Comparative Genetics and Evolution of Annexin A13 as the Founder Gene of Vertebrate Annexins

Juan-Manuel Iglesias,* Reginald O. Morgan,* Nancy A. Jenkins,† Neal G. Copeland,‡ Debra J. Gilbert,† and Maria-Pilar Fernandez*  

*Department of Biochemistry and Molecular Biology, University of Oviedo, Oviedo, Spain; and †Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland

Annexin A13 (ANXA13) is believed to be the original founder gene of the 12-member vertebrate annexin A family, and it has acquired an intestine-specific expression associated with a highly differentiated intracellular transport function. Molecular characterization of this subfamily in a range of vertebrate species was undertaken to assess coding region conservation, gene organization, chromosomal linkage, and phylogenetic relationships relevant to its progenitor role in the structure-function evolution of the annexin gene superfamily. Protein diagnostic features peculiar to this subfamily include an alternate isoform containing a KGD motif, an elevated basic amino acid content with polyhistidine expansion in the 5'-translated region, and the conservation of 15% core tetrat residues specific to annexin A13 members. The 12 coding exons comprising the 58-kb human ANXA13 gene were deduced from BAC clone sequencing, whereas internal repetitive elements and neighboring genes in chromosome 8q24.12 were identified by contig analysis of the draft sequence from the human genome project. A unique exon splicing pattern in the annexin A13 gene was corroborated by coanalysis of mouse, rat, zebrafish, and pufferfish genomic DNA and determined to be the most distinct of all vertebrate annexins. The putative promoter region was identified by phylogenetic footprinting of potential binding sites for intestine-specific transcription factors. Mouse annexin A13 cDNA was used to map the gene to an orthologous linkage group in mouse chromosome 15 (between Sdc2 and Myc by backcross analysis), and the zebrafish cDNA permitted its localization to linkage group 24. Comparative analysis of annexin A13 from nine species traced this gene's speciation history and assessed coding region variation, whereas phylogenetic analysis showed it to be the deepest-branching vertebrate annexin, and computational analysis estimated the gene age and divergence rate. The unique, conserved aspects of annexin A13 primary structure, gene organization, and genetic maps identify it as the probable common ancestor of all vertebrate annexins, beginning with the sequential duplication to annexins A7 and A11 approximately 700 MYA, before the emergence of chordates.

Introduction

Phylogenetic analysis of over 100 different species has revealed more than 500 distinct annexins representing at least 50 paralogous genes, which have been classified into five family groups corresponding to the eu-karyotic kingdoms (Morgan and Fernandez 1995; Morgan et al. 1999b; URL http://www24.brinkster.com/hofmanna/). Such broad phylogenetic distribution and numerous paralogs suggest that these proteins have important cellular roles, but the diverse expression patterns and binding properties of individual annexins have thwarted efforts to define their specific involvement in signal transduction processes related to membrane trafficking and cell differentiation. The early mortality of Anxa7 homozygous null knockout mice (Srivastava et al. 1999) further suggests some ontogenetic role in embryonic organogenesis or maintenance. Various molecular interactions with calcium, phospholipids, nucleotides, carbohydrates, cell matrix proteins, and enzymes (e.g., phospholipases, kinases, phosphatases) are individually relevant to the actions of divergent annexins. In the context of such divergence, we surmise that the original founder gene, with its highly regulated expression, tissue-specific function, and primitive species orthologs, is the most important model for investigating the origin, evolution, and basic function of this gene family.

Annexin A13 (ANXA13) harbors key information about the evolutionary origins and structure-function relationships of the A family of 12 chordate annexins. It is the earliest branching subfamily in vertebrates and is physically nonsyntetic from other human annexins (Braun et al. 1998; Morgan et al. 1998). The short isoform of the human protein has a very restricted, intestine-specific expression (Wice and Gordon 1992), and a longer isoform identified in dog exhibits differential subcellular targeting in polarized kidney epithelial cells (Fiedler et al. 1995). Expression is observed mainly in small intestine in committed proliferating crypt epithelia and in differentiated villus enterocytes, where it is selectively transported to the apical region (Noda et al. 2001). It has been proposed that annexin A13 may play a crucial role in the lipid raft-mediated delivery of apical proteins (Lafont et al. 1998) and that the different isoforms show regional and cofactor specificity in their actions (Lecat et al. 2000; Plant et al. 2000). Despite these comprehensive localization studies, further study has been impeded by the low detectability, uncertainty about

Abbreviations: ANXA13, annexin A13 gene of human; Anxa13, annexin A13 gene of rodent; anxa13, annexin A13 gene of nonmammal; BAC, bacterial artificial chromosome; CDX, caudal homeobox; dbGSS, database of Genome Survey Sequences; EST, expressed sequence tag; FREAC, forkhead related activator; HNF, hepatic nuclear factor; long interspersed nuclear elements, LINEs; NNS, nonsynonymous nucleotide substitutions; RACE, 5′ or 3′ rapid amplification of cDNA ends; RT-PCR, reverse transcription–polymerase chain reaction.

Key words: annexin gene family, gene duplication, gene organization, genetic mapping, molecular evolution, phylogenetic analysis.

Address for correspondence and reprints: Maria-Pilar Fernandez, Department of Biochemistry and Molecular Biology, Edificio Santiago Gascon, University of Oviedo, E-33006 Oviedo, Spain. E-mail: pfernandez@bioquimica.uniovi.es.

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
the relationship between subcellular location and activity, and imprecise understanding of the protein functional determinants, based on limited cDNA sequence data for human and dog.

