Volume 262, number 1, 87-92 FEBS 08198 March 1990

Structural predictions for the central domain of dystrophin

R.A. Cross, Murray Stewart and J. Kendrick-Jones

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Received 27 December 1989

The animo acid sequence of dystrophin indicates that the molecule has globular N- and C-terminal domains separated by a long central rod domain. The central rod contains multiple repeats, about 100 amino acids long and of variable length. These diverge sufficiently in sequence that, in previous studies, only 14 of the most similar repeats have been aligned and analysed in any detail. We show here that a heptad pattern of hydrophobic residues is preserved across all repeats. Using the heptad pattern together with a consensus sequence template, we identified and aligned 25 repeats in the dystrophin rod sequence. Each repeat consists of a constant-length core helix of 54 residues, coupled via a short linker to a weakly conserved variable-length helix, and then via a second linker to the next core. The variable-length helix appears truncated in repeats 10 and 13 and extended in repeats 4 and 20. The extension of repeat 20 is particularly interesting since it corresponds to a hotspot of dystrophy-inducing mutations. Detailed modelling suggests that the classical Speicher-Marchesi [(1984) Nature 311, 177–180] model for spectrin may not be appropriate to dystrophin without some modification. We propose that whilst the repeating structural motif in dystrophin is probably a bead of triple coiled coil, this bead is twice as massive as, and out of phase with, those proposed for spectrin. Our model raises the possibility that the rod domain of dystrophin may confer elasticity on the molecule. Deletions which truncate this region would then reduce the extensibility of the molecule without affecting actin crosslinking, consistent with their typically producing the relatively benign Becker phenotype of muscular dystrophy.

Dystrophin; Rod domein; Duchenne dystrophy; Triple coiled coil; Structure prediction; Heptad repeat

1. INTRODUCTION

Boys suffering from Duchenne muscular dystrophy are deficient in a protein called dystrophin. This deficiency leads to severe muscle wasting and early death, sometimes accompanied by mental retardation. The function of dystrophin in muscle and other excitable tissues is unknown, although there is evidence suggesting it is a component of the membrane-associated cytoskeleton [2–9].

A milder and rarer form of muscular dystrophy, termed Becker dystrophy, appears to be produced by mutations which alter, rather than delete, dystrophin. These Becker mutations offer a route to understanding dystrophin function. Data on the extent of Becker deletions are increasing in both quantity and resolution, but their interpretation is limited by a lack of knowledge of the structure of dystrophin. Dystrophin preparations suitable for high-resolution structural analysis have not yet been produced and, moreover, the large size and low abundance of dystrophin may make such structural investigation particularly difficult.

Analysis of the amino-acid sequence of dystrophin offers an alternative source of structural information. Initial analyses identified substantial homologies between dystrophin, α-actinin and spectrin [7,10,12]. The N-terminal domain (residues 1–238) of dystrophin is homologous to that of α-actinin, and is thought to constitute an actin-binding site [7,10,12]. A similar sequence is found in the *Dictyostelium* gelation factor [13]. The C-terminus of dystrophin is almost identical in chicken and man, and consists of 2 domains; one containing an α-actinin-like sequence of 150 residues which contains (possibly nonfunctional) Ca-binding EF hands, and a tissue-specific stretch of about 300 residues. The remaining ~2800 residues of dystrophin form the massive central rod domain, which is characterised by a repeating sequence motif about 100 residues in length. Koenig et al. [7] counted 26 such repeats and have presented a multiple pairwise alignment of the 14 of them. Here we extend their work by analysing all the rod domain repeats to determine their number, the conserved features of their sequences, and their likely conformation. The results allow us to propose an outline structural model. The model offers a testable framework within which to understand the structural effects of deletions and mutations in this region.

