Volume 335, number 2, 155–160 FEBS 13314 December 1993

The *Drosophila* ankyrin repeat protein cactus has a predominantly α-helical secondary structure

Nicholas J. Gay* and Monde Ntwasa

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Received 27 September 1993; revised version received 15 October 1993

The cactus protein is the *Drosophila* homologue of the mammalian IκB family of cytoplasmic anchor proteins. We have expressed in *E. coli* and purified a cactus fusion protein, CACT-Bgl. CACT-Bgl protein contains the six ankyrin repeat sequences which are necessary for specific binding to the *Drosophila* rel family transcription factor dorsal. We show that the purified CACT-Bgl protein can bind specifically to dorsal and, using circular dichroism spectroscopy, that the protein adopts a largely α-helical secondary structure. A further analysis of the ankyrin repeat domains of cactus, using an improved secondary structure prediction program indicates that the N-terminal of the repeat will form into a loop structure and the C-terminal section into an interrupted, amphipathic α-helix. On the basis of these findings we propose that the ankyrin repeats of cactus fold together into helical bundles interconnected by diverged loops.

Drosophila cactus protein; Ankyrin repeat; Secondary structure; α-Helix

1. INTRODUCTION

The product of the *Drosophila* cactus gene is a cytoplasmic anchoring protein required in the precellular embryo for the formation of a dorso-ventral nuclear gradient of the rel/NFκB transcription factor dorsal, a process that underlies the subsequent development of dorso-ventral pattern (for a review see [1]). It is thought that the dorsal gradient arises by a process of differential nuclear localisation. Activation of a specific receptor at ventral positions in the embryo causes dissociation of the cytoplasmic dorsal–cactus complex and allows dorsal to migrate into the ventral syncitial nuclei. The cactus product is a phosphoprotein of molecular mass 53 kDa which has both maternal and zygotic expression patterns. The predicted protein sequence contains an acidic domain in the N-terminal and in the C-terminal six copies of ankyrin (ank) repeat sequences [2,3].

Ank repeats are 33 amino acids in length of which 10 residues are highly conserved. Such repeats are also found in the mammalian cytoplasmic anchor proteins, IκB and MAD3, the cell cycle regulator cdc10, the *Drosophila* membrane receptor, notch, the mammalian oncogene bcl-3 and the prototypical molecule, erythrocyte ankyrin (see [4]). Although these molecules have diverse functions they hold in common the ability to mediate specific protein–protein interactions. For example ankyrin is able to bind specifically to a number of target molecules including the erythroid anion exchanger and tubulin while cdc10 forms part of a protein complex that regulates the expression of other cell cycle genes [5]. However, the sequences that confer binding to ankyrin repeat proteins do not appear to be conserved in the different target molecules and thus diverged features of the repeats are probably responsible for specific binding. There is also evidence in the case of the yeast transcriptional regulator SWI4 that the repeats may play a direct or indirect role in specific interactions with nucleic acids [6].

Previous studies of cactus have shown that the ankyrin repeat domains alone are sufficient and essential for binding to dorsal and that phophorylation of cactus is not necessary for such interactions [2,3]. To date there is little information as to the structure adopted by cactus or by the ankyrin repeat unit. In this paper we report the expression and the purification of a functional cactus fusion protein. We have used circular dichroism spectrocopy to show that the cactus fusion protein adopts a predominantly α-helical conformation. This finding is supported by secondary structure predictions generated using new methods based on neural networks and multiple sequence alignment [7]. Such analysis also suggests that the C-terminal section of each ankyrin repeat forms an α-helix with amphipathic properties.

2. EXPERIMENTAL

2.1. *Construction of pCACT-Bgl expression plasmid.*

A 1166 base pair *Bgl*II fragment corresponding to nucleotides 403–1569 of the cactus cDNA sequence [2,3] was cloned into pET3C [8] which had been cut with *Bam*HI and treated with alkaline phosphat-

*Corresponding author. Fax: (44) (223) 333 345.

ase. A recombinant plasmid with the fragment cloned in the correct orientation was recovered and transformed into *E. coli* strain BL21(DE3) pLysS. Such transformed cells express to high levels a protein estimated by SDS-PAGE to have a molecular mass of 50 kDa (see Fig. 1). This compares with an expected molecular mass of 45.2 kDa. However, it has been noted previously that cactus protein has an anomalously low mobility in SDS-PAGE [2]. The identity of the fusion protein was confirmed by direct protein sequence analysis. It contains 11 amino acids of the T7 gene 10 product at the N-terminal and is fused to an additional 19 residues at the C-terminus.

