Coiled-coil assembly by peptides with non-heptad sequence motifs

Matthew R Hicks[1], David V Holberton[2], Christopher Kowalczyk[1] and Derek N Woolfson[1]

Background: The seven-residue heptad repeat is the accepted hallmark of coiled coils. In extended filamentous proteins, however, contiguous patterns of heptads are often disrupted by 'skips' and 'stammers'. The structural consequences and roles of these digressions are not understood.

Results: In a cytoskeleton protein from *Giardia lamblia*, heptads flank eleven-residue units (hendecads) to give a 7-11-7 motif that dominates the sequence. Synthetic peptides made to the consensus sequence of this motif fold in solution to fully helical, parallel dimers. Both the sequence pattern and these experimental data are consistent with the coiled-coil model. We note that breaks in other extended coiled coils can also be reconciled by hendecad insertions.

Conclusions: The heptad paradigm for the coiled coil must be expanded to include hendecads. As different combinations of heptads and hendecads will give different overall sequence motifs, we propose that these provide a mechanism to promote cognate protein pairings during the folding of extended coiled coils in the cell.

Addresses: [1]Centre for Biomolecular Design and Drug Development, School of Biological Sciences, University of Sussex, Falmer BN1 9QG, UK. [2]Department of Life Sciences, University of Nottingham, Nottingham NG7 2RD, UK.

Correspondence: Derek N Woolfson
E-mail: d.n.woolfson@sussex.ac.uk

Key words: circular dichroism, hendecad, peptide models, protein folding, protein structure, undecatad

Received: 16 Dec 1996
Revisions requested: 14 Jan 1997
Revisions received: 07 Feb 1997
Accepted: 13 Feb 1997

Published: 10 Mar 1997
Electronic identifier: 1359-0278-002-00149

Folding & Design 10 Mar 1997, 2:149–158

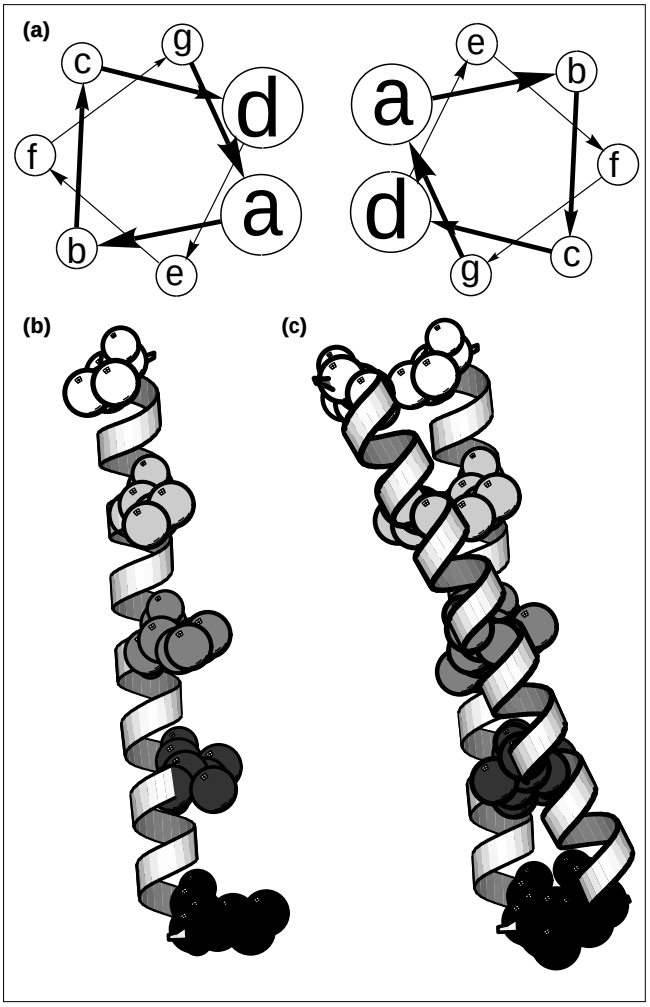
Introduction

The coiled coil is a ubiquitous protein-folding motif that guides oligomerization in a wide variety of systems, including transcription factors, muscle proteins, intermediate filaments and extracellular fibres [1–4]. The sequences of all these proteins contain patterns of seven residues (heptads) in which hydrophobic sidechains are alternately spaced three and four residues apart [5]. The seven positions of the heptad repeat are often assigned ***a***bc***d***efg, with **a** and **d** as the hydrophobic sites. When these patterns are configured into an α-helix, positions **a** and **d** are brought close together and the structure is amphipathic (Fig. 1a). In a contiguous array of heptads the **a** and **d** sites define a hydrophobic stripe, which directs oligomerization with similarly amphipathic helices. However, because the seven-residue repeat falls short of two full turns of an α-helix, the hydrophobic stripe does not continue straight up each helix, but drifts around it (Fig. 1b). Thus, in order to maintain good hydrophobic contacts, helices in coiled coils wrap around one another and produce rope-like structures [1,6–9] (Fig. 1c).

These basic tenets of coiled-coil structure have been in place for several decades [1,6], but recent high-resolution structural studies have been significant in confirming the early models, and in developing further the structural principles for coiled-coil folding. In particular, important advances have been made in understanding oligomer and partner selection in coiled coils. Many of these insights have come from studies of the leucine-zipper motif from the transcriptional activator GCN4, which is a short peptide with 4.5 contiguous heptads that folds to a stable dimeric coiled coil [3,8]. In an elegant study, Harbury and co-workers [10] demonstrated that the coiled-coil oligomer state can be selected by changes at the **a** and **d** sites of the heptad repeat alone: a dimer is formed if all **a** sites are isoleucine and all **d** are leucine, swapping these residues gives a tetramer, and a trimer results when isoleucine is substituted at all **a** and **d** positions. Sequence analyses of natural coiled coils confirm these rules and suggest additional ones that specify oligomer states in coiled coils with more heterogeneous sequences [11–13]. At present, the rules for partner selection — i.e. those that distinguish homotypic and heterotypic interactions — in short coiled coils are less clear. However, it is certain that such rules are also superimposed on the basic heptad pattern: again, residues at the **a** and **d** positions play a role in partner selection, but complementary sidechains placed at **e** and **g** also have an influence as these are also brought together upon oligomerization [14–18] (Fig. 1a).

Although the leucine zipper provides the best characterized model for the coiled coil it is a special case. The vast majority of proteins of the coiled coil family differ in three significant respects from the leucine zippers. First,

Figure 1



Helical structure in the leucine zipper from GCN4. (a) The popularized helical wheel for a dimeric coiled coil based on the heptad repeat abcdefg. By assuming 3.5 residues per turn the heptad repeat is emphasized, but this simplified picture can suggest that the hydrophobic seam set up by the a and d sites continues straight up the helix in alignment with the helical axis. (b) MOLSCRIPT [48] ribbon diagram for one chain from the leucine zipper [8]. This view reveals how residues at successive a sites, space-filled and shaded progressively less heavily, wind around the face of the structure in the opposite sense to the helical twist. (c) Both strands of the leucine zipper dimer showing the crossed, rope-like helix–helix packing necessary to maintain tight hydrophobic contacts.

