Maria-Pilar Fernandez, Montserrat Garcia[a], Silvia Martin-Almedina[b] and Reginald O. Morgan*

Novel domain architectures and functional determinants in atypical annexins revealed by phylogenomic analysis

DOI 10.1515/hsz-2016-0273
Received August 27, 2016; accepted December 11, 2016; previously published online December 20, 2016

Abstract: The fundamental cellular role and molecular interactions of annexins in vesicle trafficking and membrane remodeling remain to be further clarified in order to better understand and exploit their contributions to health and disease. We focused on distinctive features of atypical annexins from all domains of life using phylogenomic, molecular systematic and experimental approaches, to extend the current paradigm and better account for annexin diversity of structure, function and mechanistic role in membrane homeostasis. The analysis of gene duplications, organization of domain architectures and profile hidden Markov models of subfamily orthologs defined conserved structural features relevant to molecular interactions and functional divergence of seven family clades ANXA-G. Single domain annexins of bacteria, including cyanobacteria, were frequently coupled to enzymatic units conceivably related to membrane metabolism and remodeling. Multiple ANX domains (up to 20) and various distinct functional domains were observed in unique annexins. Canonical type 2 calcium binding ligands were well-preserved in roughly half of all ANX domains, but alternative structural motifs comprised of 'KGD', cysteine or tryptophan residues were prominently conserved in the same strategic interhelical loops. Selective evolutionary constraint, site-specific location and co-occurrence in all kingdoms identify alternative modes of fundamental binding interactions for annexins.

[a]Current address: Foundation for Ophtalmologic Investigation, Avda. Doctores Fernández-Vega 34, E-33012 Oviedo, Spain
[b]Current address: Cardiovascular and Cell Sciences Institute, St. George's University of London, London SW17 0RE, UK
***Corresponding author: Reginald O. Morgan,** Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Oviedo, Edificio Santiago Gascon 4.3, E-33011 Oviedo, Spain, e-mail: morganreginald@uniovi.es
Maria-Pilar Fernandez, Montserrat Garcia and Silvia Martin-Almedina: Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Oviedo, Edificio Santiago Gascon 4.3, E-33011 Oviedo, Spain

Keywords: 3D modeling; conserved motifs; functional determinants; membrane association; molecular phylogeny; profile hidden Markov models.

Introduction

Those who moil for annexins have made significant advances during four decades of multidisciplinary research in documenting the molecular properties and cellular roles of this ubiquitous and unique gene superfamily (Creutz et al., 1978; Gerke et al., 2005). Observations and empirical testing of their association with cell membrane lipids and cytoskeletal proteins are compatible with their proposed participation in intracellular transport processes leading to vesicle translocation, fusion and transcytosis (Potez et al., 2011; Tebar et al., 2014). These basic processes can be conceptually linked to an integrated physiological role in membrane remodeling and maintenance to promote cellular membrane growth, repair and to facilitate transmembrane microvesicle transport (Bouter et al., 2015; Draeger et al., 2011; Demonbreun and McNally, 2016). The prominent expression and apical trafficking of annexins in epithelial and endothelial tissues support this interpretation of a protective homeostatic role, but they have additional, distinct actions in immune responses (D'Acquisto et al., 2008), secretory, apoptotic and chemoresistance events, antithrombotic, anti-inflammatory and tumor suppressor roles, and in muscle contraction and neuronal activity. Such functional diversity argues for even broader roles as scaffolding components of multifunctional complexes involved in cellular signal transduction networks of multicellular organisms and other coordinated molecular interactions at the level of unicellular organisms.

Annexin subfamily specificity in calcium affinity is influenced by individual structural variation under different proteoglycolipid environments and intracellular pH that modulate membrane binding kinetics (Gerke et al., 2005; Potez et al., 2011). This basic paradigm defines annexins as calcium-mediated, membrane lipid-associated proteins, but does not consider calcium-independent

annexins, other conserved structural motifs, distinct functional domains, unique amino-termini and varied architectures to adequately postulate integrated roles and mechanisms for individual annexin function. A unifying concept for annexins must recognize their commonly shared characteristics as well as the unique features that define individual members. The advent of high throughput, low-cost sequencing has already supplied ample genomic, transcriptomic and proteomic data to permit the characterization and comparative analysis of some 4000 individual annexins. These data provide new evidence for 'atypical' annexins from diverse phyla in which the fundamental relationships between structure and function are expected to be preserved despite mechanistic variance.

Phylogenetic reconstruction of the annexin gene superfamily from bacterial, protist, plant, fungal and metazoan sequences (Morgan and Fernandez, 1995, 1997; Weiland et al., 2005) can chart the complete gene family history and effectively distinguish paralogs and orthologs for a comprehensive classification of individual gene subfamilies (Moss and Morgan, 2004; Clark et al., 2012). It can also provide informative evidence for positive or negative selection and the dates, rates and patterns of divergence or conservation. This is essential for comparative studies because species orthologs vary by essentially silent changes in primary structure without significant functional differences, whereas the multiple paralogous genes within single species diverge at semi-conserved sites responsible for subfamily functional specificity. The evolutionary distance between paralogs duplicated from a common ancestor depends not only on the timing and rate of divergence but also on functional adaptation pertinent to each species group, so it can be insightful to follow these features throughout an extensive period of evolutionary change that the annexin superfamily has undergone. Molecular fingerprints of each subfamily can be generated from multiple sequence alignments as profile hidden Markov models (pHMM) to predict amino acid profiles with statistical precision and to decipher conserved sequence patterns that can be used to infer functional importance. This evolutionary, structural and physicochemical property information can be represented in 3D models to substantiate those inferences and to define the spatial context of key functional domains, motifs and amino acids responsible for mediating physiological and pharmacological interactions of individual annexins. The bioinformatic development of these models is aimed at defining the common features and distinctive characteristics of all functionally unique subfamilies pertaining to this superfamily.

Results

Annexin molecular phylogenetics

The 12 annexins common to vertebrates (*ANXA1-ANXA13*, excluding unassigned *ANXA12*) extend from the earliest diverging extant species, the jawless fish lamprey and hagfish, whilst annexins A5, A6 and A10 appeared later in cartilaginous fish and *ANXA8* in coelacanth (Figure 1). The annexin A2-A1-A9 clade also separated early from a lamprey common ancestor and *ANXA9* first appeared in cartilaginous fish. Earlier metazoan origins of annexins A13 and A7 were apparent from their close association with known nonvertebrate outgroups and from other similarities in genetic linkage and gene exon organization (Fernandez and Morgan, 2003). Among the 10 remaining vertebrate annexins, *ANXA11* can be regarded as the most basal member from Agnatha with a gene organization congruous among all annexins except A7 and A13 (Moss and Morgan, 2004). Copy number variation is important for assessing associated phenotypes, so the documentation of extra paralogous gene duplications is noteworthy (Figure 1). *ANXA1* has two to five copies in certain fish, birds and amphibians, *ANXA7* and *ANXA13* duplicated in lamprey, like most other annexins in teleost fish, and a second *ANXA8* gene has emerged only in humans, consistent with a genome propensity for segmental chromosomal duplications.

The fact that major taxonomic groups 'approximate' monophyletic distributions of species groups must have been the result of evolutionary selection by which annexins have adapted and diverged within at least six major families; e.g. the ANXA family consisting of 12 gene subfamilies of more than 2000 known vertebrate annexins. Distinct clades form other families, including ANXB from nonvertebrate metazoa, ANXC with fungal-related annexins, ANXD comprising 40+ subfamilies in plants (Clark et al., 2012), ANXE in diverse protists and ANXF now represented by more than 200 bacterial annexins (this study). The loss of annexin genes from certain genomes was deduced from more extensive analyses to document the absence of ANXA3 from Archosauria (crocodiles, dinosaurs and birds) and Squamata (snakes), ANXA8 from Afrotherian mammals, Amphibia, most fish and Lepidosauria (salamanders and snakes), ANXA9 from Archosauria, ANXA7 and ANXA10 from Teleostei (bony fish). Annexin phylogeny has been partially examined for fungi (Khalaj et al., 2015), plants (Clark et al., 2012), protists (Einarsson et al., 2016) and bacteria (Kodavali et al., 2014) and these were not reexamined in detail here except for divergent examples of special interest.