2.2. *Purification of CACT-Bgl protein*

BL21(DE3)pLysS cells carrying pCACT-Bgl were grown in 2 × TY medium supplemented with ampicillin (100 $\mu g \cdot ml^{-1}$) and chloramphenicol (20 $\mu g \cdot ml^{-1}$) (1 l) until the cultures had an optical density at 590 nm of 0.5 and then were induced by the addition of 0.8 mM IPTG (isopropyl β-D-thiogalactoside). After three hours the cells were harvested (10,000 × *g*, 10 min) and resuspended in 30 ml of a buffer containing 20% glycerol, 0.1% Nonidet P40, 20 mM Tris-HCl, pH 7.6, 0.1 M KCl, 1 mM dithiothreitol (DTT), 1 mM phenyl methyl sulphonyl fluoride, 2 mM benzamidine. The cells were broken by two passages through the French press at 15,000 psi. Insoluble material was removed by centrifugation (150,000 × *g*, 1.5 h) and the fusion protein was found to purify in the soluble fraction. The soluble fraction was treated with DNase and RNase (7 $\mu g \cdot ml^{-1}$) for 15 min at room temperature and then ammonium sulphate was added to 60% saturation. The precipitate was collected by centrifugation from 12.5 ml of solution and redissolved in 5 ml of 6 M guanidine hydrochloride, 25 mM NaCl, 20 mM Tris-HCl, pH 7.8, 0.1 mM EDTA (ethylene diamine tetra-acetic acid) 0.1 mM DTT and applied to a column of Sepharose 6B-CL (1.5 meters, 175 ml) which had been equilibrated in the same buffer. Fractions containing CACT-Bgl protein were pooled, concentrated in a Centriprep concentrator (Amicon Corporation) and dialysed against a buffer containing 4 M urea, 25 mM NaCl, 20 mM Tris-HCl, pH 7.2, 0.1 mM EDTA and 0.1 mM DTT. This sample was then applied to a column of DEAE-Sepharose 6B-CL (20 ml) (Pharmacia) which had been equilibrated in the same buffer. The column was eluted with a linear gradient of NaCl. The CACT-Bgl protein eluted at approximately 0.25 M NaCl.and the peak fractions were pooled and dialysed against a buffer containing 10 mM HEPES, pH 7, 0.15 M KCl, 0.1 mM DTT and 0.1 mM EDTA.

2.3. *Binding/immunoprecipitation reactions*

E. coli extract of dorsal protein was prepared using the plasmid pARdl as described previously [9,10]. 10 μg of dorsal extract and 10 μg of purified CACT-Bgl were incubated overnight at 4°C in a buffer containing 10% glycerol, 10 mM HEPES, pH 7, 1 mM $MgCl_2$, 0.5 mM EDTA, 5 mM DTT, 0.15 M KCl. The reactions were then centrifuged (14000 × *g*, 5 min). Purified rat anti-dorsal antibody (1/25th dilution) and 75 μl of a 50% slurry of Protein G-Sepharose (Pharmacia) were added and the solution was incubated for a further 2 hours. The immunoprecipitates were then washed 5 times in the above buffer supplemented with 0.1% Triton X-100.

2.4. *Circular dichroism spectroscopy*

Spectra were recorded using Jobin-Yvon CD6 machine between 190 and 240 nm in 0.5 nm steps. The sample spectrum was scanned five times and subtracted from an identical spectrum of the buffer used. The concentration of the protein sample was determined by amino acid analysis.

2.5. *Secondary structure prediction*

Secondary structure prediction was performed by a system of neural networks. The method uses multiple sequence alignments of groups of homologous proteins and has an accuracy of at least 71.4% [7]. The alignment file contained the sequences of human mad3 [11], human and mouse kbf1 [12], human bcl3 [13], *Drosophila* notch [14], human ankyrin [15], rat V1 protein [16] and the α-latrotoxin precursor [17].

