Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain

Yizhi Tao[1], Sergei V Strelkov[1,2], Vadim V Mesyanzhinov[2] and Michael G Rossmann[1]*

Background: Oligomeric coiled-coil motifs are found in numerous protein structures; among them is fibritin, a structural protein of bacteriophage T4, which belongs to a class of chaperones that catalyze a specific phage-assembly process. Fibritin promotes the assembly of the long tail fibers and their subsequent attachment to the tail baseplate; it is also a sensing device that controls the retraction of the long tail fibers in adverse environments and, thus, prevents infection. The structure of fibritin had been predicted from sequence and biochemical analyses to be mainly a triple-helical coiled coil. The determination of its structure at atomic resolution was expected to give insights into the assembly process and biological function of fibritin, and the properties of modified coiled-coil structures in general.

Results: The three-dimensional structure of fibritin E, a deletion mutant of wild-type fibritin, was determined to 2.2 Å resolution by X-ray crystallography. Three identical subunits of 119 amino acid residues form a trimeric parallel coiled-coil domain and a small globular C-terminal domain about a crystallographic threefold axis. The coiled-coil domain is divided into three segments that are separated by insertion loops. The C-terminal domain, which consists of 30 residues from each subunit, contains a β-propeller-like structure with a hydrophobic interior.

Conclusions: The residues within the C-terminal domain make extensive hydrophobic and some polar intersubunit interactions. This is consistent with the C-terminal domain being important for the correct assembly of fibritin, as shown earlier by mutational studies. Tight interactions between the C-terminal residues of adjacent subunits counteract the latent instability that is suggested by the structural properties of the coiled-coil segments. Trimerization is likely to begin with the formation of the C-terminal domain which subsequently initiates the assembly of the coiled coil. The interplay between the stabilizing effect of the C-terminal domain and the labile coiled-coil domain may be essential for the fibritin function and for the correct functioning of many other α-fibrous proteins.

Addresses: [1]Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392, USA and [2]Howard Hughes Medical Institute, Bach Institute of Biochemistry, 33 Leninsky Prospect, 117071 Moscow, Russia.

*Corresponding author.
E-mail: mgr@indiana.bio.purdue.edu

Key words: bacteriophage T4, coiled coil, fibritin, oligomerization, X-ray crystallography

Received: 20 March 1997
Revisions requested: 21 April 1997
Revisions received: 30 April 1997
Accepted: 1 May 1997

