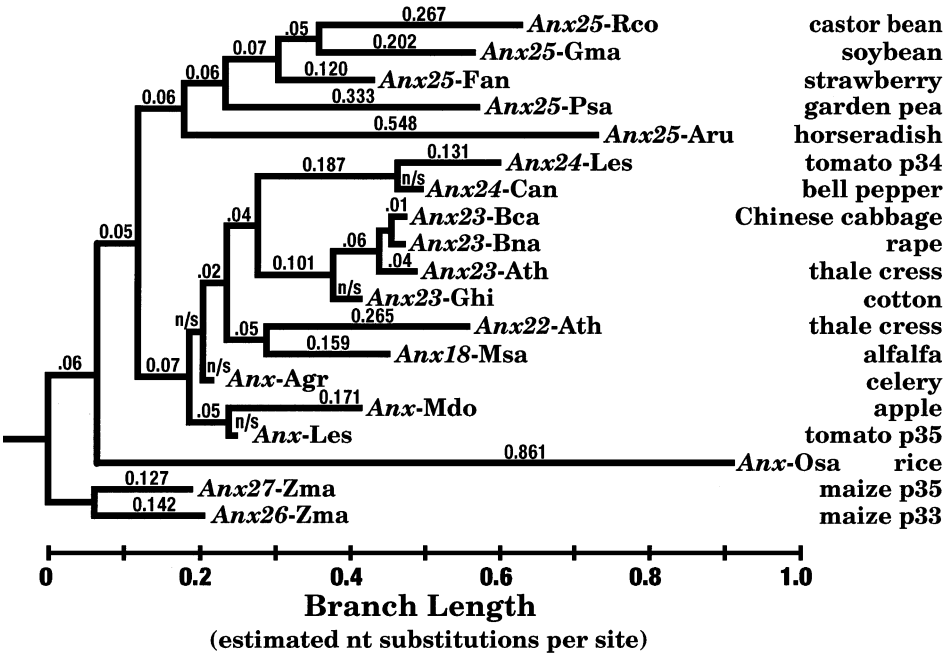

Fig. 3. Phylogenetic relationship of 19 plant annexins based on maximum likelihood analysis of cDNA sequences contained within the 933-nt homologous tetrad core region. The tree grows horizontally from its root and monocotyledon representatives at the *left.* Lengths shown on individual branches correspond to the *scale* below of estimated nt substitutions per site. Calculated confidence limits indicated that all branches were significantly positive ($P < 0.01$) except those marked as *n/s.* Annexins were named according to their proposed subfamily number (18, 22, 23, 24, 25, 26, and 27) and *genus/species abbreviation,* with common names to the *right* (see Data and Methods). Topology was based on the inclusion of three *Giardia* annexins as outgroup and was supported by subsequent analyses of peptide alignments. Celery, apple, tomato p35, and rice annexins probably represent novel subfamilies but remain unclassified because limited structural data precluded their adequate characterization. Sequences were multiply aligned using CLUSTALW (Thompson et al. 1994) and SSEARCH (Pearson 1990), consistent with the peptide alignment outlined in Fig. 1. Data were analyzed by DNAML-PHYLIP (Felsenstein 1989) with duplicate, random-order input on a DEC-VAX mainframe computer.

families. One obvious source of plant annexin cDNA fragments was the expressed sequence tag (EST) database derived from the *Arabidopsis thaliana* genome sequencing projects (Newman et al. 1994; CNRS-France unpublished). We used TBLASTN and a plant annexin consensus sequence to retrieve four candidate sequences with 95% overlapping nt identity that, within sequencing error limits, collectively defined a putative *A. thaliana* annexin, *Anx22*-Ath. Linkage of the overlapping fragments by DNASIS-CONNECT (Hitachi Software Engineering) yielded a composite cDNA sequence with an open reading frame (ORF) that spanned three segments of the tetrad core region (Fig. 1) and had 26.5% average aa identity with known annexins (Fig. 2B). Twenty other *A. thaliana* EST fragments also showed over 90% nt cross-identity in overlapping regions and significant homology with annexins but had only 68% nt identity with *Anx22*-Ath. Its composite cDNA contained a complete ORF that translated into 317 aa defining another putative annexin, ANX23-Ath (Fig. 2B).

The genomic DNA source of three new plant annexin fragments permitted some structure comparison with other known annexin genes. The *Anx25*-Aru sequence corresponded to exons 11, 12, and 13 of *ANX5*-Hsa and appeared to have no introns. The alignment of *Anx24*-Les (p34) gDNA with *ANX5*-Hsa exons 8, 9, and 10 without interruption confirmed its lack of introns also. An *Anx*-Mdo fragment contained two exons with unique splice sites (Fig. 2C) that did not coincide with the size or position of exons 4 and 5 in human annexins I, II, III, V, and VI (Fernandez et al. 1994), nor with the corresponding exons 7 and 8 of human annexin VII (Shirvan et al. 1994). This provided clear evidence for the existence of plant annexin genes distinct from those of animals.