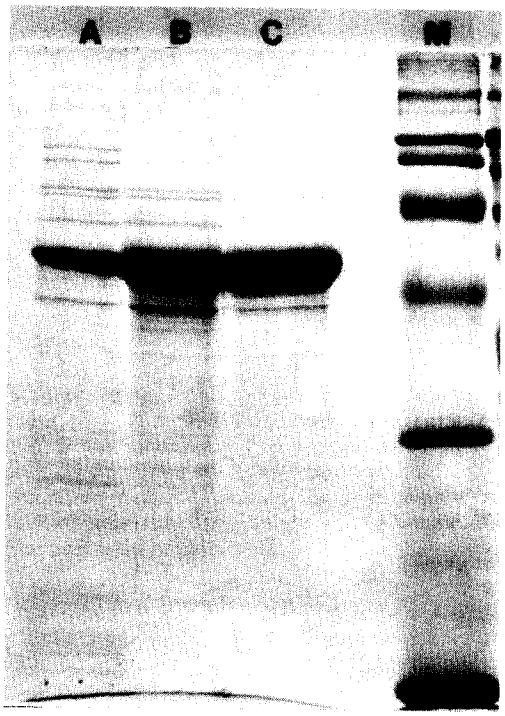


Fig. 1. Purification of CACT-Bgl protein. CACT-Bgl protein was purified as described in section 2.2. Samples of protein were separated by SDS-PAGE (12%). The gel was stained with Coomassie Blue R250. (A) Soluble fraction of induced BL21DE3(pLysS) (pCACT-Bgl) cells. (5 μg of protein). (B) CACT-Bgl protein after purification by gel filtration (10 μg of protein). (C) CACT-9Bgl protein after purification by anion exchange chromatography. (10 μg of protein) Molecular mass markers were of M_r 205, 116, 97.4, 66, 45, 29 and 14 kDa.

3. RESULTS

3.1. *Purified CACT-Bgl protein binds specifically to the dorsal protein*

As shown in Fig. 1A, CACT-Bgl protein is expressed to a high level in the soluble fraction of BL21(DE3) cells carrying pCACT-Bgl. The chromatographic fractionations described in section 2.2 further purify the protein such that it is greater than 95% pure (Fig. 1B,C). In order to determine whether the purified CACT-Bgl protein is able to bind specifically to the dorsal protein an immunoprecipitation assay was carried out as described above. The CACT-Bgl protein is purified by anti-dorsal antibody only when dorsal protein extract is included in the immunoprecipitation reactions (Fig. 2A–C). This finding is consistent with previous analyses which localised the dorsal binding domain of cactus to the region of the molecule containing the six ankyrin repeat sequences [2,3].

3.2. *The CACT-Bgl protein has a predominanly α-helical secondary structure*

The secondary structure of the purified CACT-Bgl protein has been analysed using circular dichroism spec-

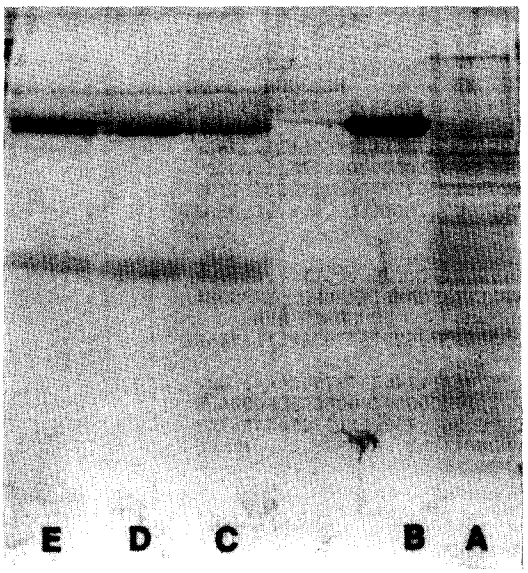


Fig. 2. CACT-Bgl protein binds to dorsal. Immunoprecipitation reactions were carried out as described in section 2.3. Samples were separated by SDS-PAGE (15%). The gel was stained with Coomassie blue R250. (A) Dorsal protein extract (2 μg); (B) CACT-Bgl protein (2 μg); (C) Immunoprecipitation reaction containing dorsal protein extract; (D) Immunoprecipitation reaction containing CACT-Bgl protein; (E) Immunoprecipitation reaction containing both dorsal and CACT-Bgl protein. CACT-Bgl protein purified in this reaction is seen to migrate slightly slower than the immunoglobulin heavy chain molecules. Molecular mass markers were of M_r 66, 45, 36, 29, 24, 20, 14 kDa.

troscopy. As shown in Fig. 3, the CD spectrum of CACT-Bgl protein has negative bands at 207 and 222 nm, characteristics which indicate the presence of α-helical structures in the protein sample. Analysis of the spectrum using the program CONTIN [18] suggests that 77% of the protein is α-helical with 23% being random coil.