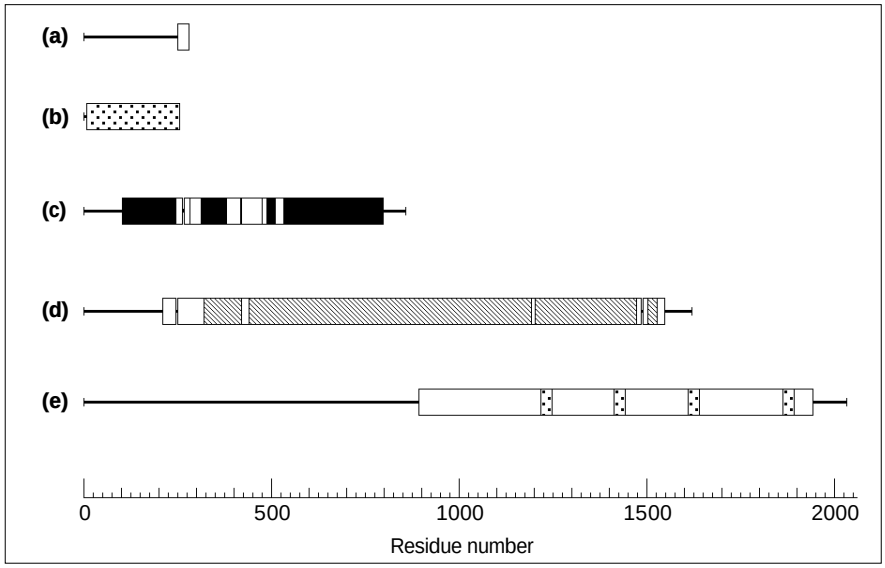
most coiled coil regions extend over hundreds of residues [1,5]. Second, although heptad repeats do dominate extended coiled-coil sequences, contiguous arrays of heptads are often broken by small non-heptad-based units [2,19]. Third, for many of these extended filamentous proteins, coiled-coil formation is only the first level of oligomer organization and higher-order assemblies are critical to the functions of these proteins [4,20]. Given these differences, it is not clear how far the rules developed from the leucine zippers extend to the folding of more extended coiled coils. This paper describes the structural consequences and discusses the possible implications of breaks in the heptad arrays of this loosely defined, but significant, class of protein structure.

The sequences of three cytoskeleton-associated proteins from *Giardia lamblia* have putative extended coiled coils. In each case, arrays of heptads are interrupted at regular intervals by non-heptad segments to produce larger motifs that dominate the protein sequences [21–23] (Fig. 2). The length of the dominant motif varies between the proteins; for the median body protein the repeat length is 24 residues, for HPSR2 it is 25 residues, and for β-giardin it is 29 residues. The study presented here focuses on HPSR2 [23]. This protein has the features of a stalked motor protein; the N-terminal 212 residues house a potential nucleotide-binding site, and the remaining 1400 residues code the putative coiled coil. It is this second region that is dominated by the 25-residue motifs; more than 40 such units are observed [23]. This 25-residue motif was initially assigned in a conventional way as three heptads followed by a four-residue skip [23] or stutter [24]. However, closer inspection revealed that the motif is best considered as two heptads flanking an 11-residue repeat. We refer to this pattern as the heptad-hendecad-heptad, or 7-11-7, motif. (We prefer to use hendecad rather than undecatad, which has appeared elsewhere, because the former is more consistent with heptad, which comes from modern Greek.) The aim of the work presented here is to determine whether this unusual sequence pattern is compatible with the coiled coil.

Using sequence analysis, circular dichroism, sedimentation-equilibrium and cross-linking studies, we have established that peptides made to the consensus sequence of the 7-11-7 motif do form helical, parallel dimers consistent with the coiled-coil model. In addition, we have found that breaks previously assigned in other extended coiled coils can also be reconciled by 11-residue motifs. More generally, many extended coiled coils read contiguously if hendecad or decad units are used in combination with classic heptad repeats. The common theme here is that the three repeats represent the most basic combinations of three-residue and four-residue spacings of hydrophobic sidechains; heptads show a 3,4 pattern, decads display a 3,4,3 combination and hendecads are 3,4,4 hydrophobic repeats. Thus, the maintenance of 3,4-based hydrophobic patterns is a major prerequisite for coiled-coil formation. Furthermore, we propose that correct and in-register coiled-coil interactions should be directed by the alignment of similar 3,4 patterns in the sequences of partnering proteins. This work has implications for the prediction of protein structure and protein–protein interactions from sequence data, and in protein-design exercises.

Figure 2

A cartoon highlighting regular sequence motifs in several coiled-coil proteins. The proteins represented are **(a)** GCN4; **(b–d)** the cytoskeleton-associated proteins from *Giardia lamblia* – (b) β-giardin, (c) the median body protein and (d) HPSR2; and **(e)** myosin heavy chain [2]. The different coiled-coil motifs that dominate these sequences are distinguished as follows: non-coiled-coil regions, horizontal lines; contiguous heptad repeats, clear blocks; 24-residue repeats, solid blocks; 25-residue repeats, cross-hatched blocks; 29-residue repeats, polka-dot-filled blocks.



Results

The 7-11-7 motif is α-helical

An alignment of the 25-residue repeats from HPSR2 was used to construct a consensus sequence for the 7-11-7 motif (Table 1). When mapped onto a helical wheel, this sequence gives an amphipathic structure typical of helices found in coiled coils (Fig. 3). To test for α-helical structure in the 7-11-7 motif, we synthesized a series of peptides based on the consensus (Table 1) and characterized their structures in solution using circular dichroism (CD). The shortest peptide, p7-11-7, corresponded to a single 7-11-7 motif. In two longer peptides, p7-7-11-7 and p11-7-7-11-7, additional heptad and hendecad units were appended to the N terminus of p7-11-7, consistent with the organization of the 7-11-7 repeats in HPSR2 [23] (Fig. 2).