DNA Maximum Likelihood Analysis of Plant Annexins

The 933-nt alignment of homologous core regions (Fig. 1) was subjected to maximum likelihood analysis by DNAML-PHYLIP to assess evolutionary interrelationships among all 19 plant annexins (Fig. 3). Initial analyses (not shown) that also included *Dictyostelium* and/or animal annexins determined that plant annexins were monophyletic with respect to other phyla. The inclusion of *Giardia* annexins as an earlier-diverging outgroup established the proper tree rooting to yield a correct orientation as shown in Fig. 3. Annexins from monocotyledons (maize and rice) represented the earliest-diverging plant annexin subfamilies. Subsequent branches formed several distinct clades, suggestive of unique subfamilies

comprising one to five different species. The maximum likelihood analysis would be proven fully consistent with later protein analyses that provided statistical support for the same.

The only previously classified plant annexin, *Anx18*-Msa, displayed significant branching ($P < 0.01$) from a node shared with *Anx22*-Ath (Fig. 3). Since neither of the *Arabidopsis* annexins showed similarity with *Anx18*-Msa in their 3′-untranslated regions (UTR) and their proteins gave no significant bootstrap probability of belonging to the ANX18 subfamily (see later), they were assigned to two new subfamilies as *Anx22*-Ath and *Anx23*-Ath, following that of *Anx21*-Gla. *Anx23*-Ath in fact appeared on a more remote branch, closely related to *Anx23*-Ghi and annexins from two *Brassica* species. The finding of conserved sequence similarity in the 3′-UTR between *Anx23*-Ath and *Anx23*-Bna supported this assignment, as did the p34 mass of *Anx23*-Ghi. *Anx24*-Can, the first plant cDNA with a complete coding region spanning the entire tetrad core, branched with *Anx24*-Les (p34), clearly separate from the *Anx23* clade. The largest subfamily grouping included the five sequences of *Anx25*-Gma, *Anx25*-Rco, a near full-length *Anx25*-Fan, *Anx25*-Psa, and *Anx25*-Aru. The three sequences from *Anx*-Les p35, *Anx*-Mdo, and *Anx*-Agr were clustered together on a branch leading to *Anx18*, *Anx22*, *Anx23*, and *Anx24*, but could not be unequivocally associated with any particular subfamily. *Anx*-Osa exhibited the longest branch length (greatest dissimilarity) among annexins and was most closely associated with the two early diverging monocot subfamilies represented by *Zea mays* annexins *Anx26*-Zma (p33) and *Anx27*-Zma (p35).

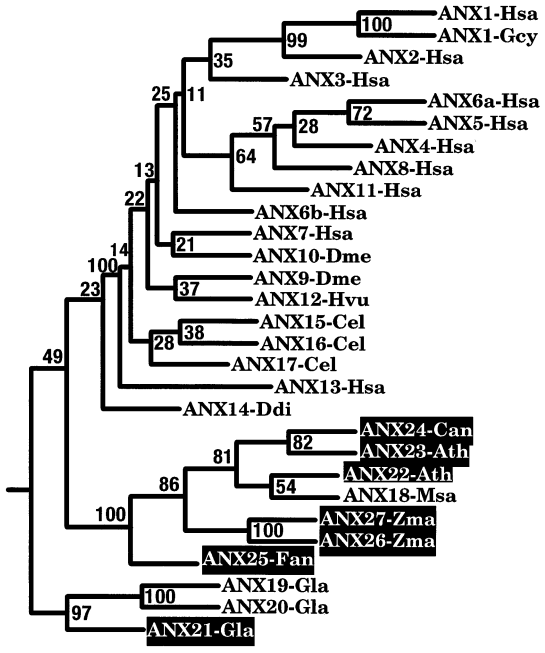

Fig. 4. Evolutionary relationships among annexin protein subfamilies. Seven new protist and plant annexin tetrads (*reverse-image labels*) were incorporated into an aa alignment of representative members from 20 known subfamilies. Pairwise distance matrices were computed by PROTDIST-PHYLIP for 100 bootstrap alignments of 29 species × 311 aa and analyzed by the least-squares method using FITCH-PHYLIP with global rearrangements (Felsenstein 1989). The consensus tree is shown with bootstrap values at each node to indicate the percentage frequency with which branches to the right occurred in the output. The tree grows from left to right from a root positioned on the branch between monophyletic protist and plant annexins. Unique designations for new subfamilies were based on consideration of the resulting bootstrap probabilities, the corresponding DNA maximum likelihood analysis (Fig. 3), and clade formation with additional peptide fragments (see later).