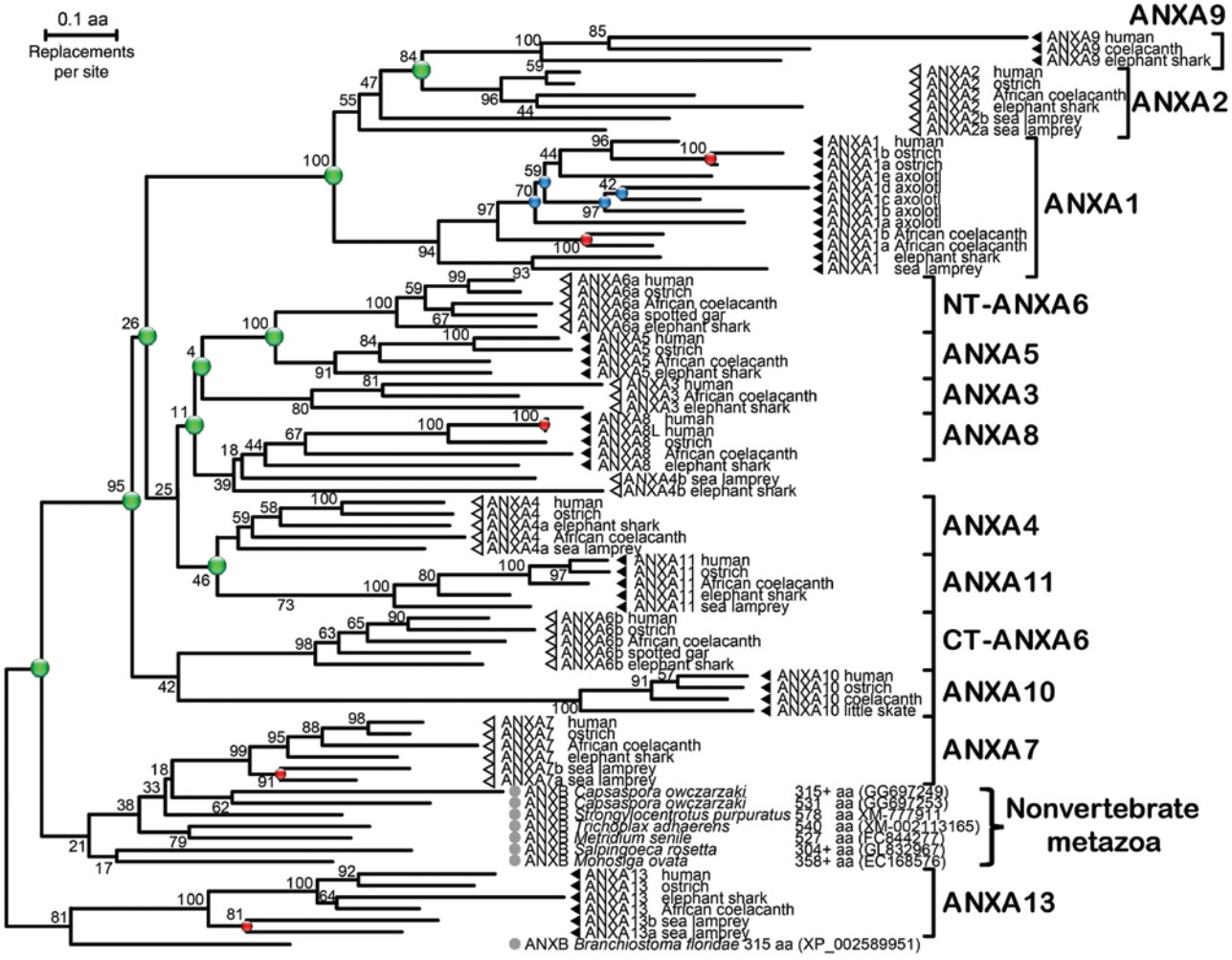


Figure 1: Molecular phylogeny of the vertebrate annexin ANXA family.
Proteins comprising the core tetrad of protein subfamilies from lamprey, shark, spotted gar, coelacanth, ostrich and human were aligned for analysis by Maximum Likelihood using RAxML v8 and MEGA v7 with 100 bootstrap pseudoalignments and gamma rate site weighting. The twelve paralogous subfamilies formed distinct clades and confirmed the origin of the earliest diverging annexins A13 and A7 from ancestors in common with nonvertebrate representatives. The remaining subfamilies with congruent gene organization duplicated from a common ancestor with the ANXA11 founder gene in Agnatha while several subfamilies such as annexins A5, A6, A9 and A10 appeared later in cartilaginous fish. The gene duplication order (green spheres) and chronology are given by the horizontal bifurcations and branch lengths, while confidence values at the nodes support the tree topology. Although several early diverging taxa for ANXA3, A4 and A5 were not well-resolved, additional intralineage, coparalogous gene duplications were evident for lamprey ANXA13 and A7, human annexin A8 and annexin A1 in coelacanth, amphibians (5 copies) and birds (red and blue spheres).

Domain architecture variations in annexin proteins

Annexin research has focused primarily on vertebrate members, plants and some earlier diverging eukaryotes in which the predominant form comprises a homologous tetrad of 68 aa ANX domains. Such an architecture presumably originated from a monomeric annexin such as those found in bacteria (Figure 2) and in principle should be capable of further duplication, although only vertebrate octad annexin A6 has provided evidence for an alternative architecture. We scrutinized thousands of nonvertebrate annexins to identify members with novel architectures ranging from one to 20 ANX domains and including other unique domains indicative of more complex multifunctional interactions or a scaffolding role. A graphical summary of annexin protein domains (Figure 2) compared numerous members with multi-domain composition and identified those with foreign domains or ANX domains lacking apparent calcium-binding sites. Numerous nonvertebrate metazoa possess an ANX octad architecture phylogenetically distinct from

