Genomic organization and chromosomal localization of the mouse synexin gene

Zhen-Yong ZHANG-KECK,* Meera SRIVASTAVA,* Christine A. KOZAK,† Hung CAOHUY,* Anat SHIRVAN,* A. Lee BURNS* and Harvey B. POLLARD*‡

*Laboratory of Cell Biology and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A. and †Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

We have isolated and characterized the gene encoding mouse synexin, which consists of 14 exons and spans approximately 30 kbp of genomic DNA. The protein's unique N-terminal domain is encoded by six exons, and the C-terminal tetrad repeat, the site of the membrane-fusion and ion-channel domain, is encoded by seven exons. The first exon encodes the 5′-untranslated region. Analysis of synexin-gene expression in different mouse tissues shows that mRNA with exon 6 is only present in brain, heart and skeletal muscle. mRNA lacking exon 6 is expressed in all tissues we have examined. The initiation site for transcription was determined by primer-extension analysis and S1 nuclease mapping. Sequence analysis of the 1.3 kb 5′-flanking region revealed that the promoter has a TATA box located at position −25 and a number of potential promoter and regulatory elements. A CCAAT motif was not observed but CCATT is located in an appropriate position for the CCAAT motif upstream from the transcription-initiation start site. In addition, the 5′-flanking region contains two sets of palindromic sequences. Finally, we have determined that the functional synexin gene (*Anx7*) is located on mouse chromosome 14 and that a pseudogene (*Anx7-ps1*) is located on chromosome 10.

INTRODUCTION

Synexin [annexin VII (Creutz et al., 1978, 1979)] is a widely distributed member of the annexin gene family (Burns et al., 1989) which drives Ca^{2+}-dependent liposome fusion (Hong et al., 1981) and chromaffin granule membrane fusion (Stutzin, 1986; Nir et al., 1987). Synexin also exhibits gating currents (Rojas and Pollard, 1987) and Ca^{2+}-channel activity (Pollard and Rojas, 1988) in bilayer membranes, and it has been proposed to mediate membrane fusion during the process of exocytotic secretion (Pollard et al., 1991; Kuijpers et al., 1992). Synexin cDNAs have been isolated from a variety of organisms, including humans (Burns et al., 1989), *Dictyostelium* (Doring et al., 1991; von Gierke, 1991) and, most recently, mouse (Zhang-Keck et al., 1993).

In the case of human synexin, an exon-edited form of synexin occurs uniquely in brain, heart and skeletal muscle (Magendzo et al., 1991), and, as will be shown below, this is also true for mouse. In the case of human synexin, polymorphisms at the level of mRNA have been traced to multiple polyadenylation sites from a single synexin gene (Magendzo et al., 1991; Shirvan et al., 1994). However, an entirely different mechanism is responsible for the apparently similar mRNA polymorphism in the mouse, where an 800 bp difference occurs in the 3′-untranslated region, associated with two nucleotide changes in one of the forms but not the other. This raises important questions, such as whether the mouse might have more than one synexin gene and whether the mechanism of tissue-specific transcriptional regulation might be fundamentally different from that of the human.

These questions can only be analysed by acquiring a detailed knowledge of the genomic structure of the mouse synexin gene, and in the present study we have used a mouse genomic DNA library to delineate the complete organization of the mouse gene. We have also analysed the 5′-flanking regions as a first step towards understanding synexin-gene regulation. The data show that the mouse synexin gene occurs as a single copy, localized to chromosome (Chr) 14, and we interpret these data to indicate that the two cDNA variants noted in our previous study may be allelic forms. These data thus disprove a multigene hypothesis for messenger polymorphisms and observed mutations. In addition, we have discovered a cassette exon, in a form and location similar to that previously found in the human, to be the basis of tissue-specific expression of synexin in brain, heart and skeletal muscle.

MATERIALS AND METHODS

Isolation of genomic clones

A mouse NIH3T3 genomic library in the λFixII vector (10^6 total recombinants; Stratagene) was screened with random-primed ^{32}P-labelled (Feinberg and Vogelstein, 1989) full-length mouse synexin cDNA (Zhang-Keck et al., 1993). Five overlapping positive clones were selected and plaque-purified by sequential platings. The phage DNA was hybridized with ten different ^{32}P-labelled oligonucleotides corresponding to various regions of the cDNA. Four of five overlapping clones (G20, G18, G10 and G2) contained the entire synexin gene and approximately 100 bp extending in the 5′ direction. In order to obtain the clones corresponding to a further upstream 5′-flanking region, this 100 bp fragment was random-primed, ^{32}P-labelled and used as a probe to rescreen the same genomic library. One additional duplicate positive clone (G5′-13) was selected and plaque-purified.

Characterization of genomic clones

Restriction-endonuclease fragments from the G2 clone (*Eco*RI 7.0, 3.5, 2.5 and *Eco*RI–*Not*I 2.0 kb), the G10 clone (*Eco*RI 10,

Abbreviations used: Chr, chromosome; poly(A)$^+$, polyadenylated; RT-PCR, reverse transcription PCR.

‡ To whom correspondence should be addressed.

3.6 and 2.4 kb), the G18 clone (*Eco*RI 4.0 and *Eco*RI–*Not*I 7.6 kb), the G20 clone (*Bam*HI 4.0, 3.0 and *Bam*HI–*Not*I 6.0 kb) and G5′-13 (*Bam*HI 6.0 kb) were subcloned into the pBluescript KS(+) vector (Stratagene). These subclones were sequenced using synthetic oligonucleotides (18-mers) corresponding to known exons or the T7 and T3 primers of pBluescript KS(+). All exons, the exon/intron junctions, the 5′-flanking region and adjacent parts of the introns were sequenced on both strands by the dideoxy-chain-termination method using Sequenase (U.S. Biochemical Corp.). Oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizer model 380B. DNA sequences were compiled by using the pcgene program (Intelligenetics). The 5′-flanking sequence was analysed for potential regulatory elements using the 'Dynamic' computer program (Smith and Waterman, 1981).

Intron size was determined by PCR. The reactions were performed in the amplification mixture recommended by the manufacturer (Perkin–Elmer Cetus) using the appropriate subclones as templates. The primer pairs used contained sequences from opposite strands of two neighbouring exons. The PCR products were analysed on a 1.0% agarose gel along with molecular-mass standards. Each intron was confirmed by either Southern-blot hybridization with oligonucleotides derived from a known sequence located between those corresponding to the primers used in the PCR reactions, or sequencing using the same PCR primers until the exon/intron junctions were reached.

Analysis of synexin-gene expression in different tissues

Poly(A)$^+$ RNA (1 μg) isolated from mouse pancreas, lung, heart, spleen, brain, liver, skeletal muscle, testes or kidney (Clontech) was used as template for reverse transcription of RNA followed by PCR (RT-PCR) according to the manufacturer's manual. The positions of the exon-3- and exon-7-specific 5′ and 3′ primers in relation to the mRNA are shown in Figure 2(c). The sizes of the RT-PCR products were analysed on 2% agarose gels. The amplified fragments from each tissue were cloned into the PCR 1000 vector (Invitrogen) and sequenced using T7 and M13 forward primers of the vector.

Gel electrophoresis was followed by transfer to nitrocellulose paper, denaturation of the DNA and hybridization with ^{32}P-labelled exon 5b.

Northern-blot analysis was carried out as follows: 5 μg of mouse liver poly(A)$^+$ RNA was transferred from 1.2% agarose–formaldehyde gels to a Nylon membrane (Clontech) and probed first with a random-primed ^{32}P-labelled mouse synexin cDNA coding region. After removal of the previous probe, the blot was rehybridized with ^{32}P-labelled exon 6.

Synexin was partially purified from heart, brain, spleen, lung, liver, skeletal muscle and kidney from 20 female mice by precipitating twice with 20% $(NH_4)_2SO_4$ (Creutz et al., 1978). The protein samples were separated by SDS/PAGE (10% gels) and then transblotted to nitrocellulose. Blots were allowed to react with a rabbit antibody against a peptide common to both human and mouse synexin, and detected with anti-rabbit IgG–horseradish peroxidase (Vector Laboratories).