2. THE HOMOLOGY OF DYSTROPHIN TO SPECTRIN AND α-ACTININ

The dystrophin repeats show homology to those in the central rod domains of spectrin and α-actinin [7,11]. The homology to α-actinin is the more striking, largely because the repeat length in dystrophin and α-actinin is variable, whereas in spectrin it is exactly 106

Correspondence address: R.A. Cross, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

residues. The observed homologies suggest that dystrophin, like spectrin and α-actinin, is composed of two chains which associate antiparallel. Work using proteolytic fragments of α-actinin has indicated that dimerisation depends mainly on interactions between the N- and C-termini of adjacent chains [14]. Electron micrographs of spectrin dimers also reveal tight association between the ends of the two chains [15]. A symmetry argument supports the view that dystrophin dimerises similarly: there are examples of patients with sizeable deletions within the dystrophin rod domain, but with barely impaired muscle function [16]. Such deletions would necessarily be assymetric in an antiparallel dimer, and would be expected to be very disruptive were specific interactions between chains to be critical for maintaining the structure of the molecule. In the following analysis we make the assumption that stabilising contacts occur principally within a single chain, rather than between chains.

3. A HYDROPHOBIC HEPTAD MOTIF IS CONSERVED IN ALL THE DYSTROPHIN REPEATS

The amino-acid residues in the dystrophin rod have a high α-helical potential [7]. In places, a heptad repeat is also present so that, in consecutive runs of seven residues *a-b-c-d-e-f-g*, residues *a* and *d* are usually hydrophobic, whilst residues at other positions are frequently hydrophilic. Fourier transforms [17,18] of the distribution of hydrophobic residues in the dystrophin rod showed significant intensity maxima at orders of 7 residues, confirming the presence of a strong heptad repeat (fig.1). The heptad pattern is characteristic of an α-helical coiled coil conformation [19], and suggests that coiled coils are the major structural element in the dystrophin rod domain. The Fourier transforms have no significant maxima at the longer intervals characteristic of interactions between coiled coils, as for example in the backbone of myosin filaments or intermediate filaments [20]. Although internal irregularities in the dystrophin sequence could potentially mask such longer repeats, this lack of longer-range repeats supports our initial assumption (section 3) that the major stabilising contacts in the dystrophin structure occur within each chain rather than between chains. The overall homology between the dystrophin repeats is too low for an unambiguous alignment to be made using multiple pairwise comparison computer methods. As a result, only 14 of the most similar repeats have so far been analysed in any detail [7]. The present analysis revealed that although the amino acid sequences of the repeats diverged considerably, the hydrophobic heptad pattern was strongly conserved between them. Using this heptad pattern together with a consensus sequence template, we identified and aligned 25 repeats (fig.2). The alignment visualises more exactly the extent of the heptad, and identifies further structural features in common between the repeats.

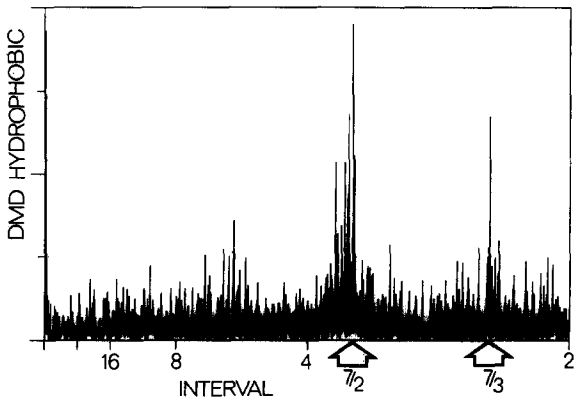

Fig.1. Fourier analysis of the periodicity of hydrophobic residues in the dystrophin rod. The analysis [17,18] reveals peaks of hydrophobic character at orders of 7 residues (7/2 and 7/3), corresponding to positions *a* and *d* in the repeating heptad *abcdefg*.

