Eur. J. Biochem. *198*, 749–760 (1991)

001429569100396M

Amino acid sequence analysis of the annexin super-gene family of proteins

Geoffrey J. BARTON[1,2], Richard H. NEWMAN[3], Paul S. FREEMONT[4] and Michael J. CRUMPTON[3]

[1] Laboratory of Molecular Biophysics, University of Oxford, England
[2] Biomedical Computing Unit, [4] Protein Structure Laboratory and [3] Cell Surface Biochemistry Laboratory, Imperial Cancer Research Fund, London, England

(Received December 3, 1990/January 24, 1991) – EJB 90 1433

The annexins are a widespread family of calcium-dependent membrane-binding proteins. No common function has been identified for the family and, until recently, no crystallographic data existed for an annexin. In this paper we draw together 22 available annexin sequences consisting of 88 similar repeat units, and apply the techniques of multiple sequence alignment, pattern matching, secondary structure prediction and conservation analysis to the characterisation of the molecules. The analysis clearly shows that the repeats cluster into four distinct families and that greatest variation occurs within the repeat 3 units. Multiple alignment of the 88 repeats shows amino acids with conserved physicochemical properties at 22 positions, with only Gly at position 23 being absolutely conserved in all repeats. Secondary structure prediction techniques identify five conserved helices in each repeat unit and patterns of conserved hydrophobic amino acids are consistent with one face of a helix packing against the protein core in predicted helices a, c, d, e. Helix b is generally hydrophobic in all repeats, but contains a striking pattern of repeat-specific residue conservation at position 31, with Arg in repeats 4 and Glu in repeats 2, but unconserved amino acids in repeats 1 and 3. This suggests repeats 2 and 4 may interact via a buried salt-bridge. The loop between predicted helices a and b of repeat 3 shows features distinct from the equivalent loop in repeats 1, 2 and 4, suggesting an important structural and/or functional role for this region. No compelling evidence emerges from this study for uteroglobin and the annexins sharing similar tertiary structures, or for uteroglobin representing a derivative of a primordial one-repeat structure that underwent duplication to give the present day annexins. The analyses performed in this paper are re-evaluated in the Appendix, in the light of the recently published X-ray structure for human annexin V. The structure confirms most of the predictions and shows the power of techniques for the determination of tertiary structural information from the amino acid sequences of an aligned protein family.

The annexins are a family of proteins that share the common features of binding both membranes and phospholipids in a Ca^{2+}-dependent manner. The annexins are found in many tissues, and their membrane binding activity has been claimed to be involved in cytoskeletal interactions [1], membrane fusion (e.g. see [2]), anticoagulation [3], intracellular signalling as a kinase substrate (e.g. see [4]), and phospholipase inhibition [5].

The annexins consist of four or eight conserved repeat units of ≈75 amino acids in length, separated by intervening sequences of variable length, and an N-terminal domain which shows the greatest variation in sequence and length between family members. The N-terminal domain which may be phosphorylated on Tyr/Ser residues, is thought to confer the functional specificity of each molecule, whilst evidence from a number of sources suggests that both the Ca^{2+} and phospholipid binding activities of the annexins reside in the repeat regions [6–9]. Accordingly, this paper concentrates on the analysis of the conserved repeat region of the annexins.

Until recently, protein crystallography had not revealed the three-dimensional structure of an annexin. In addition, annexins show no compelling similarity to proteins of known tertiary structure thus precluding the prediction of their structure by 'homology modelling' techniques. Furthermore, although they bind Ca^{2+}, the annexins do not possess amino acid patterns characteristic of the classical calmodulin-like 'EF-hand' Ca^{2+}-binding site [8]. However, by using a multiple sequence alignment of repeats from three annexins, in conjunction with pattern-matching, and secondary structure prediction techniques, Taylor and Geisow [10] proposed a structural model for an annexin repeat unit based upon the 'EF-hand' containing, bovine intestinal calcium-binding protein (ICaBP). Their speculative model suggested the detailed mode of interaction for a Ca^{2+}-annexin complex and the phospholipid head group. The lack of EF-hand Ca^{2+} ligands in the annexin loops led Taylor and Geisow [10] to base their model

Correspondence to G. J. Barton, Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

Abbreviations. AI–AX, annexins I–X; AI, annexin I; AII, annexin II; AIII, annexin III; AVI, annexin VI; AVII, annexin VII; AIX, annexin IX; AX, annexin X; AIH etc., human annexin I etc.; AIP etc., porcine annexin I etc.; AIR etc., rat annexin I etc.; AIIH etc., human annexin II etc.; AIIM etc., murine annexin II etc.; AIIB etc., bovine annexin II etc.; AIXD etc., *Drosophila* annexin IX etc.;AIH2 etc., human annexin I repeat 2; AIIx1, human, murine and bovine annexin II repeat 1; AIVx2, human, porcine and bovine annexin IV repeat 2; AIVx4, AIVH4, AIVP4 and AIVB4 together; AVIx58, human and murine annexin VI, from N-terminal of repeat 5 to C-terminal of repeat 8; AVIM58, murine annexin VI from N-terminal of repeat 5 to C-terminal of repeat 8. (For further details see first section of Methods.) ICaBP, bovine intestinal calcium binding protein.

principally on an alignment of hydrophobic patterns within the helical regions of ICaBP requiring substantial changes to the ICaBP structure to accommodate the shorter annexin loops. The predicted topology and details of tertiary structure have recently been shown to be incorrect [11, 12], suggesting that care must be taken in inferring tertiary similarities where there is no appreciable sequence similarity.

In this paper we have not attempted to predict a detailed atomic model of an annexin. Rather, we present a comprehensive analysis of the available annexin sequences, which have now expanded to 22 representatives and 88 repeat units. The application of sequence analysis techniques to such a wide ranging family of similar proteins leads to insights into the importance of specific residues to the common tertiary fold and function of the protein family. These insights have been largely confirmed by the recent publication of the X-ray structure for human annexin V [11, 12] and may provide the basis for a detailed understanding of structure/function relationships within the annexin family. A preliminary interpretation of our predictions and observations in comparison with the X-ray structure is presented in the Appendix. This analysis shows the quantity of reliable structural information that may be gleaned from aligned sequence data.

Table 1. *Annexin sequences*

Class	Species	Code	References
Annexin I	human	AIH	[39]
Annexin I	pig	AIP	[40]
Annexin I	rat	AIR	[41]
Annexin II	human	AIIH	[39]
Annexin II	mouse	AIIM	[6]
Annexin II	bovine	AIIB	[42]
Annexin III	human	AIIIH	[43]
Annexin III	rat	AIIIR	[43]
Annexin IV	human	AIVH	[44]
Annexin IV	pig	AIVP	[45]
Annexin IV	bovine	AIVB	[46]
Annexin V	human	AVH	[43, 44, 47–49]
Annexin V	rat	AVR	[43]
Annexin V	chicken	AVC	[50, 51]
Annexin VI	human	AVIH	[32, 52]
Annexin VI	mouse	AVIM	[53]
Annexin VII	human	AVIIH	[54]
Annexin VIII	human	AVIIIH	[55]
Annexin IX	*Drosophila*	AIXD	[14]
Annexin X	*Drosophila*	AXD	[14]

METHODS

Nomenclature

The terminology adopted follows that described by Crumpton and Dedman [13] with the inclusion of the recently sequenced annexins IX and X from *Drosophila* [14]. In order to allow unambiguous reference to be made to each complete protein, its individual repeat regions and combinations of repeats, the basic nomenclature is extended in this paper as follows.

a) The naming of the ten annexin classes is abbreviated to AI, AII...AX.

b) Repeat regions are labelled according to a four-character code; for example AVIH3 refers to the third repeat of human (H) annexin VI (AVI). When referring to all species, a lower case x (x) is used (e.g. AIVx4 corresponds to AIVH4, AIVP4 and AIVB4 together).

c) A repeat range (e.g. murine annexin VI repeats 5–8) is denoted by the start and end repeat numbers (AVIM58). The source organisms and references of the protein sequences analysed is illustrated in Table 1.