Protein Distance Analysis of Annexin Subfamilies

The greater conservation of protein structure in the evolution of gene families led to the use of distance comparisons between plant annexins and other known subfamilies as a means of corroborating and extending the DNA maximum likelihood results from Fig. 3. Incorporation of bootstrapping techniques provided a statistical basis for inferring subfamily identity in questionable cases of bifurcation. The first step was to take an alignment of the longest plant aa sequences with representatives of all known subfamilies shown in Fig. 1 and produce 100 bootstrap alignments using SEQBOOT-PHYLIP. A pairwise distance matrix computed for each by PROTDIST was analyzed by the least-squares method using FITCH. The resulting majority-rule consensus tree incorporated bootstrap statistics as a measure of the branching frequency in randomized samples. The results in Fig. 4 show evolutionary relationships between major representatives of the new plant annexin subfamilies and those of all previously classified annexins. They reflect the profile observed with DNA maximum likelihood analysis of plant annexins alone (Fig. 3) and clearly show plant annexins as a monophyletic group diverging after protist annexins.

High bootstrap values in Fig. 4 indicated the probability with which annexin species were evolutionarily related. Thus, the fact that ANX1-Hsa branched with sponge ANX1-Gcy in 100% of the bootstrap alignments clearly indicated that they were orthologues from different species, while the divergence of ANX2-Hsa in 99% of the analyses confirmed that it must be a closely related paralogous gene from the same species as ANX1-Hsa. The importance of the new ANX21-Gla and other *Giardia lamblia* annexins as homologues from a species estimated to have diverged from eukaryotic lineage over 1,200 Myr ago (Doolittle et al. 1996) was that they could be used as an outgroup to root the annexin family tree. Consequently, branches containing plant annexins and the later-diverging animal annexins are oriented according to their evolutionary relatedness to protist annexins and, by implication, to their common ancestral progenitor(s). While previous relationships among animal annexins were essentially maintained (Morgan and Fernandez 1995a), certain adjustments included the clear

emergence of ANX13-Hsa (Wice and Gordon 1992) as the most primitive animal annexin (with 100% bootstrap probability), with a slightly more distant appearance of ANX7-Hsa and the *Drosophila* annexins. Also, ANX8-Hsa branched prior to ANX4-Hsa in one of the two more distant clades of human annexins.

Figure 4 also showed 100% separation of the five ANX25 members from other plant annexins and of the ANX18/22 pair from ANX23/24. The relatively low bootstrap values between ANX18-Msa and ANX22-Ath, and between ANX23-Ath and ANX24-Can, implied that the individual members of these bifurcating pairs were not likely to belong to the same subfamily. We therefore regard ANX21-Gla, ANX22-Ath, ANX23-Ath, ANX24-Can, ANX25-Fan, ANX26-Zma, and ANX27-Zma as the best full-length representatives of seven novel annexin subfamilies. This was supported by comparison of protein structures (Fig. 1) where certain length differences existed among the amino-termini and carboxy-termini of the respective subfamilies. The position of ANX14-Ddi on a solitary branch intermediate between plant and animal annexins concurred with a recent phylogenetic assessment of *Dictyostelium* using 19 proteins, in which it was represented as belonging to an independent phylum of a fifth eukaryotic kingdom, Protoctista (Kuma et al. 1995).