Determination of the transcription-initiation site

The 5′ end of synexin mRNA from mouse liver was defined by primer-extension analysis and S1 nuclease mapping. Primer extension was performed using 1×10^6 c.p.m. of a ^{32}P-end-labelled 32-base oligonucleotide (5′-GCGCGGCTCCACCCAAAGCTGCAGTTCCTTTC-3′; nucleotide positions 162–193) of the sense strand complementary to synexin cDNA. Poly(A)$^+$ RNA from mouse liver (1 μg in 1 mM Tris/HCl, pH 7.5, 0.1 mM EDTA) was heated at 70 °C for 10 min in the presence of the primer. The solution was cooled immediately to 4 °C and then adjusted to 50 mM Tris/HCl, pH 7.5, containing 40 mM KCl, 6 mM $MgCl_2$, 1 mM dithiothreitol and 10 mM dNTPs. The primer was extended with Superscript RNaseH reverse transcriptase (BRL) for 1 h at 42 °C. The size of the primer-extended products was determined on a sequencing gel.

S1 nuclease mapping was carried out to confirm the transcription-initiation site of synexin mRNA determined by primer extension. A 420-base fragment from the 5′ end of the gene (nucleotide −237 to 183) was used to synthesize an anti-sense RNA probe using T7 polymerase, [α-^{32}P]UTP and an RNA transcription kit (Stratagene). The riboprobe (1×10^6 c.p.m.) was hybridized overnight at 42 °C with 40 μg of total mouse liver RNA (in 40 mM Pipes, pH 6.4, containing 1 mM EDTA, 400 mM NaCl and 80% formamide). The riboprobe was then digested with 400 units of S1 nuclease (BRL). The size of the riboprobe was determined on a sequencing gel.

Chromosomal localization

Chinese hamster × mouse somatic-cell hybrids were produced and characterized as described previously (Hoggan et al., 1988). For this study 18 hybrids were selected from a larger panel of 76. Ten were karyotyped after Giemsa/trypsin staining and eight were typed for markers on specific mouse chromosomes.

Two multilocus crosses were produced as described previously: (NFS/N or C58/J × *Mus musculus musculus*) × *M. m. musculus* (Kozak et al., 1990) and (NFS/N × *M. Mus spretus*) × C58/J or *M. spretus* (Adamson et al., 1991). DNA extracted from the progeny of these crosses has been typed for over 400 markers including the Chr-10 marker *Myb* (myeloblastosislastosis oncogene) and the Chr-14 markers *Ccnb1-rs5* (cyclin B1-related sequence 5), *Cchl1a2* (Ca^{2+} channel, L type, A1 subunit) and *Glud* (glutamate dehydrogenase). *Myb* was typed in the *M. spretus* cross after *Bam*HI digestion using as probe a 2.2 kb *Eco*RI fragment of pG4M2b (Lavu and Reddy, 1986). *Ccnb1-rs5* was typed as described previously (Hanley-Hyde et al., 1992), and *Glud* and *Cchl1a2* were typed in the *M. m. musculus* cross as described previously (Tzimagiorgis et al., 1991; Chin et al., 1991) and in the *M. spretus* cross after digestion with *Apa*I and *Eco*RI respectively. Recombination fractions and standard errors were determined as described by Green (1981).

RESULTS

Structure of the mouse synexin gene

A mouse NIH3T3 genomic library constructed in λFixII was probed with the full-length mouse synexin cDNA coding region and the 5′-untranslated region separately. Five positive overlapping clones G5′-13, G20, G2, G10 and G18 (Figure 1a) were analysed by restriction-endonuclease mapping and Southern-blot hybridization using a series of oligonucleotides that encompass the entire coding region of mouse synexin. *Eco*RI and *Bam*HI fragments, as indicated in Figure 1(a), were subcloned into pBluescript and used for DNA-sequence analysis and as a template for PCR.

The nucleotide sequence and exon/intron boundaries of the mouse synexin gene were determined by direct sequencing of the subclones using a series of 18-base oligonucleotides as primers. Exon sequences and the polyadenylation site were identified by comparing the genomic sequence with that of mouse synexin

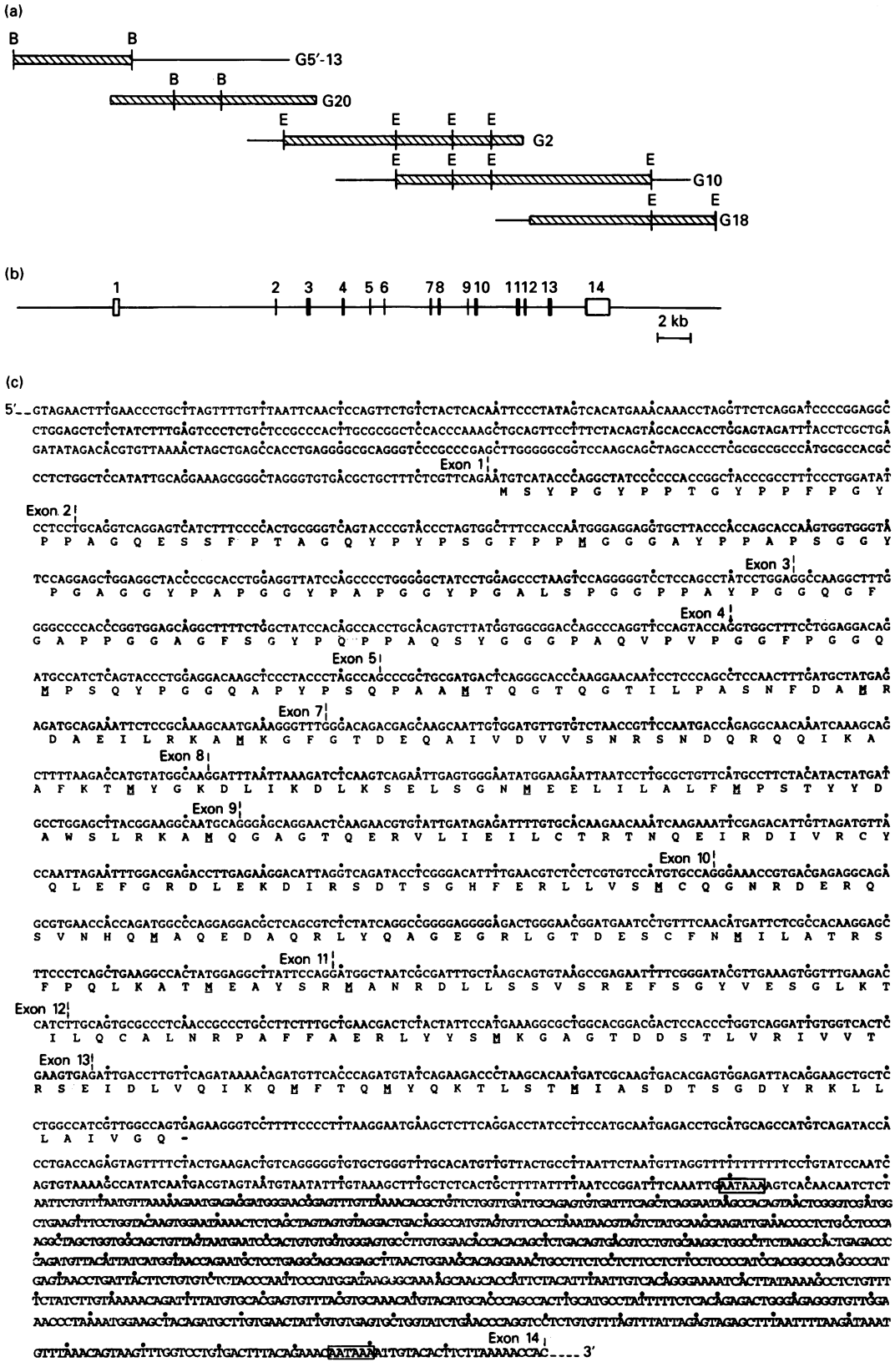

Figure 1 Physical map of the synexin gene and the nucleotide and predicted amino acid sequence of synexin cDNA

(**a**) Representative λ phage clones encompassing the synexin gene are shown as hatched bars and thin lines. The hatched bar indicates the region sequenced. Restriction sites *Bam*HI (B) and *Eco*RI (E) are drawn on the λ phage clones. (**b**) Schematic representation of the synexin gene. The direction of transcription is from left to right. Filled or open boxes represent exon or 5′ and 3′ non-coding sequences respectively. Intron and flanking regions are shown by a thin connecting line. (**c**) Nucleotide and deduced amino acid sequence of the mouse synexin cDNA. Dashed lines indicate the location of introns. Boxed sequences show the polyadenylation endonucleolytic cleavage signals.