Systematic location of repeat regions

The annexin repeat regions were delineated by applying a systematic, semi-automatic protocol. Firstly, approximate starting and ending residues for the four repeats from human annexin I were identified from the literature. Sequence fragments representing each repeat, with a three- or four-residue overlap were automatically aligned [15]. A flexible pattern [16] was then derived from the most highly conserved positions in this alignment and the pattern compared to all 20 sequences. The 88 resulting high-scoring regions identified by this approach were multiply aligned and final end points for each repeat determined by inspection.

The C-termini of repeat-1, and the N-termini of repeat-2 appear straightforward to locate in all the proteins since the repeat 1–2 linking region is very short. However, it is difficult to unambiguously define the N-termini of repeats 1 and 3, as well as the C-termini of repeats 2 and 4. These termini were therefore assigned to provide a slight overlap with the repeat-2 N-terminus and the repeat-3 C-terminus respectively.

Pairwise sequence comparison of the 88 repeats

All unique pairs of repeats (3828 pairs) were optimally aligned by the Needleman-Wunsch [17] algorithm as programmed in the AMPS package [18] using amino acid substitution scores from Dayhoff's mutation data matrix 1978 (MDM78) [19] with a constant of 8 added, and a fixed-length gap-penalty of 8.

For each pair of repeats, a conventional randomization procedure was performed with the score (z) being expressed as the number of standard deviations from the mean of scores for aligning 100 randomized pairs of sequences of the same length and composition (see legend to Fig. 1). Fig. 1 shows a dendrogram obtained from the pairwise data by applying single linkage cluster analysis to the z scores.

Multiple alignment and secondary structure prediction

Fig. 2 illustrates a multiple alignment of the 88 annexin repeats, together with conservation analysis, and summary secondary structure prediction. The multiple alignment was generated automatically [15] with a small manual modification (see below). In order to simplify the figure, the first 20 residues of repeats AVIx7 are not included. This region is described below in conjunction with the repeat 2–3 linker.

Two secondary structure prediction methods that make use of aligned sequence information were applied both to the 88-repeat alignment and to each repeat family within the alignment. The method of Zvelebil et al. [20] uses all sequences together with derived conservation information. The second approach combines the results of helix and strand predictions from the algorithms of Lim [21], Chou and Fasman [22] and Robson [23] together with turn predictions from the algorithms of Rose [24] and Wilmot and Thornton [25] (GJB, unpublished). Both methods predicted five helical regions when applied to all 88 repeats in broad agreement with the secondary structure prediction by Taylor and Geisow based

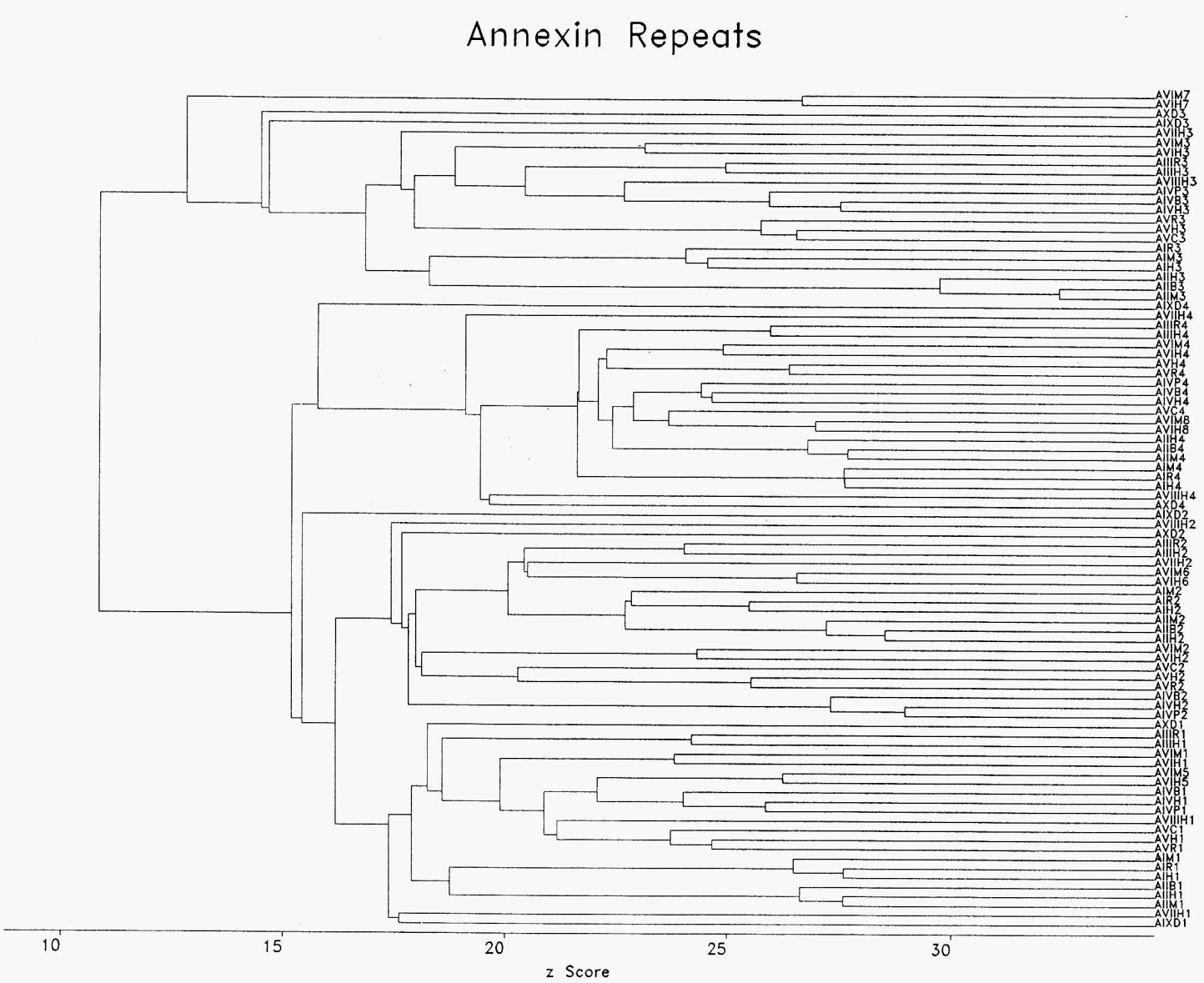

Fig. 1. *Single-linkage dendrogram from the pairwise comparison of 88 annexin repeats.* The repeat codes are shown at the right and vertical lines join repeats into clusters at progressively lower z values. z values are calculated by first obtaining the best score V for the alignment of two sequences as calculated by the Needleman and Wunsch algorithm [17]. The amino acid order of each sequence is then shuffled 100 times, and scores for each shuffled sequence pair are computed by the Needleman and Wunsch [17] algorithm. The mean ($\bar{x}$) and standard deviation (σ) of these scores is calculated, and the z score is then given by $z = (V-\bar{x})/\sigma$. The repeat codes are organized on the figure so that similar repeats are grouped together

on repeats from three annexins in groups I, II and V [10]. Fig. 2 illustrates predictions for each repeat family, and a summary prediction for all repeats.

Conservation values [20] based upon the physico-chemical properties of the amino acids as organised by Taylor [27] are shown for each repeat family (Fig. 2). The numbers run from 0 to 10 with higher numbers indicating that more properties (e.g. charged, hydrophobic) are shared at the position, with an asterisk signifying total identity (see legend for details).

Insertions and deletions

The greatest ambiguity in any sequence alignment occurs around the insertions that are required to improve the overall alignment. Even when the three-dimensional structures of two homologous proteins are compared by superposition, it can be difficult to decide precisely which amino acids have been inserted, since insertions generally occur in the more variable loop regions of the structure. Two changes were manually made to the initial multiple alignment to give the alignment shown in Fig. 2. The insertion at position 20 in repeat 3's was moved from position 18 (Gly) to avoid a deletion in the predicted α-helix a, whilst the triplet AYC (position 7–9) in AVIx2 was moved one residue towards the N-terminus allowing a further gap to be closed in α-helix a.