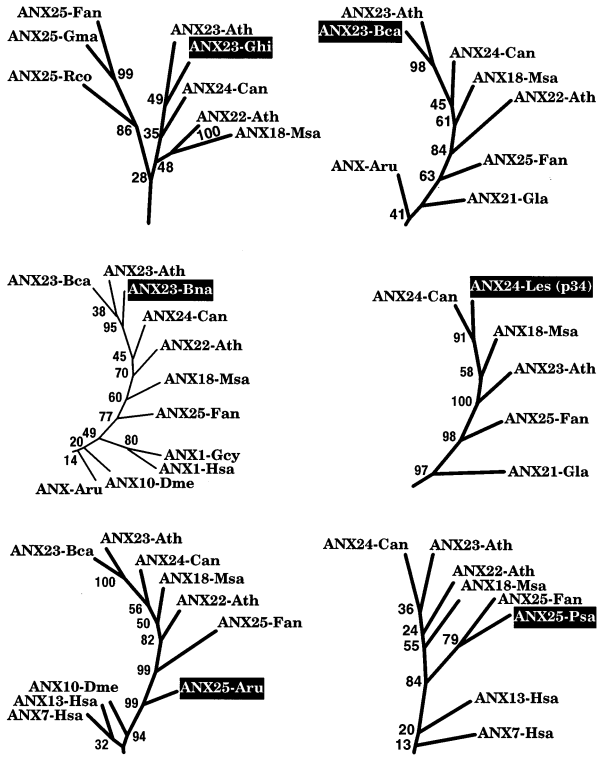

Fig. 5. Subfamily assignment of plant annexin peptide fragments. Separate partial alignments were constructed for each plant annexin test peptide (shown in *reverse highlight*) with the corresponding segments of annexin representatives from all known subfamilies and plants listed in Fig. 1. These consisted of 23 to 28 operational taxonomic units with a sequence length ranging from 14 to 81 aa, corresponding to the size of the test species. One hundred random bootstrap replicates were generated from each alignment and pairwise distance matrices were analyzed by the least-squares method using PROTDIST and FITCH programs, respectively. Only the relevant branch of each unrooted, majority-rule, consensus tree is shown with those annexins most closely related to the *highlighted* test species. *Percentages* at the branch nodes indicate bootstrap support for the proposed branching order and for the assignment of ANX23-Ghi (see text), ANX23-Bca (98%), ANX23-Bna (95%), ANX24-Les (91%), ANX25-Psa (84%/79%), and ANX25-Aru (99%/99%).

Subfamily Assignment of Plant Annexin Fragments

The relationship of ten smaller fragments of plant annexin protein, cDNA or gDNA, with established subfamilies remained to be evaluated. The procedure was essentially the same as described above for the longer sequences, except that each of the ten fragments was examined in an alignment restricted to the corresponding segment of representatives from every major subfamily and all plant annexins. Bootstrap alignments were generated by SEQBOOT and pairwise distances analyzed using PROTDIST and FITCH. The results of separate analyses (Fig. 5) permitted assignment of six additional plant annexins to the previously characterized subfamilies. Only that branch comprising the test species and most closely related annexins is shown, and the absence of protist or animal annexins simply affirmed their greater distance or dissimilarity from the test species. Annexins from cotton, Chinese cabbage, and rape were all related to ANX23-Ath with high bootstrap probability and were thus named ANX23-Ghi, ANX23-Bca, and ANX23-Bna, respectively. Although numerical results for ANX23-Ghi were somewhat less certain because of the short alignment (14 aa), its close association with both *Anx23* DNA (Fig. 3) and ANX23 proteins (Fig. 5) contrasted with the much greater distance from its nearest alternative neighbor, ANX24-Can. Similarly, ANX24-Les branched with ANX24-Can in 91% of bootstrap alignments and ANX25-Psa bifurcated on a common branch (84%) with ANX25-Fan in 79% of the samples. ANX25-Aru emerged adjacent to ANX25-Fan 99% of the time and was clearly associated with the ANX25 clade on a long branch in the DNAML analysis (Fig. 3).

Unclassified Annexin Fragments

The final four plant peptides achieved insufficient bootstrap support for unequivocal subfamily assignment and are thus likely to represent independent annexin subfamilies. Since they were not associated with any full-length representatives that could serve to define their respective subfamily characteristics, they have been identified without further classification. PROTDIST analysis of the 31-aa alignment of ANX-Agr peptides led to a bifurcation with ANX18-Msa in 42% of bootstrap replicates. Unfor-