A cross-genome survey can provide insight and perspective into protein structure-function relationships as well as gene organization, chromosomal environment, species distribution, and evolutionary origin. Phylogenetic analysis of primary sequence conservation is of fundamental value for identifying divergence patterns and functionally important protein residues because these are under selective evolutionary constraint. Computational sequence analysis calibrated to species fossil records can provide confirmatory estimates of evolutionary order, rates, and ages of gene family members. The annexin A13 gene structure is of special interest because it could clarify the initial gene duplication order of this family by its (non)congruency to the other primary clade members, annexin A7 (Shirvan et al. 1994) and annexin A11 (Bances et al. 2000). Interspecies homology is being recognized as a useful tool for identifying conserved promoter regions by phylogenetic footprinting of the regulatory elements responsible for tissue-specific expression (Wasserman et al. 2000). The map location of ANXA13 in human chromosome 8q24 was originally determined using cytogenetic techniques (Morgan et al. 1998), but comparative species maps are needed to assess chromosomal paralogy. This could help document the evolutionary history and physical integrity of this chromosomal region, confirm primitive species orthologs, and evaluate this locus for genetic traits being investigated by positional cloning. The ancient origins of vertebrate annexins demand a comprehensive analysis of all available molecular markers revealed by sequence phylogeny, gene structures, genetic maps, and divergence parameters to assess the temporal order of initial gene duplications.

Methods
Characterization of Full-length Annexin A13 cDNAs

The complete sequence for long isoform b of Homo sapiens annexin A13 cDNA (Hsa, human, gb: AJ306450) was determined by resequencing colon adenocarcinoma clones IMAGE:3349423 and IMAGE:4546150 (NIH-MGC) identified by NCBI-BLAST searches and obtained from Research Genetics (Huntsville, Ala.). Mus musculus cDNA encoding annexin A13 (Mmu, house mouse, gb:AJ306451) was obtained from RNA isolated from mouse intestine by the isothiocyanate acidic phenol method (Chomczynski and Sacchi 1987). Reverse transcription was performed with 10 μg of total RNA as template and an oligo (dT) primer. Superscript II reverse transcriptase (GIBCO-BRL) was used for the first strand synthesis reaction according to manufacturer’s recommendations and this was utilized to amplify the mouse annexin A13 coding region by polymerase chain reaction (PCR) using degenerate oligonucleotide primers deduced from the coding ends of known human and Canis familiaris cDNA sequences (Cfa, dog, gb:X80209). The 5'- and 3'-extended sequences of human and mouse cDNAs were obtained by rapid amplification of cDNA ends (RACE) according to instructions in the manual for the commercial SMART RACE amplification kit (Clontech, Cambridge, UK). All PCR products were subcloned and sequenced by the dideoxy chain termination method with α-[35S]dATP (Amersham, Little Chalfont, UK) and the Sequenase 2.0 kit (U.S. Biochemical, Cleveland, Ohio). Primer extension analysis used total RNA from mouse intestine or human HT-29 cells grown in inosine as template, following procedures described previously (Bances et al. 2000). Danio rerio annexin A13.1 cDNA (Dre, zebrfish, gb:AJ306452) was determined by resequencing WashU MPIMG clones fb40a08 and 4201386, whereas A13.2 was derived by resequencing clone 5077337 (from Ressourcen Zentrum Primar Datenbank (RZPD, Berlin, Germany).

Identification and Sequencing of Genomic Clones

Genomic sequences for human annexin A13 were obtained from two bacterial artificial chromosome clones (CIT-HSP 2334P3 and RPCI111 138L10). These were identified by TBLASTN searches of the BAC-end database (URL http://www.tigr.org) against the annexin A13 protein sequence and supplied by Research Genetics. Genomic fragments containing portions of the annexin A13 gene were generated with appropriate restriction enzymes, identified by cDNA hybridization, subcloned, and sequenced. Mouse genomic clones for annexin A13 were obtained from a mouse 129/SVJ genomic library prepared from spleen DNA ligated to λFIX II vector (Stratagene, La Jolla, Calif.). Random-primed [32P]-labeled cDNA for mouse annexin A13 was used for hybridization. Screening procedures, DNA isolation, restriction analysis, gel electrophoresis, Southern blot transfer, and hybridization were performed as described (Sambrook, Fritsch, and Maniatis 1989). Genomic fragments were subcloned and sequenced by the dideoxy chain termination method.

Human Cell Cultures

The human colon adenocarcinoma cell line HT-29 was obtained from Dr. Dario Acuña (University of Granada, Spain), and cultures were grown in Dulbecco’s modified essential medium containing 10% di-alyzed fetal bovine serum plus either 25 mM glucose (for rapid proliferation) or 2.5 mM inosine (to induce differentiation). Multipotent HT-29 cells grown in the presence of inosine and absence of glucose were cultured for at least four passages before use. Cells were harvested for RNA preparation at log phase, 3 to 4 days after plating, to perform reverse transcription–PCR (RT-PCR) reactions for extending cDNA ends by using the SMART RACE amplification kit (Clontech).