RESULTS AND DISCUSSION

Evidence for gene duplication/fusion

Fig. 1 confirms that the annexin repeats group into four distinct families corresponding to repeats 1, 2, 3 and 4. The only exception is the *Drosophila* repeat AIXD2 which shows less similarity to the repeat 2 group than does the repeat 1 group. Anomalies of this type are to be expected, since the repeats from *Drosophila* sequences AIX and AX show least overall similarity to the other equivalent repeats, as would be anticipated from the earlier species divergence. Since annexin

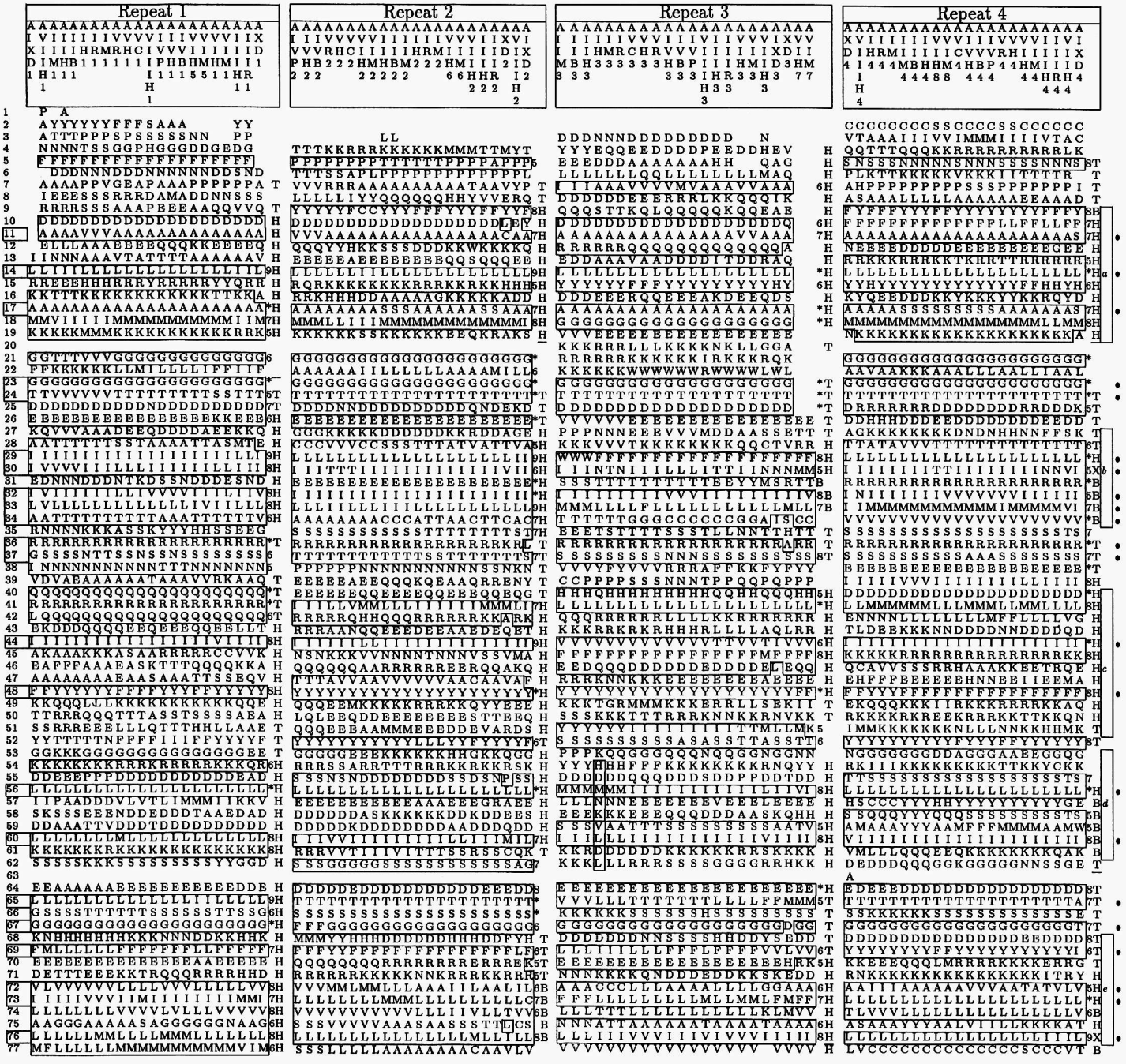

Fig. 2. *Multiple sequence alignment of the 88 annexin repeats.* The alignment was obtained automatically, followed by small manual modifications (see text). Each repeat region shows 22 sequences grouped in the same order as shown in Fig. 1, with the exception that the repeat 3 and repeat 4 groups are interchanged. Within each repeat group, the first 22 columns (boxed at top) show the repeat sequences whilst the penultimate column shows conservation values [20] for aligned positions that give values >5, with an asterisk to indicate identity. The last column shows a consensus secondary structure prediction for the repeat family (H = α-helix, B = β-strand, X = ambiguous α-/β-, T = turn, ' ' = not strongly predicted structure). The location of intron/exon boundaries as determined for AIIM [28] are shown as horizontal lines in this column. Aligned positions giving conservation values >5 are boxed. Where addition of the *Drosophila* sequences (AIX and AX) reduce the conservation value at a position below 5, the residue is shown unboxed (e.g. see position 16 in repeat 1). The location of α-helices predicted by considering all 88 repeats are shown as open boxes on the right of the figure, labelled a–e. Bullet symbols in the final column of the figure identify positions that show similar properties in all repeats, the numbers of these positions are shown boxed on the left of the figure. A vertical box surrounds the interaction site proposed by Miele et al. [33] (position 54–62)

VII shows significant differences to the rest of the family by having a longer, and generally hydrophobic N-terminal domain, one might expect the AVII C-terminal domains also to be atypical of the family. Indeed, repeats 1, 3 and 4 of AVII are the least similar members of each repeat cluster, after AIX and AX, although the AVIIH2 repeat clusters more centrally with the other repeat 2's. However, the overall similarity between the repeat units is very high, strongly supporting the inference of homology, and the small repeat-specific differences detected by the analysis in Fig. 1 are likely only to contribute to minor differences in the tertiary structure.

If the common four repeat unit arose by the recent duplication of a two repeat unit, one would expect the repeat 1 and 3 groups to be more similar to each other than to the repeats 2 and 4 and, similarly, the repeats 2 and 4 to be more similar to each other than they are to the repeats 1 and 3. This

relationship is not observed. However, the spacing between repeats 1 and 2 is short, and similar to that between repeats 3 and 4 giving limited support for the suggestion of a two-repeat ancestor.

With the exception of the *Drosophila* AIXD4, the repeat 4 group shows greatest internal similarity, having all sequences clustering above 19 SD. In contrast, the repeats 3 show greatest diversity with some repeats clustering below 13 SD. Furthermore, the repeat 3 family is less similar to repeats 1, 2 and 4 than 1, 2 and 4 are to each other. The major differences between the annexins lie in the N-terminus, arguing strongly that this region confers the functional specificity. However, the divergence of the repeat 3 family indicates that this is the most likely of the repeat units to contribute to specificity, whilst repeats 1, 2 and 4 form structural and functional regions common to all annexins.

The annexin VI repeats 5–8 are clustered in a manner consistent with the hypothesis that repeats 5–8 arose from a repeat 1–4 precursor, either by gene duplication or fusion. Thus, AVIx5 groups with the repeat 1's, AVIx6 groups with repeat 2's, AVIx7 with repeat 3's and AVIx8 groups with repeat 4's. The high pairwise similarity between all repeats strongly suggests they are homologous (i.e. derived from a common ancestor), that the repeats all share the same tertiary fold and that the alignments obtained automatically will be largely correct within the conserved secondary structural regions. However, this high similarity prevents discrimination in the present analysis of the origin of the AVIx58 unit between a duplication, or fusion event.

Location of introns

Amiguet et al. [28] have described the 22-exon structure for the AIIM gene. They observed that the intron locations do not correspond with the boundaries between the annexin repeats, although the 4/5 (position 19/20 AIIM2) and 10/11 (AIIM4) exon boundaries occur at equivalent positions within repeats 2 and 4 respectively. This fits with the gene duplication hypothesis since repeat 4 would have arisen from repeat 2; however, no other boundaries coincide.

