Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with other synexins

Zhen-Yong ZHANG-KECK, 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.

Two sets of cDNAs encoding mouse synexin were isolated from a liver cDNA library and sequenced. The coding regions of synexin clones show 99% identity. By contrast, the two mouse synexin cDNAs differ in a number of ways in both 5' and 3' non-coding regions. The two sets of cDNA encode a polypeptide of 463 amino acid residues which has a deduced molecular mass of 50 kDa. The amino acid sequence of mouse synexin shows a high degree of similarity to both the unique N-terminal domain and the highly conserved C-terminal domain of previously cloned human synexin. Northern-blot analysis using mouse liver polyadenylated RNA revealed two transcripts of 1.8 kb and 2.6 kb, corresponding to group I and group II respectively. Further hybridization analysis using specific sequences from each set of clones showed that the two sizes of

mRNAs differ in the length of the 3' non-coding region which corresponded to the cDNAs. Both mouse liver synexin and recombinant mouse synexin expressed in *Escherichia coli* reacted after Western-blot analysis with a goat antibody against bovine synexin. Only in the larger group-II cDNAs do we find point mutations leading to amino acid replacements of Ser to Ala at residue 145 in the unique N-terminal domain, and of Ala to Gly at residue 304 in the transition zone between repeats II and III. We conclude from a comparison of mouse, human and *Dictyostelium* synexins that changes occur predominantly in the hydrophobic N-terminal domain, or, in the C-terminal region at the ends of some predicted α -helices, on the hydrophobic face of the amphipathic C-helices, and within a lengthy non-helical domain connecting major repeats II and III.

INTRODUCTION

Synexin (annexin VII) is a Ca^{2+} -dependent membrane-binding protein which aggregates chromaffin granules (Creutz et al., 1978, 1979), and fuses chromaffin-granule ghosts and acidic phospholipid liposomes (Creutz, 1981; Creutz et al., 1982; Hong et al., 1981). Synexin also forms highly selective Ca^{2+} channels in phospholipid bilayers (Pollard and Rojas, 1988; Burns et al., 1989), and evidence suggests that this protein may be a mediator of membrane fusion during Ca^{2+} -dependent exocytosis (Pollard et al., 1990). In addition, we have proposed that the mechanism of such fusion could be the formation of a 'hydrophobic bridge' across fusing membranes, upon which the phospholipids could cross and mix (Pollard et al., 1990, 1991).

Insight into the molecular basis of such strong hydrophobic character has come from recent cloning and expression studies of human synexin, which have shown the protein to be in the annexin gene family. Like most other annexins, synexin contains a highly conserved C-terminal tetrad repeat with alternating hydrophobic and hydrophilic regions (Burns et al., 1989). However, the unique N-terminal domain is very large (164 amino acids in mouse synexin) and highly hydrophobic. This domain may be thus responsible for some aspect of the fusion process. Nonetheless, we have recently shown that the highly conserved C-terminal domain is probably responsible for the channel activity, and therefore also able to penetrate and span the target membrane. The evidence rests on the fact that endonexin II (annexin V), also expresses channel activity in bilayer membranes (Rojas et al., 1990), but has a unique N-terminal domain of only 19 amino acids. In further contrast with synexin, endonexin II lacks granule-aggregating and liposome-fusion ability. These results thus further emphasize the likely importance of the N-terminal domain of synexin for fusion activity and/or the ability

of synexin to self-associate. The N-terminal domain of synexin is also the site of a tissue-specific polymorphism, in that synexin from skeletal muscle, cardiac muscle and brain contains 22 additional amino acids resulting from alternative splicing of a cassette exon (Magendzo et al., 1991).

The problem remains of how to predict which parts of the synexin molecule might be responsible for binding Ca^{2+} , interacting with phospholipids, and forming the channel. Since endonexin II also forms channels, it was possible that a sequence comparison might yield specific information on the channel. However, the extent of this sequence similarity between endonexin II and synexin is only ~50%, and the channel properties are, in fact, slightly different in terms of channel kinetics, ion selectivity and voltage-dependence. Thus some ambiguity is inescapable in trying to assign functional importance on the basis of sequence-similarity arguments. Nonetheless, the crystal structure for endonexin II has recently been reported (Huber et al., 1990), and our own predictions for the location of α -helical domains in the C-terminal domain of synexin proved consistent with the reported endonexin II structure (Guy et al., 1991; Pollard et al., 1991).