Interspecific Mouse Backcross Mapping

Interspecific backcross progeny were generated by mating (C57BL/6J × Mus spretus) F1 females and C57BL/6j males as described (Copeland and Jenkins 1991). A total of 205 N2 mice was used to map the
Anxa13 locus. DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (Jenkins et al. 1982). All blots were prepared with Hybond N+ nylon membrane (Amersham). The pA13M probe was a 950-bp fragment containing the coding region of mouse annexin A13 cDNA, excised with BamHI/SalI and labeled with [α-32P]dCTP using a random primed labeling kit (Stratagene). Washing was done to a final stringency of 0.8× saline/sodium citrate/ phosphate (SSCP) and 0.1% sodium dodecyl sulfate at 65°C. Fragments of 13.5, 5.1, 4.4, 3.2, 2.6, 2.4, and 1.2 kb were detected in Sphl-digested C57BL/6J DNA, and fragments of 9.2, 4.7, 3.2, 2.6, 2.4, and 1.2 kb were detected in Sphl-digested M. spretus DNA. The presence or absence of the 9.2- and 4.7-kb Sphl M. spretus-specific fragments, which cosegregated, was followed in backcross mice. The probes and restriction fragment length polymorphisms (RFLPs) for the loci linked to Anxa13 in mouse, including syndecan2 (Sdc2), myelocyтомatosis oncogene (Myc), and thyroglobulin (Tgn) have been reported previously (Brannan et al. 1992; Spring et al. 1994). Recombination distances were calculated using Map Manager, version 2.6.5. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

Bioinformatics

Computer programs used in this study included the FASTA package (Pearson 1990) for general sequence comparison, L193 (Li 1993) for calculation of nucleotide substitutions, TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996) for maximum likelihood analysis, and PHYLIP 3.6a2.1 (Felsenstein 1989) for maximum parsimony. Public net-server programs included BLAST searches of the various sequence databases maintained at the National Center for Biotechnology Information (NCBI, Bethesda, Md; URL http://www.ncbi.nlm.nih.gov) and the European Bioinformatics Institute (EBI, Hinxton, UK; URL http://www.ebi.ac.uk). MatInspector (Quandt et al. 1995) was used for promoter site detection against the TRANSFAC 4.0 database (Heinemeyer et al. 1999), and RepeatMasker (URL http://ftp. genome.washington.edu/RM/RepeatMasker.html) identified genomic repetitive elements from REPBASE (Jurka 1998). Mouse genetic map data came from the Mouse Genome Database maintained by Jackson Laboratories (Massachusetts) at ftp://ftp.informatics.jax.org/pub. Zebrafish genetic map data (Postlethwait et al. 2000; Woods et al. 2000) were retrieved from the Stanford University database, URL http://cgmw.stanford.edu/~tallab/ Frontpage.html.

Public domain EST sequences for annexin A13 were identified and further characterized from the following species and clone source identification numbers: *Bos taurus* (Bta, cow, gb:BF604919), *D. rerio* (Dre, zebrafish A13.1 gb:BF938344 and A1437290, zebrafish A13.2 gb:B1845312), *Sus scrofa* (Ssc, pig, gb:BF198997), and *Xenopus laevis* (Xla, African clawed frog, gb:BE509061). Genomic sequence data for human chromosome 8 BAC clones RP11-562D1 and RP11-1A23 from the Human Genome Project (HGP 2001) were deciphered to identify all intact exons and splice sites for the human annexin A13 gene. Homologous *Rattus norvegicus* (Rno, Norway rat), mouse, and zebrafish exons were matched by BLAST searches of the TRACE draft sequence databases using full-length cDNAs. Genomic annexin A13 sequence from *Tetraodon nigroviridis* (Tni, pufferfish, gb:AL279602, Genoscope, France) was identified in the database of Genome Survey Sequences (dbGSS) and characterized for exon splice sites by comparison with zebrafish cDNA. *Strongylocentrotus purpurus* (Spu, purple sea urchin) clones 1012-16-M and 14-7-I from the sea urchin genome project (Cameron et al., California Institute of Technology, personal communication) were identified from dbGSS. Other genus-species names for taxa included in the phylogenetic analysis include *Ascaris suum* (Asu), *Arabidopsis thaliana* (Ath), *Crassostrea virginica* (Cvi), *Cae
orhabditis elegans* (Cel), *Dictyostelium discoideum* (Ddi), *Drosophila melanogaster* (Dme), *Giardia intestinalis* (Gin), *Hydra vulgaris* (Hvu), *Neurospora crassa* (Ncr), *Schistosoma japonicum* (Sja), and *Schistosoma mansoni* (Sma).

Results

Identification and Alternative Splicing of Annexin A13 Orthologs in Different Species

We first sought to characterize the alternate isoform of human annexin A13 cDNA (Wice and Gordon 1992) equivalent to that described for dog (Fiedler et al. 1995), as a prelude to resolving the complete gene structure. Its existence was initially reported to us by Burton M. Wice (personal communication, Washington University, St. Louis, Mo.) and validated by confirming that all of the 5’ expressed sequence tags (ESTs) for annexin A13 (seven out of a total of 21) in the current dbEST carried the cassette insert. Two clones derived from colon adenocarcinoma were procured for complete resequencing. Mouse annexin A13 cDNA was next characterized by RT-PCR from mouse small intestine RNA. Both human and mouse 5’ sequences were extended by rapid amplification of cDNA ends (5’ RACE) to ensure the full-length CDNA sequence that, together with primer extension experiments, demarcated the transcription start point and the upstream, nontranscribed region in genomic clones. Zebrafish annexin A13 cDNA was also sequenced from EST clones for use in estimating gene divergence and calibrating the gene molecular clock. The complete deduced protein sequences for human, dog, mouse, and zebrafish annexin A13 were aligned with partial sequences derived from ESTs of pig, cow, and frog and coding regions of rat and pufferfish genomic clones to assess site conservation (fig. 1, see Methods for clone names and sources).

