Mouse annexin V genomic organization includes an endogenous retrovirus

Maria-Isabel RODRIGUEZ-GARCIA, Reginald O. MORGAN, Maria-Rosa FERNANDEZ, Patricia BANCES and Maria-Pilar FERNANDEZ¹

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

Mouse annexin V genomic clones were characterized by restriction analysis, Southern blotting and DNA sequencing. The entire gene spans close to 50 kb of the mouse genome and contains 14 exons ranging in size from 31 bp for exon 2 to 482 bp for exon 13 up to the polyadenylation site. Intron sizes range from 111 bp for intron 1b to more than 17 kb for intron 2. Non-coding exon 1 is present in two alternative forms separated by approx. 7.4 kb, and the two promoters associated with exons 1a and 1b are quite distinct. The upstream promoter has a TATA box and may direct the limited, tissue-specific expression of mRNA transcripts containing exon 1a. The downstream, TATA-less promoter has high G + C content, and exon 1b predominates among abundantly expressed mRNA species. The conservation

of certain *cis*-elements, including Sp1, AP2, γ -IRE and NF-IL6, in orthologous species of annexin V genes points to their possible role in *trans*-acting protein factor binding and gene regulation. Primer-extension analysis revealed multiple origins for transcription, with principal start sites 100–150 bp upstream of the ATG start codon in exon 2. Intron 4 was longer than that previously identified in the orthologous rat gene due to the integration of an apparently complete copy of the murine endogenous retrovirus element, MuERV-L. Phylogenetic analysis of annexin V from 12 species and the presence of neighbouring loci with paralogous counterparts linked to annexin VI pointed to the common ancestry of these genes via chromosomal duplication more than 600 million years ago.

INTRODUCTION

Annexin V has been a model for investigation of the structure, function and evolution of the annexin multigene family because it is ubiquitously expressed and regulated by growth factors [1–3], its unique N-terminus is comparatively short (4–6 amino acids) and its core tetrad of internally homologous repeats harbours all of the requisite features for annexin function [4]. The key residues controlling calcium-ion channel activity and the type II calcium-binding motifs were first described for annexin V [5,6], and its avid, calcium-dependent binding to anionic membrane phospholipids has also been implicated in the inhibition of cytosolic phospholipase A₂ [7] and protein kinase C [8]. Although it is primarily intracellular and binds to cytosolic receptors [9], annexin V is also recognized to be a secreted protein and extracellular receptor for type II collagen [10], influenza viruses [11], apoptotic cells [12] and autoantibodies [13].

Comparative studies of the annexin V gene structure in chick [14,15] and human [16,17] have helped to clarify its apparent ‘housekeeping’ role in vertebrate eukaryotic cells, and characterization of the rat gene [18] has provided a novel insight into its tissue-specific regulation via a dual promoter mechanism. Genetic mapping [19,20] and molecular evolution studies [21–23] have led to the hypothesis that this gene may have been created from annexin VIa by segmental duplication of a chromosomal region involving other paralogous genes [24]. In view of the importance of the mouse as a model for human physiology, transgenic and gene knockout experiments, we have characterized the gene structure of mouse annexin V to assess further the structural conservation of this gene in both the coding and dual promoter regions, to localize its transcription initiation site(s) and to define more fully its genetic origin.

MATERIALS AND METHODS

Identification and sequencing of genomic clones

Mouse genomic clones for annexin V were identified by hybridization screening of two different mouse 129/SVJ genomic libraries prepared from liver DNA ligated to λ DASH II and λ FIX II vectors (Stratagene, La Jolla, CA, U.S.A.). Random-primed ³²P-labelled cDNA for mouse annexin V ([20]; gb:U29396, where gb refers to the GenBank/EMBL/DDJB accession number) and a 2.3 kb rat genomic probe containing exon 1a (RR2.3 provided by Dr. Yoshinori Imai and colleagues, Department of Neurochemistry, National Institute of Neuroscience, Tokyo, Japan) were used for hybridization. Screening procedures, DNA isolation, restriction analysis, gel electrophoresis, Southern blot transfer and hybridization were performed essentially as described [25]. Genomic fragments were subcloned into either pUC19 or pBluescript II SK+/- and sequencing was performed by the dideoxy chain-termination method [26] with [α -³²S]dATP (Amersham, Little Chalfont, U.K.) and the Sequenase 2.0 kit (USB, Cleveland, OH, U.S.A.).