Of the nine exon boundaries that occur within the repeat regions, only three are observed in regions predicted to be α-helix (AIIM1 55/56, AIIM3 12/13 and 43/44). This observation is consistent with the hypothesis that exon boundaries occur at the interface between structural or functional sub-domains. Furthermore, a boundary occurs in the loop between helices a and b in repeats 1, 2 and 4, but not in repeat 3, suggesting that the a – b loop in repeat 3 may be a functionally important sub-domain.

Structural and functional implications of residue conservation

The availability of 88 homologous sequences provides a wealth of information regarding residue conservation. Residues of similar structural type (e.g. hydrophobic) that are common to all repeats indicate those positions important to common features of the annexin tertiary structure, whilst the positions conserved within each repeat group, but not across all repeats, suggest the residues important to the specific functions of the individual repeats.

Table 2 lists the amino acids observed at the most highly conserved positions within each repeat, and a cross all repeats. There are 22 positions that show a high degree of residue conservation across all 88 repeats, 16 of which lie within predicted α-helices. If we consider the remaining 6 positions starting from the N-terminus, position 23 is the only locus that shows total identity in all repeats shown (Gly), suggesting an important role for this residue. Conservation of glycine may indicate a requirement for the close approach of another residue or ligand allowed by the absence of a sidechain (as in dinucleotide-binding proteins [29]). Alternatively, the greater torsional freedom of the residue may be required to effect a tight chain reversal at this position. A small residue (normally Thr) is conserved at position 24 suggesting a role in conjunction with the Gly at 23, possibly to permit tight packing of the polypeptide chain in this predicted loop region. Position 36 is Arg in all but the three repeats AIXD2 (Lys), AXD2 (Leu) and AXD3 (Ala) from *Drosophila*, a suggested function for this residue is outlined below. The Arg36 is also followed by a conserved small residue which suggests that this predicted loop packs closely against the rest of the molecule in the native annexin fold. Position 65 is more variable, but generally a residue with hydrophobic or neutral character is conserved. Finally, position 67 is glycine in all but 5 repeats. Given the special properties of glycine, the substitution of Phe at this position in repeats AIVx2 is difficult to rationalise, whereas the substitution by Asp (AXD3) and Thr (AIXD4) suggests that in these repeats the principal requirement is for a small residue at this position rather than a torsionally flexible amino acid.

The conserved positions that lie within the predicted α-helices a–e provide clues to the packing of the secondary structures in the native, folded annexin structure. Conserved hydrophobic residues at 11, 14 and 17 are consistent with a hydrophobic stripe on one edge of an α-helix that packs against the protein core (helix a). Hydrophobics at 44 and 48 (helix c), and 56 and 60 (helix d) show a similar pattern for these helices, whilst conserved hydrophobic residues at 69, 72, 73 and 76 (helix e) show the classic 'diamond' pattern of hydrophobic residues at $i, i+3, i+4, i+7$ consistent with one face of an α-helix that packs against the protein core. Of the 8 positions predicted to be part of helix b, 5 are conserved hydrophobic positions (29, 30, 32, 33 and 34). This observation strongly suggests that helix b is buried, forming interactions in the hydrophobic core of the annexin tertiary structure.

Several positions show conservation within one or more repeats, but not across all 88. The most striking example is the central residue in helix b (position 31). In repeats 1 and 3 this locus is variable; however, in the repeats 2 it is an absolutely conserved Glu, whilst in repeats 4 it is always Arg. This complementarity of charge may indicate that repeats 2 and 4 interact via position 31 in an intact annexin (see below).

Analyses of crystallographically determined protein structures suggest that glycine often terminates an α-helix [30]. There is a conserved Gly at position 18 in repeats 3, whilst in repeats 1, 2 and 4 this position is conserved hydrophobic, but not Gly. In contrast, Gly is predominantly conserved at position 21 (with the exception of AI/IIx1) in repeats 1, 2 and 4. This may indicate that helix a is shorter in the repeats 3, ending at position 17, whilst in all other repeats it ends at position 19. Secondary structure predictions support this hypothesis (Fig. 2). in addition, the hydrophobic amino acid at position 18 in 1, 2 and 4 would extend the hydrophobic face of helix a. The suggestion that positions 23–24 (Gly-Thr) are closely packed in the tertiary structure, that position(s) 17–18 pack against the hydrophobic core, and the general hydrophilic character of the residues in the region 19–22 strongly indicate that these residues (19–22) are in an exposed

Table 2. *The residues observed at each aligned position in the annexin repeats that conserve amino acid properties*
Columns headed repeat 1, 2, 3 and 4 show conserved residues for repeats 1–4, respectively. The last column shows the amino acids represented at the 22 positions with conserved properties throughout all 88 repeats. Residues shown in parentheses are only present in the *Drosophila* sequences at this position. Those residues shown in square brackets are only present in the AVIx7 repeats

No.	Repeat 1	Repeat 2	Repeat 3	Repeat 4	All
1					
2				CS	
3					
4				SN	
5	F	PTA		SN	
6			LQM(A)		
7			IAVM		
8					
9		YFC		YF	
10	D	DE(LY)	D(Q)	FL	
11	AV	AV(C)	AV	A(S)	AV
12			RQ(A)		
13				KTR	
14	IL	IL	L	L	IL
15		QRKH	YF(H)	YFH	
16	TK(A)				
17	A	AS	A	AS	AS[I]
18	MVI	MLI	G	ML	
19	KMR			K(AN)	
20	-	-		-	
21	TVG	G		G	
22		ALIM			
23	G	G	G	G	G
24	TVS	T	T	T	TVS
25	DN	QDNK(E)	D	RD(K)	
26	EK	E			
27					
28	TSAM(E)	CSTAV		ATV	
29	LI	LI	WF	L	LIWF
30	VIL	IT	INTL[M]	ITVN	VITNL[M]
31		E		R	
32	ILV	I	IV	IVN	ILVN
33	LIV	IL	MLF	MIV	MLIVF
34	TA(V)	ACT	TGCA (IS)	V	TACGV (IS)
35		ST		ST	
36	R	R(KL)	R(A)	R	R(KLA)
37	STN(G)	ST	SN	SA	SATN(G)
38	NT(I)			E	
39				IVL	
40	Q	EQ(G)	HQ	D	
41	R	ILMV	L	ML	
42	Q(L)	RQKH(A)			
43					
44	IV	IL	VT(I)	I	IVLT
45			FM	KR	
46			EDQ(L)		
47		TVAC(F)			
48	YF	Y	Y	Y[F]	YF
49					
50					
51			YITML[K]		
52		YLF	STA	YF	
53					
54	KRQ				
55		SDN(P)		ST	
56	L	L	MIL[V]	L	MIL[V]
57					
58				QSYT	
59			SVAT	AMYF	
60	LM	ILV(M)	ILV	I(V)	ILVM
61	KR				
62		SGA			
63	-	-	-		
64	AED	DE	E	ED	
65	LI	T	VLTF[M]	T(A)	VLITF(A)[M]
66	ST(G)	S			
67	G	FG	G(D)	G(T)	GF(DT)

Table 2. Continuation

No.	Repeat 1	Repeat 2	Repeat 3	Repeat 4	All
68				DE	
69	LFM	FYL	LIFV	YF(I)	LFMYI
70		QRE(K)	E(H)[RK]		
71		RKN			
72	VL	VLAIM	ACLG	IAVTL	VLAIMCGT
73	IVM	LM(C)	FLM	L	FLMIV
74	LVL	VLI(T)		VL(T)	
75	GANS	SATVC(L)	NAT		
76	LM	LM(I)	LIV	LI	LMVI
77	MFLVI				

loop in repeat 3. Further support for this hypothesis is gained by the recent proteolysis experiments of Johnsson and Weber [31]. They showed that AIIH may be nicked by trypsin at the Arg-Lys peptide bond between positions 21 and 22 of AIIH3, to leave an intact core that binds phospholipids in a Ca^{2+}-dependent manner.