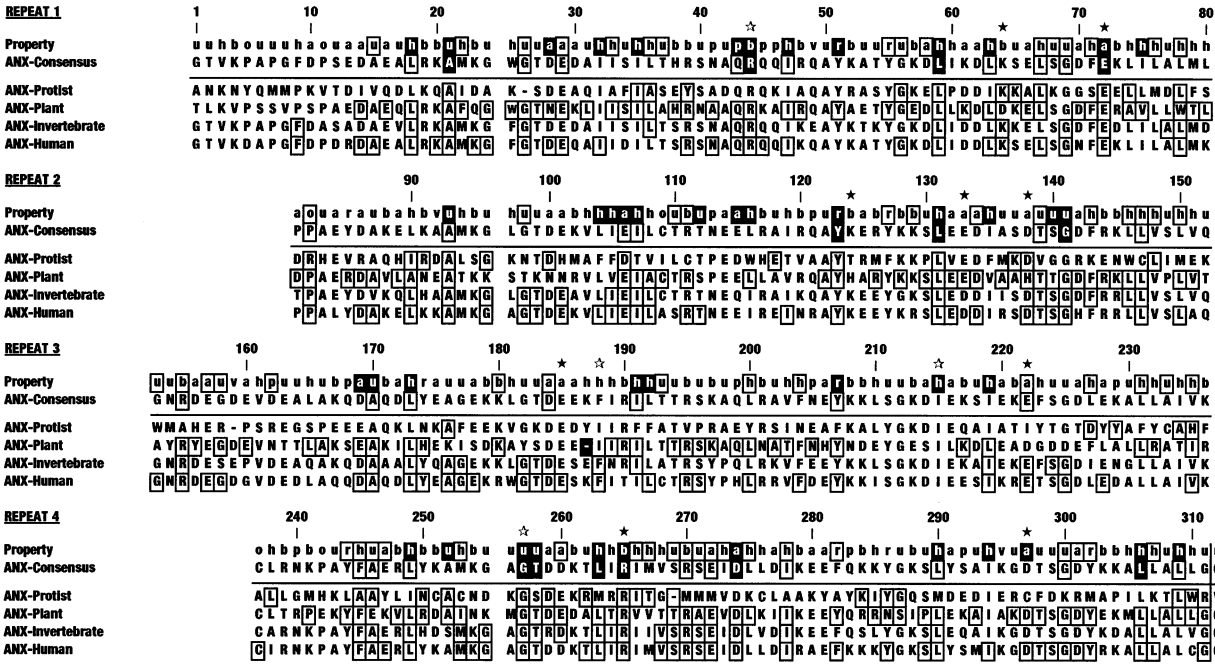

Fig. 6. Annexin protein consensus sequences and aa conservation. The most frequent aa residues at each position in a database alignment of annexin protein tetrad core regions were compiled into consensus sequences for each eukaryotic kingdom. Majority-rule consensus sequences were derived for protists ($n = 3$), plants ($n = 19$), invertebrate animals ($n = 13$), and *Homo sapiens* representatives of vertebrate animals ($n = 11$, including both tetrads of ANX6). The *top two lines* in each repeat summarize the aa property and identity observed in the master consensus of all 46 annexin taxa, while the *four lines beneath* characterize the individual kingdoms. *Letters* for aa are standard IUB symbols and *first-letter abbreviations* are used to describe their properties as acidic (*DE*), basic (*KRH*), hydrophobic (*LIVMFW*), polar (*QN*), ''r'' aromatic (*FY*), uncharged (*AGST*), or other (*CP*). *Boxed letters* signify over 80% identity at that position and *reverse-image letters* in the master consensus over 90% conservation. *Open stars* mark positions that are 100% conserved and of likely functional importance. *Solid stars* mark positions of known or perceived importance discussed in the text.

tunately, the atypical molecular mass of ANX-Agr (42 kDa) could not be compared directly with ANX18-Msa because the latter's full amino-terminal sequence is presently unknown. Analysis of the 60-aa alignment of ANX-Mdo with other annexins placed it on a bifurcation with ANX22-Ath in only 28% of bootstrap trees, distal to ANX18-Msa and ANX24-Can. ANX-Les p35 (70 aa) branched on a separate twig in 65% of bootstrap samples, near ANX27-Zma p35, ANX25-Fan, and ANX23-Ath—its closest neighbors. Finally, alignment of a rice peptide derived from the *Anx*-Osa EST sequence showed that it branched alone in 72% of bootstrap trees, adjacent to ANX14-Ddi.

The proximity of *Anx*-Agr, *Anx*-Mdo, *Anx*-Les p35, and *Anx*-Osa to each other at the nt level (Fig. 3), the inconclusive association of their peptides with other annexins, and apparent size differences (e.g., 35 kDa vs 42 kDa) pointed to the existence of two to four additional, unclassified subfamilies. Another unique EST from *Zea mays* whose translation product had 49% aa identity with the ''amino-terminus'' of ANX14-Ddi was used to retrieve similar ESTs from *A. thaliana* and *O. sativa.* These proline- and glycine-rich peptides showed 50–57% aa identity with ANX14-Ddi amino-terminus but did not faithfully reproduce its 18 tandem repeats of QGYPPQ and did not overlap with other sequences from the same species that had aligned with the annexin tetrad core region. It therefore remains uncertain whether they represent yet another plant annexin subfamily related to the unusual *Dictyostelium* annexin.

Annexin Protein Sequence Conservation in Different Kingdoms

Consensus sequences of the most frequent aa residues at each position of the tetrad core were compiled for annexins from the eukaryotic kingdoms of Protista ($n = 3$), Planta ($n = 19$), and Animalia, including invertebrates ($n = 13$) and the *Homo sapiens* representatives ($n = 11$) of vertebrates (Fig. 6). These sequences served as templates for database searches and multiple sequence alignments and as molecular models to identify conserved residues. The master sequence for all 46 taxa summarized both aa frequency and property, and certain positions conserved by over 80% or 90% marked important diagnostic references. Conserved replacements from one kingdom or subfamily to another were of special interest because they likely represented key structural and functional changes of paramount significance for evolutionary selection and divergence. One approach to viewing aa conservation in annexins is to identify constant sites

and then determine their role. The 100% conservation of Gly-257 and Arg-265 in repeats 4, the 97% identity of Arg-44 in repeat 1, and the absolute conservation of hydrophobic residues across positions 59, 131, 215, and 290 in all known annexins (open stars in Fig. 6) argue for universally key roles for these particular residues in all annexins. The Arg-265 has been suggested to serve as a voltage sensor in one ion channel model of annexins (Demange et al. 1994). A unique codon deletion at position 187 common only to plant annexins makes their repeats 3 equal in length to the 68 codons of other repeats, with unknown structural, functional, and evolutionary consequences.