The extreme amino termini of annexin A13 (fig. 1, top left) contain a conserved G2-myristoylation site known to be responsible for calcium-independent lipid association (Wice and Gordon 1992; Fiedler et al. 1995).
may signify a nuclear localization signal (PSORT II program, K. Nakai, http://www.expasy.ch), and "intrasppecies" coding triplet expansions have been associated with pathological consequences. The amino terminal location of a conserved Lys–Gly–Asp motif in the 41-aa cassette insert offers a potential KGD/RGD ligand for membrane (dis)integrin receptors (Perutelli 1995). The 311-aa homologous tetrad core region comprises variable linker segments that connect the four internally homologous 68-69 aa repeat domains (fig. 1, aligned to the right). Annexin A13 residues common to the vertebrate annexin consensus sequence include many sites with greater than 90% aa conservation that are likely to be responsible for the stability of the homologous tetrad structure and for intrinsic function. Site differences from other annexins in Cys residue conservation at core positions 22, 109, 193, and 310 could also affect the tetrad structure. Functionally important sites known to be involved in type-II and type-III calcium-binding sites or associated with putative calcium channel activity are fully preserved. Those sites at which other annexins have diverged significantly from annexin A13 (fig. 1, reverse-shaded letters) and the two single nucleotide polymorphisms R-81-H and V-267-I serve as diagnostic residues that may impart structural and functional specificity to this subfamily.

Unique Organization of the Annexin A13 Gene

The full-length cDNAs obtained for human and mouse annexin A13 could now be applied to elucidation of the structural organization of the annexin A13 gene, by selective sequencing and analysis of human and mouse genomic clones and comparison with genomic sequence data from public sequencing projects. All intact exons and splice sites were identified in the human gene (table 1). The 12 coding exons of annexin A13 span 58 kb of genomic sequence, the longest annexin gene (fig. 1). The alternate b-isoform for human and the deduced proteins for mouse and zebrafish derive from cDNA sequences obtained in this study, whereas fragments from cow, pig, rat, frog, and pufferfish were deciphered from public database sources (see Methods for genus-species abbreviations).

The presence of a 41-aa optional insert in the human protein with 83% aa identity to dog isoform A13b and a homologous, similarly located counterpart in mouse genomic clones underscores the probable functional relevance of this domain. The observations that all 5′ ESTs containing the cassette insert derived from colon adenocarcinoma and that the A13b isoform is overexpressed in inosine-treated HT-29 cells suggest that alternative splicing may vary with cell growth or differentiation state or that the two mRNA species have different stabilities. Two newly recognized features of the amino termini are of potential functional significance. There is a preponderance of amino terminal basic residues (40% or 11 of the first 28 aa in isoform b) and a marked triplet (CAY) expansion in the number of histidine residues at positions 7 and 8 of the human, mouse, and rat sequences (fig. 1). Basic residues more than 20%
core region (fig. 2) but with much smaller introns, typically less than 100 bp.

The annexin A13 gene exhibits both general similarities and significant structural differences from other annexins (fig. 2). Like the annexin A7 and A11 prototypes, it contains an alternatively spliced exon in its 5’ region and has alternative polyadenylation signals and sites in the 3’ untranslated terminus. The exon splicing pattern within the homologous tetrad is the same for annexin A11 and nine other human annexins (Bances et al. 2000), whereas annexin A7 shows partial divergence (Shirvan et al. 1994). It lacks two introns (i.e., those corresponding to annexin A11 introns 8 and 10) and displays downstream shifts of introns 11 and 13 identical to annexin A13. The earlier splice shift involves a significant change of intron phase (i.e., codon insertion position) from 0 in annexins A7 and A13 to phase 1 in annexin A11 and its congeners. Annexin A13 introns 3, 4, 7, and 10 are positioned as in all other annexins, its intron 6 placement is common to all but absent in annexin A7, and introns 9 and 11 correspond to annexin A7 introns 11 and 13. However, only annexin A13 has three unique splice sites for introns 2, 5, and 8, thus making annexin A7 compatible with a structural intermediate or common ancestor between annexins A13 and A11. Annexin A13 has the most ancestral coding sequence among human annexins, on the basis of its deeper branch position and association with invertebrate homologs in phylogenetic analysis (Morgan and Fernandez 1997; Braun et al. 1998). The uniquely divergent exon splicing pattern in its tetrad core now provides the physical basis for inferring it to be the original founder gene of the annexin A family.

### Table 1: Structural Organization of the Human Annexin A13 Gene

<table>
<thead>
<tr>
<th>No.</th>
<th>bp</th>
<th>5’ Sequence</th>
<th>3’ Sequence</th>
<th>Amino Acids</th>
<th>Splice Junctions</th>
<th>Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>GTAACATT</td>
<td>TGC CAT</td>
<td>His10/Ser16</td>
<td>0 1442 gtattg...</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>AGCCCTGAG</td>
<td>CCAAGCCT</td>
<td>Pro9/Ala19</td>
<td>0 22997 gtaggc...</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>GCTAAGGC</td>
<td>AAGAAGATG</td>
<td>Gly72</td>
<td>1 9941 gtat...</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>95</td>
<td>GAGCTATG</td>
<td>TACGCCAG</td>
<td>Lys10/Glu104</td>
<td>0 4083 gtattg...</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>171</td>
<td>GAGCTGAG</td>
<td>ACCAATAG</td>
<td>Lys106/Glu161</td>
<td>0 2666 gtattg...</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>GAAATGGC</td>
<td>TACAGGCA</td>
<td>Lys2</td>
<td>1 106 gtattg...</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>TATTGCTG</td>
<td>CTGTCTAG</td>
<td>Gin199/Ala199</td>
<td>0 1722 gtattg...</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>69</td>
<td>GATTCTGGC</td>
<td>CGTATGAG</td>
<td>Asp211/Ala222</td>
<td>0 412 gatttg...</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>GCGGGAGA</td>
<td>TATCAATT</td>
<td>Ile220/Leu256</td>
<td>0 4259 gatttg...</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>76</td>
<td>CTTTGTAG</td>
<td>TATCTGCTG</td>
<td>Val251</td>
<td>1 4152 gtattg...</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>113</td>
<td>TGAGAGG</td>
<td>AGGCCAG</td>
<td>Glu314/Val319</td>
<td>0 3284 gtattg...</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>506</td>
<td>GTTGACCTT</td>
<td>TTTGACTG</td>
<td>AATATAAACatattatatTA</td>
<td>(polyadenylation site 1)</td>
<td></td>
</tr>
<tr>
<td>566</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(polyadenylation site 2)</td>
<td></td>
</tr>
</tbody>
</table>