Table 1 Summary of the exon/intron organization of the mouse synexin gene

Adjacent exon (upper-case letters) and intron (lower-case letters) sequences are given for each junction. The numbers of amino acids bordering the splice junctions indicate the positions of the corresponding amino acids in the synexin gene deduced from the mouse synexin cDNA. The exons and introns are numbered beginning from the 5′ end of the gene.

Exon		Sequence at exon/intron junction		Intron	
No.	Size (bp)	5′ Splice donor	3′ Splice acceptor	No.	Size (kb)
1	389	GTTCAG/gtgaggact.....	.....tttgtttgtttgtcag/AATGTC MetSe	1	~ 10
2	58 (161–218)	CCTCCT/gtaagtatt.....	.....tttctaatttttccag/GCAGGT ProPro 19	2	2.0
3	202 (219–420)	CTGGAG/gtaagttag.....	.....tctaaaatctctccag/GCCAAG roGlyG 87	3	2.3
4	102 (421–522)	TACCAG/gtacgtccc.....	.....tgctttcaacatacag/GTGGCT alProG 121	4	1.8
5	65 (523–587)	AGCCAG/gtacttttc.....	.....tccactcttctttcag/ATCAAT SerGln 142	5	0.9
6	66	AGTGAT/gtacattag.....	.....ttttgttgtacatcag/CCCGCT SerAsp (66)	6	2.9
7	103 (588–690)	GGTTTG/gtaagaaa.....	..... atctatgatcttccag/GGACAG lyPheG 177	7	0.3
8	95 (691–785)	GGCAAG/gtatgcttt.....	.....atagtcattgatatag/GATTTA GlyLys 208	8	1.7
9	114 (786–899)	ATGCAG/gtaggctat.....	.....ctctcgtttcttgtag/GGAGCA MetGln 246	9	0.4
10	171 (900–1070)	TGCCAG/gtgagtata.....	.....tgccttcctcttgcag/GGAAAC CysGln 303	10	2.4
11	171 (1071–1240)	TCCAGG/gtgggtttg.....	.....gacattttgtttgcag/ATGGCT SerArg 360	11	0.29
12	76 (1241–1317)	CCATCT/gtaaggcac.....	.....catatctgccttgcag/TGCAGT hrIleL 386	12	1.5
13	113 (1318–1430)	AGTGAG/gtgaggcaa.....	.....tcctgatatttcttag/ATTGAC SerGlu 423	13	2.2
14	1353 Stop codon				
	Consensus	C a AG/gt agt..... A g	c yyyyyyyyyyn ag/GT t		

cDNA. The sizes of the introns were obtained after PCR amplification of each intron and parts of the flanking exons followed by size determination on agarose gels. The sequences of each pair of primers were complementary and identical with a sequence located 200 bp apart within mouse synexin cDNA. The PCR products were sequenced in order to obtain further confirmation of the exon/intron boundaries.

As shown in Table 1, all splice donor and acceptor sites conform to the GT-AG rule and agree with the consensus sequence compiled for the exon/intron boundaries of other genes. The surrounding nucleotides are closely related to consensus sequences found near splice junctions (Mount, 1982). The intron splice phase is type 0 (the intron occurs between codons) for introns 2, 5, 6, 8, 9, 10, 11 and 13, type I (the intron interrupts the first and second bases of the codon) for introns 3, 4, 7 and 12 (Rogers, 1985).

A schematic map (Figure 1b) and the detailed exon/intron organization of the synexin gene (Table 1) show that the gene spans approximately 30 kb of genomic DNA consisting of 28 kb of intervening sequences and 1.4 kb of coding sequences. The gene is interrupted by 13 introns ranging in size from 0.29 to 10 kb, and the lengths of the 14 exons vary from 58 (exon 2) to 202 (exon 3) bp.

We have reported previously that the isolation of two sets of mouse synexin cDNA clones corresponds to two forms of mRNA present *in vivo*. Northern blots had shown that they are different mainly in the 3′-untranslated region (Zhang-Keck et al., 1993). Here, two polyadenylation sites are found in the mouse synexin gene and the sequence is identical with the two sets of cDNA clones, the shorter 3′-untranslated region (300 nucleotides) and the longer one (800 nucleotides) corresponding to each form of mRNA. These data indicate that the two forms of synexin

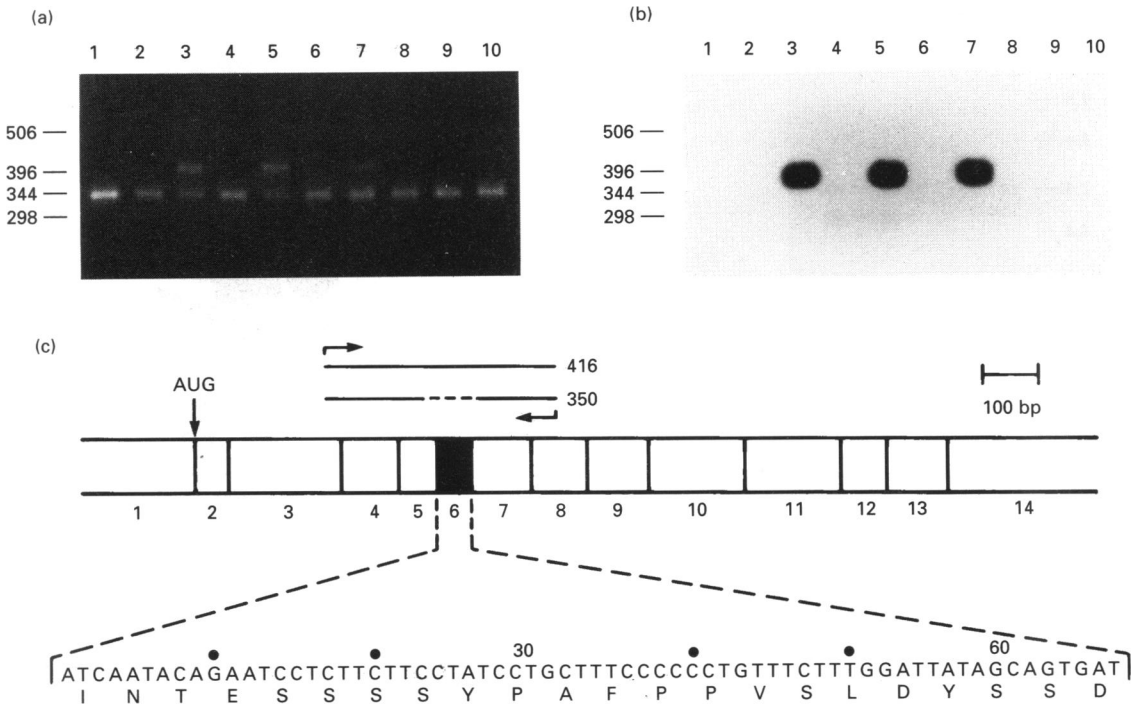

Figure 2 Southern-blot analysis of RT-PCR-amplified mouse mRNA

Poly(A)$^+$ RNA from the various mouse tissues was reverse-transcribed, PCR-amplified and separated on a 2.0% agarose gel followed by blotting on to nitrocellulose paper (see the Materials and methods section). (**a**) Agarose-gel electrophoresis of RT-PCR-amplified mouse mRNA. (**b**) Southern blot of the same gel probed with ^{32}P-labelled exon 6 (shown on diagram **c**). Lanes: 1, pancreas; 2, lung; 3, heart; 4, spleen; 5, brain; 6, liver; 7, muscle; 8, testes; 9, kidney; 10, cDNA. (**c**) Diagram of synexin cDNA indicating the location (solid block), nucleotides and predicted amino acid sequence of exon 6. The numbers represent the 14 exons. The expected length of the RT-PCR with or without the exon 6 insertion is shown by two lines on the top (416 and 350).

mRNA are transcribed by selection of alternative polyadenylation signals.