Alternatively, the aa frequency data may be used to verify the conservation of sites already proposed to be of critical structural or functional importance on the basis of experimental studies such as X-ray crystallography or site-directed mutagenesis (solid stars in Fig. 6). Residues at positions 64, 67, and 72 of repeat 1 are believed to be important for hexamer formation in the putative ion channel of ANX12-Hvu (Luecke et al. 1995), but plants (and ANX16-Cel) have 100% conserved ''acidic'' residues at position 64 while all other annexins have a basic residue. The charged residues at positions 124, 129, and 133 have been proposed to be determinants of the ion pore environment in this model, although position 124 is quite variable. The ANX5 ion channel model sees Glu-89 as the key ion selectivity filter (Demange et al. 1994), so its replacement by 100% Leu in plant annexins and 67% His in protists would be expected to affect (putative) ion channel function in the latter organisms. The accepted importance for calcium binding of the acidic GTDE motif of the ''annexin-fold'' below repeat positions 26–29 in Fig. 6 is in fact disrupted by polar and basic residues in plant and protist repeats 2, and the compact Gly is replaced by a bulky aromatic group in plant repeats 3. Nevertheless, the type II calcium-binding pairs at positions 29/66, 101/138, 185/222, and 260/297 still show 80% conservation of acidic residues in all annexin repeats. These and other observations on the consensus sequences provide a valid basis for inferences about structure–function relationships in annexins, but more experimental testing will ultimately be needed to identify key domains and to verify theories about annexin action.

Discussion

Structural divergence among individual annexins, in their 5′ regulatory regions as well as the coding tetrad, is an important determinant of diversification in their expression and function. Although such differences complicate the distinction between their shared vs individual physiological role(s), a perspective on their evolutionary divergence can provide a rational basis for comprehending such relationships. Plant annexins share with their paralogous, animal counterparts the properties of calcium-dependent phospholipid binding and tissue-specific, growth-related expression during development (Clark and Roux 1995; Raynal and Pollard 1994). The same properties may also apply to *Dictyostelium* (Doring et al. 1995) and *Giardia* (Alonso and Peattie 1992). However, the possible deficiency of calcium-binding sites in plant annexin repeats 2 and 3 and other structural changes caused by conserved aa replacements (Fig. 6) requires close scrutiny before concluding that plant annexins are functionally equivalent homologues of animal annexins. It is further apparent from the similar evolutionary distances between annexin subfamilies among protists, plants, and animals (Fig. 4) that there has been comparable divergence within each of these separate kingdoms. Hence, generalizations and distinctions about annexin structure and function should take into account the full spectrum of known genes and proteins.

Multiple annexin subfamilies exist in single plant species, including at least two genes in each of maize, tomato, and thale cress. Single subfamilies such as *Anx23, Anx24,* and *Anx25* likewise appear to be prevalent in multiple plant species. The new ANX21-Gla is as distant from the other two *Giardia* annexins as it is from plant and animal annexins (ca. 25% aa similarity) and we observed no significant cross-identity of the 5′ or 3′ flanking regions between different annexin subfamilies. This provided no evidence for gene conversion, commonly observed in plants, and indicated that the unique protist annexins were unlikely to be direct progenitors or reversion products of annexins from higher eukaryotes. For this reason, rooting of the protein family tree on the branch joining protists and plants (Fig. 4) seems justified.

The recent identification of distinct annexins in another protozoan, *Paramecium* (Knochel et al. 1996), makes it all the more perplexing that annexins have not yet been identified in the kingdom of fungi, despite recent completion of the *Saccharomyces* genome sequencing project. We used TBLASTN and TFASTA sequence comparison programs with individual and kingdom-specific consensus protein sequences described here for rigorous searches of six reading-frame back-translations of the 12-Mb yeast nt database. The absence of a convincing match was surprising in light of the seeming ubiquity of annexins, but assuming their presence in fungi, it is possible that they comprise a separate, monophyletic clade that has suffered the additional ravages of an accelerated mutation rate (Doolittle et al. 1996). Such a combination of functional divergence and sequence erosion may ultimately explain why they have eluded identification and may necessitate the analysis of three-dimensional structures and other biochemical criteria to identify fungal annexins.