3.3. *Secondary structure prediction indicates that 20 amino acids in the C-terminus of the cactus ankyrin repeats will form into amphipathic α-helices*

A secondary structure prediction of cactus performed by a neural network system also indicates that the protein has a highly helical structure. The region of cactus corresponding to CACT-Bgl protein is predicted to be 53% helical, 39% loop and only 8% β-sheet (data not shown). In the ank repeat region alone this figure is 52% helix, 42% loop with no regions reliably predicted as β-sheet (Fig. 4). A closer inspection reveals that 9 amino-terminal residues of each repeat are generally predicted to form loop structures while the C-terminal 20 are forecast to form into a helical segment with an interruption around residue 9. Projection of the residues in the C-terminus of repeats 2 and 4 onto a helical wheel plot (Fig. 5A,B) and of the C-terminal consensus of all 6 repeats (Fig. 5C) shows that if these segments form into a helices they would have a strongly amphipathic character. An analysis of positional preferences within the N and C caps of the proposed α-helical section of the repeats indicates that conserved residues at N-CAP+ 1 (Pro), N-CAP +4 (Leu), C-CAP-3 (Leu) and C-CAP (Gly) are strongly preferred in α-helical segments of known structure [19].

The sequence labelled GAP in Fig. 4 which is absent in mammalian IκB molecules may constitute an additional, degenerate ankyrin repeat. The C-terminal residues 117–131 are significantly related to residues 5–20 of the repeat consensus. However, the N-terminal sequence does not match the consensus and the secondary structure prediction is distinct from the consistent pattern described above for the other repeats.

4. DISCUSSION

The results reported above are the first direct measurements of the secondary structure of cactus protein. Michaely and Bennett [4] have proposed previously that ankyrin repeats form into an α–β type structure. Their model was based on a consensus sequence derived from a number of ankyrin repeats which takes as the N-terminal residue that designated in this and other work as residue 7. This phasing of the repeat is supported by the position of intron/exon boundaries in ankyrin but in cactus the splice sites are not consistently placed in each repeat (repeat 1 – residue 10; repeat 3 – residue 1; repeat 4 – residue 7; repeat 5 – residue 4 [2,3]). Furthermore, an examination using the DIAGON comparison program confirms our assignment for the N-terminal of the repeat, at least in the case of cactus. We do not favour the model of Michaely and Bennet for the following reasons. Firstly, our CD spectra do not contain significant spectral elements that correspond to β-sheet structure but instead suggest the protein consists solely of helix and loop structures. Secondly, the N-terminal sections of the repeats are not predicted by the secondary structure program to form into extended structures. Rather they are predicted to have random coil or loop

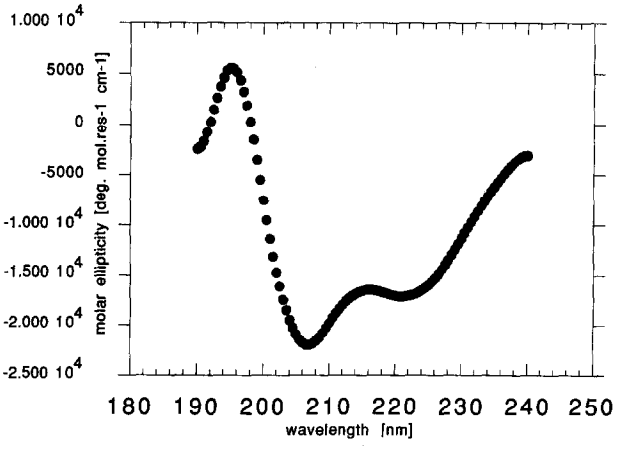


Fig. 3. Circular dichroism spectrum of CACT-Bgl protein.