The triplet Gln-Arg-Gln (40 – 42) is conserved in all repeat 1's with the exception of AIXD1 (Gln-Arg-Leu), but is not present in repeats 2 – 4, suggesting a repeat-1-specific function for these residues. In contrast, position 64 is a conserved Asp/Glu in all but the AIIx1 repeats. It is interesting to note that the AIIx1 repeats also differ from other repeats in conservation at position 21 as discussed above. Further repeat-specific conservation serves to extend the hydrophobic faces of helices c and d in repeats 2 (position 47) and repeats 4 (position 60). In addition to the hydrophobic diamond, the helix e region has further conserved positions of hydrophobic/neutral character at 74, 75 and 77 (repeats 1), 74 and 77 (repeats 2), 75 (repeats 3) and 74 (repeats 4). This variability in hydrophobicity suggests that, in addition to having one face that packs against the protein core, helix e is partially buried by other parts of the structure.

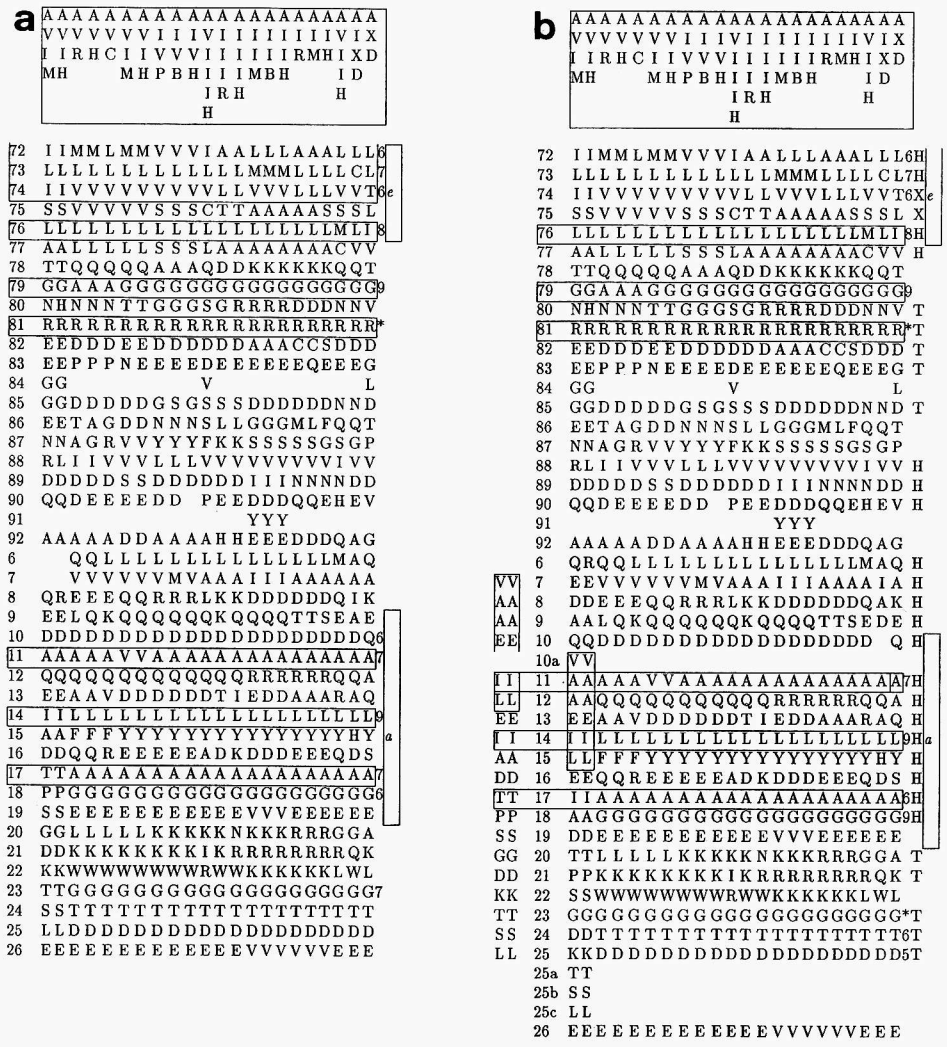

Fig. 3. *Alternative alignments in the repeat 2 – 3 link region.* (a) Multiple alignment of the repeat 2 – 3 link including AVIx, without the deletable VAAEIL segment. Numbering is continued from repeat 2 to overlap with the repeat 3 numbering. (b) As for (a) but with the VAAEIL segment included in the AVIx sequences. The alignment to the right of the numbering was obtained automatically, the alternative arrangement for the AVIx is shown to the left of the position numbers (see text)

The repeat 2–3 link and discussion of the annexin VI deletable segment

The linking sequences between repeats 2 and 3 are illustrated in Fig. 3. There are between 16 and 18 amino acids in the link between the C-terminus of predicted helix e (repeat 2) and the N-terminus of predicted helix a (repeat 3). Two positions are highly conserved, at 79 (predominantly Gly) and 81 (Arg). The region from 77 to 82 could therefore be regarded as a repeat-2-specific extension to the basic annexin repeat. The remainder of the linker is quite variable in characteristics with the exception of a normally hydrophobic amino acid being conserved at position 88. In most of the repeat 2–3 links, the region from 88 to the start of helix a is predicted as α-helical. The variability in this region, coupled with this prediction, is consistent with a helical loop that is not tightly packed into the annexin core.

The annexin VI repeat 6–7 region exists in two forms [32]. Fig. 3a illustrates a multiple alignment of the shorter form (missing VAAEIL between positions 12 and 13) with the 20 repeat 2–3 regions. Hydrophobic/neutral residues are conserved at positions 11, 14 and 17 of predicted α-helix a, consistent with the conservation seen in repeats 3. Similarly, the Asp at position 10 is also conserved in the repeats 7, whilst the putative helix-terminating Gly at position 18 is substituted by an effective alternative helix-termination residue, Pro. Surprisingly, the Gly at position 23, which is absolutely conserved in all other repeats, is substituted by Thr. This suggests either that the Gly is not in a conformation disallowed to Thr, or that the local conformation of this region in the repeats 7 is different to the other repeats.

Fig. 3b illustrates two alternative alignments of the longer annexin VI form. In the main alignment, the result obtained automatically [15] is shown. The VAAEIL segment is aligned as part of helix a and preserves the conserved positions at 11, 14 and 17. The Gly at position 23 is also conserved, though Ala is substituted for the putative helix-termination site at position 18. This alignment suggests that the helix a-b loop would be extended by three residues in repeats 7 at the beginning of helix b (positions 25a–25c). This insertion might be compensated by extending helix a by two residues to terminate at position 21 (Pro). Although this alignment suggests structural changes in a highly conserved loop, the loop already accommodates as single insertion relative to repeats 1, 2 and 4 and proteolysis indicates that the loop is exposed [31], hence further changes might be easily accommodated. An alternative alignment, that assumes the helix a–b region is conserved between the short and long annexin 6 forms, is also plausible. The positions at 11, 14 and 17 are again conserved, Asp/Glu predominate at position 10, and a predominately hydrophobic position in the repeat 6–7 linker (position 7) is also conserved. This alignment appears more favourable since it concentrates the structural changes required by the insertion of VAAEIL into the already variable repeat 2–3 linker.

Suggested homology with uteroglobin

Miele et al. [33] have suggested that the annexins may be related to the steroid-induced rabbit secretary protein, uteroglobin. They base this conclusion on the observation of nine residues at the end of helix 3 of uteroglobin (MQMKKVLDS) that are similar to part of AIH3 (HDMNKVLDL). Further support for this relationship was purported from the observation that both AIH and uteroglobin inhibit phospholipase A_2, as do both synthetic nonapeptides. The implications drawn from this study were that uteroglobin monomer and the annexin repeat share similar tertiary structures and that they inhibit phospholipase A_2 by direct interaction with the HDMNKVLDL region.