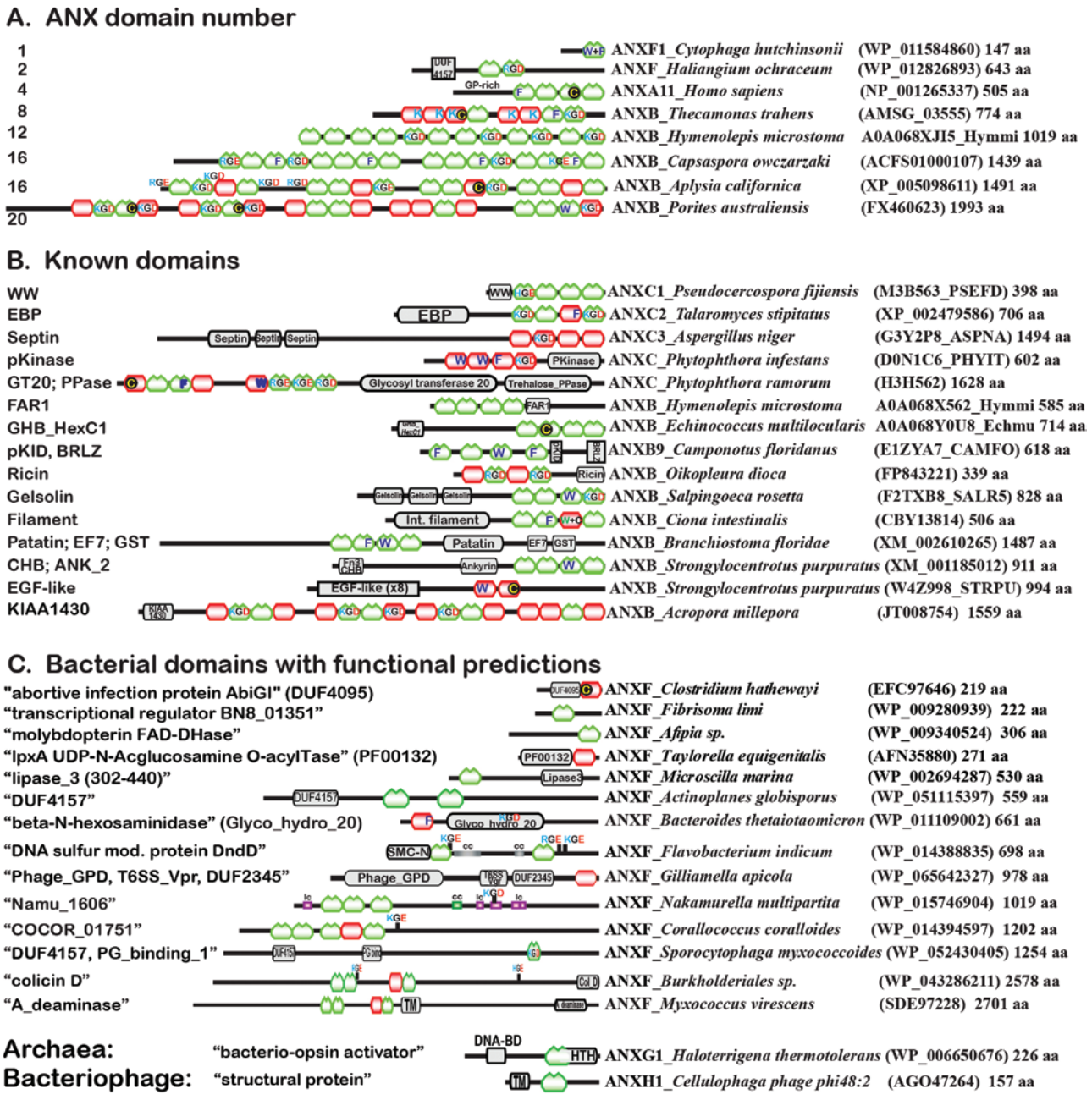


Figure 2: Diversity in the organization of protein functional domains and structural motifs of atypical annexins.
A curated library of annexins derived from protein, transcriptomic and genomic sequence databases was searched with locally reconstructed pHMM models providing molecular fingerprints of the annexin domain and established subfamilies. Confirmed annexins were analyzed for annexin domain number (A), further scrutinized by the SMART (Letunic et al., 2015), PFAM (Finn et al., 2016) and ELM (Dinkel et al., 2016) servers to identify additional distinct domains (B), and extended to the annotations accompanying bacterial annexin GenBank entries (C). The schematic shows linear protein length, the absence (red) or presence (green) of type 2 calcium-binding motifs in ANX domains, and the inclusion of other prominently conserved or strategically located amino acids with hydrophobic, aromatic, bulky properties (Trp-W, Phe-F), redox potential (Cys-C), positively charged basic residues (Arg-R, Lys-K) or the disintegrin motif 'KGD'.

vertebrate annexin A6. They were thus formed by independent convergent evolution and this amplification extended to octad in the protist *Thecamonas trahens* and 20 ANX domains in the coral *Porites australiensis*. Single domain annexins thus amplified to multidomain annexins within bacteria (Figure 2C) and, in the absence of large-scale lateral transfer verified by phylogenetic analyses (Garcia-Diaz, 2011), gave rise to eukaryotic annexins

in algae and protozoa via other processes (e.g. symbiosis, genetic recombination).

The extended ANX domain architectures of non-vertebrate annexins were further accompanied by various unrelated domains with alternate binding properties (Figure 2B). The binding properties and general function of these supplementary domains are either uncharacterized (e.g. KIAA1430, PF00132, Domains of Unknown Function DUF234, 515, 975, 1104, 3106, 4095, 4116 and especially DUF4157), partially defined (viz. septin, gelsolin, EGF-like) or with intrinsic enzymatic activity (i.e. adenosine deaminase, protein kinase glycosyl transferase, hydrolase or lipase). Other domains detected by comparison with PFAM or SMART pHMM models included WW, EBP, FAR1, GHB_HexC1, pKID, BRLZ, ricin, intermediate filament, patatin, GST, CHB, ankyrin, SMC_N or colicin_D. The incorporation of such diverse domains with potential, varied roles is reminiscent of the unique amino terminal domain of certain vertebrate annexins that have similarly been proposed to realize distinct, essential functions (Bena et al., 2012; Yuan et al., 2016). More complete names and documentation are available from the corresponding source databases PFAM (Finn et al., 2016) and SMART (Letunic et al., 2015).

Conserved structural motifs were also prevalent within many ANX domains where aa replacements made them incapable of binding calcium, and such altered features are of growing interest to resolve the full spectrum of molecular interaction mechanisms as described in the Eukaryotic Linear Motifs database (Dinkel et al., 2016). The presence of 'KGD' disintegrin binding motifs, aromatic hydrophobic Trp/Phe residues and Cys residues have been well-recognized to be conserved in a strategic location (D–E-interhelical loop) of the ANX domain, so the observation here that these structural features now extend to protists and bacteria (Figure 2) accentuates their probable importance in the fundamental binding interactions of annexins independent of the type 2 calcium-binding site. The composite design of selected bacterial annexins (Figure 2C) has provided extra information for tentative annotations, functional predictions, even implicating some directly in membrane-related enzymatic activity.

Identification of conserved structural motifs within the ANX domain

It is clear that the calcium-binding capacity of annexin domains responsible for union with membrane phosphatidylserine represents a single property that may be neither universal nor essential for their function. This is exemplified by the vertebrate annexins ANXA9 and ANXA10 and similar homologs throughout eukaryotic and prokaryotic phyla and implicates other known domains and structural motifs in the translocation and association to membrane scaffold constituents, cytoskeletal proteins and extracellular matrix components. A statistical analysis of multiple sequence alignments for individual subfamilies was given by the construction of profile hidden Markov models, to define conserved structural motifs within the ANX domain that could be relevant to their functional specificity (Figure 3).

We compiled pHMM matrices for annexins A9 and A10 and visualized the information content in their domains 3 and 4 as sequence logos. The obliteration of calcium coordinating residues in all 100 ANXA9 representatives was accompanied by the emergence of HGD/KGD tripeptides conserved in many D–E interhelical loops. Note that this motif and a YKKKY motif in domain 4 (residues 312–316) are characteristics shared with and inherited from annexins A1 and A2 (Morgan et al., 2004) and that annexins A4 and A5 also manifest KGD ligands in the same D–E loop but prior to the acidic calcium-binding residue. The pHMM molecular profile of annexin A10 was strikingly different in that A–B and D–E loop functional residues were replaced by conserved basic residues and Cys in domain 3 and by the aromatic, hydrophobic, bulky amino acid Phe in domain 4. Elevated levels of basic, Cys and Phe/Trp were found to be characteristic of most annexins A10 (highlighted in Figure 3) and presumably substitute for the eroded calcium-coordinating ligands to determine the specific, distinct binding targets with which ANXA10 interacts, perhaps including the nuclear ribonucleoprotein bodies, paraspeckles (Quiskamp et al., 2014).

Annexin A9 and A10 molecular interactions and functional roles

The pHMM sequence logos generated for annexins A9 and A10 (Figure 3) identified conserved motifs with the highest probability of a functional role determined by evolutionary selection. This information must be revisualized in a 3-dimensional context (Figure 4A) to properly interpret how molecular interactions might transpire at the cellular or vesicular membrane, cytoskeleton or extracellular matrix. Physiological receptors have been sought in extensive studies of annexin A9 through immunohistochemical colocalization with cytoskeletal proteins such as keratin (Garcia-Diaz, 2011) and periplakin (Boczonadi and Maatta, 2012) and for annexin A10 among other