*The 12 coding exons of the 58 kb human annexin A13 gene (ANXA13) are listed by number, total bp length, and 5’/3’ terminal sequences (left, uppercase). Splice junctions (center) are described by the corresponding amino acid positions and codon splice phase, either 0 (between codons) or 1 (between codon bases 1 and 2). Introns (right, lowercase) are summarized by total bp length from contig analysis of human genome public draft sequence, and 5’/3’ sequence reads from PCR of purchased BACs. The start of exon 1 corresponds to the most 5’ cDNA identified by primer extension analysis and 5’ RACE of RNA from human intestinal HT-29 cells. The lengths of final exon 12 extend from the 3’ coding region to either the primary or secondary polyadenylation sites following 2 polyadenylation signals (underlined).
The controlled expression of annexin A13 in intestine and its limited expression in other tissues suggest the additional presence of repressor–enhancer element(s) in its promoter, but detailed site analyses reveal no obvious candidate(s). An analysis of the annexin A13 promoter region for repetitive elements by RepeatMasker or RepBase detected a LINE4a element spanning 1 kb upstream of bp –737 relative to the putative transcription start point of exon 1 (fig. 3). This is evidently a 5’-truncated repeat of an active LINE1 element still capable of retrotransposition in mammalian genomes (HGP 2001). Its possible presence in the mouse gene and the consideration that a remnant of the extinct LINE2 element exerts a key repressive role in the annexin A6 promoter (Donnelly, Hawkins, and Moss 1999) suggest that this region warrants attention for the promoter function of annexin A13. Although the intestine-specific expression of annexin A13 is well-documented (Wice and Gordon 1992), it is present in dog kidney MDCK cells (Fiedler et al. 1995), and the aforementioned transcription factors may have a broader role in other gastrointestinal organs (especially liver and pancreas). We therefore examined 21 human ESTs, 19 mouse ESTs, and 15 zebrafish ESTs for annexin A13 in dbEST to determine the true extent of its tissue distribution. Annexin A13 was represented by one to three ESTs from each of the normalized libraries of pancreas, prostate, ovary, embryonic liver pool, and lung-testis–B-cell pool, lung, testis, stomach, olfactory epithelium, retina, and brain. Despite this apparent tissue diversity in normalized cDNA libraries, expression levels are known to be low outside of intestine, and relative expression levels of the two isoforms await quantitative assessment.

Genomic Environment of ANXA13

The draft sequence containing and surrounding annexin A13 in the human genome (HGP 2000) permitted the identification of internal repetitive elements and external gene neighbors that might bear on this gene’s regulation, function, and evolutionary history. Two HGP BAC

conservation with the human counterpart and confirmed the retention of certain elements that have been implicated in gastrointestinal gene expression (fig. 3). The identification of cis-element binding sites for putative transcription factors in the promoters of highly expressed annexins has generally been of uncertain value in the absence of direct functional studies, but such predictions can be valuable guides for promoter analysis of a tightly regulated, tissue-specific gene such as annexin A13. We used MatInspector (Quandt et al. 1995) to detect transcription factor binding sites with high compatibility (>80%) to sequence matrices. Quite remarkably, the numerous potential binding sites for CDX2, hepatic nuclear factor HNF1, related winged helix/forkhead transcription factors FREAC and XFD, Kruppel factor GKF, and GATA3, read like a model gene promoter for gastrointestinal-specific gene expression. In particular, the homeodomain transcription factor and tumor suppressor CDX2 has a well-defined synergistic role with HNF1 in gut organogenesis and functional maintenance (Mitchelmore et al. 2000), including a temporal and spatial expression very similar to annexin A13 (Sibley et al. 2000). Annexin A13 clearly has an AT-rich promoter (65% from bp –1,000 to +100 in human) and an unusually low CpG dinucleotide content (0.8%) suggestive of methylation-mediated decay, both compatible with the putative binding sites for these particular transcription factors. This contrasts with other annexin promoters characterized as GC-rich and replete with Sp1 binding sites (Fernandez et al. 1994; Bances et al. 2000; Carcedo et al. 2001).

The controlled expression of annexin A13 in intestine and its limited expression in other tissues suggest the additional presence of repressor–enhancer element(s) in its promoter, but detailed site analyses reveal no obvious candidate(s). An analysis of the annexin A13 promoter region for repetitive elements by RepeatMasker or RepBase detected a LINE4a element spanning 1 kb upstream of bp –737 relative to the putative transcription start point of exon 1 (fig. 3). This is evidently a 5’-truncated repeat of an active LINE1 element still capable of retrotransposition in mammalian genomes (HGP 2001). Its possible presence in the mouse gene and the consideration that a remnant of the extinct LINE2 element exerts a key repressive role in the annexin A6 promoter (Donnelly, Hawkins, and Moss 1999) suggest that this region warrants attention for the promoter function of annexin A13. Although the intestine-specific expression of annexin A13 is well-documented (Wice and Gordon 1992), it is present in dog kidney MDCK cells (Fiedler et al. 1995), and the aforementioned transcription factors may have a broader role in other gastrointestinal organs (especially liver and pancreas). We therefore examined 21 human ESTs, 19 mouse ESTs, and 15 zebrafish ESTs for annexin A13 in dbEST to determine the true extent of its tissue distribution. Annexin A13 was represented by one to three ESTs from each of the normalized libraries of pancreas, prostate, ovary, embryonic liver pool, and lung-testis–B-cell pool, lung, testis, stomach, olfactory epithelium, retina, and brain. Despite this apparent tissue diversity in normalized cDNA libraries, expression levels are known to be low outside of intestine, and relative expression levels of the two isoforms await quantitative assessment.