However, insight into synexin function might be gained by a detailed comparison of human synexin with synexin from another species. In the present paper we report the cloning, sequencing and expression of cDNAs encoding the mouse synexin gene from a mouse liver cDNA library. Nucleotide sequence differences in the coding region of synexin cDNAs revealed sites where amino acid replacement can occur while preserving function. These substitutions were in the hinge region between the unique N-terminal domain and the conserved tetrad repeat, at the ends of predicted helices, on the hydrophobic face of the amphipathic C helices, and in a lengthy, non-helical domain connecting the major repeats II and III. These data serve to establish the specific

importance of most residues within the helical domains of synexin, and suggest that the additional variations observed among members of the annexin gene family may be of fundamental importance.

MATERIALS AND METHODS

Isolation of mouse synexin cDNAs

A mouse liver cDNA library in the cloning vector λ zap (10^6 total recombinants; Stratagene) was screened with a 990 bp *Pst*I–*Hind*III fragment of human synexin cDNA (nucleotide positions 116–1107; Burns et al., 1989). Duplicate filters were hybridized in 50% (v/v) formamide/5 $\times$ SSPE/5 $\times$ Denhart's solution/0.1% SDS/100 μ g/ml sheared salmon sperm DNA/random primed probe (10^6 c.p.m./ml) at 42 °C for 12–24 hr. Washes in 2 $\times$ SSC/0.1% SDS were performed at room temperature and 65 °C for 30 min. The resulting 13 positive clones were plaque-purified and analysed (Sambrook et al., 1989 2nd edn.).

All DNA sequence data were determined by the dideoxy-chain-termination method using Sequenase (U.S. Biochemicals) for both strands. Sequencing primers were synthesized on an Applied Biosystems model 380B DNA synthesizer.

Northern-blot analysis

A 5 μ g sample of mouse liver polyadenylated [poly(A)⁺] RNA was transferred from 1.2%-agarose/formaldehyde gels to nylon membrane (Clontech) and probed first with random primer-labelled *Pst*I–*Bgl*II mouse synexin cDNA coding region (nucleotide positions 216–794). After removing the previous probe, the blot was re-hybridized with the mouse 3' non-coding region (nucleotide positions 1900–2300) and amplified by PCR from the full-length mouse synexin cDNAs (Kawasaki and Wang, 1989). The hybridization solution contained 10^6 c.p.m./ml probe, 50% formamide (v/v), 5 $\times$ SSC, 10 mM NaHPO₄, pH 6.5, 2 $\times$ Denhart's reagent, 0.1% SDS and 50 μ g/ml denatured salmon sperm DNA. The blot was hybridized at 42 °C for 12 h and washed in 6 $\times$ SSPE at 60 °C and bands were revealed by autoradiography.

Expression of mouse synexin cDNA in *Escherichia coli*

The coding region of mouse cDNA was amplified by PCR with 5' and 3' primers containing *Sal*I and *Bam*HI restriction sites respectively. The product was then cloned into the *Sal*I and *Bam*HI sites of the pET12 plasmid (Studier, 1990) and expressed in the BL21 strain of *E. coli* and induced with 0.5 mM IPTG. The construct was verified by the dideoxy-chain-termination method (Sambrook et al., 1989).

Protein isolation and immunoblotting

Synexin was partially purified from six female mouse livers by precipitating twice with 20% (satd.) (NH₄)₂SO₄ (Creutz et al., 1978).

The protein samples were separated by SDS/10%-PAGE under reducing conditions and then transblotted to nitrocellulose. Blots were reacted with a goat anti-(bovine synexin) antibody diluted 1:200 in PBS containing 5% powdered milk and detected with rabbit anti-(goat IgG)-horseradish peroxidase (Vector Laboratories).

RESULTS

Isolation and characterization of mouse synexin cDNA

We first verified the existence of synexin in mouse liver by immunological cross-reactivity with anti-synexin antisera. Immunoblot analysis using goat anti-(bovine synexin) antibody detected a 50 kDa synexin band in extracts from mouse liver and human lung (Figure 1). These data indicate that the synexin gene is expressed in mouse liver and that mouse and human synexin share epitopes with bovine synexin.

Figure 1 Immunoblot analysis of mouse liver extract with anti-(bovine synexin) polyclonal antibody

Purified human lung synexin (lane 1) and mouse liver extract (lane 2) were electrophoresed on an SDS/10%-PAGE gel and transferred to nitrocellulose. Positions of molecular-mass (*M*) markers are shown.