The alignment was achieved in a number of steps. Initial DIAGON [21] matrix comparison plots of the sequence indicated that there was a core region of each repeat, of length 54 residues, that seemed to be more strongly conserved than the rest of the repeat. These core regions often contained two tryptophan residues separated by 34 residues (although in several instances the tryptophan was substituted by another large hydrophobic residue such as tyrosine). We used the repeats 3,7,8,13,17,18,22 and 23, which contained both tryptophans, to produce a consensus sequence for the core region and then used this consensus sequence to check the alignment of the cores of repeats 2,4,5,9, 10,12,16,19,21,24 and 25, in which one tryptophan had been substituted. This alignment preserved the phase of the heptad pattern in each core, although this was not implicit in the analysis. We then used a consensus sequence derived from these 19 aligned cores to probe the sequence for any other cores, and found an additional 6 (1,6,11,14 and 15 in fig. 2). In all, 25 core regions were identified and aligned in this way and in all of them the phase of the heptad pattern was preserved. Lastly we observed that the second (variable length) helix in each repeat also exhibited a heptad motif, and we refined our alignment by bringing the heptad motif in these regions into register. To do this 1–3 padding characters were inserted within a 12–14 residue zone at the C-terminal margins of the cores. These regions contain glycine, proline and clusters of like-charges, suggesting they may form loops or turns [22]. Surface loops in other proteins are typically subject to residue insertions or deletions. Within this putative loop, the exact position of the padding characters did not influence the quality of the alignment, and the position shown was chosen for clarity.

The final alignment (fig.2) indicated that each repeat consisted of two helical regions, which we have

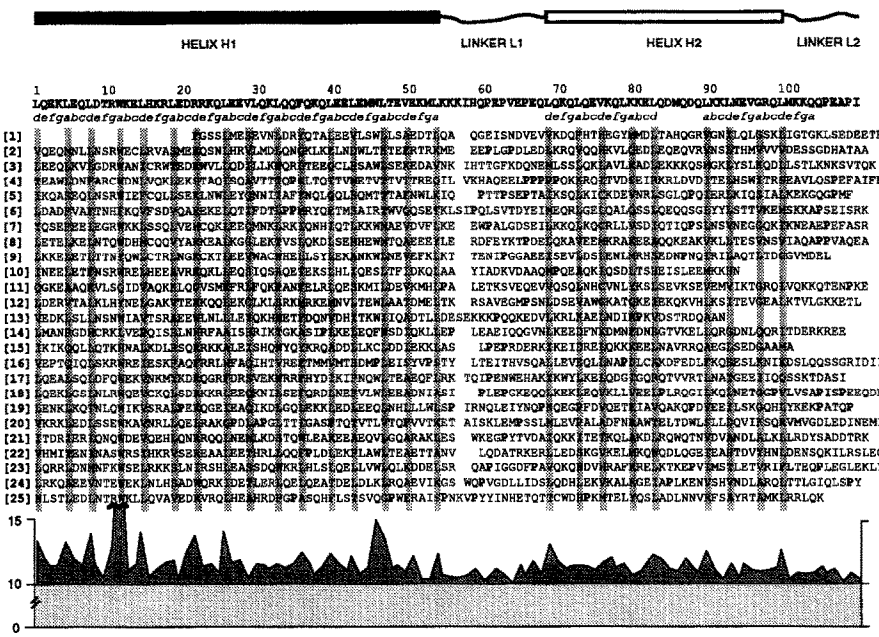

Fig.2. Alignment of the repeating sequences. Initial DIAGON analysis revealed that the repeat region runs from residue 400 to residue 2390. The repeat boundaries have been defined using protein-structural criteria, and differ in some cases from those interpolated from the exon boundaries of the dystrophin gene [7]. Whilst structural boundaries in a gene may correspond to those in a protein, often they do not [26]. The diagram (top) is aligned to correspond with the major predicted structural features of each repeat, namely a constant-length helix H1 and a variable-length helix H2, separated by linkers L1 and L2. The vertical shaded stripes mark the positions of the hydrophobic residues which form a repeating heptad motif. Proline residues within the sequence are shown in bold type, and their tendency to cluster in the turn regions can be seen. Within the linker L1 1–3 gaps have been inserted in order to optimise the alignment of the heptad across the H2 helices. A consensus sequence for the repeats (bold type, at top) was produced using an algorithm which searched for the residue at each position which maximised the similarity score summed across all the repeats. The graph (below) displays the degree of similarity for the residues at each position with this consensus sequence. Scores above 10 indicate a degree of conservation above the expectation for a random sequence. It is striking that in almost every case, the most conserved residues are those which form the repeating hydrophobic heptad.

designated H1 and H2, joined by two linkers, designated L1 and L2 (fig.2). Helix H1 was the more strongly conserved, heptad-rich core of the repeat, whereas L1 and L2 were putative loops or turns in the polypeptide chain, rich in proline and glycine. Helix H2 varied in length between repeats, and exhibited a weaker heptad pattern.