Furthermore the two sets of synexin cDNA differ from each other at two nucleotides in the open reading frame of the N-terminal domain and the second repeat of the C-terminal domain. At amino acid position 145, the codon TCG (encoding serine) is found in set-II cDNA, whereas the codon GCG (encoding alanine) is found in set-I cDNA and in the synexin gene. In set-I cDNA at amino acid 304 and in the synexin gene, a GGA codon (encoding glycine) is found, whereas set-II cDNA bears GCA (encoding alanine) at this position. These data indicate that set-I cDNA is identical with the synexin gene. The differences between the gene sequence and the set-II cDNA sequences may represent allelic differences among heterozygotes (Winkel et al., 1991).

Analysis of the amino acid sequence of synexin predicts two principal domains: an N-terminal domain consisting of the first 164 amino acids and a C-terminal domain consisting of the remainder. The C-terminus comprises four repeats, and each repeat is approximately 70–80 amino acids in length. A projection of the introns on the predicted synexin amino acid sequence is shown in Figure 1(c). Exon 1 contains the entire 5′-untranslated region of the mRNA. The N-terminal domain is encoded by sequences contained within six exons (2, 3, 4, 5, 6 and 7) and the C-terminal domain is encoded by the last seven exons and by 19 amino acids in exon 7. The four repeats in the C-terminus are not encoded in separate exons. Exon 14 encodes the 40 amino acids at the C-terminus of the protein, the translation stop codon and the 3′-untranslated region of the mRNA. These data suggest that there is no strong correlation between the exon/intron arrangement of the synexin gene and the predicted structural domains of the proteins.

Tissue-specific expression of mouse synexin gene

Using oligonucleotides hybridizing to exons 3 and 8 sequences, RT-PCR products from poly(A)$^+$ RNA were generated from a variety of mouse tissues and these were analysed on 2% agarose gel (Figure 2a). Two bands were obtained. The smaller bands (350 bp) are found in pancreas, lung, heart, spleen, brain, liver, muscle, testes and kidney and the larger bands (416 bp) are present in heart, brain and skeletal muscle. Sequence analysis of these two bands revealed that the 350 bp band has exactly the same sequence as the cDNA in the corresponding region, whereas the 416 bp band contains an extra 66 bp in-frame insertion. The insertion sequence is shown in Figure 2(c). To confirm this observation further the gel was transblotted and hybridized to the 66 bp oligonucleotide (Figure 2b). As expected, the 66 bp is specific only to heart (lane 3), brain (lane 5) and muscle (lane 7). To determine if the extra 66 bp sequence exists in mature mRNA and is not the result of amplified unprocessed heterogeneous nuclear RNA produced by RT-PCR, the Northern blot was hybridized to the mouse cDNA coding region and an oligonucleotide for the 66-base sequence (Figure 3). Two types of mouse synexin mRNA differing in the 3′-non-coding region (Zhang-Keck et al., 1993) were expressed in all tissues examined (Figure 3a). The hybridized bands of mRNA in heart, brain and skeletal muscle migrate slightly faster than those in spleen, lung, liver, kidney and testes. The same blot was stripped and reprobed with the labelled 66-base oligonucleotide. As shown in Figure

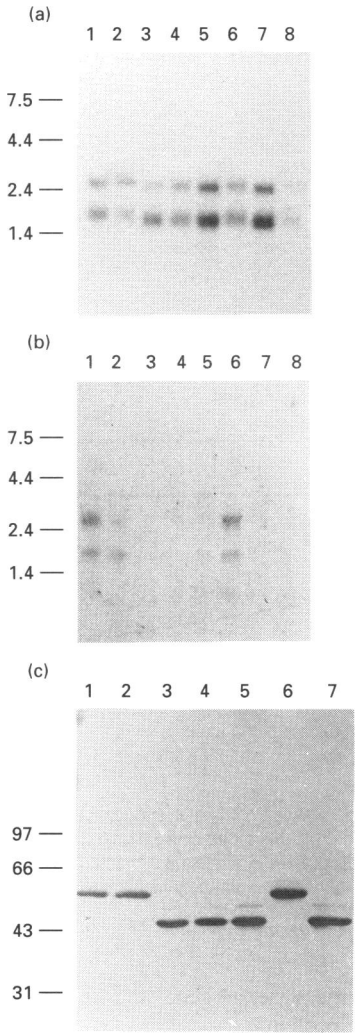

Figure 3 Northern- and Western-blot analysis

(**a**) Northern-blot analysis of mouse poly(A)$^+$ RNA from various tissues. The blot was hybridized to the mouse synexin coding region. (**b**) The same blot was hybridized to labelled exon 6 after stripping. The size markers on the left are in kb. (**c**) Western-blot analysis of partially purified synexin from various mouse tissues. The numbers (kDa) indicate the positions of molecular-mass markers. Lanes: 1, heart; 2, brain; 3, spleen; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, testis.

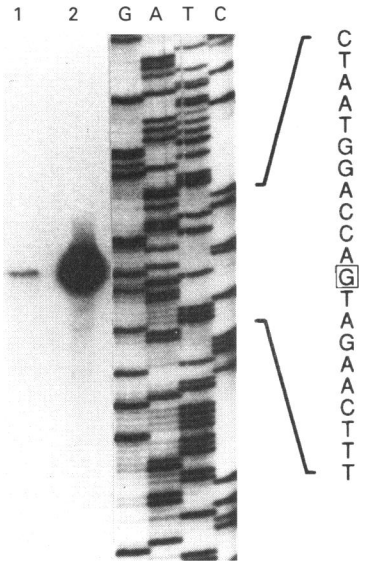

Figure 4 Mapping of the synexin-gene transcription-start site by S1 nuclease protection and primer extension

Mouse liver total RNA (20 μg) was annealed to a ^{32}P-labelled antisense RNA probe (prepared by T7 polymerase transcription) and digested with S1 nuclease as described in the Materials and methods section. The same oligonucleotide served as a primer for reverse transcriptase using 1 μg of mouse liver poly(A)$^+$ RNA. Lane 1, extension product; lane 2, protected fragment obtained after S1 nuclease treatment. GATC lanes are the sequence ladder derived from a fragment of mouse synexin genomic DNA (containing 5′-flanking sequences) using the same oligonucleotide primer. Brackets (right) enclose a portion of the sequencing ladder. The boxed guanosine residue indicates the transcription-start point.

3(b), only the RNA bands in heart, brain and skeletal muscle are detected. These results indicate that there are two isoforms of synexin RNA which are expressed in a tissue-specific manner regardless of which polyadenylation site is being used.

A mini-exon, which we termed exon 6, corresponding to the 66-base insertion was found in the synexin gene (Table 1). Exon 6 is 66 bp in size and encodes 22 amino acids which are inserted between Gln-142 and Pro-143. The nucleotide sequence of exon 6 and the deduced amino acid sequence are shown in Figure 2(c). Tissue-specific expression of mouse synexin isoforms were also determined by Western-blot analysis (Figure 3c). As expected, the protein bands from heart, brain and skeletal muscle are larger than those isolated from spleen, lung, liver and kidney. The size difference of approximately 2000 Da is consistent with the difference in molecular mass predicted from the proteins encoded by the two isoforms of synexin mRNA. These data suggest that alternative splicing, resulting in the addition or elimination of this exon, might play an important role in the physiological function of synexin in different tissues.