Figure 2 Schematic diagram of mouse synexin clones

(a) A summary of the two groups of cDNA isolated. The coding sequences are shown as box, and the different 5' and 3' ends of the coding sequences are denoted by lines and dashes respectively. Arrows indicate the positions of the substitution of the nucleotides between the two groups. The poly(A) tails are indicated. (b) The members of cDNA isolates were divided into two groups. The positions of overlapping cDNA clones were established by sequencing both ends.

Figure 4 Northern-blot analysis of mouse liver poly(A)⁺ RNA separated on a 1.2%-formaldehyde/agarose gel followed by blotting on to a nylon filter (see the Materials and methods section)

(a) Lane 1, the nylon filter was hybridized to the mouse synxin coding region using probe A (shown in b); lane 2, the nylon filter was hybridized to the 3' uncoding region of mouse synxin using probe B (shown in b). The same nylon filter was used for both hybridizations after stripping. Size markers are shown at the right.

cDNA, with the exception of two amino acid substitutions, which reflect the two nucleotide changes. The two predicted proteins thus differ by the substitution of an alanine for a serine residue at position 145 and a glycine for an alanine residue at position 304, for pMBC 10.1 and pMBC 2L respectively.

Northern-blot analysis

Northern-blot analysis (Figure 4) with the mouse cDNA coding region (Figure 4b) revealed the presence of two synxin transcripts expressed in mouse liver (Figure 4a, lane 1). The approx. 1.8 and 2.6 kb size of hybridized bands correspond to the lengths of the group I [1861 bp excluding the poly(A) tail] and the group II clones [2636 bp excluding the poly(A) tail] respectively. To confirm further the relationship between the two mRNAs, the same blot was stripped and reprobed with a labelled 400 bp PCR fragment (positions 1900–2300) (Figure 4b) corresponding to the 3' non-coding region present only in group II cDNA. As shown in Figure 4(a), lane 2, only the 2.6 kb transcript

Figure 5 Immunoblot analysis of mouse synxin cDNA expressed in *E. coli* with goat anti-(bovine synxin) antibody

Lane 1, *E. coli* extract; lane 2, extract of *E. coli* induced with IPTG; lane 3, extract of *E. coli* carrying the mouse synxin gene; lane 4, extract of *E. coli* carrying the synxin gene induced with IPTG. The positions of molecular-mass (*M*) markers are shown.

was detected after autoradiography, consistent with the lack of these sequences in the 3' non-translated regions of the group I clones. These data show that the two groups of cDNAs represent two different lengths of mRNAs in the mouse liver.

Expression of mouse synxin in *E. coli*

To further assess the identity of the cDNA isolated from the mouse liver library, the cDNA was subcloned into the pET12 *E. coli* expression vector and the protein was induced with 0.5 mM IPTG. Western-blot analysis using goat anti-(bovine synxin) antibody revealed a major band and several minor bands after induction (Figure 5, lanes 3 and 4), which was absent in *E. coli* controls (lanes 1 and 2). The size of the major immune reactive band is in good agreement with the native synxin isolated from mouse liver (Figure 1). The minor bands probably correspond to degradation products of the induced synxin protein, since they are not found in the control lane.

DISCUSSION

The results of this analysis clearly show that mouse synxin cDNAs can be separated into two classes, based on both size and on the consistent substitution of two amino acids at positions 145 and 304. In the shorter length, group I cDNA class, residue 145 is Ala, and residue 304 is Gly. In the longer-length, group II, class, residue 145 is Ser and residue 304 is Ala. These amino acid replacements each correspond to single base-pair substitutions, which occur in the proximal end of the unique N-terminal domain, and in the region connecting repeats II and III of the C-terminal domain. Consistently, the basis for the difference in size of the two classes rests on the fact that the larger class contains 893 bp of unique non-coding sequence after the 311 bp 3' non-coding region found in both groups. The longer group II cDNAs also have the expected two polyadenylation signals,

Figure 6 (a) Schematic diagram of the mouse synexin protein structure and (b) a comparison of the amino acid sequence among mouse, human, *Dictyostelium* and bovine synexins