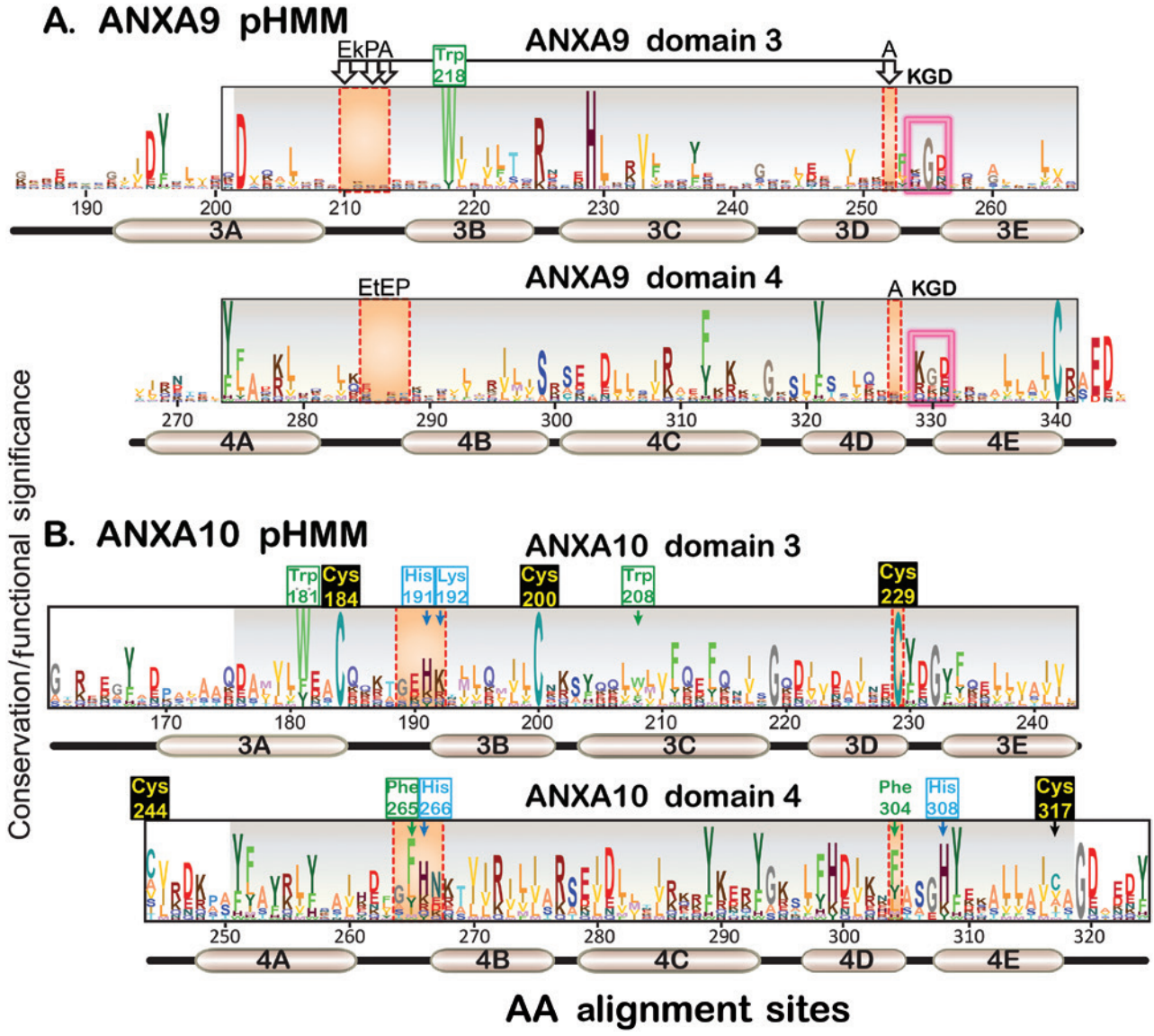


Figure 3: Profile hidden Markov models of ANX domains 3 and 4 in annexins A9 and A10.
Multiple sequence alignments of 100 ANXA9 and 200 ANXA10 proteins were used to build pHMM probability matrices using HMMBUILD v3.1b2 for all amino acids at all positions, then converted to sequence logos using the SKYLIGN server utility. Individual letter height reflects the probability of occurrence (i.e. conservation by selection pressure) for each amino acid whilst the total column height can be used to infer information content or functional importance of each site. (A) The ANXA9 logo established the absence of canonical type 2 calcium-binding sites (i.e. GxGT…38 aa…D/E) from all domains (orange rectangles) and the emergence of 'KGD' disintegrin motifs instead within the D–E alpha helical connecting loops. (B) The ANXA10 profile was likewise unique, showing prominent conservation of Cys residues with notable replacement of the acidic calcium-binding ligand at position 229 in domain 3, and a predominance of hydrophobic, aromatic bulky residues with replacement of the acidic calcium-binding ligands for Phe at aa 265 and 304 in domain 4.

cysteine-containing proteins such as stomach mucin 5AC (Martin-Almedina, 2010) although no bona fide receptors have yet been proven. For ANXA9, the absence of calcium-dependent membrane binding (Goebeler et al., 2003; Garcia-Diaz, 2011) due to replacement of the responsible residues makes the introduction of conserved KGD motifs later in the D–E loop a significant finding. This is a well-known (dis)integrin binding motif, also conserved and interactive for related clade members ANXA1 and A2 (Cheng et al., 2012; Rankin et al., 2013), at distinct locations for annexins A3, A4 and A5 (Saint-Guirons et al., 2007; Setti et al., 2013) and with variants able to interact with other domains including C2 (Simões et al. 2005) or integrins (Wang and Kirsch, 2006). The case of annexin A10 is rather different in that the sites specifically involved in calcium-binding have been substituted by residues with

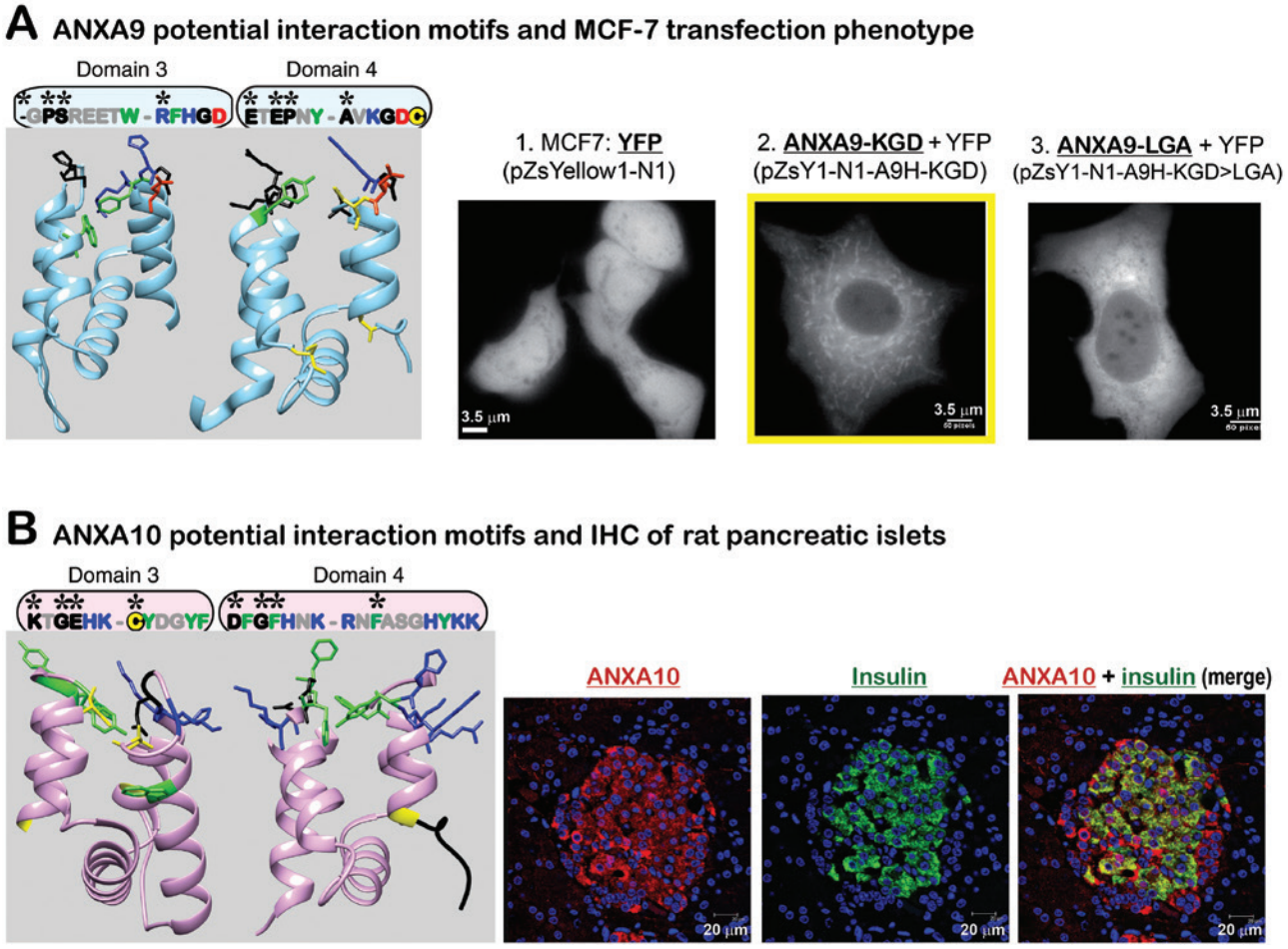
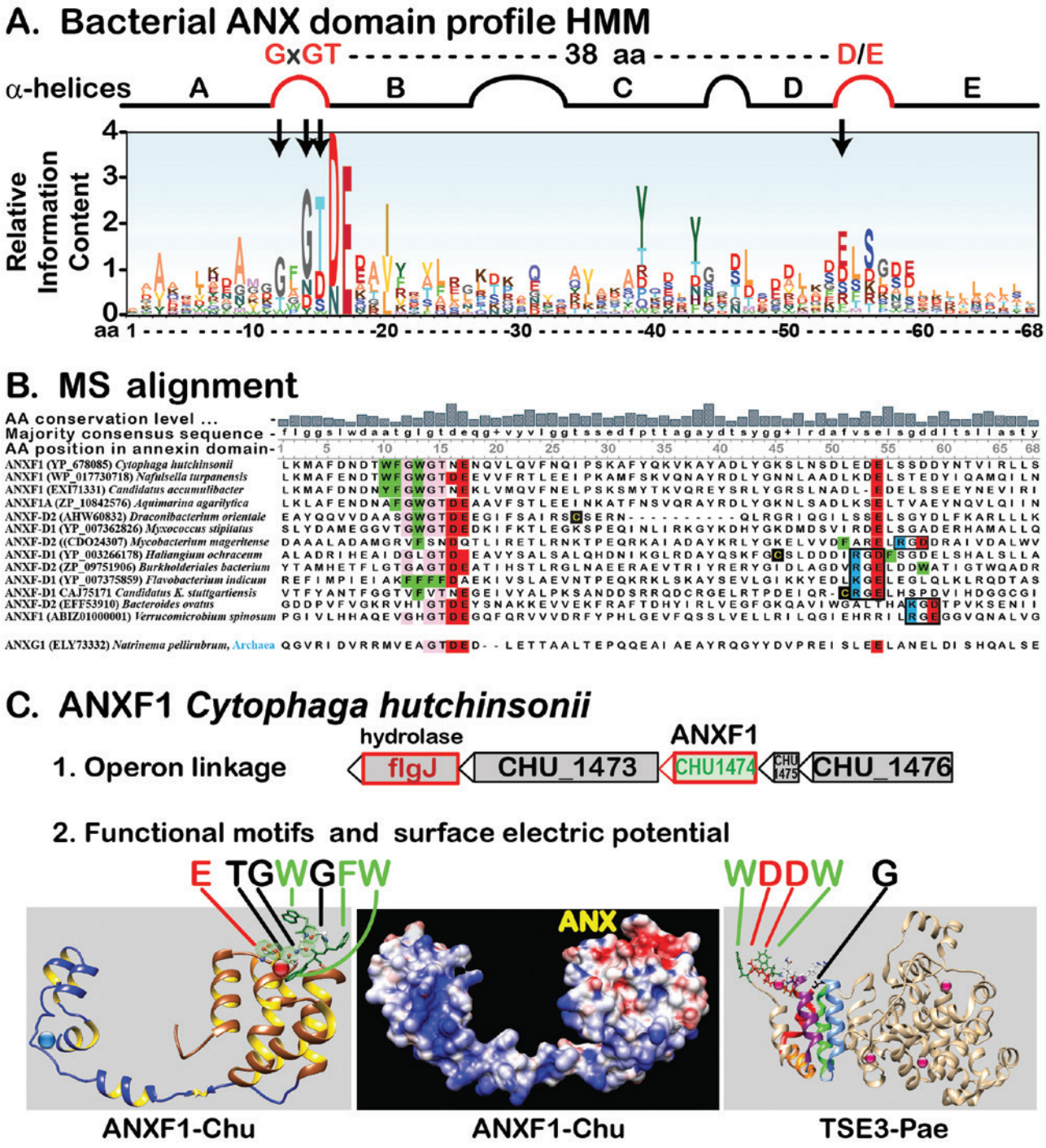