Primer-extension analysis

Mouse total RNA from murine NIH 3T3 cells was isolated by the isothiocyanate/acidic phenol method [27] and 50 μ g were used as template. The antisense primer was 100 pmol of 5'-end ³²P-labelled oligonucleotide 5M4, identical with the 20 bp non-coding strand at the 3' end of mouse annexin V exon 1b (5'-CAGGCAGTAGAGCGCAGGCT). These were preheated together at 90 °C for 2 min and annealed at 53 °C for 7 h. Extension proceeded at 37 °C for 2 h in the presence of 400 units of Moloney murine leukaemia virus reverse transcriptase

Abbreviations used: ANX/Anx, human/non-human annexin gene locus or DNA suffixed with the subfamily number; ANX, annexin protein; EST, expressed sequence tag; gb, GenBank/EMBL/DDJB accession number; tsp, transcription start point(s).

¹ To whom correspondence should be addressed (e-mail mpff@dwarf1.quimica.uniovi.es).

The nucleotide sequence data reported in this paper have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases and have been assigned the accession numbers AJ230107–AJ230124.

(Bethesda Research Laboratories, Gaithersburg, MD, U.S.A.) and the reaction products were subjected to electrophoresis through a 6% polyacrylamide denaturing gel that included a sequence ladder for estimation of molecular masses.

Computer sequence analysis

Sequence comparisons were performed with the Wisconsin Package version 8.0 (Genetics Computer Group, Madison, WI, U.S.A.). Putative regulatory elements in gene promoter regions were identified using SIGNAL SCAN version 4.0 software [28] with the currently available transcription factor databases. Molecular phylogenetic analysis of annexins was performed after multiple sequence alignment [29] by the maximum likelihood method using the PUZZLE program version 4.0 [30]. The resulting alignment and tree were input into the ANCESTOR parsimony program [31] to reconstruct ancestral protein sequences at each node, including that for the putative common ancestor of annexins V and VI.

RESULTS

Isolation, sequencing and characterization of mouse annexin V

Six genomic clones isolated from 129/SVJ mouse liver were characterized according to size and relative position by restriction mapping, subcloning and Southern blot hybridization to restriction fragments of the cDNA (Figure 1A). Four overlapping clones λ A5M-F10, λ A5M-F17, λ A5M-D10 and λ A5M-D14 spanned over 29 kb of the mouse genome and contained most of the 3'-coding portion of annexin V from 5 kb upstream of exon 3 through final exon 13. The extensive length of intron 2 (> 17 kb) necessitated rescreening the library with a 190 bp, *EcoRI*–*KpnI* fragment from the 5'-end of mouse cDNA to isolate

clones containing exons 1b and 2. The resulting genomic clone λ A5M-D13 contained 1 kb of the downstream promoter B, exons 1b and 2 and approx. 12 kb of intron 2. Mouse clone λ A5M-70 containing upstream promoter A and exon 1a was identified by hybridization with the 2.3 kb rat genomic probe derived from the corresponding region of the rat gene [18]. Its 16 kb size extended from 10 kb upstream of the alternative promoter A to 6 kb downstream of exon 1a.

Conservation of exon–intron splice junctions

The mouse annexin V gene spans at least 49 kb of the mouse genome and is composed of 14 exons and 13 introns with the distribution shown in Figure 1(B). The complete sequence of all exons and the size, location and partial sequences of flanking introns were obtained using vector and cDNA sequence-specific oligonucleotide primers to characterize exon–intron splice junctions. Complete sequence data, including 5'-upstream regions, have been deposited in the EMBL databank under accession numbers AJ230107–AJ230124 and results are summarized in Table 1. The sizes of introns not fully sequenced were determined by electrophoretic analysis of digested clones or amplified intron fragments obtained by PCR and Southern blot analysis. Exons range in size from 31 bp for exon 2 to 482 bp for exon 13 up to the polyadenylation site and the 13 interspersed introns range from 111 bp for intron 1b to > 17 kb for intron 2. All splice junctions conform to the 5'-donor/3'-acceptor sequences 'GT/AG' observed for U2-type introns [32] and the majority nucleotide consensus observed in other annexin genes, i.e., 3'-exon/intron/5'-exon 'ag/GTaaag...y_n-ncAG/ggt', where y stands for pyrimidine and n for any nucleotide. Introns 3, 6, 9 and 10 interrupted exons after the initial codon base (i.e. phase 1), whereas the remainder spliced between codons (phase 0). The absence of symmetrical exons in adjacent or equivalent repeat positions precluded shuffling, and exon boundaries showed no