It is interesting that the progressive evolution of an-

nexin subfamilies (Fig. 4) coincides with the known speciation order (Doolittle et al. 1996) and that annexin subfamilies appear to have been successively lost and created in the separate eukaryotic kingdoms of Protista, Protocista, Planta, (Fungi?) and Animalia. The almost complete separation of annexin subfamilies between invertebrate and vertebrate animals, except for *Geodia* annexin I, emphasizes this point. Divergence of annexins through gene duplication and random drift has apparently been accompanied by selection based on functions related to the specific needs of individual organisms and cell types. The multiplicity of annexin genes in unicellular eukaryotes such as *Giardia* and *Paramecium* further implies that their association with membrane-related processes and cell growth or differentiation may even subserve a vital subcellular role.

The placement of 19 plant annexins into seven (plus several unnamed) subfamilies provides a limited perspective on what may be a diverse collection of subfamilies when one considers the vast differences in genome size and morphology within the plant kingdom. Actin genes, for example, number one to a few in protists or fungi and are single-digit in animals, but comprise ten genes in the 70-Mb genome of *Arabidopsis*, dozens in other plants, and over 100 in petunia (McDowell et al. 1996). Despite the impression of uniformity in plant annexin protein structures (Fig. 1), other subfamilies with longer amino-termini such as the 42-kDa annexin of celery (Seals et al. 1994), the 44–51-kDa annexins of *Paramecium* (Knochel et al. 1996), and perhaps even 68–70-kDa species similar to annexin VI (Clark and Roux 1995) all remain to be characterized. The annexin consensus sequences developed here can be applied to the design of degenerate oligonucleotides for PCR to amplify and identify annexin core regions contained within conserved segments for various species. Full characterization of amino-termini remains crucial for subfamily identification. More complete gene structures, such as that of *Anx*-Mdo (Clark and Roux 1995), will help to decipher annexin genomic elements (Fernandez et al. 1994; Morgan and Fernandez 1995b) and the mechanisms by which they control gene expression of different annexins. Further molecular biology studies of plant annexins should reveal the scope of annexin prevalence and involvement in plant (patho)physiology and may offer insight into how mammalian annexins contribute to human health and disease.

Note Added in Proof

A protein corresponding to our ANX23-Ath consensus sequence was recently characterized by Gidrol et al. (1996).

Acknowledgments. This work was supported by grant PB92-1000 from DGICYT of Spain.

References

Alonso RA, Peattie DA (1992) Nucleotide sequence of a second α-giardin gene and molecular analysis of the α-giardin genes and transcripts in *Giardia lamblia.* Mol Biochem Parasitol 50:95–104

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Andrawis A, Solomon M, Delmer DP (1993) Cotton fiber annexins: a potential role in the regulation of callose synthase. Plant J 3:763–772

Battey NH, Blackbourn HD (1993) The control of exocytosis in plant cells. New Phytol 125:307–338

Battey NH (1996) cDNA isolation and gene expression of maize annexins. Direct submission, db:X98244 and X98245

Clark GB, Roux SJ (1995) Annexins of plant cells. Plant Physiol 109: 1133–1139

Clark GB, Dauwalder M, Roux SJ (1992) Purification and immunolocalization of an annexin-like protein in pea seedlings. Planta 187: 1–9

Demange P, Voges D, Benz J, Liemann S, Gottig P, Berendes R, Burger A, Huber R (1994) Annexin V the key to understanding ion selectivity and voltage regulation? Trends Biochem Sci 19:272–277

Doolittle RF, Feng DF, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Doring V, Veretout F, Albrecht R, Muhlbauer B, Schlatterer C, Schleicher M, Noegel AA (1995) The *in vivo* role of annexin VII (synexin): characterization of an annexin VII-deficient *Dictyostelium* mutant indicates an involvement in Ca^{2+}-regulated processes. J Cell Sci 108:2065–2076

Felsenstein J (1989) PHYLIP—phylogeny inference package. Cladistics 5:164–166

Fernandez MP, Morgan RO, Fernandez MR, Carcedo MT (1994) The gene encoding human annexin V has a TATA-less promoter with a high G + C content. Gene 149:253–260

Fiedler K, Simons K (1995) Annexin homologues in *Giardia lamblia.* Trends Biochem Sci 20:177–178

Fujiyama K, Takemura H, Shinmyo A, Okada H, Takano M (1990) Genomic DNA structure of two new horseradish-peroxidase-encoding genes. Gene 89:163–169