(a) The structure is divided into N-terminal and C-terminal domains, which are further divided into four repeating units. Amino acid residue numbers for the group I mouse sequence are indicated. (b) * Indicates sequence identity (upper level) and * indicates conservative substitution (lower level). The definition of 'homology' is that of the Swiss-Prot data bank, in which amino acids in the following groups are 'homologous': [S, T, A, G, P]; [N, D, E, Q]; [R, K, H]; [M, I, L, V]; and [F, Y, W]. The amino acid sequence of bovine synexin is derived from polypeptide sequencing (Creutz et al., 1988; Burnsetal, 1989). The location of α -helices in the repeating units are boxed and indicated as A, B, C, D and E. Arrows indicate amino acid differences between mouse and human synexin.

located at the end of the 311 bp of non-coding sequence shared with the shorter, group I cDNAs, and at the 3' end of the complete 1204 bp sequence. The 5' non-coding sequences of the two groups are also different. These data suggest that the two groups of cDNAs represent different alleles of the same gene or that two synexin genes exist in the mouse. Presently we are doing experiments to address this question.

Two mouse synexin messages can be detected by Northern analysis as two differently sized synexin mRNAs of 1.8 and

2.6 kb respectively. In the case of human synexin, two mRNAs can also be detected. However, the differences in size within the human synexin cDNAs is based on differences in the choice of poly(A) sites in the 3' non-coding region. This contrasts with the case for mouse, where the differences are also based on nucleotide substitutions in the coding region. By contrast, in the case of the mouse liver synexin, only the larger group II cDNAs (containing Ser-145 and Ala-304) also carry the extended 3' non-coding sequence.

Figure 7 Predicted synexin structure

The cylinders A, B, and D, E represent a series of α -helices which are approximately perpendicular to helix C in the crystal structure. The amino acid sequence of each helix is shown in Figure 6. The asterisk (*) represents the location of the 'endonexin fold'. The number symbol (#) indicates the region between helices D and E. In the nomenclature of Guy et al. (1991) these are S₁ and S₂ respectively. Each large solid circle (●) represents a difference between mouse, human and *Dictyostelium* amino acids where the mouse amino acid position is indicated. The target represents the difference seen between the two mouse synexin proteins. The four individual repeats are indicated by roman numerals.

This study may also have important implications for our understanding of synexin structure and function. With respect to the N-terminal domain, which shows 86% sequence identity with the equivalent human domain, there are 20 amino acid substitutions (Figure 6b). There is also a three-amino-acid segment deleted in the mouse. However, these changes do not modify the basic predicted β -sheet/ β -turn motif characteristic of the entire N-terminal domain. Furthermore, eight of the 23 changes are clustered at the end of the N-terminus (137–164). This cluster is twice as populated with mutations than might be expected, since eight of 23 mutations (~35%) occur in 28 of 164 residues, representing 17% of the N-terminal sequence. This region is also of special interest because residue 143 (145 in human synexin) is the exact location of the 66 bp cassette exon, which is edited into human synexin from skeletal and cardiac muscle and brain (Magendzo et al., 1990). These data thus indicate that this cluster of mutations, located close to the junction between the unique N-terminal and the conserved C-terminal domains, has been identified in other annexins as a hinge region with functional importance. Parenthetically, the equivalent region of gene products for lipocortin I (annexin I) contains tyrosine-21, the site of phosphorylation by epidermal-growth-factor receptor kinase (De et al., 1986), whilst the equivalent region of the gene product for calpactin I heavy chain (annexin II) binds p11, the S-100 homologue (Saris et al., 1986). Although exon structures vary among these annexins in this region, these results do serve to delineate further the region of residue 145 in synexin as a domain with regulatory potential, and certainly deserving of further scrutiny.

We now turn our attention to the C-terminal tetrad repeat domain, and will simplify our analysis by representing the α -helical domains and their connecting loops in an open format which we have derived from the crystal structure of the synexin homologue endonexin II (Huber et al., 1990; also see Figure 7). Indeed, predictions of synexin structure by computation and model-building (Guy et al., 1991; Pollard et al., 1990) justify the assumption that the endonexin II and synexin structures are homologous. As shown in Figure 7, one can view individual repeats as an assembly of anti-parallel A- and B-, and D- and E-helices, oriented normal to a backbone C-helix. The tetrad repeat is formed by four such assemblies in series, and the length of each cylinder is proportional to the number of amino acids in the helix. In Figure 7, the asterisk (*) represents the locations of the GXGTDE motif in the highly conserved 'endonexin fold' between helices A and B, which has been frequently viewed as a

possible site for Ca²⁺ binding. Recent X-ray-crystallographic results also support this concept (Huber et al., 1990). In Figure 7, the 14 solid circles (●) are locations of differences between amino acids in mouse and man, and the numbers above or below the circles represent the amino acid position in the mouse sequence. The 'target' symbol is the location of the only difference between the two mouse cDNAs in the C-terminal domain, and occurs at the C-terminal end of the E-helix in the second repeat.