Miele et al. [33] presented three pair alignments of uteroglobin to annexin repeats to support their argument (Fig. 1, I, II and III of [33]). Alignment I was with AIH2, II with AIH3, and III with AIIH2. All three alignments are insignificant, both by conventional statistical criteria with z values of 0.9 and 0.4, and also by structural criteria since values below $z = 6.0$ do not reliably indicate that the alignment is correct within secondary structural regions [15]. Furthermore, the alignments shown by Miele et al. [33] are totally inconsistent with each other. Since it is known that the three annexin repeats are unambiguously homologous, the individual pair alignments should show equivalent regions of each repeat aligned with the same region in uteroglobin. However, the MQMKKVLDS of uteroglobin is aligned with positions 54–62 (Fig. 1 of this paper) of AIH3, but with positions 67–75 of AIH2 and, by the inclusion of two gaps, with a region spanning from position 71 into the repeat 2–3 link at position 91 of AIIH2. Clearly, any conclusions drawn from such inconsistent data must be treated with caution.

Miele et al. [33] also reported that peptides MQMKKVLDS (uteroglobin) and HDMNKVLDL (AIH3) both inhibit phospholipase A_2 and suggested that these were also regions of the intact molecules that mediated inhibition by direct interaction with the enzyme. The region 54–62 of the annexin repeat corresponds to predicted helix d, which has conserved hydrophobic amino acids on one face at positions 56 and 60 (Fig. 2). If this region is responsible for phospholipase A_2 inhibition, then the residues predicted to be exposed (i.e. not 56 and 60) must be responsible for the interaction. Furthermore, if this is to be a common mode of phospholipase A_2 inhibition by annexins, then the exposed positions must show conservation of physical properties across all repeats. Inspection of Fig. 2 shows that such conservation is not observed. Indeed, the peptide HDMNKVLDL is atypical, even within the repeat 3 family, and the region 54–62 is one of the least conserved of all overlapping nine-residue segments from the alignment shown in Fig. 2 (data not shown). Fig. 4 illustrates an alignment of the two available uteroglobin sequences, together with packing information obtained from the crystallographically determined structure of the uteroglobin dimer [34]. The nine residues 39–47 are largely buried in the intact uteroglobin dimer with residues 40, 41, 44 and 45 contributing to the dimer interface, residues 41, 42, 45 and 46 packing against helix 1, and residues 39–41 packing against helix 2. In the light of these data, it seems unlikely that this buried region is directly responsible for phospholipase A_2 inhibition in the manner suggested by Miele et al. [33].

Recently, in a number of different phospholipase A_2 assays, van Binsbergen et al. [35] found no evidence of inhibition by the Miele et al. [33] peptides, either by preincubation with the enzyme, or by direct interaction with phospholipids. Our own studies using a lipid-monolayer system similar to that described by Verger et al. [36] show the annexin peptide HDMNKVLDL to inhibit phospholipase A_2, but not activity for the uteroglobin peptide [37].

Although there is no significant sequence similarity between uteroglobin and the annexin repeats, five helices are predicted for the annexins in common with uteroglobin, and both systems share similar numbers of amino acids (uteroglobin 70, annexin repeat 75). Given this weak association,

No.	UTGRB	UTGRBL	HELIX	D	P	G	O	Y
1	G	G						
2	I	I		D				
3	C	C		D				
4	P	P						
5	R	G		D	P			
6	F	F		D				
7	A	A						
8	H	H						
9	V	V			P			
10	I	I	1			G		
11	E	E				G		
12	N	N			P		O	
13	L	L					O	
14	L	L				G	O	
15	L	L				G		
16	G	G						
17	T	T						
18	P	P						Y
19	S	S						
20	S	S			P			
21	Y	Y			P			Y
22	E	G						Y
23	T	T	2					
24	S	S			P			Y
25	L	L						Y
26	K	K						
27	E	E						
28	F	F		D	P			
29	E	Q		D				
30	P	P		D				
31	D	D						
32	D	D						
33	T	A		D				
34	M	M		D				
35	K	K						Y
36	D	D						
37	A	A		D				
38	G	G						Y
39	M	M						Y
40	Q	Q	3	D				Y
41	M	M		D			O	Y
42	K	K					O	
43	K	K						
44	V	V		D				
45	L	L		D			O	
46	D	D					O	
47	S	T						
48	L	L		D				
49	P	P						
50	Q	Q						
51	T	T						
52	T	T		D				
53	R	R				G		
54	E	E				G		
55	N	N		D				
56	I	I				G		
57	M	I	4			G		
58	K	K						
59	L	L		D				
60	T	T		D		G		
61	E	E						
62	K	K		D				
63	I	I		D				
64	V	V		D				
65	K	K						
66	S	S		D				
67	P	P						
68	L	L	5	D				
69	C	C		D				
70	M	M						

Fig. 4. *Alignment of uteroglobin sequences from rabbit and hare together with outline of tertiary packing.* The fourth column gives the location of helices I – V as defined by the DSSP programme [56] from the crystallographic coordinates of the dimer [34]. The next column (D) gives residues that make van-der Waals contact in the uteroglobin dimer. The last four columns give residues that contact between helices 1 and 2, 1 and 4, 1 and 3 and 2 and 3

could the uteroglobin dimer structure be used as a model on which to base a tertiary structure prediction of the annexin repeat? The available sequence data cannot rule out the possibility of such tertiary structural similarity; however, two pieces of evidence suggest that uteroglobin is an inappropriate model. Even if we assume both proteins to have five helices, helix e in the annexin repeat is predicted to be about nine residues long; in contrast, only a short 3_{10}-helix is observed in the equivalent position in uteroglobin. In the light of this observation, it might be argued that the structural similarity only extends to helix d. However, helix b is predicted to be mostly buried, whilst in the uteroglobin structure, it is helix 3 that is least exposed to solvent. Together, these data suggest that the annexin repeat unit packs differently to the uteroglobin protein fold.

In conclusion, there is no compelling evidence from this sequence analysis that uteroglobin and the annexins share similar tertiary structures, or that uteroglobin represents a derivative of a primordial one-repeat structure which underwent two duplications to give the present day four-repeat annexins.

The repeat 3 a – b predicted loop

Several features of the loop joining predicted helices a and b in repeat 3 emerge from our analysis and are summarised here. (a) The loop is one residue longer than in repeats 1, 2 and 4. (b) It has a different pattern of conserved glycine residues, suggesting a shorter helix a. (c) The major difference between AVIx7 and all other repeats lie in the a – b loop; insertion/deletion of the VAAEIL segment may cause conformational changes in this region (Fig. 3). (d) In AIIM, the a – b loop of repeat 3 is the only a – b loop that is not split by an exon boundary.

These observations show that the a – b loop of repeat 3 is characteristically different from the a – b loops in repeats 1, 2 and 4 and suggests that the loop may perform a specific function in the intact annexin. Further support for this hypothesis comes from recent tryptophan fluorescence studies [38] that indicate the Trp residue at position 22 of AIB3 and AVB3 associates closely with the phospholipid in membrane-bound annexins. These data suggest that the uniqueness of the repeat 3 a – b loop may be due to a specific phospholipid binding function. However, the data are clearly insufficient to exclude the possibility that other regions of the annexin four-repeat unit also interact with phospholipid. It is intriguing to note that a recent study by Johnsson and Weber [31] shows the position 21/22 peptide bond of AIIP3 to be susceptible to proteolysis by trypsin, suggesting that it is in an exposed position on the molecule. If the a – b loop is indeed an important phospholipid interaction site, the observation that the nicked protein has similar Ca^{2+}-dependent phospholipid binding characteristics to the native form suggests that this region of the structure is stabilised in a form maintaining the function by non-covalent interactions.