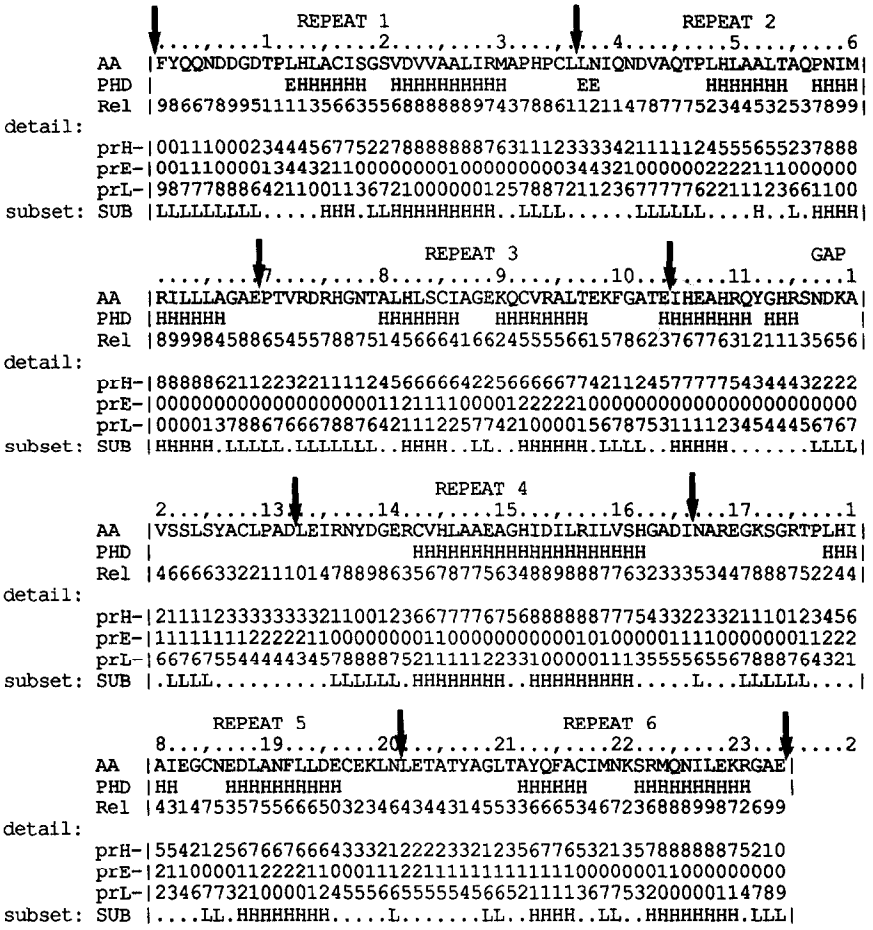
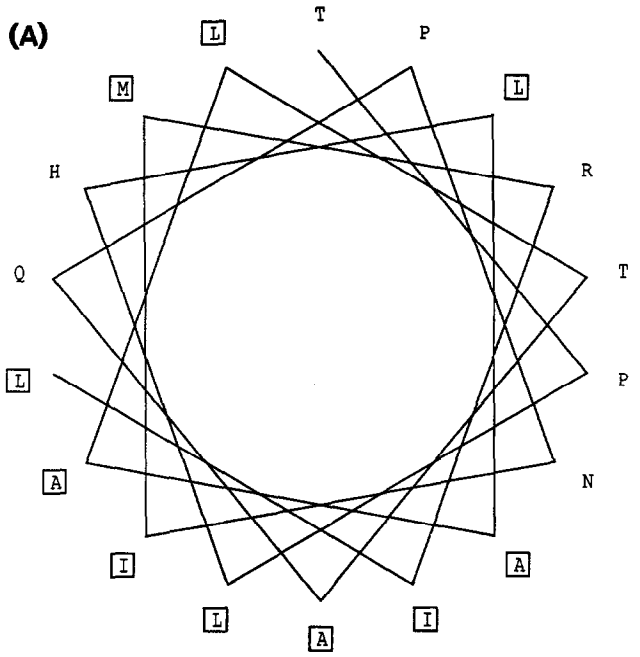
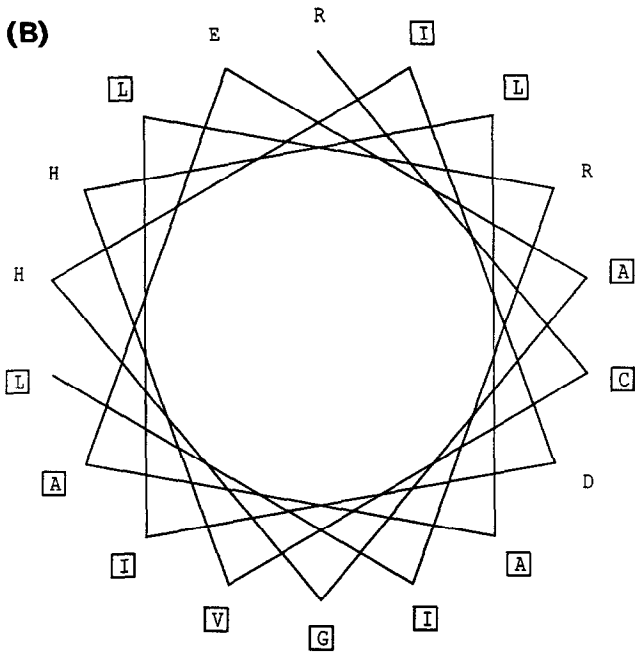