Figure 1 Schematic organization of the mouse annexin V gene

(A) Restriction map and positions of the isolated clones which span the annexin V genomic region of mouse chromosome 3. *ANX/Anx*, human/non-human annexin gene locus or DNA suffixed with the subfamily number and genus-species abbreviation. Clone labels and sizes are indicated and letters B, E, H and P represent *Bam*HI, *Eco*RI, *Hind*III and *Pst*I enzyme cut sites, respectively. (B) The exon–intron organization of the mouse annexin V gene is shown on the same scale as its rat, human and chick orthologues, which are aligned at the translation start sites in exon 2. Unique inserts in introns 4, 5 and 6 of the mouse gene (arrows) include a 6.5 kb MuERV-L element, a poly(A)₂₄ tract, a 5-fold tandem repeat (TR)₅ and a poly(GT)₂₃ Z-DNA segment that are further detailed in the text.

Table 1 Primary structure of the major organizational components of the mouse annexin V gene

The 14 exons and intervening introns spanning the 50 kb mouse annexin V gene on chromosome 3 are described by partial nucleotide sequences, codon phase and numbered amino acid residues at each experimentally determined splice junction. The bp lengths of fully sequenced exons and mapped introns are included. The ATG start codon in 3' exon 2 and polyadenylation signal in 3' exon 13 are underlined. All introns conformed to the GT/AG rule for U-type introns [32].

Exon				Splice junction		Intron			
No.	5' Sequence	bp	3' Sequence	Phase	Amino acids	No.	5' Sequence	kb	3' Sequence
1a	AGCAGCTGAAC-	70	-TCTTCATGTGG	—	—	1A	gtaaga-	≥ 7.5	-ctgcag
1b	AGGTCGGGCTA-	122	-TCTACTGCCTG	—	—	1B	gtaagt-	0.111	-ctgcag
2	CTCCAGTCCAC-	31	-CAGCATC <u>ATG</u>	0	Met ¹ /Ala ²	2	gtaagc-	≥ 17	-tcacag
3	GCT ACG AGA-	85	-GGC TTG G	1	G/Iy ³⁰	3	gtaacg-	1.2	-tttcag
4	GT ACC GAT-	95	-TTT GGC AGG	0	Arg ⁶¹ /Asp ⁶²	4	gtaaga-	8.9	-aggtag
5	GAC CTT GTG-	114	-GCT CTT AAG	0	Lys ⁹⁹ /Gly ¹⁰⁰	5	gtaaaa-	1.6	-caacag
6	GGA GCT GGT-	91	-GAA GAA G	1	G/Iu ¹³⁰	6	gtaacc-	2.9	-tttcag
7	AA TAT GGT-	80	-CTC CTT CAG	0	Gln ¹⁵⁶ /Ala ¹⁵⁷	7	gtactg-	0.4	—
8	GCG AAT AGA-	57	-GAT GCT CAG	0	Gln ¹⁷⁵ /Ala ¹⁷⁶	8	gtgagt-	3.5	-ttacag
9	GCA TTG TTC-	94	-AGA AGA G	1	V/al ²⁰⁷	9	gtaagt-	1.1	-ttctag
10	TG TTT GAC-	96	-GCT GTC G	1	V/al ²³⁹	10	gtaggt-	1.3	-tcctag
11	TG AAG TCT-	59	-GCC ATG AAG	0	Lys ²⁵⁸ /Gly ²⁵⁹	11	gtgagt-	0.2	-ttaaag
12	GGT GCT GGG-	123	-ATG ATC AAG	0	Lys ²⁹⁹ /Gly ³⁰⁰	12	gtagaa-	0.6	-ttccag
13	GGC GAC ACA-	464	-GCGACAATAAA

Figure 2 Intronic repetitive elements

The partial sequence of a genomic fragment from intron 4 was identified as the 5'-LTR/*gag* portion of a murine endogenous retrovirus with leucine tRNA primer, MuERV-L [34]. The *gag* and *pol* genes encode polyproteins involved in the regulation of viral replication and are flanked on both ends by long terminal repeats (LTR). Other repetitive elements include a poly(A)₂₄ segment at the 3' end of intron 4, a (29-mer)₅ repeat in intron 5 and a continuous Z-DNA polypurine-pyrimidine tract, poly(TG)₂₃ in intron 6.

correspondence between homologous repeats of the protein tetrad repeat structure.