CONCLUSIONS

In this paper a detailed sequence analysis of 88 conserved repeat regions from the annexin super-gene family has been performed. The general conclusions are as follows.

a) The annexin repeats clearly fall into four families corresponding to repeats 1, 2, 3 and 4. Repeats 5 – 8 of annexin VI fall into the same four families.

b) The analysis does not discriminate between gene duplication and fusion hypotheses for the origin of repeats 5–8 in annexin VI.

c) The repeat 3 family shows greatest diversity and is less similar to the repeat 1, 2 and 4 families than these families are to each other, suggesting that a contribution to annexin specificity is made by this repeat.

d) The repeat 4 family shows least diversity.

e) The secondary structure of the annexin repeat is predicted to consist of five helices.

f) The physico-chemical properties at 22 aligned positions are highly conserved in all repeats. Of these, 16 are in predicted α-helices (Table 2 and Fig. 2). Only the glycine at position 23 is totally conserved across all repeats.

g) Hydrophobic amino acids are conserved in patterns consistent with one face or edge of an α-helix in predicted helices a, c, d and e. Predicted helix b has five out of eight residues conserved as hydrophobic in all repeats, suggesting this helix is buried in the native annexin tertiary structure.

h) Position 31 in buried helix b is variable in repeats 1 and 3, absolutely conserved Glu in repeats 2 and absolutely conserved Arg in repeats 4, suggesting repeats 2 and 4 may interact via a salt bridge at this position.

i) Patterns of glycine conservation suggest that helix a is shorter in repeats 3 than in repeats 1, 2 and 4.

j) There is no compelling evidence from this analysis that uteroglobin and the annexins share similar tertiary structures, or that uteroglobin represents a derivative of a primordial one-repeat structure which underwent two duplications to give the present day annexins.

k) The loop between predicted helices a and b of repeat 3 shows features distinct from the equivalent loop in repeats 1, 2 and 4, suggesting an important structural and functional role for this region of the molecule.

The X-ray structure of human annexin V has recently been solved [11, 12] and confirms the majority of the observations and predictions made in this paper. A preliminary comparison of our predictions with the results of crystallography is contained in the Appendix.

We thank Drs C. J. Rawlings, J. Fox and Prof L. N. Johnson for their encouragement. GJB is a Royal Society University Research Fellow.

REFERENCES

1. Glenney, J. (1986) *J. Biol. Chem. 261*, 7247–7252.
2. Shahid, M. A., Geisow, M. J. & Burgoyne, R. D. (1989) *Nature 340*, 313–315.
3. Funakoshi, T., Heimark, R. L., Hendrickson, L. E., McMullen, B. A. & Funikawa, K. (1987) *Biochemistry 26*, 5572–5578.
4. Creutz, C. E., Zaks, W. J., Hamman, H. C., Crane, S., Martin, W. H., Gould, K. L., Oddie, K. M. & Parsons, S. J. (1987) *J. Biol. Chem. 262*, 1860–1868.
5. Flower, F. J., Wood, J. N. & Parente, L. (1984) *Adv. Inflammation Res. 7*, 61–69.
6. Saris, C. J. M., Tack, T., Kristensen, J. R., Glenney, J. R. & Hunter, T. (1986) *Cell 46*, 201–212.
7. Geisow, M. J., Fritsche, U., Hexham, J. M., Dash, B. & Johnson, T. (1986) *Nature 320*, 636–638.
8. Kretsinger, R. J. & Creutz, C. E. (1986) *Nature 320*, 573–575.
9. Weber, K. & Johnson, N. (1986) *FEBS Lett. 203*, 95–98.
10. Taylor, W. R. & Geisow, M. J. (1987) *Protein Eng. 1*, 183–187.
11. Huber, R., Romsich, J. & Paques, E.-P. (1990) *EMBO J. 9*, 3867–3874.
12. Huber, R., Schneider, M., Mayr, I., Romsich, J. & Paques, E.-P. (1990) *FEBS Lett. 275*, 15–21.
13. Crumpton, M. J. & Dedman, J. R. (1990) *Nature 345*, 212.
14. Johnston, P. A., Perin, M. S., Reynolds, G. A., Wasserman, S. A. & Sudhof, T. C. (1990) *J. Biol. Chem. 265*, 11382–11388.
15. Barton, G. J. & Sternberg, M. J. E. (1987) *J. Mol. Biol. 198*, 327–337.
16. Barton, G. J. & Sternberg, M. J. E. (1990) *J. Mol. Biol. 212*, 389–402.
17. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol. 48*, 443–453.
18. Barton, G. J. (1990) *Methods Enzymol. 183*, 403–428.
19. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of protein sequence and structure*, vol. 5 (Dayhoff, M. O., ed.) pp. 345–358, National biomedical research foundation, Washington DC.
20. Zvelebil, M. J. J. M., Barton, G. J., Taylor, W. R. & Sternberg, M. J. E. (1987) *J. Mol. Biol. 195*, 957–961.
21. Lim, V. I. (1974) *J. Mol. Biol. 88*, 873.
22. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol. 47*, 45–148.
23. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
24. Rose, G. D. (1978) *Nature 272*, 586–591.
25. Wilmot, A. C. M. & Thornton, J. M. (1988) *J. Mol. Biol. 203*, 221–232.
26. Reference deleted.
27. Taylor, W. R. (1986) *J. Theor. Biol. 119*, 205–218.
28. Amiguet, P., D'Eustachio, P., Kristensen, T., Wetsel, R. A., Saris, C. J. M., Hunter, T., Chaplin, D. D. & Tack, B. F. (1990) *Biochemistry 29*, 1226–1232.
29. Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (1986) *J. Mol. Biol. 187*, 101–107.
30. Richardson, J. S. & Richardson, D. C. (1988) *Science 240*, 1648–1652.
31. Johnsson, N. & Weber, K. (1990) *Eur. J. Biochem. 188*, 1–7.
32. Crompton, M. R., Owens, R. J., Totty, N. F., Moss, S. E., Waterfield, M. D. & Crumpton, M. J. (1988) *EMBO J. 7*, 21–27.
33. Miele, L., Cordella-Miele, E., Facchiano, A. & Mukherjee, A. B. (1988) *Nature 335*, 726–730.
34. Bally, R. & Delettre, J. (1989) *J. Mol. Biol. 206*, 153.
35. van Binsbergen, J., Slotboom, A. J., Aarsman, A. J. & de Haas, G. H. (1989) *FEBS Lett. 247*, 293–297.
36. Verger, R. & de Haas, G. H. (1973) *Chem. Phys. Lipids 10*, 127–136.
37. Newman, R. H., Freemont, P. S., Barton, G. J. & Crumpton, M. J. (1990) *Biochem. Soc. Trans. 18*, 1233–1234.
38. Meers, P. (1990) *Biochemistry 29*, 3325–3330.
39. Huang, K.-S., Wallner, B. P., Mattaliano, R. J., Tizard, R., Burne, C., Frey, A., Hession, C., McGray, P., Sinclair, L. K., Chow, E. P., Browning, J. L., Ramachandran, K. L., Tang, J., Smart, J. E. & Pepinsky, R. B. (1986) *Cell 46*, 191–199.
40. Sakati, T., Iwagami, S., Tsuruta, Y., Suzuki, R., Hojo, K., Sato, K. & Teraoka, H. (1988) *Nucleic Acids. Res. 16*, 11818.
41. Tamaki, M., Nakamura, E., Nishikubo, C., Sakati, T., Shin, M. & Teraoka, H. (1987) *Nucleic Acids Res. 15*, 7637.
42. Kristensen, T., Saris, C. J. M., Hunter, T., Hicks, L. J., Noonan, D. J., Glenney, J. R. & Hunter, T. (1986) Biochemistry *25*, 4497–4503.
43. Pepinsky, R. B., Tizard, R., Mattaliano, R. J., Sinclair, L. K., Miller, G. T., Browning, J. L., Chow, E. P., Burne, C., Huang, K.-S., Pratt, D., Wachter, L., Hession, C., Frey, A. Z. & Wallner, B. P. (1988) *J. Biol. Chem. 263*, 10799–10811.
44. Grundmann, U., Abel, K.-J., Bohn, H. H., Lobermann, H., Lottspeich, F. & Kupper, H. (1988) *Proc. Natl Acad. Sci. USA 85*, 3708–3712.
45. Weber, K., Johnsson, N., Plessmann, U., Van, P. H., Soling, H.-D., Ampe, C. & Vandekerckhove, J. (1987) *EMBO J. 6*, 1599–1604.
46. Hamman, H. C., Gaffey, L. C., Lynch, K. R. & Creutz, C. E. (1988) *Biochem. Biophys. Res. Commun. 156*, 660–667.
47. Kaplan, R., Jaye, M., Burgess, W. H., Schlaepfer, D. D. & Haigler, H. T. (1988) *J. Biol. Chem. 263*, 8037–8043.