Figure 4: External ligands of atypical annexins A9 and A10.
(A) 3D structural models of annexin domains 3 and 4 for human ANXA9 highlight the conserved residues of probable functional importance for external binding interactions. The color-matched residues in single letter code and stick graphic representations signify basic (blue), hydrophobic (green), acidic (red) and redox cysteine (yellow) properties and (non-functional) calcium-binding residues (black). Despite the absence of significant membrane binding and the lack of known true receptors, the strategically located 'KGD' motif of ANXA9 could favor calcium-independent binding of phosphates, fatty acids and glycoproteins in the cytoskeleton, plasma membrane, cytosol or extracellular matrix. Fluorescence microscopy of cultured mammary cancer MCF7 cells transfected with (1) Control plasmid pEYFP-N1 (Clontech) encoding an enhanced yellow-green variant of the *Aequorea victoria* green fluorescent protein; (2) Plasmid construct pA9H-YFPN1 containing human ANXA9 cDNA sequence cloned into the pEYFP-N1 control plasmid with its KGD motif intact (domain 4, residues 329-331 aa); and (3) Plasmid construct pA9H(KGD_LGA)-YFPN1 containing the site-directed mutations K329L and D331A converting KGD → LGA of human ANXA9 cDNA sequence also cloned into the pEYFP-N1 control plasmid. Fluorescence microscopy was at 1000 × magnification, 50 pixels per 3.5 µm. (B) Putative binding ligands in ANXA10 consist of positively charged basic residues, hydrophobic or cysteine residues in both domains. The immunofluorescence detection of annexin A10 (red), insulin (green) and merged images in rat pancreatic islet tissue demonstrated specific binding to the cell surface. In image 1, the primary antibody anti-ANXA10 reacted with secondary antibody rabbit anti-IgG conjugated to Alexa fluor 555® (red); image two corresponds to incubation with an anti-insulin primary antibody (Abcam) and guinea-pig anti-IgG conjugated to Alexa fluor 488® (green); cellular nuclei were stained with DAPI (blue). The colocalization of annexin A10 and insulin (yellow stain) confirmed the elevated expression of annexin A10 in pancreatic β-cells, although it also appeared to be associated with other (endocrine) cell types in islets of Langerhans and, to a lesser extent, with some exocrine pancreatic cells. Scale 20 µm.

alternate, distinctive properties, specifically aromatic and hydrophobic Phe/Trp and Cys with redox binding potential, surrounded by adjacent basic residues. The ANXA9 KGD motif and ANXA10 Phe/Cys residues shown exposed in the ANX domain D–E loop (Figure 4A,B) are hypothesized to represent alternative binding ligands for glycoprotein and membrane lipid targets distinct from calcium-mediated PS binding.

The proposal of novel binding mechanisms and binding partners for annexins A9 and A10 is expected to be accompanied by altered subcellular protein distribution so the role of its KGD motif was investigated experimentally.



This was verified by transfecting MCF7 cells with plasmid constructs containing ANXA9 coding sequence, with either the intact KGD or a mutated sequence (KGD-LGA) fused to yellow fluorescent protein. Figure 4A shows a visible change in the distribution of the fluorescence in the cell cytoplasm, giving the intact KGD plasmid a fibrillar and granular distribution (Figure 4A, panel 2) vs. a homogeneously dispersed fluorescence in the cytoplasm of cells transfected with the sequence site-mutated construct. (Figure 4A, panel 3). In contrast, the Figure 4B panel shows annexin A10 associated with secretory vesicles in rat pancreatic islets and other cells of pancreatic tissue visualized by fluorescence confocal microscopy. The association of endogenous annexin A10 with secretory cells of the stomach (Martin-Almedina, 2010) concurs with its strong expression in pancreatic islet endocrine