The identification of dual promoters in the rat annexin V gene [18] predicted at least two distinct transcription start sites and mRNA species differing in that portion of their untranslated regions encoded by either exon 1a or 1b. cDNA species containing alternative exon 1a were not detected by us nor observed as cDNA transcripts in the expressed sequence tag database (dbEST) [33]. The mouse annexin V coding region extends from bp 29 of exon 2 to bp 60 of exon 13, both largely untranslated exons. The previously reported cDNA [20] was identical with exon sequences beyond bp 57 of exon 1b, and both were consistent with the consensus sequence from 52 mouse annexin V expressed sequence tags detected by a BLAST search of the dbEST. The gene organization is congruent to that of its human and chick homologues downstream of promoter B [15,16], as promoter A has not yet been identified in these latter genes. One minor difference occurs in the truncation of the two final codons from mouse exon 2 which reduces the N-terminal length by 2 amino

acids. The rat gene [18] shares the same deletion in its exon 2 and its structural similarity with the mouse gene also extends to the homologous upstream promoters A.

Repetitive elements

Two marked size differences between the species gene maps (Figure 1B, Table 1) were the extensive and inconstant length of intron 2 in various species and the 6.5 kb greater length of intron 4 in mouse compared with the closely related rat gene. Additional sequencing of the 3'-end of mouse intron 4 yielded a 0.8 kb segment with 93.5 % nucleotide identity with the 5' long terminal repeat and *gag* portion of a murine endogenous retrovirus element, MuERV-L [34], and 59.4 % identity with the *gag*-like *Fv1* gene for Friend virus susceptibility [35]. The 6.5 kb total insert size accounts for the equivalent size differential between the mouse and rat introns 4 (Figure 2) and restriction analysis matched the pattern expected for a complete retroviral element. Sequence evidence was also obtained for the presence of other repetitive elements in mouse introns 4, 5 and 6. Microsatellites included a 24 bp stretch of polyadenine near the 3'-end of intron 4, a 5-fold repeated 29-mer (AGCAGGGAGTGGCTTGCTCTAGTAGGTT) and a 23 dinucleotide TG-repeat of alternating purine-pyrimidine Z-DNA.

Upstream promoter A

Sequence data from the most 5' genomic clone included 512 bp of a putative upstream promoter A (Figure 3) homologous with the rat gene [18]. An alignment of the mouse and rat promoters A revealed 82.2 % identity over 601 nucleotides, including the 52–70 bp proposed exons 1a. Perfect sequence conservation of specific sites with the potential to bind transcription regulatory factors was observed for several AP1 and GR (glucocorticoid regulatory) upstream elements, E2A and Hinf-A elements in exon 1a, and coincident TATA box elements. The low G+C content of promoter A (i.e. 40 %) compared with promoter B (64 %) was reflected in a lack of Sp1 sites common to annexin V

Figure 3 DNA sequence of the upstream promoter A region

Partial genomic DNA sequences are aligned for promoters A from the mouse and rat [18] annexin V genes. Possible regulatory element binding sites for known transcription factors are underlined and identified by standardized abbreviations above. The proposed *tsp* at alignment position +1 is upstream of the longest rat cDNA and 30 bp downstream from the TATA box. The 747 bp contiguous sequence from promoter A to 5'-intron 1a has been assigned accession number AJ230107.

downstream, TATA-less promoters [15,16]. The proposed transcription start point (*tsp*) was therefore assigned in consideration of the longest cDNA obtained for the rat gene (Figure 3), the expectation that TATA-dependent transcription should be initiated within 30 bp downstream of the TATA box [36], and alignment similarity with unrelated, but characterized TATA-containing promoters in the eukaryotic promoter database [37].