Genomic Environment of ANXA13

The draft sequence containing and surrounding annexin A13 in the human genome (HGP 2000) permitted the identification of internal repetitive elements and external gene neighbors that might bear on this gene’s regulation, function, and evolutionary history. Two HGP BAC
clones have been positioned in contig CTG17737 by the Washington University Genome Sequencing Center and in the NCBI annotation contig NT_008157. Introns of this 58-kb gene sequence contain numerous repetitive elements (fig. 4), including 17 Alu and 22 mammalian interspersed repeat short-interspersed repetitive elements, 11 L1, five L2, and one L3 long interspersed nuclear elements (LINEs), 11 MalR and two endogenous retrovirus long terminal repeats, and 11 simple microsatellite repeats. Their significance in gene function, diagnostics, and evolution remains to be determined. BLAST searches of mapped genes against this contig localized other genes in the immediate proximity of chromosome 8q24.12, including a zinc finger homeodomain transcription factor, ZHX1, squalene epoxidase, SQLE, and several predicted open reading frames for unidentified genes MGC3067, FLJ10204, and KIAA0493 (fig. 4). These genes await functional characterization to determine whether they share regulatory or functional features in common with annexin A13. As closely linked genes, they can serve to verify orthologous regions in other species genomes and may ultimately help to identify an invertebrate ancestor of annexin A13.

Genetic Linkage Mapping of Mouse Anxa13

Our previous map of human annexin A13 to chromosome 8q24.1-q24.2 (Morgan et al. 1998) was confirmed in the human genome sequence map, approximately 125 Mb from the p-telomere corresponding to band 8q24.12 (HGP December 12, 2000 draft). The orthologous mouse locus was localized by interspecific backcross analysis using progeny derived from matings of ([C57BL/6J × M. spretus] F1 × C57BL/6J) mice. This interspecific backcross mapping panel has been typed for over 2,600 loci that are well-distributed among all the autosomes as well as the X chromosome (Cope-land and Jenkins 1991). C57BL/6J and M. spretus DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative RFLPs using a mouse cDNA probe. The 9.2- and 4.7-kb SpH1 M. spretus RFLPs were used to follow the segregation of the Anxa13 locus in backcross mice. The mapping results indicated that Anxa13 is located in the central region of mouse chromosome 15 linked to syndecan 2 (Sdc2), myelocytosis (Myc), and thyroglobulin (Tgn). Although 93 mice were analyzed for every marker and are shown in the segregation analysis (fig. 5A), up to 151 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. The ratios of the total numbers of mice exhibiting recombinant chromosomes to the total numbers of mice analyzed for each pair of loci and the most likely gene order are: Sdc2—20:134—Anxa13—7:132—Myc—6:151—Tgn. The recombination frequencies, expressed as genetic distances in centimorgans (cM) ± standard error, are Sdc2—14.9 ± 3.1—Anxa13—5.3 ± 2.0—Myc—4.0 ± 1.6—Tgn. This placed the gene in the middle of an extensive human-mouse homology group corresponding to human 8q22-q24 and mouse chromosome 15 ca. 10 to 45 cM from the centromere (fig. 5A).

Annexin A13 Synteny and Interspecies Chromosome Homology

The ANXA13 chromosomal region (8q24.12) is responsible for the contiguous gene syndrome Langer-Giedion (8q24.11-q23.13), various cancers in which the nearby MYC gene (8q24) is overexpressed, and colon adenocarcinoma in which the linked gene WISP1 (8q24.1-q23.3) has been implicated. The significant expression of annexin A13 in HT-29 transformed colon cells (Wice and Gordon 1992) and high incidence (35%) of ESTs derived from colon adenocarcinoma further suggested that its functional activity might be altered by chromosomal aberrations affecting cell growth and differentiation. The extensive homology region with mouse chromosome 15 (fig. 5B) makes it of interest in finding out whether the evolutionary history of the ancestral chromosome can be traced to earlier diverging species such as fish. Our full-length zebrafish annexin A13.1 cDNA (gb:AJ306452) was used to identify 12 zebrafish clones in dbEST, among which one homologous clone f840a08 (gb:AI461284) was found to have been localized to linkage group 24 in the current zebrafish map.
(Postlethwait et al. 2000; Woods et al. 2000). The c-myc locus is presently the only other gene in the same linkage group with a known ortholog in the extensive human-mouse homology group containing annexin A13 (fig. 5B). This suggests a more limited, albeit uncertain, extent of genetic linkage surrounding annexin A13 in earlier diverging species, although details should eventually clarify the formative genomic history of this chromosomal region. Further comparisons will be particularly useful for determining the pattern and extent of ancient polyploidization in fish, as we have detected duplicate cDNAs from zebrafish for most annexins except A7, A8, A9, and A10 (not shown).