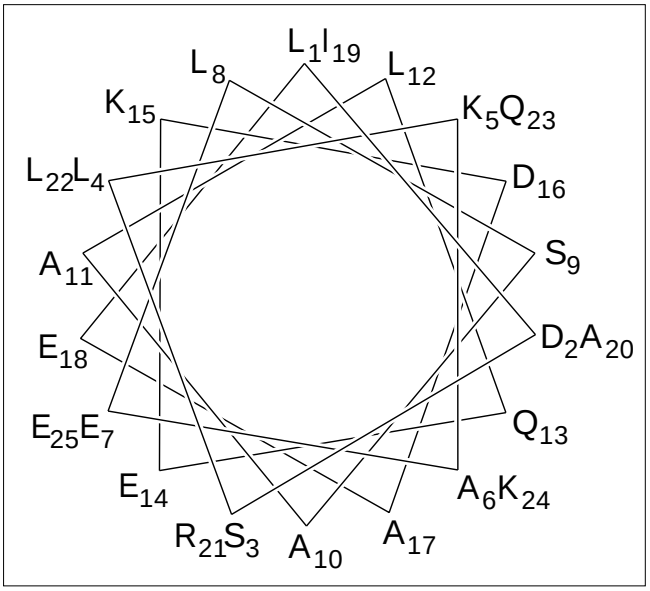
In CD studies, the shortest peptide, p7-11-7, gave a spectrum consistent with partial folding into an α-helix in aqueous solution (Fig. 4a); the spectrum showed a positive maximum at ≈190 nm and a double negative minimum at 208 and 222 nm. The negative molecular ellipticity at 222 nm ($[\theta]_{222}$) can be used to quantify the degree of α-helical structure present [25]. However, due to sidechain contributions to the signal, $[\theta]_{222}$ values corresponding to 100% helix can vary between systems [25–27]. Nevertheless, estimates of helicity can be made if the peptide is induced to fold completely and a benchmark set [25,28]. To promote helix and establish these markers for the 7-11-7-based peptides, CD spectra were recorded in aqueous 2,2,2-trifluoroethanol (TFE) solutions. TFE induces hydrogen-bonded structure, but it also disrupts hydrophobic interactions. It is anticipated, therefore, that the TFE-induced helical states described here are monomeric. The TFE titration for p7-11-7, which is typical of that shown by the peptides, displays a clear sigmoidal folding transition (Fig. 4b,c). This particular transition has a midpoint at ≈10% TFE and a plateau at ≈30% TFE (Fig. 4c), after which point further increases in TFE

Table 1

7-11-7 consensus sequence for the 7-11-7 motif and sequences of the 7-11-7-based synthetic peptides.

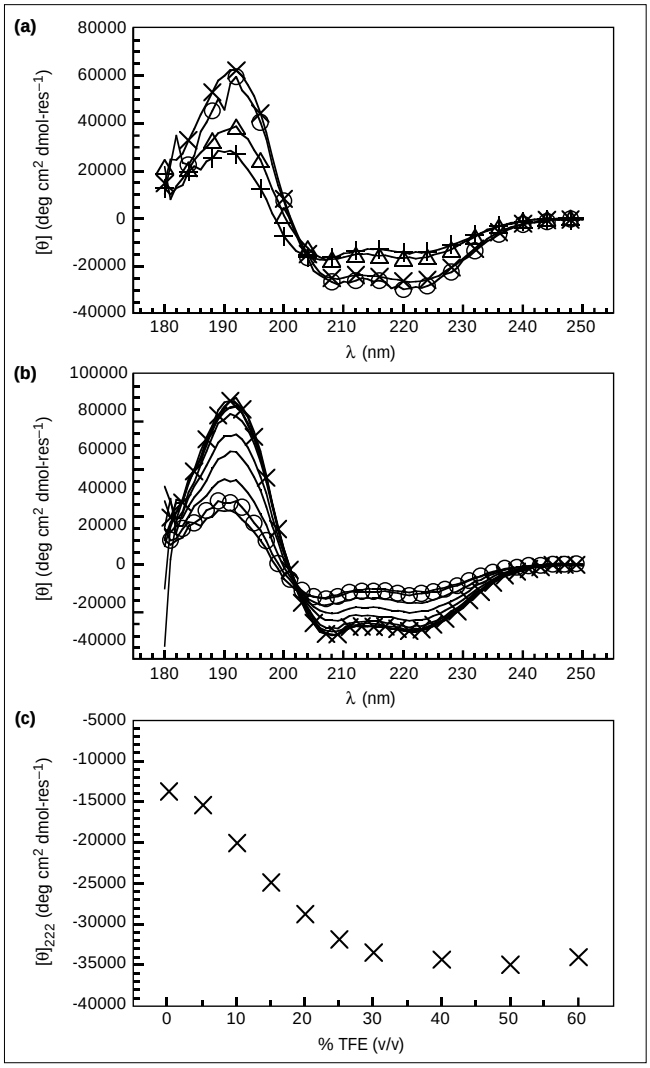
Peptide	Sequence
Repeat	***abcdefghijk abcdefg abcdefg abcdefghijk abcdefg ab***
Consensus	LDSLKAE LSAALQEKDAE IARLQKE
p7-11-7	Ac-LDSLKAE LSAALQEKDAE IARLQKE GY-NH_2
p7-7-11-7	Ac-IARLQKE LDSLKAE LSAALQEKDAE IARLQKE GY-NH_2
p11-7-7-11-7	Ac-LSAALQEKDAE IARLQKE LDSLKAE LSAALQEKDAE IARLQKE GY-NH_2
pC7-11-7	Ac-CGG LDSLKAE LSAALQEKDAE IARLQKE GY-NH_2

Figure 3



The consensus sequence for the 7-11-7 motif from HPSR2 configured into a regular 3.6 residues per turn α-helical wheel. Hydrophobic and hydrophilic residues partition to produce an amphipathic structure. The predominance of leucine in the hydrophobic stripe (the putative core of the coiled coil) and the appearance of an isoleucine at an a position of one of the regular heptads is characteristic of dimeric coils. The single polar residue that resides on the apolar face is a lysine, which is known to be tolerated in the interfaces of coiled-coil structures [13]. The one-letter amino acid code is used.

Figure 4



CD spectra for the 7-11-7-based peptides. **(a)** Spectra recorded in 25 mM potassium phosphate, pH 7 at 0°C and at 15 μM peptide: p7-11-7 (+), p7-7-11-7 (△), p11-7-7-11-7 (×), and $(pC7\text{-}11\text{-}7)_2$ (○). Estimates of the helicity in each of these peptides using the water–TFE comparison were ≈40, 74, 80 and 87%, respectively. **(b)** Spectra of p7-11-7 recorded at various concentrations of TFE: 0% (○) and 60% (×) TFE (v/v); all intervening spectra, recorded at 5, 10, 15, 20, 25, 30, 40, and 50% TFE, are shown as solid lines. The titration shows a well defined isodichroic point at 202 nm indicative of a two-state transition. **(c)** The CD signal for p7-11-7 followed at 222 nm, $[\theta]_{222}$, as a function of TFE concentration. The plateau value for $[\theta]_{222}$ in this titration was ≈–34 000 deg cm^2 dmol-res^{-1}, which corresponds well with complete helix formation in other small coiled coils [3]. Note that to maintain clarity, only a fraction of the real data points are shown in parts (a) and (b); the remaining points are linked by straight interpolation.

concentration have little affect on the CD spectrum. The limiting value of $[\theta]_{222}$ of ≈ –34 000 deg cm^2 dmol^{-1} is closely similar to that measured for other fully helical coiled-coil peptides [3]. A comparison of the $[\theta]_{222}$ values from the water and TFE-induced spectra yields an estimate of ≈40% helix in the water-only environment. Thus, for p7-11-7 at a concentration of 15 μM, ≈11 of the 25 residues adopted α-helical structure in water at 0°C.