Fig. 4. Secondary structure prediction for the ank repeat domain of the cactus protein. The position of the six repeats is indicated by arrows above the sequence. The second line (PHD) indicates the detailed structure prediction: H = helix; E = extended structure. Blank spaces are predicted as loop. The third line (rel) gives a value for the reliability of the prediction (e.g. 5 = 50–60% confidence). Lines four to seven provide a breakdown of the prediction for each secondary structure and a subset of the prediction which has a reliability of 50% or greater (L = loop). The interval between repeat 3 and 4 marked 'gap' may constitute a seventh repeat (see section 3.3).

conformations, usually with high reliability values (Fig.4). Thirdly, if each repeat formed into an α–β structure two reverse turns would be needed in each unit, a requirement that would leave insufficient sequence to form a strand of equivalent length to the helical region. We prefer a model in which the C-terminal sections of the repeats form into helical bundles, the helices being interconnected by diverged loops. This hypothesis is consistent with the amphipathic property of the putative α-helical sections of the repeats and with the reported interdependence of the repeats for function and for folding [20,21]. As noted above, structure prediction suggests that the helical segment is interrupted at position 10 (Fig. 5) where glycine is found in the consensus. Such an interruption would introduce a flexible region into the helix and would not necessarily affect helix–helix packing. In addition, glycine is sometimes present in continuous helical segments. For example, in glutathione reductase a glycine residue is observed in the middle of a helix which packs together with its counterpart to form part of the subunit interface of the homodimer [22]. In cactus and other known cytoplasmic anchor proteins each of which have six ankyrin repeats it is possible that two three helix bundles are formed and that each bundle interacts with one subunit of the dimeric rel transcription factors. Triple helical bundles have been characterised in the structures of Protein A [23], of spectrin [24] and of the synthetic peptide, sercoil [25]. In the latter case, the helices have a hydrophobic surface similar to though more restricted than that of the cactus ankyrin repeat consensus (Fig.5) and the bundles are formed from two peptide molecules running parallel and one anti-parallel. A second possibility is that all six helical regions pack together into a hexagonal array, a hypothesis that is consistent with the finding in ankyrin that individual repeats do not fold independently but in specific ordered groups of six [19]. Such a model requires that the seventh partial repeat folds independently and is not essential for cactus function.

In view of the diversity of target molecules for ankyrin repeat proteins, it is unlikely that the conserved sequences are involved in specific binding. Instead they





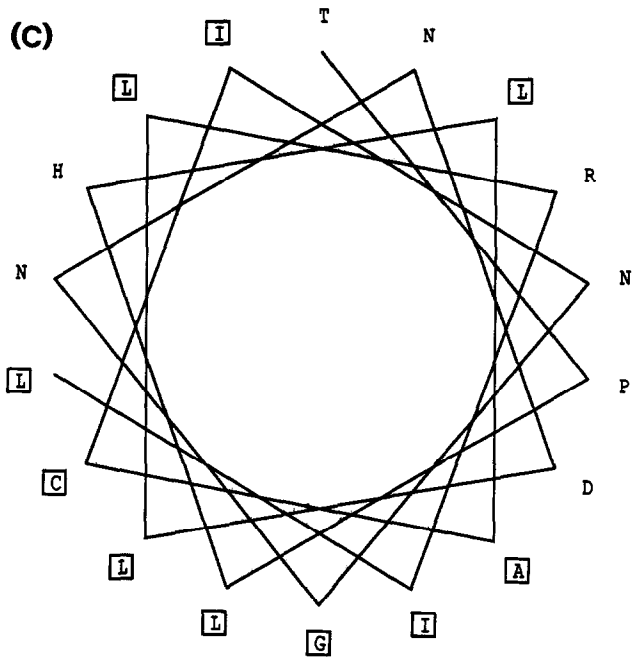


Fig. 5. The C-terminal section of the cactus ankyrin repeats would form into amphipathic α-helices. (A) Helical wheel plot of residues 10–27 of repeat 2 (see Fig. 4); (B) Helical wheel plot of residues 10–27 of repeat 4 (see Fig. 4); (C) The consensus sequence for residues 10–30 of the cactus anykrin repeat sequence projected onto a helical wheel plot. Boxed residues are hydrophobic. Consensus used = TPLHLACINGNNDLLRIL (see [2,3]). N = any amino acid. Note: The glycine residue at position 10 in the consensus may cause an interruption in the helical segment.