Initiation of transcription

The observed rarity of cDNA species containing exon 1a meant that the principal initiation site(s) for annexin V transcription were controlled by the downstream promoter B and initiated at exon 1b for the majority of mRNA species. Localization of the transcription start point(s) (*tsp*) was confirmed by primer-extension analysis. Total RNA isolated from mouse NIH 3T3 cells was used as a template for the extension of an end-labelled primer, corresponding to positions 53–72 on the non-coding strand of the published cDNA sequence [20]. Four major *tsp* were localized 150, 138, 130 and 100 bp upstream from the ATG initiation codon (Figure 4). Minor bands were occasionally observed in primer-extension experiments, but they are believed to represent pauses of the reverse transcriptase caused by

Figure 4 Multiple transcription-initiation sites inferred from primer-extension analysis

The products from primer extension of total RNA from NIH 3T3 cells (lane P) were size-estimated by simultaneous electrophoresis of the reaction products from a related sequencing reaction (lanes ACGT). The latter utilized a plasmid subclone of the mouse annexin V genomic region containing exon 1b and the same 5M4 oligonucleotide primer used for primer extension. The correspondence between the sequence ladder (lanes ACGT) and the primer extension stops for *tsp1* to 4 (arrows) was used to position transcription starts at 150, 138, 130 and 100 bp, respectively, upstream of the ATG translation start codon.

secondary structures in the GC-rich untranslated region. Since band compression due to the high GC content impeded precise transcript sizing, the actual start site may lie slightly more downstream to coincide with human, rat and chick annexin V genes. The proposed primary *tsp1* (designated position +1) thus lies 150 bp upstream of the ATG start codon and follows two Sp1 binding sites conserved in orthologous genes. Alternate *tsp* may initiate at bp +13, +21 and +51 bp of exon 1b, as we have detected a mouse EST (gb:AA002439) coinciding with *tsp3* at bp 21 of exon 1b, and both 5'-rapid amplification of cDNA ends and multiple ESTs indicated frequent starts at *tsp4* at bp 51 of exon 1b.

Downstream promoter B

An *XbaI*–*EcoRI* digestion fragment from the λ gDA5M-13 genomic clone that included exon 1b and 510 bp of 5' flanking region was subcloned and fully sequenced. Its alignment showed 80% identity with 615 nt of the homologous region in the rat gene and 63% identity with the human gene (Figure 5). As primer extension had indicated multiple apparent start sites, mouse exon 1b was designated to begin near *tsp1*, coincident with the proposed promoters B and exons 1b of the rat and human. Inspection of the 5' flanking sequence revealed a CpG island with an average of 65.6% G+C content, 11.2% CG dinucleotides and several putative, regulatory Sp1 sites spanning the primary *tsp*, all comparable with orthologous promoters B. Other potential *cis*-acting regulatory elements in the promoter region included matches to known DNA consensus sequences for NF-IL6, GR, c-myc, MyoD, NF κ B and AP2 that are also preserved in other species of annexin V promoters B [15,16,18].

Figure 5 DNA sequence of the downstream promoter B region

Partial genomic DNA sequences are aligned for promoters B from the mouse, human [16] and rat [18] annexin V genes. Possible regulatory element binding sites for known transcription factors are underlined and identified by standardized abbreviations above. The proposed *tsp* at alignment position +1 was based on the most 5' primer-extension analysis and extends into exon 1b (122 bp) and intron 1b (111 bp) complete). The *tsp2-4* sites were identified by primer-extension analysis, 5'-RACE-PCR and the most 5' expressed sequence tags from dbEST. The 702 bp contiguous sequence from promoter B to 5'-intron 2 has been assigned accession number AJ230108.

Phylogenetic origin

Annexin V consistently associates with the 5' tetrad of annexin VI in phylogenetic analyses [20–24], and the fact that their chromosomal loci have many of the same paralogous neighbours (e.g. *FGF2*, *CCNA* and *IL2*) has led to the suggestion that they may have shared a common ancestor [24]. A more detailed analysis was therefore performed on a sequence alignment of 22 tetrads from various species to confirm their branching order and estimate the times of these events. Sequences were analysed by the maximum likelihood method and possible trees were resolved by the quartet puzzling algorithm of the PUZZLE program [30]. These results facilitated reconstruction of the protein sequence for their putative common ancestor using the ANCESTOR program [31] and this is included in the protein evolutionary tree along with the probabilities for the given branching order (Figure 6). Values above 80 % reflect a strong likelihood that the individual subfamilies and early-diverging species of annexin V evolved in the order shown. The results imply that annexin VIa shared its most recent common ancestor with annexin V, and a more distant or indirect ancestry with its complementary 3' tetrad VIb.