We checked the alignment by calculating scores for the similarity of each residue in each repeat to the consensus residue at that position. Average scores at each position across all 25 repeats are plotted beneath the alignment in fig.2, whilst table 1 lists the sum of these scores for each repeat. Generally (and at every residue position within the cores), the similarity to the consensus residue was consistently higher than would be expected for a random alignment of the sequences. It was striking that the level of conservation was highest for the hydrophobic residues which form the heptad pattern, suggesting that the heptad motif has indeed been selectively conserved.

4. CONSERVED BREAKS IN THE PHASE OF THE HEPTAD PATTERN

Key determinants of the tertiary structure adopted by the dystrophin rod domain are the lengths of regions of α-helical secondary structure, and the extent and position of loops or turns in the path of the chain. Small areas with low α-helix forming potential, thought to represent turns, have been identified in spectrin and α-actinin. Speicher and Marchesi [1] identified two such regions per spectrin repeat, and proposed that the chain undergoes a reverse turn at each, producing a structure 1/3 the length of an equivalent α-helix, in agreement with electron microscopic measurements [25]. Koenig et al. [7] have proposed that the dystrophin rod folds analogously. In contrast, Davison and Critchley [10] have proposed a modified model in which an additional turn occurs within the structural repeat of spectrin and α-actinin.

For dystrophin as noted above, the H1 core helices are flanked by L1 and L2, regions in which the heptad pattern is interrupted, and helix-breaking residues are common. We searched for other points of weakness in the predicted α-helical coiled coil conformation of dystrophin. There are signs of such weakness at position *d* of the third heptad in helix H1, where frequently a basic residue is found (fig.2). However the continuity of the phase of the heptad repeat argues against a break in the coiled coil at this point. Analogous regions containing a basic residue in one of the heptad hydrophobic positions are found in myosin rod [23] and tropomyosin

Table 1

Calculated similarity of the core of each repeat to the core of the consensus sequence

Repeat	Similarity score	Probability of a chance match
[1]	–	–
[2]	601	$5.1 \cdot 10^{-5}$
[3]	665	$<1.0 \cdot 10^{-10}$
[4]	592	$2.4 \cdot 10^{-4}$
[5]	607	$1.7 \cdot 10^{-5}$
[6]	582	$1.2 \cdot 10^{-3}$
[7]	691	$<1.0 \cdot 10^{-10}$
[8]	667	$<1.0 \cdot 10^{-10}$
[9]	643	$8.0 \cdot 10^{-9}$
[10]	658	$2.0 \cdot 10^{-10}$
[11]	577	$2.5 \cdot 10^{-3}$
[12]	656	$3.0 \cdot 10^{-10}$
[13]	667	$<1.0 \cdot 10^{-10}$
[14]	570	$6.4 \cdot 10^{-3}$
[15]	632	$1.0 \cdot 10^{-7}$
[16]	582	$1.2 \cdot 10^{-3}$
[17]	688	$<1.0 \cdot 10^{-10}$
[18]	723	$<1.0 \cdot 10^{-10}$
[19]	646	$3.9 \cdot 10^{-9}$
[20]	576	$2.8 \cdot 10^{-3}$
[21]	656	$3.1 \cdot 10^{-10}$
[22]	675	$<1.0 \cdot 10^{-10}$
[23]	670	$<1.0 \cdot 10^{-10}$
[24]	668	$<1.0 \cdot 10^{-10}$
[25]	602	$4.3 \cdot 10^{-5}$

The similarity score for each repeat is the sum of the scores for the similarity of each residue to the corresponding residue in the consensus sequence. The probability that the observed match could have arisen by chance was calculated acording to [27]. In 18 of the 25 repeats this probability was estimated at less than 1 in 10^4. However, in 6 repeats (numbers 4,6,11,14,16 and 20), the degree of homology was less marked. These were usually the repeats that lacked the fiducial tryptophans and which also sometimes contained proline residues. On the basis of overall homology alone the alignment of these repeats could be problematic, but the clear presence of a conserved heptad pattern within each permitted an alignment to be completed