may be a consequence of a common structural requirement such as helix–helix packing. It is interesting to note in our model that diverged loop sequences are predicted to inter-connect conserved helical regions. This arrangement is analagous in some respects to that of immunoglobulin molecules in which specific binding is mediated by hypervariable loops supported upon a common structural framework.

Acknowledgements: N.J.G. is supported by a University Reseach Fellowship of the Royal Society and MN by the Canon Collins Educational Trust for South Africa. We thank Dr. J.E. Walker for critical reading of the manuscript.

REFERENCES

[1] Steward, R. and Govind, S. (1993) Curr. Opinion Genet. Dev. Biol. 3, 556–561.
[2] Kidd, S. (1992) Cell 71, 623–635.
[3] Geisler, R., Bergmann, A., Hiromi, Y. and Nusslein-Volhard, C. (1992) Cell 71, 613–621.
[4] Michaely, P. and Bennett, V. (1992) Trends Cell Biol. 2, 127–129.
[5] Lowndes, N., McInerny, C., Johnson, A., Fantes, P. and Johnston, L. (1992) Nature 355, 449–453.
[6] Sidorova, J. and Breeden, L. (1993) Mol. Cell. Biol. 13, 1069–1075.
[7] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7558–7562.
[8] Rosenberg, A., Lade, B., Chui, D., Lin, S., Dunn, J. and Studier, W.F. (1987) Gene 56, 125–135.
[9] Rushlow, C.A., Han, K., Manley, J. and Levine, M. (1989) Cell 59, 1165–1177.
[10] Ip, Y.T., Kraut, R., Levine, M. and Rushlow, C.A. (1991) Cell 64, 439–446.
[11] Haskill, S., Beg, A.A., Tompkins, S.M., Morris, J.S., Yurochko, A.D., Sampson-Johannes, A., Mondal, K., Ralph, P. and Baldwin, A.S. (1991) Cell 65, 1281–1289.
[12] Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M.B., Kourilsky, P., Baeuerle, P. and Israel, A. (1990) Cell 62, 1007–1018.
[13] Ohno, H., Takimoto, G. and McKeithan, T.W. (1990) Cell 60, 991–997.

[14] Fehon, R., Johansen, K., Rebay, I. and Artavanis-Tsakonas, S. (1991) J. Cell Biol. (1991) 113, 657–669.
[15] Lux, S., John, K. and Bennett, V. (1990) Nature 344, 36–42.
[16] Taoka, M., Yamakuni, T., Song, S.Y., Yamakawa, Y., Seta, K., Okuyama. and Isobe, T. (1992) Eur. J. Biochem. 207, 615–620.
[17] Volkova, T.M., Galkina, T.G., Kudelin, A.B., Nazimov, I.V. and Grishin, E.V. (1991) Bioorg. Khim. 17, 437–441.
[18] Provencher, S.W. and Glockner, T. (1981) Biochemistry 20, 33–37.
[19] Richardson, J.S. and Richardson, D.C. (1988) Science 240, 1648–1655.
[20] Nolan, G.P. and Baltimore, D. (1992) Curr. Opinion Genet. Dev. 2, 211–220.
[20] Michaely, P. and Bennett, V. (1992) Mol. Biol. Cell. 3, S-264.
[21] Deisenhofer, J., Jones, T.A., Huber, R., Sjodahl, J. and Sjoquist, J. (1977) Hoppe-Seyler Z. Physiol. Chem. 359, 975–982.
[22] Karplus, P.A. and Schulz, G.E. (1987) J. Mol. Biol. 195, 701–729.
[23] Speicher, D.W. and Manhesi, V.T. (1991) Nature 311, 177–179.
[24] Lovejoy, B., Choe, S., Cascio, D., McKorie, D.K., DeGrado, W.F. and Eisenberg, D. (1993) Science 259, 1288–1293.