Evolutionary Separation of Annexins A7 and A11 from A13

Phylogenetic analysis was used to confirm annexin A13 orthology for the nine representatives described in figure 1, whereas sea urchin and urochordate annexins(s) achieved only weak association with invertebrate ANXB members intermediate between annexins A13 and A7 (results not shown). However, the coincidence of exon splice sites for sea urchin and urochordate genomic sequences with annexins A7 and A13 are highly suggestive of an ancestral relationship. Whether they represent direct orthologs or a vertebrate outgroup must be determined from structural and phylogenetic analysis of their respective coding sequences. Bootstrap parsimony analysis of protein alignments comprising full-length representatives from all eukaryotic phyla portrayed the vertebrate annexin A7-A11 bifurcation as the most evolved pair in a protist annexin-rooted tree (fig. 6A). The annexin A13 orthology group branched from a more basal position amidst nonvertebrate members of the animal ANXB family but later than ANXC-fungi, in agreement with a previous analysis (Braun et al. 1998). This implies that the Pro-Gly-rich amino termini of A7 and A11 may not have been inherited from either ANXC1 or C. elegans nex-2 (i.e., they evolved by independent, convergent evolution) or that ANXA13 selectively lost this feature. We infer that annexin A13 possesses a higher proportion of ancestral characters, and its exon splice pattern, closely related to annexins A7 and A11 (fig. 2), is consistent with the sequential, progressive evolution of all human annexins from this primary progenitor. Phylogenetic and molecular dating of annexin A13-A7-A11 separations within this base clade of the A family annexins further substantiated the fundamental importance of this concept for the origin of human annexins. Evolutionary distances between these genes were estimated by protein maximum likelihood using TREE-PUZZLE to reveal relatively longer interspecies branch lengths for annexin A13 orthologs and roughly equivalent distances of annexins A7 and A11 from their common ancestor (fig. 6B). The former observation reflects a more rapid evolutionary rate for annexin A13, and the latter is consistent with rapid, successive duplication of annexins A7 and A11 from A13.

Individual gene evolution rates were calculated by measuring nonsynonymous nucleotide substitutions (NNS) between species orthologs (Li 1993) calibrated against the assumed species separation times of 360 Myr for amniotes from amphibia and 450 Myr for tetrapods from fish, given by fossil and molecular evidence (Kumar and Hedges 1998). Annexin A13 has evolved at a rate of 0.468 NNS per site per 1,000 Myr, almost double the rates for annexins A7 and A11. Gene rates calibrated to the tetrapod-fish separation were considered most reliable for statistical and temporal reasons, although we note that these values were 15% to 25% higher than those obtained from more recent interspecies divergences of the same annexins, e.g., rodent versus mammal (Morgan et al. 1998). The accelerated evolution of duplicated genes in fish (Robinson-Rechavi and Laudet 2001) is not apparent from the branch lengths in figure
Annexin A7 and A11 were thus estimated to have diverged from annexin A13 approximately 714 and 737 MYA, i.e., more or less contemporaneously and before chordate emergence, allowing for a minimum 10% error margin in gene rates because of standard error of the NNS estimates and uncertainty in species separation times. The apparent absence of annexins A13, A7, and A11 from invertebrate genomes of insects, worms, and molluscs by phylogenetic, gene structure, and chromosomal linkage criteria precludes their origins much earlier than 800 MYr. However, the full diversity of invertebrate annexin genes is presently uncertain, and we do not exclude the possible silencing of protostome orthologs or the extinction of their primary hosts.

**Discussion**

The history of multigene families is intimately tied to the dynamics of chromosomal rearrangement and speciation. Phylogenetic and molecular dating analyses of the vertebrate annexin A family place its inception and ensuing expansion to 12 member genes in a period 500 to 800 MYr (Morgan et al. 1998), when the ancestral genome was undergoing amplification by polyploidization and selective chromosomal rearrangement (Pébusque et al. 1998; Martin 2001). Although certain annexin genes may indeed have been created as a consequence of chromosome duplications (Morgan et al. 1999a, 1999b), the available data refute any simple notion that vertebrate annexins derived entirely by large-scale genome duplication(s) of some invertebrate genome. Comparative gene structures and genetic maps provide crucial molecular markers to corroborate phylogenetic analyses, showing that at least five annexins in plathelminths (e.g., *Schistosoma, Taenia*), five in nematodes (e.g., *Caenorhabditis, Brugia*), and three in insects (e.g., *Drosophila, Bombyx*) all represent unique members of the B family structurally distinct from vertebrate annexins (Morgan and Fernandez 1997). The lack of a strong hierarchical signal in these ancient duplications and the incomplete spectrum of chordate invertebrate sequences presently preclude resolution of their most proximal invertebrate relatives by phylogenetic analysis alone (fig. 6). Our previous gene structure analysis confirmed that annexin A11 is congruent to the nine more divergent vertebrate annexins (Bances et al. 2000) and is otherwise most closely related to annexin A7 and the *Drosophila* annexins. Annexin A13, however, possesses a unique, intron-poor gene structure and retains more ancestral characters, to give it a more deeply rooted association with invertebrate annexins in phylogenetic analysis (fig. 6; Braun et al. 1998). Partially characterized echinoderm and urochordate annexins (Shen et al. 1994; this study, and our unpublished observations) predict the existence of novel subfamilies in invertebrate chordates with exon splicing congruent to that of higher vertebrates. These ongoing sequencing projects should provide useful genetic data for tracing the true species distribution and orthologous invertebrate origins of annexins A13, A7, A11, and the remaining nine descendant members constituting the annexin A family.