Percent helicities for the remaining peptides were determined by comparing $[\theta]_{222}$ values measured from spectra recorded in aqueous buffer and 50% TFE. In the series, p7-11-7, p7-7-11-7 and p11-7-7-11-7, helicity in water increased with increasing peptide length (Fig. 4a). For example, at peptide concentrations of 15 μM, the percent helicities were 40, 74 and 80%, respectively. For the longest peptide, p11-7-7-11-7, this high helicity of ≈80% corresponded to ≈36 of the 43 residues of the regular motif folding to helix. Thus, additional heptad and hendecad repeats induce additional folding in the 7-11-7-based peptides. Increasing peptide concentrations also increased helicity in the peptides. This effect was more dramatic for the longer peptides, indicating further that association constants became more favourable, and helical structures became more stable, with increasing peptide length. For example, p7-11-7 showed a slight increase in helicity from ≈40 to 44% in water at 0°C when peptide concentration was increased from 15 to 100 μM. In comparison, for p11-7-7-11-7 over the same concentration range, the helicity increased from ≈80 to 99% in water. Thus, the longest peptide, p11-7-7-11-7, which has fewer

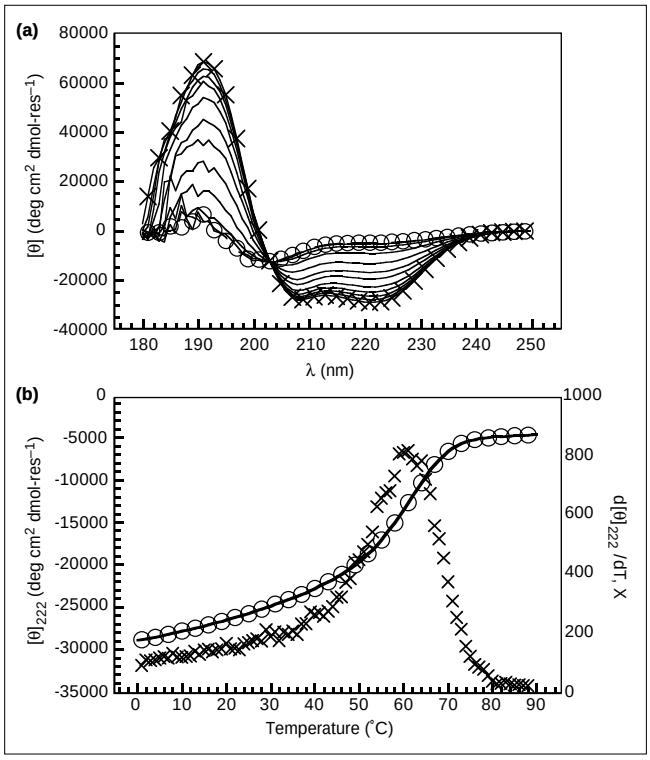
than 50 residues and corresponds to less than two complete 7-11-7 motifs, was fully folded in water at moderate peptide concentrations.

The 7-11-7 motif forms cooperatively folded units

To quantify the stabilities of the helical states of the 7-11-7-based peptides in water, CD spectra were recorded as a function of temperature. The midpoints (T_M values) of the resulting transitions were used to compare the stabilities of the different peptides. For p11-7-7-11-7 at 100 μM, spectra recorded between 0 and 90°C changed from those characteristic of an α-helix to those more representative of a random coil (Fig. 5a). These spectra had an isodichroic point at 203 nm consistent with a simple two-state transition between the helical and unfolded forms. Moreover, when followed by the $[\theta]_{222}$ signal, the unfolding curve was sigmoidal, indicative of a cooperative transition (Fig. 5b). The T_M, which was taken as the point of maximum slope of the unfolding curve, was 62 ± 1°C for p11-7-7-11-7 at 100 μM. Similar unfolding transitions for the shorter peptides were apparent although the peptides were much less stable: T_M values were ≈0 and 29°C for p7-11-7 and p11-7-11-7, respectively. The value for p7-11-7 was estimated on the basis that its helicity approximately doubled in 50% TFE (see above). In addition, this peptide did show the latter half of a cooperative thermal-unfolding transition in water.

The stability of p11-7-7-11-7 is comparable to that of the archetypal coiled-coil peptide GCN4p1 [3]. Moreover, the unfolding transitions of the 7-11-7-based peptides were sharp and sigmoidal. Such transitions are characteristic of well defined cooperatively folded units and are distinct from the relatively broad and flat transitions associated with free-standing α-helices [29] and ill defined molten-globule species [30,31].

Figure 5



Thermal unfolding and stability of the 7-11-7-based peptides. **(a)** Spectra following the thermal unfolding of peptide p11-7-7-11-7 at 100 μM peptide concentration in pH 7, 25 mM phosphate buffer: 0°C (×) and 90°C (○); all intervening spectra, recorded at 10, 20, 30, 40, 50, 55, 60, 65, 70, 75 and 80°C, are shown as solid lines. Note the distinct isodichroic point at 203 nm. For clarity, only a fraction of the real data points are shown. **(b)** $[\theta]_{222}$ as a function of temperature for the thermal unfolding of p11-7-7-11-7 (○) and the first derivative (×) of this plot used to calculate T_M values. Similar behaviour was displayed by the shorter peptides, p7-11-7 and p7-7-11-7, although T_Ms were reduced to ≈0 and 29°C, respectively.