We next compared exon 6 of the mouse synexin gene with that of the human synexin gene (Magendzo et al., 1991). The position of the exon shows an identical arrangement. The amino acid sequence of mouse exon 6 is different from that of human exon 6 in five amino acid residues. Mouse exon 6 has E, S, A, P and D at residues 4, 6, 11, 13 and 22 respectively instead of D, F, V, S and E. The nucleotide sequences of mouse and human exon 6 share 89% homology.

Identification of the transcription-initiation site and sequence analysis of the 5′-flanking region

The initiation site for transcription was determined by S1 nuclease protection and primer-extension experiments (Figure 4). For S1 nuclease protection, 20 μg of mouse liver total RNA was hybridized with a ^{32}P-labelled antisense RNA probe (prepared by T7 polymerase transcription to 420 bases upstream from the primer shown in Figure 5) and digested with S1 nuclease. When compared with the sequencing reaction obtained by using DNA from the same region (position −237 to 183) and the same oligonucleotide primer, a single major protected band ended at G, 388 nucleotides upstream from the translation-initiation ATG codon (Figure 4, lane 2). To map the transcription-initiation site more precisely, primer-extension analysis of mouse liver poly(A)$^+$ RNA was performed. The same 32-base oligonucleotide primer located between position 151 and 183 upstream of the initiation codon ATG as in S1 nuclease protection was used. The size of the reverse-transcribed fragment was consistent with the S1 nuclease protection band (Figure 4, lane 1).

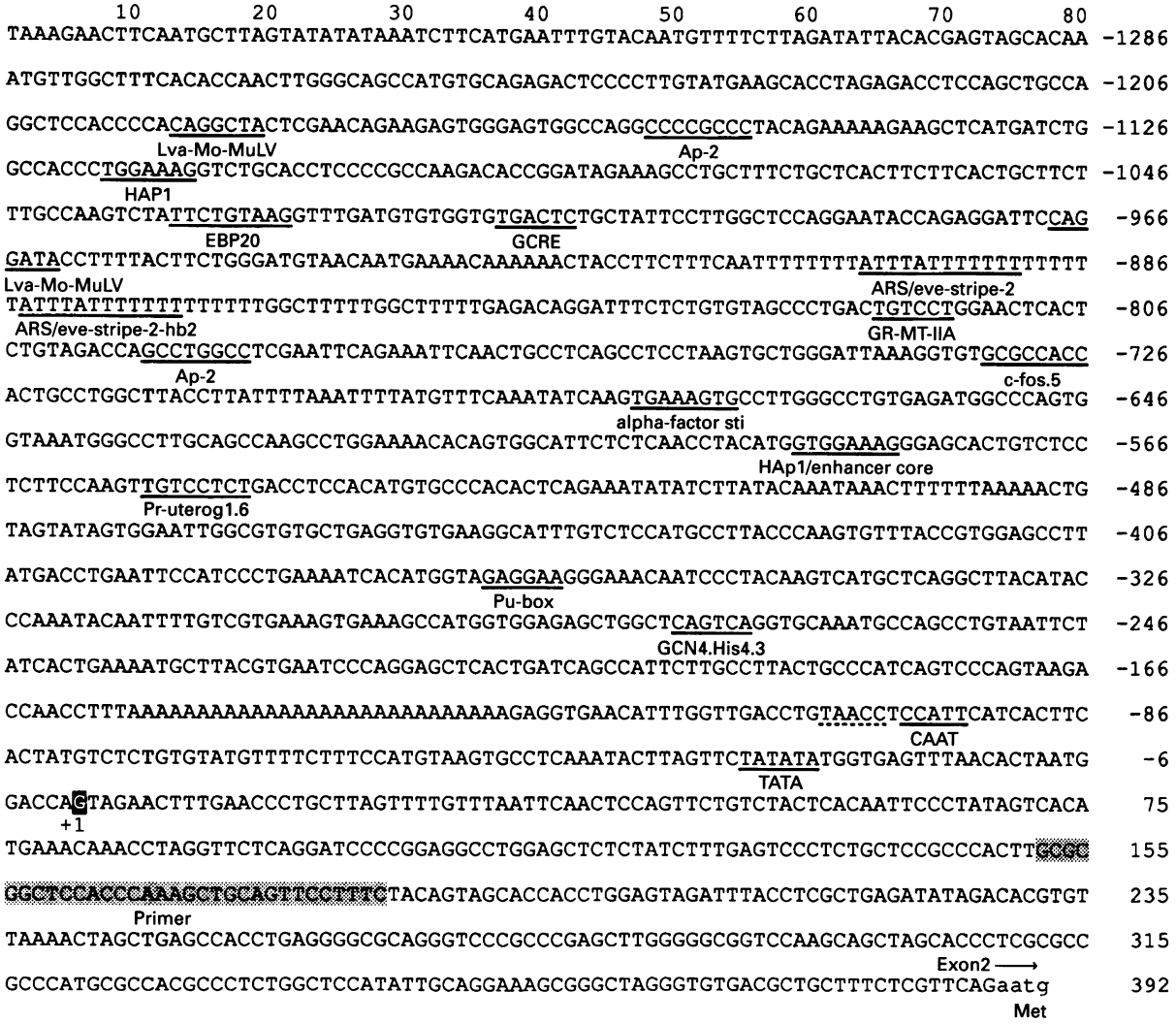

Figure 5 Nucleotide sequence of the 5′-flanking region of the mouse synexin gene

The transcription-initiation site is indicated by reverse contrast and numbered as +1. Nucleotide numbers relative to the transcription-initiation site are shown on the right. Sequences complementary to the primer used for S1 nuclease and primer-extension analysis are shaded. Nucleotides occurring in exon 1 and 5′ untranslated region are shown in upper-case letters, those occurring in exon 2 are shown in lower-case and amino acids encoded by exon 2 are shown below the nucleotide sequence. Potential binding sites for TATA, CCAAT, enhancers (Ap-2, HAP-1, GCN4.His4.3, Lva-Mo-MuLV, EBP20, c-fos.5 and enhancer core sequence), repressor (alpha-factor sti) factor, the progesterone response element (Pr-uterog 1.6), the heavy-metal-ion-binding site (GR-MT-IIA), tissue-specific binding factor (Pu-box), glucocorticoid responsive element (GCRE), and ARS/eve-stripe-2-hb2 elements are underlined and labelled. A CCAAT-binding transcription factor (in the reverse orientation TAACC) is shown with a dashed line underneath.

The sequence of 1.3 kb of the 5′-flanking region was determined (Figure 5) and analysed by the 'Dynamic' computer program (Smith and Waterman, 1981) to identify potential regulatory elements. In the immediate transcription-initiation-site region there are two sequences that might serve as transcription-factor-binding sites. TATA sequences (Breathnach and Chambon, 1981) are located 25 bases upstream of the transcription-start site. A CCATT and a potential reversed sequence TAACC are located at −90. Although, on the basis of the reported consensus CCAAT, the sequence is not ideal (Breathnach and Chambon, 1981), the location relative to the transcription-initiation sites and to the TATA sequence suggests it may be functional.

In addition to the potential regulatory elements found near the transcription-initiation site, the 5′-flanking sequence contains a number of potential transcription-factor- and hormone-receptor-binding sites. Potential binding sites exhibiting a perfect match to known consensus sequences include two ARS WTTTAYRTTTW (Houten and Newlon, 1990), two eve-stripe-2-hb2 TTATTTTTTT (Stanojevic et al., 1991), two GCRE TGACTC (Hope and Struhl, 1985), CAGTCA (Arndt and Fink, 1986), two Ap2 GSSWGSCC (Mitchell et al., 1987), two Lva-Mo-MuLv CAGGATA (Speck and Baltimore, 1987), two HAP1 TGGAAAG (Pfeifer et al., 1987), one EBP20 TKNNGYAAK (Johnson et al., 1987), one alpha-factor sti (Kronstad et al., 1987), one GR-MT-IIA TGTCCT (Karin et al., 1984), one c-fos.5 GCGCCACC (Fisch et al., 1987), one enhancer-core GTGGWWWG (Weiher et al., 1983), one Pr-uterog 1.6 TGTCCTCT (Bailly et al., 1986) and one Pu-box GAGGAA (Klemsz et al., 1990). The imperfect sequences closely resembling the reported binding sites have not been listed in Figure 5. However, they might have a potential regulatory function, as it has been shown that many binding sites deviate to a certain extent from reported consensus sequences.