Figure 5: Characterization of bacterial annexins.
(A) The Skylign sequence logo of a profile hidden Markov model was built from an alignment of 68 aa in 350 individual ANX domains. The sequences constituting this model scored well above the background threshold E-value and excluded up to 25% potentially false positives. New bacterial and archaeal candidates could therefore be confidently assigned by significant matches to this molecular fingerprint using HMMSEARCH. The relative probabilities of individual amino acids are reflected in the letter heights and total column height is inferred to be proportional to site selection pressure and functional importance. The conserved residues containing carbonyl and carboxyl oxygens implicated in type 2 calcium ion coordination are marked by arrows below the interhelical A–B and D–E loops. (B) Partial alignment of the ANX domain in bacterial annexins demonstrating the presence of type 2 calcium-binding sites GxGT…38 aa…D/E canonical for vertebrate annexins. Amino acids are shown in single letter code and source species are identified on the left. Additional features in the selected examples reinforced the observation of frequent inclusion of hydrophobic, aromatic, bulky residues (Trp, Phe in green background) adjacent to the A–B interhelical loops, the strategic location of KGD/RGE motifs in the D–E interhelical loops and only occasional insertion of Cys residues. Putative archaeal annexins ANXG1 from *Natrinema pellirubrum* (Figure 5B) and *Haloterrigena thermotolerans* (Figure 2) showed weak matches to the pHMM model from bacteria (E = 9.7e-06). (C) 1. The annotated chromosomal linkage map for the operon containing the ANXF1 locus Chu-1474 of *Cytophaga hutchinsonii* associates it with a known cellulose hydrolase, FljG (DOE Joint Genome Institute and MicrobesOnline). 2. A predicted 3D structural model of the ANXF1-Chu annexin (I-Tasser) is shown in ribbon style to highlight the location of both calcium-binding and hydrophobic residues (stick format) as potential interactors with the cell membrane. The middle figure displays surface electric potential charge distribution with mottled red (acidic, negative) expected for the ANX domain and predominantly blue (basic, positive) in the accessory left domain of unknown function and with smoother surface topology. The rightmost image of TSE3 from *Pseudomonas aeruginosa* shows putative ANX and muramidase domains, the former lacking amino acid homology with other annexins in type 2 Ca-binding sites of loops A–B and D–E.

secretory cells (Figure 4B, right). The pattern was similar to that of insulin (middle) but the merged images established that annexin A10 was not restricted to β-cells, but to other islet cells and some exocrine cells were also slightly stained (right image). It may have a related role in intestinal M-cells which endocytose external antigens (Nakato et al., 2009).

Bacterial annexins

Annexin genes originated in eubacteria at least 2 billion years ago, based on the presence of a 137 aa protein containing a single ANX domain in *Tolypothrix bouteillei* (Cyanobacteria) and a 2882 aa protein with 4 ANX domains in a close relative *Obscuribacter phosphatis* (Melainabacteria). More than 200 bacterial annexins containing over 400 ANX domains have been identified from database searches with pHMM models built from ANX domain alignments of unicellular eukaryotes and the full complement of new bacterial annexins (Figures 2 and 5). The latter pertain to at least nine bacterial phyla (Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Firmicutes, Gemmatimonadetes, Planctomycetes, Proteobacteria, Verrucomicrobia). The majority (> 60%) of bacterial annexins are small proteins (< 200 aa) containing a single ANX domain in diverse bacteria, similar to possible archaeal annexins (Figures 2 and 5B) but others possess up to five or more ANX domains and total lengths up to 3000 aa. The pHMM built from an alignment of representative domains confirmed the presence of a canonical type 2 calcium-binding signature GxGT…38 aa…D/E (Figure 5A) although this motif, like the KGD and aromatic hydrophobic residues Trp/Phe, were actually represented in a minority of the aligned sequences (Figure 5B). These characteristics are sufficient to firmly establish the presence of annexins in bacteria, considering also the motif location and the inclusion of multiple and alternate functional domains. The putative archaeal annexins possessed canonical calcium-binding ligands GxGT…D/E, an amino acid distribution significantly matching the pHMM profile (E = 4.9e-06), an α-helical secondary structure and related to several apparent orthologs, collectively indicative but inconclusive for definitive classification. Public annotations also specify an HTH_10 (helix-turn-helix) domain which overlaps with the predicted ANX domain, so it would be prudent to consider all the evidence until functional proof becomes available.

The fact that some ANX domains matched the pHMM model with high threshold but did not contain predicted calcium-binding sites raised the same questions posed by vertebrate annexins A9 and A10 regarding the binding mechanisms and targets involved in their actions. The proposed functions of some bacterial annexins are proffered in the GenBank/UniProt annotations, taking into account all known structural and functional features such as evidence for involvement in enzymatic activity (Figure 2C). The possibility that the ANX domain serves primarily a carrier role to target the cell membrane with enzymes destined to degrade, repair or traverse this protective barrier emphasizes the need to investigate all potential accessory domains. The first discovered bacterial annexin from *Cytophaga hutchinsonii* (Morgan et al., 2006) resides in the middle of an operon characterized by

known cellulose hydrolase activity in FLGJ (Figure 5C) and key ANX domain residues probably confer the capacity for calcium and hydrophobic dependent membrane binding. The best model predicted by I-Tasser (C-score −2, TM ~5) showed a C-terminal ANX domain with surface distribution of negative charge potential (red) typical of vertebrate annexins and apparently separated by a flexible hinge region from the N-terminal positively charged (blue) domain which lacked obvious surface binding pockets or protrusions (middle image). The present lack of evidence for enzymatic activity in the latter unkown domain of CHU_1474 annexin leads us to instead postulate a cooperative role in operon control of the FLGJ-based hydrolase activity via scaffolding or complex formation. There are other examples where ANX and hydrolase catalytic domains (Glyco_hydro_20) coexist in the same molecule (Figure 2C, Bacteroides). A contrasting example of TSE3 from *Pseudomonas aeruginosa* (Figure 5C right image) shows an experimentally determined annexin-like structure with N-terminal muramidase activity able to degrade membrane peptidoglycan (Lu et al., 2014). Despite the presence of a 5 α-helix bundle and confirmed calcium-binding activity, we note that the acidic Asp-Asp calcium binding residues in loop A–B and the uncharged Gly residue 38 aa downstream in loop D–E do not conform to the primary structure motif 'GxGT…E' typical of bacterial and other annexins as represented in the pHMM model (Figure 5A). This demonstrates the need to precisely evaluate various structural features to properly define annexins and to infer their apparently diverse functions.

Discussion

The annexin gene superfamily comprises several thousand known representatives in nearly all eukaryotic phyla, ranging from 12 gene subfamilies common to vertebrates and with homology traceable to hundreds of unicellar eukaryotes and prokaryotes, including at least 13 bacterial classes and possibly extremophilic archaea. The elucidation of 3D structures for calcium-binding mammalian annexins (Liemann and Huber, 1997) identified a unique type 2 calcium coordination site involving discontiguous loops of a five α-helical structure repeated within each of the four core ANX domains. The formation of a calcium bridge with membrane anionic phospholipids such as phosphatidylserine (PS) or phosphatidylinositol 4,5-bisphosphate has long been considered a common property of annexins, but other non-homologous calcium-binding domains such as C2 and EF-hand also exist in bacteria and appear to serve principally as transporters or scaffolds for other proteins with independent activity within membrane-associated complexes (Zhang and Aravind, 2010). A modified paradigm would resolve the dilemma of whether 'functional activity' must be embodied solely in the ANX domain for tetrad annexins with short amino termini such as A3, A4 or A5 or plants) or in the unique amino terminus of other annexins like ANXA1, especially when calcium-independent annexins exist in all phyla. The novel additional features found in the more elaborate architectures of nonvertebrate annexins and membrane-directed enzymatic activity in bacteria and pathogenic fungi argue for multifunctional modules in primitive annexins followed by a separation of roles in favor of vesicular transport and scaffolding functions for vertebrate annexins. The unification of all functional features implies some central role of membrane association via diverse mechanisms (including the cytoskeleton, extracellular matrix and nucleus) whilst other associated interactions may ultimately be necessary for functional execution in membrane remodeling or repair originating in other internal domains or external multimolecular complexes (Draeger et al., 2011).

The varied composition of annexins with extensive architectures of ANX and other domains (Figure 2), calcium-binding kinetics ranging from null to high affinity (Potez et al., 2011), additional ligand motifs and distinct physicochemical properties collectively contribute to a broad potential spectrum of molecular interactions, mechanisms and functional specificity for different annexins. The presence of strategically placed KGD ligands in all phyla from bacteria to vertebrates implies its key functional role, although its potential binding partners (e.g. C2 domains, integrins or SUMO) (Simões et al., 2005; Cheng et al., 2012; Caron et al., 2013; Rankin et al., 2013) and subcellular localization remain to be determined. Consistent with these observations, we propose a less restrictive definition of annexin as a unique structural domain comprising 68 amino acids in a five α-helical arrangement with two external loop regions, but containing 'distinct functional ligands' able to interact with membrane lipids, cytoskeletal proteins and/or extracellular matrix components. These basic properties of a single ANX domain can further altered or enhanced with to 20 homologous ANX domains, other distinct domains, unique amino-termini and even enzymatic activity in a single protein. To the extent that nonvertebrate and prokaryotic annexins may have related yet remote structures and corresponding functions, the validation of true binding partners for atypical, calcium-independent, vertebrate annexins A9 and A10 remains of special interest to better comprehend their unique molecular interactions, physiological roles and pathological

consequences in humans (Morgan and Fernandez, 1998; Morgan et al., 1999).