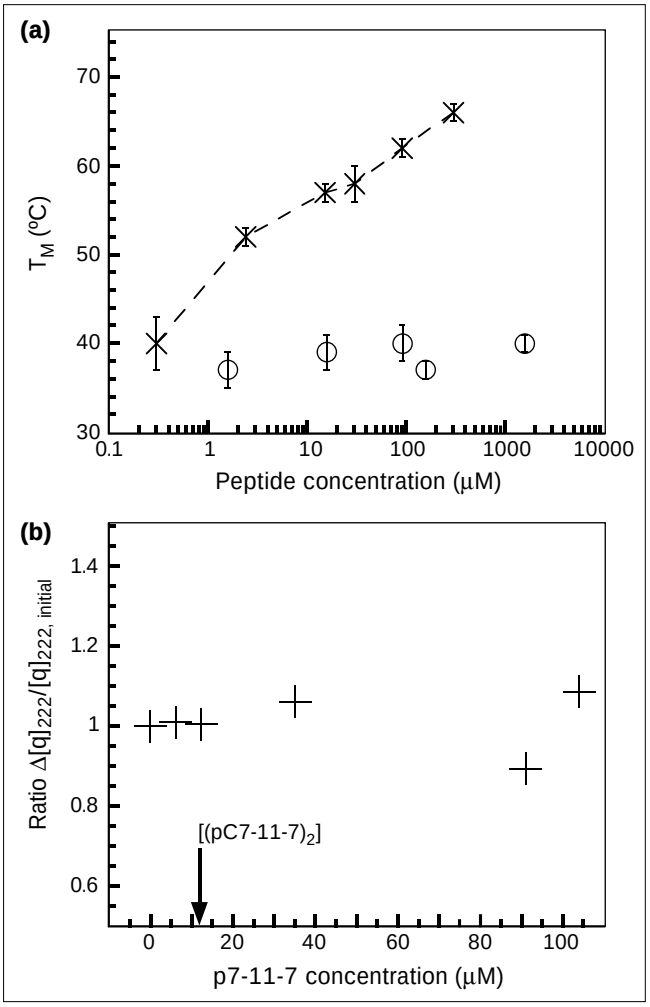
The 7-11-7 motif forms dimers

As stated above, the percent helicities for the 7-11-7-based peptides increased with increases in peptide concentrations. This is significant as it indicates that the helical folded states of the peptides are oligomers consistent with the coiled-coil model. A method, SCORER, is available that distinguishes parallel dimeric and trimeric coiled coils from sequence data alone [13]. Using this metric, and on the basis of its regular heptads, the HPSR2 protein predicted strongly as a dimer; SCORER values were positive and in the range 1.44–3.03 depending on the stringency of the heptad-selection protocol. To test this prediction, we probed the oligomerization state of the 7-11-7 motif using sedimentation-equilibrium analysis in the analytical ultracentrifuge.

Sedimentation-equilibrium experiments were conducted at 20°C and at starting peptide concentrations of between 100 and 500 μM. After sedimentation, single-species fits to the data reported molecular weights significantly above that expected for a monomer (expected $M_R \approx 5$ kDa, observed mean = 7.2 ± 0.7 kDa). Fitting the same data to monomer/dimer or monomer/trimer models suggested that the preferred oligomer state of the p11-7-7-11-7 was dimer. Monomer/dimer fits were more stable and consistently returned dissociation constants in the μM range consistent with the CD data; for example, at a peptide concentration of 3 μM and 20°C, p11-7-7-11-7 was ≈80% folded as judged by a thermal unfolding curve measured by CD.

To probe the preferred oligomer state of the 7-11-7 motif further, we compared the concentration dependencies of the folding of cross-linked and non-cross-linked 7-11-7-based peptides. Typically, in coiled-coil systems, peptide dissociation and helix unfolding are linked to give overall unfolding transitions of the type $F_n \rightleftharpoons nU$. Thus, from Le Châtelier's principle, increases in peptide concentration

Figure 6



Oligomerization of the 7-11-7-based peptides. (a) Concentration dependence of the T_Ms for p11-7-7-11-7 (×) and the covalent dimer $(pC7\text{-}11\text{-}7)_2$ (○). (b) The concentration dependence of the difference in $[\theta]_{222}$ between two sets of samples. The first contained a fixed concentration of $(pC7\text{-}11\text{-}7)_2$, whereas the second had none of this peptide. In both sets, the concentration of p7-11-7 was increased according to the values shown on the x-axis. The results are plotted as a ratio of the difference in $[\theta]_{222}$ values between the first and second samples, all divided by the starting $[\theta]_{222}$ value for the first sample. In effect, the difference value gives helicity from $(pC7\text{-}11\text{-}7)_2$ plus any changes that come from dimer–p7-11-7 interactions. Normalizing this difference with the starting ellipticity for $(pC7\text{-}11\text{-}7)_2$, therefore, reports on any changes alone. Thus, in this scheme, if the $(pC7\text{-}11\text{-}7)_2$ acted as a template for the folding of a trimer of the 7-11-7 motif, a limiting ratio of ≈1.5 would be expected. The near constant value of 1 therefore implies no significant interaction between $(pC7\text{-}11\text{-}7)_2$ and p7-11-7.

will shift the equilibrium to the left and stabilize the helical folded state; in other words, increasing peptide concentration is expected to increase T_M. The concentration dependence of T_M for p11-7-7-11-7 is shown in Figure 6a. The plot shows a marked increase in protein stability with increased peptide concentration. It is not possible to model this curve with any confidence and discriminate between the alternative unfolding processes: $F_2 \rightleftharpoons 2U$, $F_3 \rightleftharpoons 3U$ and so on. However, we tested for the possibility of parallel dimer formation (the simplest oligomer arrangement) using the method of O'Shea *et al.* [3]: a fourth peptide, pC7-11-7, was synthesized (Table 1) that allowed two p7-11-7 monomers to be linked through N-terminal cysteines to give a covalent dimer $(pC7\text{-}11\text{-}7)_2$. Using the water–TFE comparison, this peptide folded with ≈87% helix in water (Fig. 4a). The shortfall from full helix was consistent with the N- and C-terminal linkers failing to adopt regular structure in water. Furthermore, the T_M for this peptide showed no concentration dependence and remained steady at ≈39°C in the μM to mM range (Fig. 6a). Thus, $(pC7\text{-}11\text{-}7)_2$ behaved as a fully helical, cooperatively folded monomer in solution. The simplest model compatible with these data is that the two 7-11-7 motifs fold to α-helices which are stabilized by parallel helix–helix interactions. In turn, this suggests that the preferred oligomer state of the 7-11-7 motif is a parallel dimer.

To be sure of this, we conducted a further experiment: increasing concentrations of p7-11-7 were added to a fixed concentration of $(pC7\text{-}11\text{-}7)_2$. CD spectra were recorded at 5°C, where p7-11-7 was only partly folded and $(pC7\text{-}11\text{-}7)_2$ fully structured. The differences in the $[\theta]_{222}$ values between these spectra and controls, which contained corresponding amounts of the p7-11-7 monomer but no $(pC7\text{-}11\text{-}7)_2$, are plotted in Figure 6b. $[\theta]_{222}$ remained unchanged over the whole range of p7-11-7 concentrations and corresponded to the value expected for the $(pC7\text{-}11\text{-}7)_2$ dimer alone. Thus, the dimer did not act as a template to promote further folding of p7-11-7 as might have been expected if the preferred oligomer state of the 7-11-7 motif were trimer or some state other than parallel dimer.