The fact that branch tips do not vertically align at the right indicates some (minor) rate variation along individual lineages, but the horizontal scale may approximate a linear rate of amino acid replacement over time, based on previously calculated estimates [24]. The branching order of taxa within individual subfamilies corresponds to the expected divergence order of the corresponding species, namely, fish, amphibians, avians and mammals respectively. If we assume that fish annexins V (carp gb:C88438; zebra fish gb:Y16043) separated from tetrapod orthologues 450 million years ago, then the annexin V gene itself can be estimated graphically to have shared a common ancestor with annexin VIa about 600 million years ago, perhaps due to a chromosomal duplication [24]. The available results do not clarify the mechanism or timing by which the annexin VI octad was created through the fusion of two precursor tetrads, although a tandem duplication event has been proposed [38]. The annexin V mutation rate was recalculated from a previous estimate [20] by including additional, earlier diverging species and their accepted divergence times [39]. Linear regression analysis of amino acid changes from the human sequence versus divergence times indicated an average of 1 % amino acid replacements every 12.5 million years, which was ostensibly constant in all species ($r = 0.99$). The 10 % amino acid difference between the common ancestor of annexins V and VIa, and the ancestral node from which fish annexin V derived, was thus calculated to an additional 125–150 million years before the divergence of fish, or about 600 million years ago, consistent with the graphic estimation above.

DISCUSSION

The DNA sequence and structural organization of the mouse annexin V gene show clear homology with the orthologous genes of the chick, human and rat [15,16,18] and permit their comparative assessment. Overall gene sizes appear to have increased in the listed order, so that the rat gene is most closely related (Figure 1), except for the sizable insertion of an apparently complete murine endogenous retrovirus, MuERV-L, in mouse. Bénit et al. [34] recently described the first of an estimated 100–200 copies of MuERV-L in the mouse genome with complete open reading frames for the *gag* and *pol* genes, and we have confirmed its very high sequence similarity to the *gag* and 5'-long terminal repeat segment and equivalent total size to the copy found in mouse annexin V intron 4. Both of these MuERV-L copies in mouse have 60 % nucleotide similarity to the highly conserved *gag* portion of the HERV-L in the human. If MuERV and HERV lineages separated with these two species about 80 million years ago [39] and evolved at equal rates, then the 4.3 % nucleotide difference between the two MuERV-L *gag* genes of mouse annexin V in chromosome 3 [18] and the originally described mouse homologue [34] implies that they diverged about 8.6 million years ago. This was presumably subsequent to the divergence of mice from rats [39] and is consistent with the absence of these retroviral elements from the latter species. Functions for other, internally repetitive, intron elements detected in mouse annexin V are yet to be investigated, but poly(A) and poly(GT) Z-DNA tracts may contribute significantly to chromatin structure and recombination [40], and their expected propensity to polymorphic change suggests their utility as distinguishing genomic markers.

The regions and extents of sequence conservation in annexin V from various species are likely to reflect the importance of those regions in the structure, function and regulation of this gene. The confirmation of a highly conserved, TATA-containing promoter upstream in both the mouse and rat genes suggests that it should also be present and functional in annexin V genes of other