The most interesting features of the 5′-flanking region include

a repeat $[AT_3AT_{13}/AT_3AT_{13}]$ containing the palindrome $[T_{13}AT_3AT_{13}]$ and a repeat consisting of $GGCT_5/GGCT_5$, both located about 850 bp upstream of the transcription-initiation site. The first AT palindromic sequence could be either for two ARS-binding sites, two eve-stripe-2-hb2 binding sites or for overlapping above four binding sites. Yet, none of these binding sites are large enough to explain the full-length palindromic sequence. There is no known transcription-factor-binding site that matches the second $GGCT_5$ repeat sequence.

Chromosomal localization of the mouse synexin gene

Eighteen hamster × mouse somatic-cell hybrids were typed after *Pst*I digestion for the presence or absence of mouse synexin using the cDNA coding region as a probe. *Pst*I digestion produced fragments of 13, 8.1 and 1.7 kb in mouse DNA (Figure 6a, lane 4). All three fragments were present together in three hybrids and absent together in 15 hybrids. This pattern showed a perfect correlation with the presence or absence of Chr-14 in these hybrids (results not shown) indicating that all mouse synexin-related sequences map to Chr-14. There were at least three discordant hybrids for all other chromosomes.

In order to position the synexin fragments on Chr-14, the progeny of two multilocus crosses were typed for restriction-enzyme polymorphisms of synexin using synexin cDNA as probe. In the first cross, *Pst*I digestion produced fragments of 13.0, 8.1 and 1.7 kb in parental NFS/N DNA (Figure 6b, lane 3) and fragments of 8.0 and 1.7 in *M. m. musculus* (Figure 6b, lane 1). The 13 kb *Pst*I fragment was present in 57 of 119 backcross progeny consistent with single-gene segregation. Comparison of the pattern of inheritance of this fragment with markers previously mapped in this cross confirmed Chr-14 linkage and positioned the gene for synexin, *Anx7* (annexin VII), in the centromeric region of this chromosome (Table 2).

A second cross was typed for a *Bam*HI polymorphism of synexin. *Bam*HI produced a single fragment of 18.8 kb in NFS/N mice (Figure 6c, lane 5) and two fragments of 19.2 and 10.8 kb in *M. spretus* mice (Figure 6c, lane 1). The 19.2 kb *M. spretus*

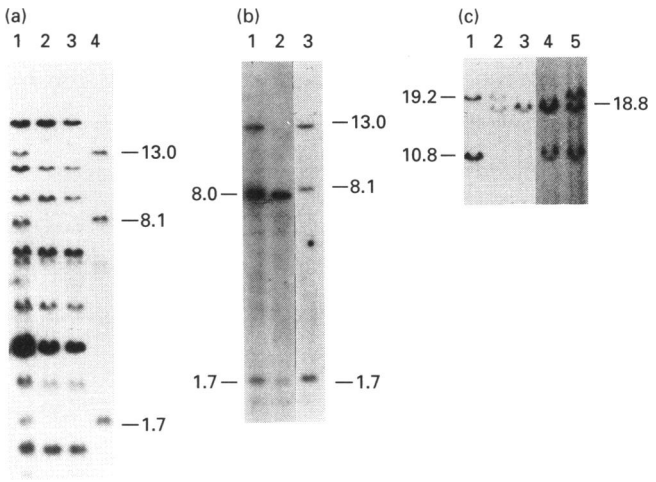

Figure 6 Southern-hybridization analysis of mouse somatic-cell hybrids and the progeny of two multilocus crosses using a probe derived from the coding region of mouse synexin cDNA

*Pst*I-digested DNA in (**a**) is from somatic-cell hybrids (lanes 1–3) and mouse (lane 4). *Pst*I-digested DNA in (**b**) and *Bam*HI-digested DNA in (**c**) are from the progeny of two multilocus crosses. (**b**) Lanes 1 and 2, progeny of cross (NFS × *M. m. musculus*) × *M. m. musculus*; lane 3, NFS/N. (**c**) Lane 1, *M. spretus* mice; lanes 2–5, progeny of the cross (NFS × *M. spretus*) × C58/J.

fragment and the 18.8 kb NFS/N fragment were typed in progeny derived from (NFS/N × *M. spretus*) F1 females crossed to C58/J and *M. spretus* respectively, and also identified the Chr-14 *Anx7* locus (Table 2). The data from the two crosses taken together provide the following gene order and distances in cM: *Ccnb1-rs5*–0.9–*Anx7*–4.2–*Cchl1a2*–5.7–*Glud*. The additional 10.8 kb *Bam*HI *M. spretus* fragment was typed in progeny of the F1 crossed to C58/J and was mapped to Chr-10; no recombinants were detected with *Myb* in 56 mice indicating that, at the upper

Table 2 Inheritance of the *Anx7* restriction fragments with *Ccnb1-rs5*, *Cchl1a2* and *Glud* in two multilocus crosses

Mice heterozygous (+) or homozygous (−) for parental alleles are indicated. Results for percentage recombination are means ± S.E.M.

	Inheritance of parental type					
	Ccnb1-rs5	*Anx7*	*Cchl1a2*	*Glud*	Cross* 1	Cross† 2
Non-recombinants	+	+	+	+	47	45
	−	−	−	−	57	39
Single recombinants	+	+	+	−	7	1
	−	−	−	+	2	2
	+	+	−	−	3	2
	−	−	+	+	1	3
	+	−	−	−	0	0
	−	+	+	+	1	1
Double recombinants	−	−	+	−	1	0

Recombination	Locus pair	Recombinants/total	Recombination (%)
	Ccnb1-rs5, Anx7	2/212	0.9 ± 0.6
	Anx7, Cchl1a2	9/212	4.2 ± 1.4
	Cchl1a2, Glud	12/212	5.7 ± 1.6

* Cross 1, (NFS/N or C58/J × *M. m. musculus*) × *M. m. musculus*.
† Cross 2, (NFS/N × *M. spretus*) × C58/J or *M. spretus*.

limit of the 95% confidence interval, these genes are within 5.2 cM. These data suggest that the synexin copy on Chr-10 represents a species-specific pseudogene, designated *Anx7-ps1*.

DISCUSSION

In this study we present the genomic cloning, characterization, genetic mapping and expression of the mouse synexin gene. The gene spans approximately 30 kb and is split into 14 exons. The coding sequence of the gene is approximately 1.3 kb, representing 4.3% of the gene, which is similar to 4.5% of the murine calpactin I heavy-chain gene [*Anx2* (Arrignet et al., 1990)]. In the annexin family, it has been reported that the human, rat and mouse lipocortin I genes have an identical intron/exon arrangement (Kovacic et al., 1991; Horlick et al., 1991) and that the structures of the murine and human calpactin I heavy-chain genes [annexin II, *Anx2* (Arrignet et al., 1990; Spano et al., 1990)] are identical except for a shift of one codon in the location of the 12th intron. In addition, a comparison between the lipocortin I (*Anx1*) and calpactin I (*Anx2*) genes shows that the intron/exon structures are conserved, except for a slight shift in introns 2 and 7, and 12 as indicated above. Mouse and human synexin have an identical intron/exon arrangement (Shirvan et al., 1994), but synexin differs substantially from lipocortin I and calpactin I. As shown in Figure 7, the much larger unique N-terminal domain of synexin is encoded by exons 2, 3, 4, 5, 6 and part of 7. By contrast, the much shorter unique N-terminal domains of the lipocortin and calpactin I genes are encoded only by exon 2 and part of 3. With respect to the C-terminal domain, which is the common structure for this gene family, there are three classes of relationships for splice junctions. One class occurs in the repeat I and II. The positions of introns 7, 8, 9 and 10 in the synexin gene are virtually identical with introns 3, 4, 5 and 7 in the lipocortin I and calpactin I genes, except for one codon shift in intron 7 between lipocortin I and calpactin I. These identical intron locations among the annexin gene family members may suggest that the sequence for this portion of the protein originated from the same ancestral intron-containing gene. The second class occurs in repeats III and IV. The positions of introns 11, 12 and 13 in the synexin gene are shifted by 7, 1 and 23 codons respectively relative to introns 9, 10 and 12 in lipocortin I and calpactin I genes. The third class is represented by an intron insertion in repeats II, III and IV. These are the locations of introns 6, 8 and 11 in the calpactin and lipocortin genes, but they are absent from the synexin gene. Synexin introns 7–13 do not fall systematically between domains and there is no striking correlation between the boundaries of the exons and the boundaries of the repeats I–IV. It is therefore apparent that annexin gene family members are not only internally divergent, but that otherwise functional units at the gene product level are not encoded in separate exons, as is the case for the immunoglobulin gene family (Tonegawa, 1983). It has been hypothesized that lipocortin genes arose by duplication of a single domain (Kovacic et al., 1991), but it is clear that the duplication process may include divergent remodelling of this ancestral gene with subsequent deletion and insertion of introns.