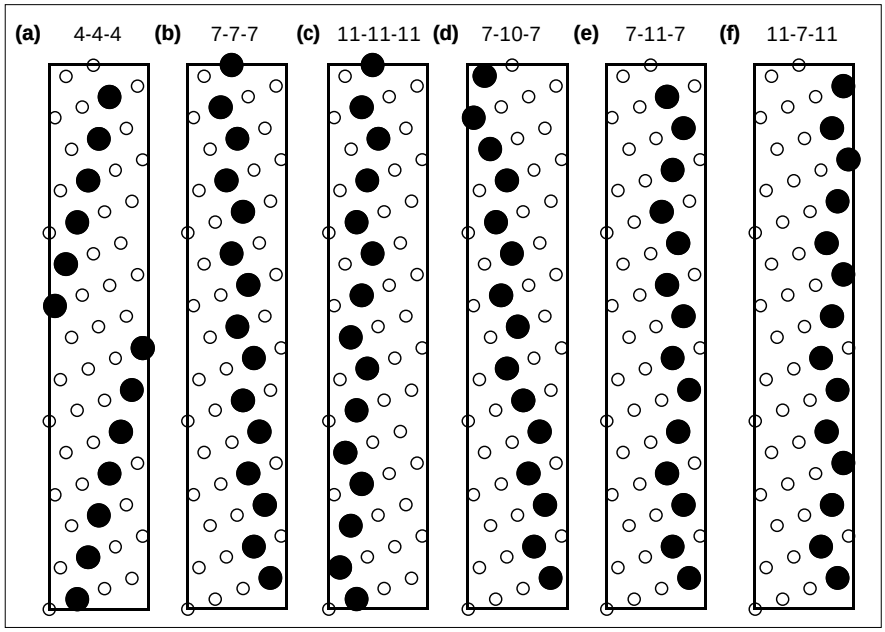
Discussion

In summary, synthetic peptides corresponding to the 7-11-7 motif form fully α-helical, stable, cooperatively folded, parallel dimers in solution. These results indicate that the stalk of intact HPSR2 forms a two-chain coiled coil in which repeated hendecads are fully accommodated within an otherwise heptad-based sequence.

11-residue repeats have been recognized in other proteins

Hendecad repeats appear in a number of recently reported proteins from a variety of sources, notably bacteria, other parasites, viruses, plants and fish [32–36]. In three cases, hendecads occur contiguously and coiled-coil structures are proposed [32,33,35]. Moreover, in the crystal structure of a fragment from influenza haemagglutinin, hendecads are accommodated in a coiled coil which is a predominantly heptad-based trimer [34]. Residues of the hendecad that contribute to the hydrophobic, helical interface are spaced three, four and four residues apart, or at positions **a**, **d** and **h** of an ***a*bc*d*efg*h*ijk** repeat. This pattern is an

Figure 7



Helical net diagrams showing (a) patterns of hydrophobic residues typical of helices in globular proteins [37], and (b–f) patterns of core residues for the various 3,4-patterns compatible with the coiled-coil model. Open circles mark the positions of the C_α atoms, while filled circles show the relative positions of highlighted residues in the relevant repeat. In these diagrams, the positions of the C_α atoms in an α-helix have been projected onto the outside of a cylinder wrapped around the structure, and the cylinder opened out for viewing in two dimensions.

extension of the 3,4-hydrophobic pattern of the classic heptad, ***a***bc***d***efg. Combinations of three- and four-residue intervals are, in fact, a prerequisite for coiled-coil-type helix packing and distinguish it from ridges-into-grooves packing observed in globular proteins [1,6,37] (Fig. 7a,b). A 3,4,4 pattern is apparent also in the hendecads of HPSR2 [23], giving an overall hydrophobic pattern in the 7-11-7 repeat of 3,4,3,4,4,3,4 (Table 1).

Two other proteins known from *Giardia* have repeats with non-integral numbers of heptads. In both of these cases, hydrophobic residues are spaced at combinations of three and four residues and, in this respect, conform with the coiled-coil model. For example, β-giardin exhibits a 29-residue repeat that has been assigned as four heptads followed by a 'skip' residue [21]. 29-residue repeats recognized for some time in certain myosin sequences have been assigned in a similar way [2,19]. However, searches for 3,4 patterns revealed that these 29-residue units are better viewed as 11-7-11 motifs, with 3,4,4,3,4,3,4,4 spacings of hydrophobic residues. A similar unbroken 3,4 pattern can be discerned in the sequence of SF-assemblin, the flagellar rootlet protein from *Spermatozopsis similis*, which aligns with β-giardin [38,39]. CD studies of the intact SF-assemblin protein confirm that the 11-7-11-type patterns in β-giardin and SF-assemblin fold to predominantly α-helical structures [39]. The second example from *Giardia*, the median body protein, displays a 24-residue repeat [22], wherein hydrophobic residues are spaced at 3,4,3,4,3,3,4 in a heptad-decad-heptad (or 7-10-7) motif. This unit extends further the set of basic units compatible with the coiled coil.

Structural consequences of non-heptad repeats

In the haemagglutinin structure, hendecads are separated by three heptads and have the effect of straightening the twist (or supercoil) of the coiled coil [34]. The twist arises because, when superimposed on a regular α-helix with 3.6 residues per turn, a heptad repeat falls short of two complete helical turns. The seam set up by hydrophobic residues placed at positions **a** and **d** of the heptad therefore moves around the face of the helix in the opposite sense to the helical twist of the backbone [1,8,40] (Figs 1b,7b). In order to maintain good hydrophobic contacts, the helices of coiled-coil oligomers supercoil around one another to form a rope (Fig. 1c), and the sense of the rope is left handed. A hendecad repeat, on the other hand, spans a little over three turns of a regular helix, and the hydrophobic stripe therefore slants in the opposite sense to that produced by a heptad (Fig. 7c), and a right-handed supercoil is predicted [32,33,35].

Extension of these arguments to the combinations of 7-, 10- and 11-residue patterns observed in the *Giardia* sequences produces significantly different geometries for the coiled-coil hydrophobic stripes (Fig. 7). Combining heptad and decad repeats accentuates the left-handed stripe (compare Fig. 7 parts b and d). In contrast, hendecads introduce a right-handed twist. With one hendecad every two heptads, as in the 7-11-7 motif, the hydrophobic seam is straightened (Fig. 7e). When two hendecads combine with a single heptad, as in the 11-7-11 repeat, the seam slants in the opposite direction to that set up by heptads alone (Fig. 7b,f). These changes in the hydrophobic seam are not expected to be abrupt, but smooth, as

seen in the haemagglutinin structure [34] and proposed for a related motif in the nematode myosin structure [19].

Possible roles for the different coiled-coil twists

We propose that the different sequence motifs that result from combinations of heptad, decad and hendecad repeats and their associated hydrophobic patterns provide a simple, but elegant, mechanism for coiled-coil recognition.

With long coiled coils, and a straightforward sequence pattern of repeating heptads, the potential for mismatched protein pairings or out-of-register alignment in oligomers may be high. The first of these problems may be particularly acute in regions of the cell rich in coiled-coil proteins. This is the case in the cytoplasm of eukaryotes, where a large number of coiled coils are observed in cytoskeleton-associated proteins. In the assembly of shorter coiled coils of the leucine-zipper class, mismatches are usually avoided by complementary residue patterns superimposed on the **a**, **d**, **e** and **g** sites of the heptad repeat [3,8–10,16,41]. However, even for these relatively simple structures, alternative topologies have been observed for individual sequences [10,42–44]. Rules similar to those observed in the leucine zippers are almost certainly at play in the longer coiled coils of filamentous proteins [11–13], but it is not clear whether they carry all of the information necessary to achieve faultless chain docking over a thousand residues or more. For such proteins, the problems with promiscuous and fruitless interactions may be compounded further: in the case of the coiled-coil proteins from *Giardia*, all three have similar pIs, all are predicted to be dimers [13], and all are cytoskeleton-associated. We believe that the 24-, 25- and 29-residue patterns distinguished in these proteins direct homodimer assembly and reduce the possibility of promiscuous protein pairings: mismatched pairs of proteins with different repeats would be disfavoured because their hydrophobic seams would not be aligned, complete complementarity would not be possible and a free-energy penalty would be incurred. For similar reasons, the motifs may also ensure in-register alignment of extended coiled coils.