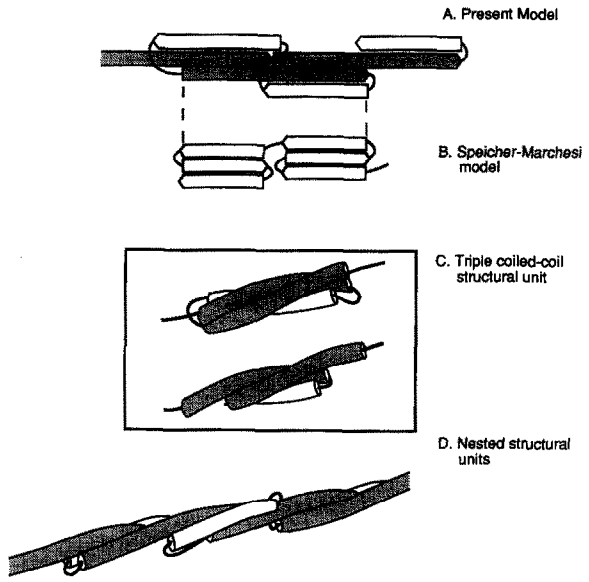

Fig.3. Predicted structure of the dystrophin rod. (A) The expected topology of the structural repeat in the dystrophin rod, given the predicted lengths of the helices H1 and H2 of each repeat. The model distinguishes 2 types of helix, H1 (shaded) and H2 (unshaded). Its main feature is the effectively continuous rod formed by the H1 helices, which interact strongly with one another. The H2 helices wrap back to allow this interaction to occur. (B) The Speicher and Marchesi [1] model for spectrin is aligned for comparison. In dystrophin the H1 and H2 helices probably wrap together to form beads of triple coiled coil, as shown in (C). Comparison with known coiled coils suggests that the pitch here would be as shown, about twice the length of H1. If H2 is sufficiently long then almost complete overlap of successive H1s could occur, as in the upper example. The length of most copies of H2 is such that overlap of H1s will be incomplete, as in the lower example. (D) Illustration of how successive copies of this structural bead may nest together. In this case the entire red domain would have a total predicted length of about 100 nm, but note that the structure could have some elasticity (see text).

[18], which are known to form unbroken coiled coils. It seems likely that the long side-chains of arginine or lysine are sufficiently flexible that they can effectively maintain the hydrophobic core between the two chains in the coiled coil while still allowing their charged groups access to the solvent. Fig.2 also shows that proline residues occasionally interrupt the coiled coil towards the centre of H1. These will locally open out the α-helix, and could introduce a kink into the axis of the coiled coil. They do not however interrupt the phase of the heptad pattern, and their distorting effects are expected to be local. Such short-range distortion occurs, for example, in the membrane-spanning helices of the purple membrane protein bacteriorhodopsin, into which single proline residues introduce a kink.

Another potential turn occurs towards the centre of the H2 helices of dystrophin, where there is a consistent 3 residue insertion, producing a shift in the phase of the heptad motif (fig.2). The unusually short repeats 10, and 13 terminate in this area, and in some of the longer repeats, prolines are interposed at this position, further suggesting that the predicted α-helical secondary structure is interrupted at this point. If the α-helix were to be rigid, such insertions might be expected to introduce a substantial deformation into a coiled coil. However, known coiled coils appear to be sufficiently flexible to be able to accomodate skip residues without substantial weakening of the coiled coil structure [23]. Indeed the function of the skip insertions into the myosin rod has been suggested to be to relieve an accumulated stress inherent to long stretches of coiled coil. We feel that the skip residues in dystrophin may produce a localised distortion of the coiled coil, but probably not a turn in the chain at this position, as suggested by [11] to occur in spectrin and α-actin.