48. Mauer-Fogy, I., Reutelingsperger, C. P. M., Pieters, J., Bodo, G., Stratowa, C. & Hauptmann, R. (1988) *Eur. J. Biochem. 174*, 582–592.
49. Iwaski, A., Suda, M., Nakao, H., Nagoya, T., Saino, Y., Arai, K., Mizoguchi, T., Sato, F., Toshizaki, H., Hirata, M., Miyata, T., Shidara, Y., Murata, M. & Maki, M. (1987) *J. Biochem. (Tokyo) 102*, 1261–1273.
50. Fernandez, M. P., Selmin, O., Martin, G. R., Yamada, Y., Pfaffle, M., Deutzmann, R., Mollenhauer, J. & von der Mark, K. (1988) *J. Biol. Chem. 263*, 5921–5925.
51. Moss, S. E. & Crumpton, M. J. (1989) *Trends Biochem. Sci. 14*, 325.
52. Sudhof, T. C., Slaughter, C. A., Leznicki, I., Barjon, P. & Reynolds, G. A. (1988) *Proc. Natl Acad. Sci. USA 85*, 664–668.
53. Moss, S. E., Crompton, M. R. & Crumpton, M. J. (1988) *Eur. J. Biochem. 177*, 21–27.
54. Burns, A. L., Magendzo, K., Shirvan, A., Srivastava, M., Rojas, E., Alijani, M. R. & Pollard, H. B. (1989) *Proc. Natl Acad. Sci. USA 86*, 3798–3802.
55. Hauptman, R., Mauer-Fogy, I., Krystek, E., Bodo, G., Andree, H. & Reutelingsperger, C. P. M. (1989) *Eur. J. Biochem. 185*, 63–71.
56. Kabsch, W. & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

APPENDIX

Shortly after we submitted this paper, two papers reporting the three-dimensional structure of human annexin V were published by R. Huber and co-workers [1, 2]. It is relatively unusual for the results of sequence analysis and prediction from the amino acid sequence to be followed so quickly by experimental X-ray data. As a consequence, this study presents a rare opportunity to assess the accuracy of structural inference from sequence information and predictive schemes.

Accordingly, this Appendix gives a brief discussion of the main observations and predictions made in the main paper with reference to the X-ray structure. Unfortunately, a more detailed evaluation of the observed residue preferences in the annexin repeats must await the release of the annexin V coordinates.

Delineation of repeat units

Table A1 illustrates the correspondence between the predicted repeat boundaries and those observed by Huber et al. [1]. Three residues from the C-terminal of repeat 1 were incorrectly assigned to repeat 2, two from the C-terminal of repeat 2 were assigned to repeat 3, whilst four residues from repeat 3 were assigned to repeat 4. This is a very close correspondence, given that small variations in the loop region repeat boundaries are to be expected between different members of the annexin family.

Secondary structure prediction

The consensus secondary structure prediction taken over all 88 repeats, as illustrated in Fig. 2, suggested the location of five distinct helices. This prediction is in overall agreement with the X-ray structure. Table A2 summarises a comparison of the consensus secondary structure prediction and a consensus secondary structure derived from the X-ray data by defining a helical residue wherever one occurs at an equivalent position in all four repeats.

The errors in prediction are confined to differences of up to two residues in the location of the helix terminii. The restriction to such small errors is a very encouraging result, since the exact definition of helix end points can be difficult, even from a high-resolution crystal structure (e.g. see [3]) and secondary structure prediction techniques normally give poor definition of structure ends.

Helix a of repeats 3 was predicted to be shorter than the equivalent helix in repeats 1, 2 and 4 on the basis of glycine conservation at position 18 rather than 21 in the a–b loop of repeats 3. Such a difference is not observed in the X-ray structure.

Conservation analysis and repeat packing

Inspection of the stereo plots of Huber et al. [1, 2] suggest that all the residues predicted to belong in the protein core (positions 11, 14, 17, 29, 30, 32, 33, 34, 44, 48, 56, 60, 69, 72, 73, 76 of Fig. 2) do take part in helix–helix interaction in the annexin tertiary structure.

The most striking feature identified by conservation analysis was the total identity at position 31, of Glu in repeats 2 and Arg in repeats 4 in the middle of hydrophobic helix b. The suggestion that helix b is buried in the annexin structure and that repeats 2 and 4 interact via a salt bridge between these two residues is confirmed by the X-ray structure.

Discussion

The main paper came to 11 conclusions, of which six (Conclusions e, g, h, i and j) concern structural features of the proteins that can be verified by the X-ray structure of human annexin V. The accuracy of these conclusions in the light of

Table A1. *X-ray repeat positions*

Repeat	Position numbering	
	Fig. 2	human annexin V
1	7–6 (of repeat 2)	17– 88
2	7–79	89–159
3	3–5 (of repeat 4)	168–246
4	6–78	247–317

Table A2. *Comparison of consensus secondary structure prediction over 88 annexin repeats with consensus structure from the four repeats of human annexin V*
Difference column gives residues over (+) or under (−) predicted at the N and C termini of each helix

Helix	Consensus secondary structure		Difference
	prediction	X-ray	N:C
a	9–19	7–18	−2:+1
b	27–34	26–36	−1:−2
c	40–51	40–51	0:0
d	53–61	55–61	+2:0
e	68–76	67–78	−1:−2

Table A3. *Summary of comparison between the predictions and observations based on sequence data and the X-ray structure of human annexin V*

Conclusion	Comments
e	Correct prediction of five helices to within two amino acids at each terminus
g	Correct identification of conserved hydrophobics as core residues and correct prediction that helix b is buried
h	Correct prediction of a salt bridge at position 31 between repeats 2 and 4
i	Incorrect prediction that helix a is shorter in repeats 3
j	Correct deduction that uteroglobin is not structurally related to the annexins
k	Correct identification of the a–b loop in repeats 3 as performing a different role to the equivalent loops in repeats 1, 2 and 4. However, did not determine that Ca^{2+} sites are only present in repeats 1, 2 and 4

the crystal structure is summarised in Table A3. Five of the six conclusions are confirmed by the X-ray structure, with the only incorrect suggestion being that helix a in repeats 3 is shorter.

When the annexin V coordinates are released, it may be possible to characterise the uncertainties in the predictions in a form that can be applied to other protein families where no tertiary structure is available. For example, the precise role of the differing glycine patterns in the a–b loops may be clear from the X-ray coordinates and provide rules that could aid the definition of similar helix end points. Although the secondary structure predictions in the main paper have now been superseded by the X-ray crystallographer's results, the analysis of residue conservation (Table 2 and Fig. 2) will be invaluable as a guide to interpretation of annexin-specific sequence preferences in conjunction with the X-ray structure. For example, a detailed evaluation of the coordinates together with the conservation data may explain why annexins V and VII can apparently form channels, whilst other annexins appear unable to do so [4].

The main paper did not attempt to predict the tertiary structure of an annexin repeat. However, given the accurate identification of the secondary structural elements, and the location of helix–helix packing residues, it should be possible to propose consistent tertiary models using techniques such as Cohen's combinatorial method (e.g. see [5]). These findings bode well for future attempts to infer structural features from aligned sequence data.

References

1. Huber, R., Romsich, J. & Paques, E.-P. (1990) *EMBO J. 9*, 3867–3874.
2. Huber, R., Schneider, M., Mayr, I., Romsich, J. & Paques, E.-P. (1990) *FEBS Lett. 275*, 15–21.
3. Richardson, J. S. & Richardson, D. C. (1988) *Science 240*, 1648–1652.
4. Rojas, E., Pollard, B., Haigler, H. T., Parra, C. & Lee Burns, A. (1990) *J. Biol. Chem. 265*, 21207–21215.
5. Cohen, F. E. & Kuntz, I. D. (1987) *Proteins Struct. Funct. Genet. 2*, 162–166.