Figure 6 DNA and protein phylograms of annexin V and VI evolution

Maximum likelihood trees for annexin tetrad alignments of nucleotides (A) and amino acids (B) were computed by the PUZZLE program [30] and fully resolved by quartet puzzling with annexin XI as the primary outgroup. The protein alignment and corresponding tree were used for parsimony analysis in the ANCESTOR program [31] to derive the putative common ancestor of annexins V and VI (ANX6-CA), subsequently introduced as the outgroup. Sequence alignments consisted of 933 nucleotide or 311 amino acid sites from the homologous tetrad core regions of up to 22 taxa, and site rate variation was approximated by a discrete gamma distribution. Percentage values at each node reflect the branching frequency (probability), and branch lengths (extent of evolution) correspond to the horizontal scale. Genus-species abbreviations are Bta (*Bos taurus*, cow), Cca (*Cyprinus carpio*, carp), Cpy (*Cynops pyrrhogaster*, newt), Dre (*Danio rerio*, zebra fish), Gga (*Gallus gallus*, chicken), Hsa (*Homo sapiens*, human), Mau (*Mesocricetus auratus*, hamster), Mmu (*Mus musculus*, mouse), Ocu (*Oryctolagus cuniculus*, rabbit), Rra (*Rattus rattus*, rat), Ssc (*Sus scrofa*, pig), and Tma (*Torpedo marmorata*, marbled electric ray).

species, even though cDNA corresponding to alternative, non-coding exon 1a has so far been demonstrated only in lung, skeletal muscle, brain and placenta from rat [18]. The equally strong sequence conservation, elevated G + C content and putative regulatory elements, such as Sp1, in common among downstream promoters B of annexin V from four different species argue for its role as the core promoter responsible for the ubiquitous expression of annexin V. All 52 mouse annexin V ESTs detected in the dbEST corresponded to mRNA derived from non-coding exon 1b in the downstream promoter B. Exons 2–13 are common to both types of mRNA transcripts, since exons 1a and 1b share the same 3' splice site with exon 2, despite being separated by 7.4 kb. Our evidence for multiple transcription start sites also corroborates results for the human gene [16], and the availability of genomic sequences from various species should now facilitate the localization of regulatory signals controlling such diversity. It should be noted that cDNAs from various species support the assignment of the primary *tsp1* near primer-extension ends, as observed in the present study, although transcription start sites may actually occur more frequently at downstream sites, as reflected in the observed EST distribution.

The proposed origin of the annexin VI octad by tandem duplication, followed by fusion of the resulting tetrads [38], is intuitively consistent with known genetic mechanisms, but is inconsistent with the structural differences and greater evolutionary separation of its component tetrads VIa and VIb compared with annexins V and VIa. Phylogenetic analysis rather favours a recent common origin for annexins V and VIa, coinciding with an independent fusion event involving an as yet unidentified tetrad progenitor of annexin VIb. A complete knowledge of all extant annexins, their comparative genetics and evolutionary history will be required to resolve such issues. The possible role of a chromosomal segment duplication in the creation of annexins V and/or VI, together with other paralogous genes [24], will require more detailed genetic mapping in human

and mouse to eventually ascertain the true extent of physical linkage among these genes. In view of the structural coincidence of exon splice sites between vertebrate species of most annexin genes, rigorous phylogenetic analyses will also be required to delineate the order and timing of gene duplications that gave rise to individual members of this multigene family.

Dr. Yoshinori Imai kindly provided a rat genomic clone containing exon 1a for use as probe. This work was supported by grant PM95-0152 from Dirección General de Enseñanza Superior (D.G.E.S.) of Spain and by a fellowship from the Fondo de Investigaciones Sanitarias (F.I.S.) to M.-I. R.-G.

REFERENCES

- Schlaepfer, D. D. and Haigler, H. T. (1990) *J. Cell Biol.* **111**, 229–238
- Brasemann, S., Bergers, G., Wrighton, C., Graninger, P., Surpirturiga, G. and Busslinger, M. (1992) *J. Cell Sci.* **16** (Suppl.), 97–109
- Hofmann, C., Gropp, R. and von der Mark, K. (1992) *Dev. Biol.* **151**, 391–400
- van Heerde, W. L., de Groot, P. G. and Reutelingsperger, C. P. M. (1995) *Thromb. Haemostasis* **73**, 172–179
- Liemann, S., Benz, J., Burger, A., Voges, D., Hofmann, A., Huber, R. and Gottig, P. (1996) *J. Mol. Biol.* **258**, 555–561
- Swairjo, M. A., Concha, N. O., Kaetzel, M. A., Dedman, J. R. and Seaton, B. A. (1995) *Nat. Struct. Biol.* **2**, 968–974
- Mira, J. P., Dubois, T., Oudinet, J. P., Lukowski, S., Russo-Marie, F. and Geny, B. (1997) *J. Biol. Chem.* **272**, 10474–10482
- Dubois, T., Mira, J. P., Feliars, D., Solito, E., Russo-Marie, F. and Oudinet, J. P. (1998) *Biochem. J.* **330**, 1277–1282
- Ohsawa, K., Imai, Y., Ito, D. and Kohsaka, S. (1996) *J. Neurochem.* **67**, 89–97
- Mollenhauer, J. and von der Mark, K. (1983) *EMBO J.* **2**, 45–50
- Huang, R. T., Lichtenberg, B. and Rick, O. (1996) *FEBS Lett.* **392**, 59–62
- Reutelingsperger, C. P. M. and van Heerde, W. L. (1997) *Cell. Mol. Life Sci.* **53**, 527–532
- Rodríguez-García, M. I., Fernandez, J. A., Rodríguez, A., Fernandez, M. P., Gutierrez, C. and Torre-Alonso, J. C. (1996) *Ann. Rheum. Dis.* **55**, 895–900
- Pfämmüller, E., Turnay, J., Bertling, W. and von der Mark, K. (1993) *FEBS Lett.* **336**, 467–471
- Fernandez, M. P., Fernandez, M. R. and Morgan, R. O. (1994) *Gene* **141**, 179–186