5. STRUCTURAL MODELS

Based on our analysis, it is possible to advance plausible models for the structure of the dystrophin rod do-

main. The dominant stabilising contacts in the structure are likely to be between the H1 helices, because they are more strongly conserved between repeats, of constant length, and have an unbroken heptad repeat. Moreover, the presence of this heptad pattern and its conservation between repeats (fig.2) strongly suggests that the H1 helices interact by forming coiled coils. The H2 helices are typically shorter, of variable length, and contain a skip in the heptad pattern. They probably therefore interact less strongly with a partner helix, and so would have a secondary role in the generation of tertiary structure. In order for the H1 helices to overlap maximally, the H2 helices would have to fold back. They might then interact with the two parallel H1 helices to form a 3-chain coiled coil of increased stability. Two and three chain coiled coils are expected to have the same heptad repeat [18].

The topology of this model is illustrated in fig.3A. The Speicher and Marchesi [1] model for spectrin is aligned for comparison in fig.3B. In the dystrophin model, the degree of overlap of successive H1 helices would be limited by the length of H2. Fig.3C illustrates the extreme cases of this type of arrangement. In the upper example, the connecting H2 helix is short enough to limit considerably the overlap of the two H1 helices. In the other, a longer H2 allows the two H1 helices to overlap fully. In both cases the skip residues in the centre of H2 correspond to the centre of the bead of α-helical triple-coiled coil. The length of most copies of H2 is such that the partially overlapping structure would be formed. Successive beads of triple coiled coil could nest together to form a continuous rod, and alternate H2 helices might then occupy alternate grooves in the coiled coil formed by the H1 helices, as shown in fig.3D. This nested structure might appear stiff, but studies on other coiled coils (reviewed by Stewart et al. [24]) indicate that these are flexible enough to be consistent with the electron microscopic appearance of the spectrin molecule [25]. An interesting possibility is that this model structure might have elastic properties. The net formed by spectrin molecules certainly has considerably elasticity, and it is possible that dystrophin molecules too may be elastic. Stretching the structural bead shown in fig.3C would reduce the overlap between successive H1 helices and tend to dissociate the H2 helices from the triple coil. An opposing return force would come from the tendency of the H1 helices to redevelop maximum overlap.

As shown in fig.3, this outline model for the dystrophin rod differs in several respects from that of Speicher and Marchesi [1] for spectrin, which has previously been taken to apply to dystrophin. First, the H1 helices are continuous instead of having a central break. Second, in our dystrophin model the H1 α-helices wrap around one another to form a coiled coil, probably reinforced by helix H2 binding to produce a triple coil. The hydrophobic heptad motif described here strongly suggests that such a structure forms. Third, the repeating structural unit in our model, the bead of triple-coiled coil, is twice as massive as, and is out of phase with, that in the Speicher-Marchesi model. Common features between the two models include their total predicted lengths (which are approximately equal), and the positions of two of the turns or backfoldings in the primary chain.

6. STRUCTURAL CORRELATES OF THE MUTATIONAL HOTSPOT

Fig.2 illustrates that although the helix H2 is typically 40 to 50 residues long, there is a considerable variation around this length. The H2 helices of repeats 10 and 14 appear truncated, whilst those of repeats 4 and 20 are extended by about 40 residues. In all these cases however the heptad motif is conserved. The extension of repeat 20 is particularly interesting since it corresponds in position to a hotspot of dystrophy-inducing mutations in the dystrophin gene [16]. Parts of this repeat are unusually proline-rich, and strongly hydrophobic, suggesting they may fold into a globular subdomain, rather than having an extended rod structure. Possibly this region represents a binding site for some other component of the cortical cytoskeleton.

7. CONCLUSIONS

We have analysed in detail the ~100-residue repeats in the sequence of the dystrophin rod domain, to identify patterns of secondary structure that can be intergrated into a plausible structural model. The model we propose derives support principally from our observation of an extensive heptad pattern, present in every repeat. It relies also on extrapolation from the properties of known coiled coils, and from the known properties of spectrin and α-actinin. Once dystrophin is isolated in workable quantities, it should be possible to test the model biochemically and by electron microscopy.

Acknowledgements: We wish to thank Drs Roger Staden and Andrew McLachlan for their very helpful advice on sequence analysis.