Genomic sequence data are being generated at an accelerating rate, offering an untapped information resource for investigating gene family molecular evolution and hence the relationships between structure and biological roles. The prolific divergence of annexin subfamilies within individual phyla provides clear evidence of their adaptability, which includes both redundancy and diversity in structure and function. If there is a functional relationship between primitive annexins from bacteria and later evolved vertebrate annexins it is partially distorted by this adaptive evolution within different species groups, but could consolidate a common basic cellular role in membrane homeostasis aimed at supporting cell growth, division, adhesion, signaling and wound healing. Finally, it should be recognized that the regulation of genome expression may represent an even more important determinant of function than the protein products themselves. Thus noncoding RNA and epigenetic influences on annexin expression (Wang et al., 2013; Bae et al., 2015) also warrant greater attention and may be particularly relevant in special cases such as the copy number variants of annexins A8 in human or A1 in species with special capabilities such as limb regeneration (Saxena et al., 2016).

Conclusion

The molecular evolutionary study of the annexin family broadens our perspective on the functional constraints in gene and protein design and delivers insight into the molecular origins of functional diversification. It does not presume a universal mechanism nor function but rather aims to integrate all structural and biological information into a flexible, coherent working model applicable to individual annexin subfamily functions. The integration of physicochemical data, including post-translational modifications, into 3D models participating in validated pathway schemes enables molecular dynamics and docking simulation studies that can predict or corroborate external evidence for bonafide molecular interactions. The combined objective of evolutionary modeling is to provide a molecular rationale and mechanistic predictor that can relate patterns of structural change to modifications in annexin gene expression or protein interaction, ergo phenotype. The comparative bioinformatic analysis of emerging genome sequences and the associated epigenetic control of RNA transcriptomics are likely to contribute additional genetic insight into viable annexin variants of pharmacogenomic interest, and continuing studies are needed to more precisely define the molecular interaction mechanisms involved.

Materials and methods

Bioinformatics

Annexin molecular phylogeny was analyzed by the Maximum Likelihood algorithm using the programs RAxML (Stamatakis, 2014) and Mega (Tamura et al., 2013) on an alignment of selected vertebrate annexins (311 aa × 82 taxa). Parameters included a minimum of 100 bootstraps, gamma rate correction (Gamma + P-Invar with alpha shape estimate 1.8) and WAG substitution matrix. Confidence levels > 70 at the branch bifurcation nodes indicated statistical support for the displayed topology. Domain architectures initially detected by SMART (Letunic et al., 2015) and PFAM (Finn et al., 2016) were validated and mapped by comparison of full-length annexin proteins with pHMM models from the databases or rebuilt from enhanced alignments and setting the E-value threshold < 0.1. These models were progressively refined as sequence alignments were amplified, to ensure greatest accuracy and precision in the detection of new or distant homologs. Sequence logos were created from pHMM models using the Skylign server (Wheeler et al., 2014). Other putative structural motifs were subjected to comparison with the Eukaryotic Linear Motif (ELM) server (Dinkel et al., 2016). Programs from the HMMER package (Finn et al., 2015) were used to build and align profile hidden Markov models, which were used to search multiple protein databases for entries with specific domains. Three dimension protein models were computed by the I-Tasser server (Roy et al., 2010) for the predicted conformations of annexins A9, A10 and bacterial ANXF1 while electric potential calculations were obtained from PDB2PQR (Dolinsky et al., 2007). The visualization program for 3D structure and analysis of optimal models with the highest C-score (– 5 to 2) was UCSF CHIMERA (Pettersen et al., 2004).

Plasmid constructs, cell transfection and fluorescence microscopy visualization

The full coding sequence for the human ANXA9 was amplified by PCR using primer specific sequences and additional sequence introduced for cloning recognition into the *Hin*d III and *Bam*H I sites of plasmid pEYFP-N1 (Clontech Labs, Mountain View, CA). The site specific change of the KGD motif to LGA was also performed by PCR and specific primers containing the sequence change. Both constructs pA9H-YFPN1 and the mutated version pA9H(KGD_LGA)-YFPN1 were confirmed by sequencing. Both plasmids were transfected into cultured mammary cancer MCF7 cells using Fugene (Roche Diagnostics Canada, Laval, QC) and the expression of ANXA9 was visualized 24 h after transfection,by the emitted fluorescence using an inverted microscope Zeiss Axiovert 100 (Carl Zeiss, Göttingen, Germany) equipped with an objective Zeiss 40x/0.75Ph2 Plan-Neo-Fluar and coupled to a monocromator Polychrome IV and 12 bits CCD camera image (TILL-Photonics, Martinsreid, Germany). Monocromator and camera control, data collection and analysis were performed with the program TillVision (Till-Photonics).

Tissue immunofluorescence microscopy

Tissue slices of 5 μm were placed onto transparent slides SuperFrost® Plus (VWR International, VWR Scientific Inc., West Chester, PA). Following dewaxing, rehydration and unmasking of antigens the sections were incubated for 30 min with Image-iT® FX (Invitrogen Life Technology, Carlsbad, CA). After a 10 min wash with phosphate-buffered saline samples were blocked for 30 min with horse serum diluted to 2.5% in PBS. This was followed by overnight incubation at 4°C with primary antibodies (anti-ANXA10 produced in lab) and anti-insulin (AbCam, Cambridge, UK), after three 10 washes with PBS, a 1 h incubation with fluorescent secondary antibodies (Molecular Probes, Eugene, OR) under darkness. Following three new washes of 10 min sections were nuclear stained with DAPI (Sigma-Aldrich, St. Louis, MO, USA) for 15 min. Upon completion of final washes, coverslips were applied with mounting media Vectashield over the slides containing tissue slices and the preparations were stored in the dark at 4°C until visual observation of the samples. This was performed by confocal microscopy with Ultra-Espectral Leica TCS-SP2-AOBS. Image processing utilized the software LCS (Leica Confocal Software, Leica Microsystems, Wetzlar, Germany) and Image J.

Acknowledgments: Spanish grant support was received from the *Ministerio de Educación y Ciencia* (Ref. BFU2007-67876) and *Gobierno Principado de Asturias* (SV-PA-13-ECO-EMP-69 and GRUPIN14-097).

References

Bae, J.M., Kim, J.H., Rhee, Y.Y., Cho, N.Y., Kim, T.Y., and Kang, G.H. (2015). Annexin A10 expression in colorectal cancers with emphasis on the serrated neoplasia pathway. World J. Gastroenterol. *21*, 9749–57.

Bena, S., Brancaleone, V., Wang, J.M., Perretti, M., and Flower, R.J. (2012). Annexin A1 interaction with the FPR2/ALX receptor: identification of distinct domains and downstream associated signaling. J. Biol. Chem. *287*, 24690–24697.

Boczonadi, V. and Maatta, A. (2012). Annexin A9 is a periplakin interacting partner in membrane-targeted cytoskeletal linker protein complexes. FEBS Lett. *586*, 3090–3096.

Bouter, A., Carmeille, R., Gounou, C., Bouvet F., Degrelle, S.A., Evain-Brion, D., and Brisson, A.R. (2015). Review: Annexin-A5 and cell membrane repair. Placenta 36 (Sup. 1) 29, S43–S49.

Caron, D., Maaroufi, H., Michaud, S., Tanguay, R.M., and Faure, R.L. (2013). Annexin A1 is regulated by domains cross-talk through post-translational phosphorylation and SUMOYlation. Cell. Signal. *25*, 1962–1969.

Cheng, T.Y., Wu, M.S., Lin, J.T., Lin, M.T., Shun, C.T., Huang, H.Y., Hua, K.T., and Kuo, M.L. (2012). Annexin A1 is associated with gastric cancer survival and promotes gastric cancer cell invasiveness through the formyl peptide receptor/extracellular signal-regulated kinase/integrin β-1-binding protein 1 pathway. Cancer *118*, 5757–5767.

Clark, G.B., Morgan, R.O., Fernandez, M.P., and Roux, S.J. (2012). Evolutionary adaptation of plant annexins has diversified their molecular structures, interactions and functional roles. New Phytol. *196*, 695–712.