Gidrol X, Sabelli PA, Fern YS, Kush AK (1996) Annexin-like protein from *Arabidopsis thaliana* rescues *delta-oxyR* mutant of *Escherichia coli* from H_2O_2 stress. Proc Natl Acad Sci USA 93:11268–11273

Gould SJ (1993) The book of life. Ebury Hutchinson, Random Century House, London

Greenwood M, Tsang A (1991) Sequence and expression of annexin VII of *Dictyostelium discoideum.* Biochim Biophys Acta 1088: 429–432

Knochel M, Kissmehl R, Wissmann J-D, Momayezi M, Hentschel J, Plattner H, Burgoyne RD (1996) Annexins in *Paramecium* cells. Involvement in site-specific positioning of secretory granules. Histochem Cell Biol 105:269–281

Kuma K, Nikoh N, Iwabe N, Miyata T (1995) Phylogenetic position of *Dictyostelium* inferred from multiple protein data sets. J Mol Evol 41:238–246

Lim CO, Kim HY, Kim MG, Lee SI, Chung WS, Park SH, Hwant I, Cho MJ (1996) Expressed sequence tags of Chinese cabbage flower bud cDNA. Plant Physiol 111:577–588

Luecke H, Chang BT, Mailliard WS, Schlaepfer DD, Haigler HT (1995) Crystal structure of the annexin XII hexamer and implications for bilayer insertion. Nature 378:512–515

McDowell JM, Huang S, McKinney EC, An Y-Q, Meagher RB (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana.* Genetics 142:587–602

Morgan RO, Fernandez MP (1995a) Molecular phylogeny of annexins

and identification of a primitive homologue in *Giardia lamblia*. Mol Biol Evol 12:967–979

Morgan RO, Fernandez MP (1995b) A BC200-derived element and Z-DNA as structural markers in annexin I genes; relevance to Alu evolution and annexin tetrad formation. J Mol Evol 41:979–985

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 106:1241–1255

Park YS, Kwak JM, Kwon OY, Kim YS, Lee DS, Cho MJ, Lee HH, Nam HG (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. Plant Physiol 103:359–370

Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 183:63–98

Pirck M, Hirt H, Heberle-Bors E (1994) The cDNA sequence encoding an annexin from *Medicago sativa*. Plant Physiol 104:1463–1464

Proust J, Houlne G, Schantz ML, Schantz R (1996) Characterization and gene expression of an annexin during fruit development in *Capsicum annum*. FEBS Lett 383:208–212

Raynal P, Pollard HB (1994) Annexins: the problem of assessing the biological role for a gene family of multifunctional calcium- and phospholipid-binding proteins. Biochim Biophys Acta 1197:63–93

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Nagasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, Minobe Y (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J 6:615–624

Seals DF, Parrish ML, Randall SK (1994) A 42-kilodalton annexin-like protein is associated with plant vacuoles. Plant Physiol 106:1403–1412

Shen B, Carneiro N, Torres-Jerez I, Stevenson R, McCreery T, Helentjaris T, Baysdorder C, Almira E, Ferl R, Habben J, Larkins B (1994) Partial sequencing and mapping of clones from two maize cDNA libraries. Plant Mol Biol 26:1085–1101

Shi J, Dixon RA, Gonzales RA, Kjellbom P, Bhattacharyya MK (1995) Identification of cDNA clones encoding valosin-containing protein and other plant plasma membrane-associated proteins by a general immunoscreening strategy. Proc Natl Acad Sci USA 92:4457–4461

Shirvan A, Srivastava M, Wang MG, Cultraro C, Magendzo K, McBride OW, Pollard HB, Burns AL (1994) Divergent structure of the human synexin (annexin VII) gene and assignment to chromosome 10. Biochemistry 33:6888–6901

Smallwood MF, Keen JN, Bowles DJ (1990) Purification and partial sequence analysis of plant annexins. Biochem J 270:157–161

Smallwood MF, Gurr SJ, McPherson MJ, Roberts K, Bowles DJ (1992) The pattern of plant annexin gene expression. Biochem J 281:501–505

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Townson SM, Hanson GR, Upcroft JA, Upcroft P (1994) A purified ferredoxin from *Giardia duodenalis*. Eur J Biochem 220:439–446

van de Loo FJ, Turner S, Somerville C (1995) Expressed sequence tags from developing castor seeds. Plant Physiol 108:1141–1150

Wice BM, Gordon JI (1992) A strategy for isolation of cDNAs encoding proteins affecting human intestinal epithelial cell growth and differentiation: characterization of a novel gut-specific *N*-myristoylated annexin. J Cell Biol 116:405–422

Wilkinson JQ, Lanahan MB, Conner TW, Klee HJ (1995) Identification of mRNAs with enhanced expression in ripening strawberry fruit using polymerase chain reaction differential display. Plant Mol Biol 27:1097–1108