-
- 16 Fernandez, M. P., Morgan, R. O., Fernandez, M. R. and Carcedo, M. T. (1994) *Gene* **149**, 253–260
 - 17 Cookson, B. T., Engelhardt, S., Smith, C., Bamford, H. A., Prochazka, M. and Tait, J. F. (1994) *Genomics* **20**, 463–467
 - 18 Imai, Y. and Kohsaka, S. (1995) *Eur. J. Biochem.* **232**, 327–334
 - 19 Tait, J. F., Frankenberry, D. A., Shiang, R., Murray, J. C., Adler, D. A. and Disteche, C. M. (1991) *Cytogenet. Cell Genet.* **57**, 187–192
 - 20 Rodriguez-Garcia, M. I., Kozak, C. A., Morgan, R. O. and Fernandez, M. P. (1996) *Genomics* **31**, 151–157
 - 21 Morgan, R. O. and Fernandez, M. P. (1995) *Mol. Biol. Evol.* **12**, 967–979
 - 22 Morgan, R. O. and Fernandez, M. P. (1997) *J. Mol. Evol.* **44**, 178–188
 - 23 Morgan, R. O. and Fernandez, M. P. (1997) *Cell. Mol. Life Sci.* **53**, 508–515
 - 24 Morgan, R. O., Bell, D. W., Testa, J. R. and Fernandez, M. P. (1998) *Genomics* **48**, 100–110
 - 25 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
 - 26 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
 - 27 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
 - 28 Prestidge, D. S. (1991) *Comput. Appl. Biosci.* **7**, 203–206
 - 29 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
 - 30 Strimmer, K. and von Haessler, A. (1996) *Mol. Biol. Evol.* **13**, 964–969
 - 31 Zhang, J. and Nei, M. (1997) *J. Mol. Evol.* **44** (Suppl.), S139–S146
 - 32 Sharp, P. A. and Burge, C. B. (1997) *Cell* **91**, 875–879
 - 33 Boguski, M. S., Lowe, T. M. and Tolstoshev, C. M. (1993) *Nat. Genet.* **4**, 332–333
 - 34 Bnit, L., de Parseval, N., Casella, J. F., Callebaut, I., Cordonnier, A. and Heidmann, T. (1997) *J. Virol.* **71**, 5652–5657
 - 35 Best, S., Le Tissier, P., Towers, G. and Stoye, J. P. (1996) *Nature (London)* **382**, 826–829
 - 36 Sawadogo, M. and Sentenac, A. (1990) *Annu. Rev. Biochem.* **59**, 711–754
 - 37 Bucher, P. (1996) The eukaryotic promoter database-EPD, European Bioinformatics Institute, Hinxton, U.K.
 - 38 Smith, P. D., Davies, A., Crumpton, M. J. and Moss, S. E. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2713–2717
 - 39 Li, W. S. (1997) *Molecular Evolution*, Sinauer Associates, Sunderland, MA
 - 40 Herbert, A. and Rich, A. (1996) *J. Biol. Chem.* **271**, 11595–11598
-

Received 23 June 1998/23 September 1998; accepted 16 October 1998