Creutz, C.E., Pazoles, C.J., and Pollard, H.B. (1978). Identification and purification of an adrenal medullary protein (synexin) that causes calcium-dependent aggregation of isolated chromaffin granules. J. Biol. Chem. *253*, 2858–2866.

D'Acquisto, F., Perretti, M., and Flower, R.J. (2008). Annexin-A1: a pivotal regulator of the innate and adaptive immune systems. Br. J. Pharmacol. *155*, 152–169.

Demonbreun, A.R. and McNally, E.M. (2016). Plasma membrane repair in health and disease. Curr. Topics Membr. *77*, 67–96.

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., et al. (2016). ELM 2016 – data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res. *44*, D294–D300.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., and Baker, N.A. (2007). PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. *35*, 522–525.

Draeger, A., Monastyrskaya, K., and Babiychuk, E.B. (2011). Plasma membrane repair and cellular damage control: the annexin survival kit. Biochem. Pharmacol. *81*, 703–712.

Einarsson, E., Astvaldsson, A., Hultenby, K., Andersson, J.O., Svard, S.G., and Jerlstrom-Hultqvist, J. (2016). Comparative cell biology and evolution of annexins in Diplomonads. mSphere *1*, e00032–15.

Fernandez, M.P. and Morgan, R.O. (2003). Structure, function and evolution of the annexin gene superfamily. In: Annexins, J. Bandorowicz-Pikula, ed. (Georgetown, TX, USA: Eurekah.com and Kluwer Academic/Plenum Publishers), pp. 21–37.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server: 2015 update. Nucleic Acids Res. *43*, W30–W38.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Garcia-Diaz, M. (2011). Annexins – Molecular evolution, prediction of novel functional motifs and experimental studies of atypical models. Ph.D. thesis, University of Oviedo, Oviedo, Spain.

Gerke, V., Creutz, C.E., and Moss, S.E. (2005). Annexins: linking Ca^{2+} signalling to membrane dynamics. Nat. Rev. Mol. Cell Biol. *6*, 449–461.

Goebeler, V., Ruhe, D., Gerke, V., and Rescher, U. (2003). Atypical properties displayed by annexin A9, a novel member of the annexin family of Ca^{2+} and lipid binding proteins. FEBS Lett. *546*, 359–364.

Khalaj, K., Aminollahi, E., Bordbar, A., and Khalaj, V. (2015). Fungal annexins: a mini review. Springerplus *4*, 721.

Kodavali, P.K., Dudkiewicz, M., Pikula S., and Pawlowski K. (2014). Bioinformatics analysis of bacterial annexins – putative ancestral relatives of eukaryotic annexins. PLoS One *9*, e85428.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. *43*, D257–D260.

Liemann, S. and Huber, R. (1997). Three-dimensional structure of annexins. Cell. Mol. Life Sci. *53*, 516–521.

Lu, D., Shang, G., Zhang, H., Yu, Q., Cong, X., Yuan, J., He, F., Zhu, C., Zhao, Y., Yin, K., et al. (2014). Structural insights into the T6SS effector protein Tse3 and the Tse3-Tsi3 complex from *Pseudomonas aeruginosa* reveal a calcium-dependent membrane-binding mechanism. Mol. Microbiol. *92*, 1092–1112.

Martin-Almedina, S. (2010). Annexin A10, a new structural model with gene expression and function associated with processes of cellular secretion. Ph.D. thesis, University of Oviedo, Oviedo, Spain.

Morgan, R.O. and Fernandez, M.P. (1995). Molecular phylogeny of annexins and identification of a primitive homologue in *Giardia lamblia*. Mol. Biol. Evol. *12*, 967–979.

Morgan, R.O. and Fernandez, M.P. (1997). Distinct annexin subfamilies in plants and protists diverged prior to animal annexins and from a common ancestor. J. Mol. Evol. *44*, 178–188.

Morgan, R.O. and Fernandez, M.P. (1998). Expression profile and structural divergence of novel human annexin 31. FEBS Lett. *434*, 300–304.

Morgan, R.O., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Balsara, B.R., Testa, J.R., and Fernandez, M.P. (1999). Novel human and mouse annexin A10 are linked to the genome duplications during early chordate evolution. Genomics *60*, 40–49.

Morgan, R.O., Martin Almedina, S., Pruneda, L., and Fernandez, M.P. (2004). Evolutionary models of annexin A1 divergence. Annexins *1*, 143–151.

Morgan, R.O., Martin-Almedina, S., Garcia, M., Jhoncon-Kooyip, J., and Fernandez, M.P. (2006). Deciphering function and mechanism of calcium-binding proteins from their volutionary imprints. Biochim. Biophys. Acta *1763*, 1238–1249.

Moss, S.E. and Morgan, R.O. (2004). The annexins. Genome Biol. *5*, 219.

Nakato, G., Fukuda, S., Hase, K., Goitsuka, R., Cooper, M.D., and Ohno, H. (2009). New approach for M-cell-specific molecules screening by comprehensive transcriptome analysis. DNA Res. *16*, 227–235.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin T.E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Potez, S., Luginbühl, M., Monastyrskaya, K., Hostettler, A., Draeger, A., and Babiychuk, E.B. (2011). Tailored protection against plasmalemmal injury by annexins with different Ca^{2+} sensitivities. J. Biol. Chem. *286*, 17982–17991.

Quiskamp, N., Poeter, N., Raabe, C.A., Hohenester, U.M., König, S., Gerke, V., and Rescher, U. (2014). The tumor suppressor annexin A10 is a novel component of nuclear paraspeckles. Cell. Mol. Life Sci. *2*, 311–329.

Rankin, C.R., Hilgarth, R.S., Leoni, G., Kwon, M., Den Beste, K.A., Parkos, C.A., and Nusrat, A. (2013). Annexin A2 regulates β1 integrin internalization and intestinal epithelial cell migration. J. Biol. Chem. *288*, 15229–15239.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Saint-Guirons, J., Zeqiraj, E., Schumacher, U., Greenwell, P., and Dwek, M. (2007). Proteome analysis of metastatic colorectal cancer cells recognized by the lectin *Helix pomatia* agglutinin (HPA). Proteomics *7*, 4082–4089.

Saxena, S., Purushothaman, S., Meghah, V., Bhatti, B., Poruri, A., Meena Lakshmi, M.G., Sarath Babu, N., Murthy, C.L., Mandal, K.K., Kumar, A., et al. (2016). Role of annexin gene and its regulation during zebrafish caudal fin regeneration. Wound Repair Regen. *24*, 551–559.

Setti, A., Sankati, H.S., Devi, T.A., Sekhar, A.C., Rao, J.V., and Pawar, S.C. (2013). Structural insights into the extracellular segment of integrin β5 and molecular interaction studies. J. Recept. Signal Transduct. Res. *33*, 319–324.

Simões, I., Mueller, E.C., Otto, A., Bur, D., Cheung, A.Y., Faro, C., and Pires, E. (2005). Molecular analysis of the interaction between cardosin A and phospholipase D(alpha). Identification of RGD/KGE sequences as binding motifs for C2 domains. FEBS J. *272*, 5786–5798.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Tebar, F., Gelabert-Baldrich, M., Hoque, M., Cairns, R., Rentero, C., Pol, A., Grewal, T., and Enrich, C. (2014). Annexins and endosomal signaling. Methods Enzymol. *535*, 55–74.

Wang, X., Zhang, S., Zhang, J., Lam, E., Liu, X., Sun, J., Feng, L., Lu, H., Yu, J., and Jin, H. (2013). Annexin A6 is down-regulated through promoter methylation in gastric cancer. Am. J. Transl. Res. *5*, 555–562.

Wang, W. and Kirsch, T. (2006). Annexin V/beta5 integrin interactions regulate apoptosis of growth plate chondrocytes. J. Biol. Chem. *281*, 30848–30856.

Weiland, M.E., McArthur, A.G., Morrison, H.G., Sogin, M.L., and Svärd, S.G. (2005). Annexin-like alpha giardins: a new cytoskeletal gene family in *Giardia lamblia*. Int. J. Parasitol. *35*, 617–626.

Wheeler, TJ., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics. *15*, 7.

Yuan, S., Qu, L., and Shou, C. (2016). N-Terminal polypeptide of annexin A2 decreases infection of *Mycoplasma hyorhinis* to gastric cancer cells. PLoS One *11*, e0147776.

Zhang, D. and Aravind, L. (2010). Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes. Gene *469*, 18–30.