**Fig. 6.—**Evolutionary origins of annexins A13, A7, and A11. **A**. The 311-aa homologous tetrad regions of 33 eukaryotic annexins were aligned (CLUSTALW), bootstrap replicated 1,000 times (SEQBOOT), analyzed by protein parsimony (PROTPARS), and resolved to a majority-rule consensus tree (CONSENSE) using programs from PHYLIP. *G. intestinalis* ANXE1 was defined as outgroup (Morgan and Fernandez 1997), and other taxa with genus-species abbreviations defined in Methods include the eight plant annexins from *Arabidopsis*, two ANXC1 members of the fungi-mycetozoa clade, two platyhelminth annexins from *Schistosoma*, five nematode annexins from *Caenorhabditis* and *Ascaris*, three insect annexins from *Drosophila, Hydra* ANXB12, and a novel mollusc annexin from *Crassostrea*. Tree topology received the statistical bootstrap support shown as percentages at each bifurcation. **B**. Protein maximum likelihood analysis was performed on 311-aa homologous portions of species orthologs of annexins A7, A11, and A13 using TREE-PUZZLE 5.0 (JTT substitution model, eight discrete gamma rates with alpha 1.3, no clock assumption). The solid circle marks the point (714 to 737 MYr) when annexins A7 and A11 separated from their common ancestor with annexin A13, whereas branch numbers give the maximum likelihood distance to the respective branch tips. The assumed species separation times are underlined: for fish-tetrapod 450 MYr, amniote-amphibia 360 MYr, and rodent-primate 110 MYr (Kumar et al. 1998). The number of nonsynonymous nucleotide substitutions (NNS) between orthologous taxa was calculated (Li 1993) to estimate individual gene evolution rates for annexins A13, A11, and A7 (0.468, 0.251, and 0.253 NNS × 10⁻⁹ per site per year, respectively) based on interspecies divergence divided by twice the host species separation times. Paralogous gene divergence times were similarly calculated using the NNS distances divided by twice the average gene evolution rate to calculate the separation times of annexin A13 from A7 (0.514/2 × 0.360 = 714 MYr) and from A11 (0.532/2 × 0.361 = 737 MYr).
Although the sequential formation of annexins A13, A7, and A11 could be loosely inferred from phylogenetic analysis of coding regions, the comparative gene structures establish their direct relatedness and probable order of duplication (fig. 2). The homology of the human, rodent, and fish annexin A13 genes (fig. 2) affirms a stable architecture over the latter half of this gene’s lifetime, even though chromosome regional homology may be less extensive in fish versus mammal (fig. 5). Molecular dating implies that annexins A13, A7, and A11 all existed before chordate genome expansion, so that the greatest subsequent upheaval may have been in chromosomal rearrangements, e.g., an intrachromosomal duplication of annexin A7 to A11, syntenic in human chromosome 10 and mouse chromosome 14 (Fernandez et al. 1996; Morgan et al. 1998). The available genetic linkage data do not, however, provide direct mechanistic evidence of a segmental chromosome duplication event involving annexins A13 and A7. The extrapolation of genetic maps and gene structures to other species should help to confirm true orthologs in simpler animals and facilitate the study of annexin genetic function.

The presence of a coding cassette exon in annexin A13 (fig. 2) and the conservation of DNA cis-elements for intestine-specific trans-acting protein factors in the human and rodent promoters (fig. 3) are significant findings. Both are relevant to the intestine-specific regulation of this gene’s expression and, together with subcellular localization studies (Lafont et al. 1998; Massey-Harroche, Mayran, and Maroux 1998), can help define the cell conditions under which this gene is active. The specific role of promoter regulatory elements for caudal homeobox (CDX), hepatic nuclear factor (HNF), forkhead (FREAC, XFD) transcription factors, and the upstream LINE1 element should be especially pertinent to the role of annexin A13 in gut tissue development and maintenance. The limited expression of annexin A13 outside of intestine, its temporal pattern during tissue development, and comparisons with other species such as fish merit further study to determine how gene regulation is coupled to protein function.

Gene evolution and function are ultimately determined at the level of protein interactions, so that the identification of diagnostic residues for annexin A13 (fig. 1) can be instructive for functional studies. The findings that the myristoylation site and cassette isoform are conserved in other species are relevant to the proposed apical transport function of annexin A13 in lipid rafts (Plant et al. 2000), whereas the excess of basic residues and a potential KGD ligand for integrins pose new considerations for nuclear and membrane actions, respectively. The apparent interspecies conservation and amino terminal accessibility of this ligand are analogous to a similarly conserved and accessible KGD-RGD motif in the carboxy termini of annexins A1-A2-A9 (Morgan and Fernandez 1998). Although the functional significance of this putative ligand requires experimental verification, such a new line of investigation could help to clarify the mechanism and specificity of annexin membrane–binding interactions and their sporadic, extracellular localization. Because the intrinsic, common function of all annexins is, however, determined by the homologous tetrad core, its widespread amino acid conservation leaves open the mystery about essential domain interactions. In this regard, annexin A13 overexpression or knockout, especially in a primitive species lacking redundant annexins, offers the enhanced prospect of detecting an altered phenotype. A comprehensive view of the evolutionary genetics of the annexin gene superfamily should eventually provide an indication as to how the member genes contribute individually and collectively to the pathophysiologies and phenotypes of the species that contain them.

**Supplementary Material**

The sequence data described in this paper have been submitted to the GenBank/EMBL/DDBJ data libraries under the following accession numbers: AJ306450, human annexin A13 isoform b cDNA; AJ306451, mouse annexin A13 isoform a cDNA; AJ306452, zebrafish annexin A13.1 cDNA.

**Acknowledgments**

We thank Burton Wice of Washington University, St. Louis, Mo., for annexin A13 cDNA probe and information on human isoform b, Wen-Hsiung Li of University of Chicago, Chicago, Ill., kindly provided the program to calculate nt substitutions. The comments of the anonymous referee were particularly helpful. Deborah B. Householder gave excellent technical assistance. This study was supported by grant PB98-1529 from D.G.I. of Spain and by the National Cancer Institute under contract with ABL.

**LITERATURE CITED**