REFERENCES

[1] Speicher, D.W. and Marchesi, V.T. (1984) Nature (Lond.) 311, 177–180.
[2] Hoffman, E.P., Brown, R.H. and Kunkel, L.M. (1987) Cell, 51, 919–928.
[3] Hoffman, E.P., Knudson, C.M., Campbell, K.P. and Kunkel, L.M. (1987) Nature (Lond.) 330; 754–758.
[4] Hoffman, E.P., Hudecki, M.S., Rosenberg, P.A., Pollina, C.M. and Kunkel, L.M. (1988) Neuron, 1, 411–420.
[5] Arachata, K., Ishiura, S., Ishiguro, T., Tsukhara, T., Suhara, Y., Eguchi, C., Ishihara, T., Nonaka, I., Ozawa, E. and Sugita, H. (1988) Nature (Lond.) 333, 861–863.

[6] Bonilla, E., Samitt, C.E., Miranda, A.F., Hays, A.P., Salvati, G., DiMauro, S., Kunkel, L.M., Hoffman, E.P. and Rowland, L.P. (1988) Cell 54, 447–452.
[7] Koenig, M., Monaco, A.P. and Kunkel, L.M. (1988) Cell 53, 219–228.
[8] Watkins, S.C., Hoffman, E.P., Slayter, H.S. and Kunkel, L.M. (1988) Nature (Lond.) 333, 863–866.
[9] Zubrzycka-Gaarn, E.E., Bulman, D.E., Karpati, G., Burghes, A.H.M., Belfall, B., Klamut, H.J., Talbot, J., Hodges, R.S., Ray, P.N. and Warton, R.G. (1988) Nature (Lond.) 333, 466–469.
[10] Davison, M.D. and Critchley, D.R. (1988) Cell 52, 159–160.
[11] Davison, M.D., Baron, M.D., Critchley, D.R. and Wooton, J.C. (1989) Int. J. Biol. Macromol. 11, 81–90.
[12] Hammonds, R.G. (1987) Cell, 51, 1.
[13] Noegel, A., Rapp, S., Lottspeich, F., Sleicher, M. and Stewart, M. (1989) J. Cell Biol. 109, 607–618.
[14] Blanchard, A., Ohanian, V. and Critchley, D. (1989) J. Muscle Res. Cell Mot. 10, 280–289.
[15] Goodman, S.R., Krebs, K.E., Whitefield, C.F., Riederer, B.M. and Zagon, I.S. (1988) CRC Crit. Rev. Biochem. 23, 171–234.
[16] Love, D.R., Forrest, S.M., Smith, T.J., England, S., Flint, T., Davies, K.E. and Speer, A. (1989) Br. Med. J. 45, 659–680.
[17] Stewart, M. and McLachlan, A.D. (1975) Nature (Lond.) 257, 331–333.
[18] McLachlan, A.D. and Stewart, M. (1975) J. Mol. Biol. 97, 293–304.
[19] Crick, F.H.C. (1953) Acata Crystallog. 6, 689–697.
[20] Stewart, M., Quinlan, R.A., Moir, R.D., Clarke, S.R. and Atkinson, S.R. (1989) In: Cytoskeletal and Extracellular Proteins: Structure, Interactions and Assembly (Aebi and Engel eds.) pp. 150–159, Springer, Heidelberg.
[21] Staden, R. (1982) Nucl. Acids Res. 10, 4731–4751.
[22] Cohen, C. and Parry, D.A.D. (1986) Trends Biol Sci. 11, 245–248.
[23] McLachlan and Karn, (1983) J. Mol. Biol. 164, 605–626.
[24] Stewart, M., Mclachlan, A.D. and Calladine, C.R. (1987) Proc. R. Soc. Lond. Ser. B 229, 381–413.
[25] Tyler, J.M., Anderson, J.M. and Branton, D. (1980) J. Cell. Biol. 85, 489–495.
[26] Rogers, J. (1989) Trends Genet. Sci. 5, 213–216.
[27] McLachlan, A.D. (1971) J. Mol. Biol. 61, 409–424.