The present study demonstrates that synexin mRNA is expressed in a wide variety of mouse tissues and that the mRNA that contains 66 additional bases is expressed specifically in the brain, heart and skeletal muscle. This is true regardless of which polyadenylation site is being used (Figures 2 and 3). The sequence of the 66-base insertion was further located in the mouse synexin gene as exon 6. These data strongly suggest that the tissue-specific larger form of synexin mRNA is expressed by an alternative splicing process, as is also the case for human synexin (Magendzo et al., 1991). The location of exon 6 is between Gln-142 and Pro-143, which is within the highly hydrophobic N-terminal domain of synexin. The insertion of the 22 amino acids encoded by the 66 bp of exon 6 may be important for the structure of the N-terminus, because it inserts three acidic residues (one E and two Ds) into a highly hydrophobic domain. It follows that it is possible that regulation of synexin protein activity is achieved by alternative splicing to produce two proteins with different N-terminal structures. The region of the insertion has been identified in other annexins as a hinge region with functional importance, such as the site of phosphorylation by epidermal growth factor receptor kinase for lipocortin I (De et al., 1986) and the site of p11 binding for calpactin I heavy chain (Saris et al., 1986). As yet, however, the physiological function of exon 6 is unclear.

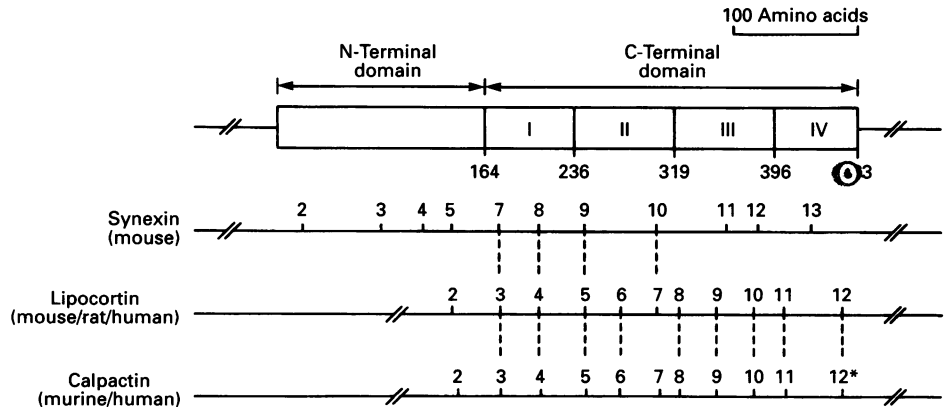

Figure 7 Comparison of locations of exon/intron splice sites for the annexin family

The locations of intron/exon splice sites are compared with the domain structure of gene products for the mouse synexin, rat lipocortin, human lipocortin, murine calpactin and human calpactin. The four individual repeats are indicated by roman numerals. Residue numbers are shown at the beginning of each domain of mouse synexin, except for residue 1. The intron/exon splice sites are indicated by vertical lines and are numbered. A dashed line indicates that the annexins have virtually identical intron/exon splice site at the same location. The asterisk indicates the location of the difference between murine and human calpactin.

In the more distal regions of the 5′-flanking sequence, there are a number of transcription-factor-binding sites which may play a regulatory role. These potential binding sites could be classified as follows. Potential enhancer-factor-binding sequences include Ap-2, HAP-1, GCN4.His4.3, Lva-Mo-MuLV, EBP20 and the enhancer core sequences. The potential repressor-factor-binding sequences include alpha-factor sti. There exist a potential hormone-receptor-binding site for progesterone, a potential heavy-metal-ion-binding site (GR-MT-IIA) for cadmium, a potential enhancer sequence (c-fos.5) for epidermal growth factor and a potential activation sequence (Pu-box), which binds a tissue-specific DNA-binding factor.

A striking feature of the synexin 5′-flanking sequence is a region 850 bp upstream of the transcription-initiation site that contains a repeated sequence, AT_3AT_{13}/AT_3AT_{13}, with an internal palindrome, $T_{13}AT_3AT_{13}$, and a repeated sequence, $GGCT_5/GGCT_5$. Analysis by the computer program 'Dynamic' showed that two ARS (WTTTAYRTTTW) and two eve-stripe-2-hb2 (TTATTTTTTT) match part of the AT-rich palindromic sequences. There is no known regulatory-factor-binding sequence matching the $GGCT_5$ set of repeated sequences. It has been reported that ARS CS is the yeast autonomously replicating sequence, and that the eve-stripe-2-hb2 element has the properties of a genetic on/off switch involved in the regulation of a segmentation strip by overlapping activators and repressors in the *Drosophila* embryo. Even so, these two binding sequences only partially match the AT palindromic sequences, and the functions of the respective binding proteins do not appear to be synexin functions. It also has been reported that two motifs rich in A and T residues appear to be major contributors to the muscle-specific transcriptional activity of the mouse II B myosin heavy-chain promoter when transfected into quail myogenic and non-myogenic cells (Takeda et al., 1992). These observations suggest an important functional role for these AT-rich sequence motifs in the regulation of genes of the myosin heavy-chain family. It is therefore possible that these palindromic and repeated sequences are important to the mouse synexin gene, and that investigations of these potential regulatory elements will help elucidate the regulation of synexin-gene expression. Studies to determine whether these elements alone play a role in the regulation of synexin-gene expression are being undertaken currently in our laboratory.

A comparison of the 3′-non-coding sequences between the two cDNA sets and genomic DNA revealed identical sequences. This suggested that the expression of two types of mRNA is due to alternative choices of polyadenylation site in the 3′-untranslated region. These two types of mRNA appear to be expressed at equal levels in all of the tissues we tested (Figure 3a). The mechanism of alternative polyadenylation is not yet understood in synexin. In the case of human CD59 antigen (Tone et al., 1992), it has been suggested that alternative polyadenylation may be a result of the high degree of secondary structure around the typical polyadenylation signal AAUAAA. Indeed, many eukaryotic genes generate multiple transcripts from a single gene through the use of different sites of polyadenylation. For example, studies with the actin 5C gene of *Drosophila melanogaster* have indicated that the various 3′ ends are differentially represented throughout development (Bond and Davidson, 1986). In the case of the chicken vimentin gene (Capetanaki et al., 1983), there is also some tissue specificity with respect to the use of different sites. The downstream polyadenylation sites are also involved in neural tissue for three other genes (Bermingham and Scott, 1988). The significance of these observations is not known, and the only example in which the alternative use of polyadenylation sites has a known functional significance is the immunoglobulin heavy-chain gene. Kemp et al. (1983) showed that mRNA at the second polyadenylation site can be cleaved and repolyadenylated in the production of secretory and membrane immunoglobulin heavy chains.