From comparison of human and mouse synexin, three classes of changes can be seen to occur. One class occurs at the ends of helices. Specific instances are A188S (i.e. A-188 → Ser) (beginning of I-B), P233S (beginning of II-A), A304G (end of II-E), R352K (at the end of III-B), and S403A (end of IV-A). The second class occurs in a limited region between helices in the lengthy and poorly conserved segment linking regions II and III. These are N309R and I312V. The third class is represented by changes in the middle of helices, which occur six times. In fact, of these six, four occur in nearly identical locations on the same face of the C-helix of repeats I and II. These are Q197K and M205S on repeat I and D269E and L276S for repeat II. In both cases the changes occur in a hydrophilic pocket on the hydrophobic face of the amphipathic C-helix. The other two instances include T434A, on the C helix of repeat IV, and K455R, in the E-helix of repeat IV. Of the 15 total differences between mouse and human synexin, ten are also different in *Dictyostelium* synexin (Doring et al., 1991; Gerke, 1991). Of these ten, only three are conservative substitutions, and all occur in repeat IV. These are T434 (A, human; S, *Dictyostelium*), S442 (G, human; A, *Dictyostelium*) and S447 (G, human (h); A, *Dictyostelium*). Presumably the seven remaining positions may be less critical for function, although we must temper this conclusion by the knowledge that synexin has multiple functions. Of the remaining five differences, three are common between mouse (m) and *Dictyostelium* (d) [namely Q (m,d)197K(h); S(m,d)403A(h); K(m,d)455R(h)] and two are common between human and *Dictyostelium* [namely P(h,d)233S(m) and I(h,d)312V(m)].

We thank Dr M. Srivastava, Dr. A. Shirvan and Dr. K. Magendzo for excellent advice during the course of our work. We also thank C. Cultraro and H. Caohuy for technical assistance. This work was partially supported by the Cystic Fibrosis Foundation.

REFERENCES

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

- Creutz, C. E. (1981) *J. Cell. Biol.* **91**, 247–256
- 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
- Creutz, C. E., Scott, J. H., Pazoles, C. J. and Pollard, H. B. (1982) *J. Cell Biochem.* **18**, 87–97
- Creutz, C. E., Snyder, S. L., Husted, L. D., Beggerly, L. K. and Fox, J. W. (1988) *Biochem. Biophys. Res. Commun.* **152**, 1298–1303
- 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
- Gerke, V. (1991) *J. Biol. Chem.* **266**, 1697–1700
- Guy, H. R., Rojas, E. M., Burns, A. L., Rojas, E. and Pollard, H. B. (1991) *Biophys. J.* **59**, 372a
- Hong, K., Duzgunes, N. and Papahadjopoulos, D. (1981) *J. Biol. Chem.* **256**, 3641–3644
- Huber, R., Schneider, M., Mayr, I., Romisch, J. and Paques, E.-P. (1990) *FEBS Lett.* **275**, 15–21
- Kawasaki, E. S. and Wang, A. M. (1989) in *PCR Technology Principles and Applications for DNA Amplification* (Erich, H. A., ed.), Stockton Press, New York
- Magendzo, K., Shirvan, A., Cultraro, C., Srivastava, M., Pollard, H. B. and Burns, A. L. (1991) *J. Biol. Chem.* **266**, 3228–3232
- Pollard, H. B. and Rojas, E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2974–2978
- Pollard, H. B., Burn, A. L. and Rojas, E. (1990) *J. Membr. Biol.* **117**, 101–112
- Pollard, H. B., Rojas, E., Pastor, R. W., Rojas, E. M., Guy, H. R. and Burns, A. L. (1991) *Ann. NY. Acad. Sci.* **635**, 328–351
- Rojas, E., Pollard, H. B., Haigler, H. T., Parra, C. and Burns, A. L. (1990) *J. Biol. Chem.* **265**, 21207–21215
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Saris, C. J. M., Tack, B. F., Kristensen, T., Glenney, J. R., Jr. and Hunter, T. (1986) *Cell* **46**, 201–212
- Studier, F. W. (1992) *Int. Symp. Genet. Indust. Microorg.* 6th, Strasbourg, France, in the press