In the case of fibrous proteins, the roles of different coiled-coil twists could extend to the assembly of higher-order structures. Indeed, a similar proposal has been made for the 29-residue repeats present in the myosins [2,20].

Apart from the regular repeating heptad-hendecad and heptad-decad structures considered above, there are many examples of isolated pattern faults (usually referred to as skips, stutters or stammers) in filamentous protein sequences [19,45]. These have been assigned in order to maximize the number of classic heptad repeats in coiled-coil segments. In most of these cases, however, the sequences local to the fault can be reconciled with 3,4-based patterns of hydrophobic residues, predominantly hendecads. For example, an 11-residue insertion occurs in the gp17 tail fibre of bacteriophage T7 [45], and units that can be viewed as hendecads have been noted in the Ompα rod from *Thermotoga maritima* [46]. For the latter, Lupas and colleagues make the point that insertions might be accommodated within a coiled coil in the same way as skips in the myosins [19], and as observed in haemagglutinin [34]. The experimental data presented here demonstrate directly that sequences rich in hendecads can fold effectively to fully helical oligomers consistent with coiled coils. Furthermore, our work suggests that digressions from the heptad repeat may play significant roles in coiled-coil recognition and filament protein assembly. Accordingly, we suggest that more examples will be found in protein sequences. To illustrate, we have identified two contiguous hendecads in a region previously assigned as a linker between the heptad-based coils 1A and 1B of the aligned human lamin sequences [4]; the insertion begins at Ile63 of lamins A and C and at Val64 of lamin B1. This feature is not shared by similarly placed linkers in other intermediate filament proteins, thus it may help locate lamins to the nuclear envelope by preventing interactions with the filament network of the cytoplasm [4].

Materials and methods

Helical wheel and helical net diagrams

Unless stated otherwise, these were constructed using α-helix parameters typical of those found in recently determined high-resolution structures of coiled coils [9]: 3.6 residues per turn, a rise per residue of 1.5 Å and a helical radius of 2.3 Å.

Peptide synthesis

Peptides were synthesized on an Applied Biosystems 432A Peptide Synthesizer using solid-phase methods and Fmoc chemistry. Peptides were purified using reversed-phase HPLC and their identities confirmed by MALDI-TOF mass spectrometry: p7-11-7, expected M_R 3031, observed 3029 (electrospray result); p7-7-11-7, expected 3869, observed 3869; p11-7-7-11-7, expected 5025, observed 5022. The pC7-11-7 peptide was air-oxidized overnight in 100 mM Tris at pH 8.5. The cysteine-linked dimer, $(pC7\text{-}11\text{-}7)_2$, was purified from the reaction using HPLC and its identity confirmed by electrospray mass spectrometry (expected M_R 6493, observed 6496). All peptides were buffer exchanged into MQ water, or 25 mM potassium phosphate at pH 7 prior to dilution for CD and other experiments.

Circular dichroism

CD spectra were measured on a JASCO J-715 spectropolarimeter fitted with a Peltier temperature controller. Typically, spectra were recorded by stepping through the range 250–180 nm in 1 nm intervals and integrating the signal for 4 or 16 s at each point. Sample concentrations were determined from the absorbance at 280 nm measured in 6 M GdnHCl [47]. Rectangular quartz cells (Hellma) of 1 mm and 1 cm path lengths were used throughout. All samples (both in water and TFE) were 25 mM in potassium phosphate and at pH 7. After baseline correction, ellipticities in mdeg were converted to molecular ellipticities (deg cm^2 $dmol\text{-}res^{-1}$) by normalizing for the concentration of peptide bonds. Percent helicities were calculated by comparing $[\theta]_{222}$ values measured in water and 50% TFE. Thermal denaturation curves were determined by following the CD signal at 222 nm; signal was averaged for 16 s at 1° increments in the range 0–90°C. A ramping rate of 1° min^{-1} was used for all experiments; however, in preliminary experi-

ments where this rate was varied, no significant changes in the melting curves were observed. The midpoints of these curves, T_M, were taken as the maximum of the first derivative of the $[\theta]_{222}$ versus T plots. With some of the more dilute samples, denaturation curves were smoothed or data from several runs averaged; the errors quoted in figures are derived on the basis of these manipulations.

Analytical ultracentrifugation

Sedimentation-equilibrium experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge fitted with an AN60 titanium rotor. Measurements were taken at 40 000 rpm and 20°C. Data were analyzed by nonlinear least-squares curve fitting (MicroCal Origin) to obtain molecular weights and equilibrium constants. Values of 1 g ml^{-1} and 0.745 were taken for the buffer density and the vbar, respectively.

Acknowledgements

We thank Olwyn Byron and Neil Errington, from the National Centre for Macromolecular Hydrodynamics at the University of Leicester, for running the analytical ultracentrifugation; John Armstrong and Pehr Harbury for valuable discussions; George Banting, Jonathan Cox and Tim Clackson for comments on the manuscript; Constantinos Brikos for lessons in ancient and modern Greek; and Graham Bloomberg for preparation and mass spectrometry of peptides p7-11-7 and pC7-11-7. We are grateful to the Wellcome Trust for equipment and project grants to DN Woolfson and DV Holberton, respectively. DV Holberton is a Wellcome Senior Fellow and MR Hicks holds a BBSRC studentship.