Besides alternative polyadenylation, the two types of synexin cDNA also show differences in the 5′-non-coding regions and in two amino acid substitutions. A comparison of the 5′-non-coding region of the cDNA and the 1.3 kb 5′-flanking genomic DNA revealed sequence identity with only the 5′-non-coding region of set-I cDNA. These data are consistent with the hypothesis that the different mRNA species at the 5′-non-coding region might be transcribed from different promoter regions and produced by differential splicing. For example, the sequence of the set-II 5′-non-coding region might have existed further upstream of the 1.3 kb 5′-flanking region of the synexin gene reported here.

A comparison of the coding-region sequences between the cDNA and the genomic copy has shown that the coding-region sequence of the genomic DNA is identical with that of the set-I cDNA. However, as we know that only one copy of the synexin gene is expressed in the mouse, the data suggest that either the set-I and -II clones, which differ from each other by nucleotide substitutions, represent two individual alleles or this region may represent a hot spot for mutation during transcription. In support of the former hypothesis we note that the original synexin cDNA clones came from a library produced from a non-inbred mouse.

REFERENCES

Adamson, M. C., Silver, J. and Kozak, C. A. (1991) Virology **183**, 778–781

Arndt, K. and Fink, G. R. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 8516–8520

Arrignet, P., D'Eustachio, P., Kristen, T., Wetsel, R. A., Saris, C. J. M., Hunter, T., Chaplin, D. D. and Tack, B. F. (1990) Biochemistry **29**, 1226–1232

Bailly, A., Page, C. L., Rauch, M. and Milgrom, E. (1986) EMBO J. **5**, 3235–3241

Bermingham Jr., J. R. and Scott, M. P. (1988) EMBO J. **7**, 3211–3222

Bond, B. J. and Davidson, N. (1986) Mol. Cell. Biol. **6**, 2080–2088

Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383

Burns, A. L., Magendzo, K., Shirvan, A., Srivastava, M., Rojas, E., Alijani, M. R. and Pollard, H. B. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 3798–3802

Capetanaki, Y. G., Ngai, J., Flytzanis, C. N. and Lazarides, E. (1983) Cell **35**, 411–420

Chin, H., Kozak, C. A., Kim, H.-L., Mock, B. and Mcbride, O. W. (1991) Genomics **11**, 914–919

Creutz, C. E., Pazoles, C. J. and Pollard, H. B. (1978) J. Biol. Chem. **253**, 2858–2866

Creutz, C. E., Pazoles, C. J. and Pollard, H. B. (1979) J. Biol. Chem. **254**, 553–558

De, B. K., Misono, K. S., Lukas, T. J., Mroczkowski, B. and Cohen, S. (1986) J. Biol. Chem. **261**, 13784–13792

Doring, V., Schleicher, M. and Noegel, A. A. (1991) J. Biol. Chem. **266**, 17509–17515

Feinberg, A. P. and Vogelstein, B. (1989) Anal. Biochem. **137**, 266–269

Fisch, T. M., Prywes, R. and Roeder, R. G. (1987) Mol. Cell. Biol. **7**, 3490–3502

Green, E. L. (1981) Genetics and Probability in Animal Breeding Experiments, pp. 77–113, MacMillan, New York

Hanley-Hyde, J., Mushinski, J. F., Sadofsky, M., Huppi, K., Krall, M., Kozak, C. A. and Mock, B. (1992) Genomics **13**, 1018–1030

Hoggan, M. D., Halden, N. F., Buckler, C. E. and Kozak, C. A. (1988) J. Virol. **62**, 1055–1056

Hong, K., Duzgunes, N. and Papahadjopulos, D. (1981) J. Biol. Chem. **256**, 3641–3644

Hope, I. A. and Struhl, K. (1985) Cell **43**, 177–188

Horlick, K. R., Cheng, I. C., Wong, W. T., Wakeland, E. K. and Nick, H. S. (1991) Genomics **10**, 363–374

Houten, J. V. V. and Newlon, C. S. (1990) Mol. Cell. Biol. **10**, 3917–3925

Johnson, P. F., Landschultz, W. H., Graves, B. J. and McKnight, S. L. (1987) Genes Dev. **1**, 133–146

Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M. and Beato, M. (1984) Nature (London) **308**, 513–519

Kemp, P. J., Morahan, G., Cowman, A. F. and Harris, A. W. (1983) Nature (London) **301**, 84–86

Klemsz, M. J., McKercher, S. R., Celada, A., Beveren, C. V. and Maki, R. A. (1990) Cell **61**, 113–124

Kovacic, R. T., Tizard, R., Cate, R. L., Frey, A. Z. and Wallner, B. P. (1991) Biochemistry **30**, 9015–9021

Kozak, C. A., Peyser, M., Krall, M., Mariano, T. M., Kumar, C. S., Pestka, S. and Mock, B. A. (1990) Genomics **8**, 519–524
Kronstad, J. W., Holly, J. A. and Mackey, V. L. (1987) Cell **50**, 369–377
Kuijpers, A. J. G., Lee, G. and Pollard, H. B. (1992) Cell Tissue Res. **269**, 323–330
Lavu, S. and Reddy, E. P. (1986) Nucleic Acids Res. **14**, 5309–5320
Magendzo, K., Shirvan, A., Cultraro, C., Srivastava, M., Pollard, H. B. and Burns, A. L. (1991) J. Biol. Chem. **266**, 3228–3232
Mitchell, P. J., Wang, C. and Tjian, R. (1987) Cell **50**, 847–861
Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
Nir, S., Stutzin, A. and Pollard, H. B. (1987) Biochim. Biophys. Acta **903**, 309–318
Pfeifer, K., Arcarigiolo, B. and Guarente, L. (1987) Cell **49**, 9–18
Pollard, H. B. and Rojas, E. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2974–2978
Pollard, H. B., Rojas, E., Pastor, R. W., Rojas, E. M., Guy, H. R. and Burns, A. L. (1991) Ann. N.Y. Acad. Sci. **635**, 328–351
Rogers, J. (1985) Nature (London) **315**, 458–459
Rogers, J. and Pollard, H. B. (1987) FEBS Lett. **217**, 25–31
Saris, C. J. M., Tack, B. F., Kristensen, T., Glenney, Jr. J. R. and Hunter, T. (1986) Cell **46**, 201–212
Shrivan, A., Srivastava, M., Wang, M. G., Cultaro, C., Magendzo, K., McBride, O. W., Pollard, H. B and Burns, A. L. (1994) Biochemistry, in the press.
Smith, T. F. and Waterman, M. S. (1981) J. Mol. Biol. **147**, 195–197
Spano, F., Rangei, G., Palla, E., Colella, C. and Melli, M. (1990) Gene **95**, 243–251
Speck, N. A. and Baltimore, D. (1987) Mol. Cell. Biol. **7**, 1101–1110
Stanojevic, D., Small, S. and Levine, M. (1991) Science **254**, 1385–1387
Stutzin, A. (1986) FEBS Lett. **197**, 274–280
Takeda, S., North, D., Iakich, M. M., Russell, S. D. and Whalen, R. G. (1992) J. Biol. Chem. **267**, 16957–16967
Tone, M., Walsh, L. A. and Waldmann, H. (1992) J. Mol. Biol. **227**, 971–976
Tonegawa, S. (1983) Nature (London) **302**, 575–581
Tzimagiorgis, G., Adamson, M. C., Kozak, C. A. and Moschonas, N. K. (1991) Genomics **10**, 83–88
von Gierke, V. (1991) J. Biol. Chem. **266**, 1697–1700
Weiher, H., Honig, M. and Gruss, P. (1983) Science **219**, 626–631
Winkel, J., Van de, G. J., Krnst, L. K., Anderson, C. L. and Chiu, J.-M. (1991) J. Biol. Chem. **266**, 13449–13455
Zhang-Keck, Z.-Y., Burns, A. L. and Pollard, H. B. (1993) Biochem. J. **289**, 735–741

Received 22 November 1993/18 February 1994; accepted 22 February 1994