References

1. Crick, F.H.C. (1953). The packing of α-helices: simple coiled-coils. *Acta. Crystallogr.* **6**, 689–697.
2. Dibb, N.J., Maruyama, I.N., Krause, M. & Karn, J. (1989). Sequence-analysis of the complete *Caenorhabditis elegans* myosin heavy-chain gene family. *J. Mol. Biol.* **205**, 603–613.
3. O'Shea, E.K., Rutkowski, R. & Kim, P.S. (1989). Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538–542.
4. Quinlan, R., Hutchison, C. & Lane, B. (1995). Intermediate filament proteins. *Protein Profile* **2**, 801–952.
5. McLachlan, A.D. & Stewart, M. (1975). Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293–304.
6. Pauling, L. & Corey, E.B. (1953). Compound helical configurations of polypeptide chains: structure of proteins of the α-keratin type. *Nature* **171**, 59–61.
7. Wilson, I.A., Skehel, J.J. & Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373.
8. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544.
9. Harbury, P.B., Kim, P.S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83.
10. Harbury, P.B., Zhang, T., Kim, P.S. & Alber, T. (1993). A switch between two-, three- and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407.
11. Conway, J.F. & Parry, D.A.D. (1990). Structural features in the heptad substructures and longer range repeats of two-stranded α-fibrous proteins. *Int. J. Biol. Macromol.* **12**, 328–334.
12. Conway, J.F. & Parry, D.A.D. (1991). Three-stranded α-fibrous proteins: the heptad repeat and its implications for structure. *Int. J. Biol. Macromol.* **13**, 14–16.
13. Woolfson, D.N. & Alber, T. (1995). Predicting oligomerization states of coiled coils. *Protein Sci.* **4**, 1596–1607.
14. O'Shea, E.K., Rutkowski, R., Stafford, W.F., III & Kim, P.S. (1989). Preferential heterodimer formation by isolated leucine zippers from *fos* and *jun*. *Science* **245**, 646–648.
15. O'Shea, E.K., Rutkowski, R. & Kim, P.S. (1992). Mechanism of specificity in the *fos–jun* oncoprotein heterodimer. *Cell* **68**, 699–708.
16. Vinson, C.R., Hai, J.W. & Boyd, S.M. (1993). Dimerization specificity of the leucine-zipper-containing bZIP motif in DNA binding – prediction and rational design. *Genes Dev.* **7**, 1047–1058.
17. O'Shea, E.K., Lumb, K.J. & Kim, P.S. (1993). Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658–667.
18. Nautiyal, S., Woolfson, D.N., King, D.S. & Alber, T. (1995). A designed heterotrimeric coiled coil. *Biochemistry* **34**, 11645–11651.
19. McLachlan, A.D. & Karn, J. (1983). Periodic features in the amino acid sequence of nematode myosin rod. *J. Mol. Biol.* **164**, 605–626.
20. Offer, G. (1990). Skip residues correlate with bends in the myosin tail. *J. Mol. Biol.* **216**, 213–218.
21. Holberton, D., Baker, D.A. & Marshall, J. (1988). Segmented α-helical coiled-coil structure of the protein giardin from the *Giardia* cytoskeleton. *J. Mol. Biol.* **204**, 789–795.
22. Marshall, J. & Holberton, D.V. (1993). Sequence and structure of a new coiled-coil protein from a microtubule bundle in *Giardia*. *J. Mol. Biol.* **231**, 521–530.
23. Marshall, J. & Holberton, D.V. (1995). *Giardia* gene predicts a 183 kDa nucleotide-binding head-stalk protein. *J. Cell Sci.* **108**, 2683–2692.
24. Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.
25. Kallenbach, N.R., Pingchiang, L. & Hongxing, Z. (1996). CD spectroscopy and the helix-coil transition in peptides and polypeptides. In *Circular Dichroism and the Conformational Analysis of Biomolecules*. (Fasman, G.D., ed.), pp. 201–259, Plenum, New York.
26. Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R.L. (1993). Aromatic side-chain contributions to far-ultraviolet circular dichroism of helical peptides and its effect an measurements of helix propensity. *Biochemistry* **32**, 5560–5565.
27. Woody, R.W. & Dunker, A.K. (1996). Aromatic and cystine side-chain circular dichroism in proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*. (Fasman, G.D., ed.), pp. 109–157, Plenum, New York.
28. Jasanoff, A. & Fersht, A.R. (1994). Quantitative determination of helical propensities from trifluoroethanol titration curves. *Biochemistry* **33**, 2129–2135.
29. Scholtz, J.M., *et al*., & Bolen, D.W. (1991). Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. USA* **88**, 2854–2858. [Published erratum appears in *Proc. Natl. Acad. Sci. USA* 1991 **88**, 6898.]
30. Yutani, K., Ogasahara, K. & Kuwajima, K. (1992). Absence of the thermal transition in apo-α-lactalbumin in the molten globule state. A study by differential scanning microcalorimetry. *J. Mol. Biol.* **228**, 347–350.
31. Raleigh, D.P. & DeGrado, W.F. (1992). A *de novo* designed protein shows a thermally induced transition from a native to a molten globule-like state. *J. Am. Chem. Soc.* **114**, 10079–10081.
32. Peters, J., Baumeister, W. & Lupas, A. (1996). Hyperthermostable surface-layer protein tetrabrachion from the archaebacterium *Staphylothermus marinus* – evidence for the presence of a right-handed coiled-coil derived from the primary structure. *J. Mol. Biol.* **257**, 1031–1041.
33. Werner, E., Holder, A.A., Aszodi, A. & Taylor, W.R. (1996). A novel 11-residue coiled-coil motif predicts a histidine zipper. *Protein Pept. Lett.* **3**, 139–145.
34. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43.
35. Dure, L.I. (1993). A repeating 11-mer amino-acid motif and plant desiccation. *Plant J.* **3**, 363–369.
36. Chao, H., Hodges, R.S., Kay, C.M., Gauthier, S.Y. & Davies, P.L. (1996). A natural variant of type-I antifreeze protein with four ice-binding repeats is a particularly potent antifreeze. *Protein Sci.* **5**, 1150–1156.
37. Chothia, C., Levitt, M. & Richardson, D. (1981). Helix-to-helix packing in proteins. *J. Mol. Biol.* **145**, 215–250.
38. Weber, K., Geisler, N., Plessmann, U., Bremerich, A., Lechtreck, K.F. & Melkonian, M. (1993). SF-assemblin, the structural protein of the 2-nm filaments from striated microtubule-associated fibers of algal flagellar roots, forms a segmented coiled coil. *J. Cell Biol.* **121**, 837–845.
39. Lechtreck, K.F., Frins, S., Bilski, J., Teltenkotter, A., Weber, K. & Melkonian, M. (1996). The cruciated microtubule-associated fibers of the green-alga *Dunaliella bioculata* consist of a 31-kDa SF-assemblin. *J. Cell Sci.* **109**, 827–835.
40. Seo, J. & Cohen, C. (1993). Pitch diversity in α-helical coiled coils. *Proteins* **15**, 223–234.
41. Gonzalez, L.J., Woolfson, D.N. & Alber, T. (1996). Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **3**, 1011–1018.
42. Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J. & Wu, C. (1993). Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. *Science* **259**, 230–234.
43. Gonzalez, L., Plecs, J.J. & Alber, T. (1996). An engineered allosteric switch in leucine-zipper oligomerization. *Nat. Struct. Biol.* **3**, 510–515.

44. Gonzalez, L.J., Brown, R.A., Richardson, D. & Alber, T. (1996). Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nat. Struct. Biol.* 3, 1002–1010.
45. Cohen, C. & Parry, D.A.D. (1990). α-Helical coiled coils and bundles – how to design an α-helical protein. *Proteins* 7, 1–15.
46. Lupas, A., Muller, S., Goldie, K., Engel, A.M., Engel, A. & Baumeister, W. (1995). Model structure of the omp-α rod, a parallel four-stranded coiled coil from the hyperthermophilic eubacterium *Thermotoga maritima*. *J. Mol. Biol.* 248, 180–189.
47. Gill, S.C. & Von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino-acid-sequence data. *Analyt. Biochem.* 182, 319–326.
48. Kraulis, P.J. (1991). MOLSCRIPT – a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.