Structure 15 June 1997, 5:789–798
http://biomednet.com/elecref/0969212600500789

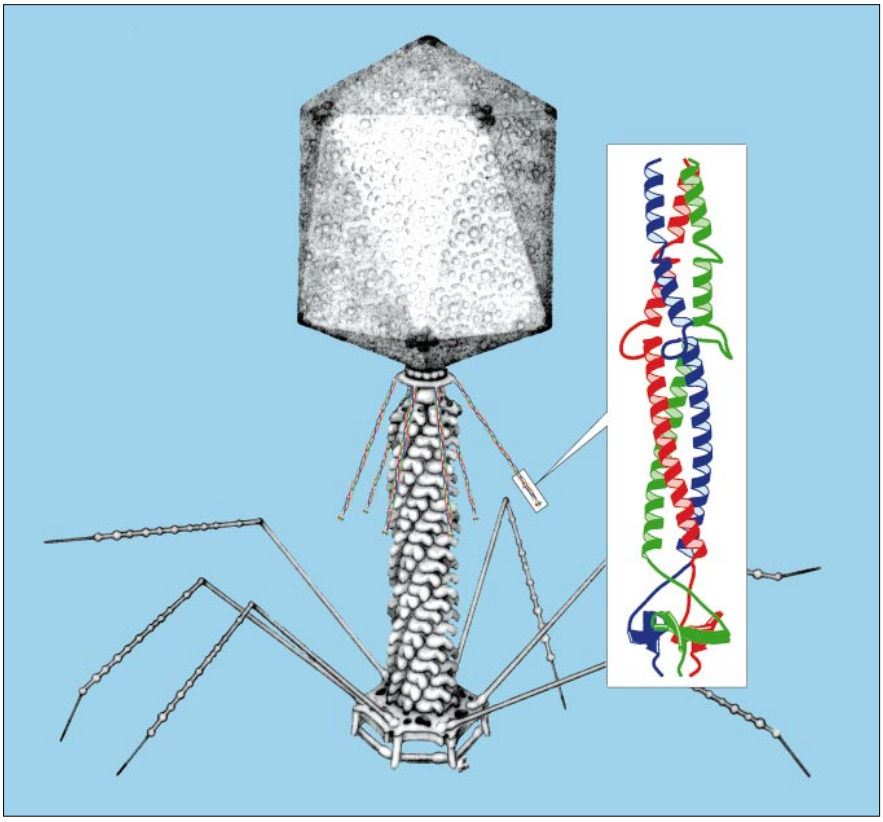
Introduction

Phage T4 is a bacterial virus that consists of a prolate head, containing double-stranded DNA, and a sixfold-symmetric tail, through which the viral DNA is extruded during infection [1] (Fig. 1). In addition, there is a set of six long and six short tail fibers, which are responsible for receptor recognition and initialization of the infectious process. The fibers are connected to a complex substructure, called the baseplate, located at the distal end of the tail.

Fibritin is a homotrimer of the protein gpwac, a 52 kDa product of the gene *wac* (whisker antigen control) that is expressed late in the virus life cycle. After the phage head is joined to a fiberless tail, six fibritin molecules attach to the neck of the virion and form the collar with six fibers ('whiskers'). Interaction of the whiskers with the distal and proximal parts of the long tail fibers stimulates assembly of the long tail fibers and their subsequent attachment to the tail baseplate [2,3]. Viruses without or with fewer than six long tail fibers are either non- or less infectious [4]. Additional to their role in virus assembly, the whiskers are a rudimentary environment-sensing device. The whiskers control the retraction of the long tail fibers in response to certain adverse environments, therefore preventing infection [5]. It is unclear whether long tail fiber assembly and retraction involve the same interaction of the whiskers with the fibers.

Recombinant gpwac protein, expressed in *Escherichia coli*, assembles into filamentous molecules with a length of 530 Å. Sequence analysis of the gpwac protein has predicted that there are three major structural regions — the

Figure 1



Bacteriophage T4, with the fibritin whiskers shown in color. Adapted from a drawing by Eiserling and Black [1]. On the right is an enlarged ribbon diagram of fibritin E.

N-terminal, the central and the C-terminal domains. The central domain, which is 80% of the entire sequence, contains the heptad repeat amino acid sequence pattern $(abcdefg)_n$, where *a* and *d* are preferably occupied by hydrophobic residues [6]. Such a motif is characteristic of an α-helical coiled coil [7]. In gpwac this periodicity is interrupted at several places, which suggests the existence of twelve coiled-coil segments connected by linker regions. Deletion of the N-terminal domain abolishes attachment of the protein to the virion neck, whereas certain mutations in the C-terminal domain, or its partial deletion, result in an insoluble expression product. However, a partial deletion of the central domain or fusion of a foreign peptide to the C terminus affects neither fibritin assembly nor attachment of the whiskers to the virion neck [8]. An antigen-display system has been developed, in which the extended fibritin gene, with foreign genetic material appended to the C terminus, is re-introduced into T4 phage [9].

Our structural study of fibritin had two main objectives. Firstly, to study fibritin's function as a specialized chaperone during T4 phage assembly. Secondly, to use fibritin as a convenient model for studying the mechanisms that determine protein folding, oligomerization and stability of coiled-coil proteins. As full-length fibritin could not be crystallized, a series of mutant proteins A, B, C, D, E, F and M were engineered that had the N-terminal domain and progressively more of the α-helical (central) domain deleted. Fibritin E consists of the final 119 residues of the wild-type fibritin (from Glu368 to Ala486), comprising the last three putative coiled-coil segments and the C-terminal domain [10]. Here, we present the X-ray structure of fibritin E at 2.2 Å resolution. It is the first structural protein of bacteriophage T4 for which the structure has been solved to atomic resolution, and it provides some insights into how fibritin can act as a specialized chaperone during phage assembly.

Results and discussion

The fibritin E structure was determined using the multiple isomorphous replacement (MIR) method (Table 1). The final atomic model included 113 out of the 119 residues, lacking three residues at the N terminus and three at the C terminus. A subunit of fibritin E consists of an α-helical region and a C-terminal region (Fig. 2). Three identical fibritin E subunits associate into a trimer, forming a parallel α-helical coiled coil along the crystallographic threefold axis (Fig. 1). In the C-terminal domain, there is a small β sheet formed by threefold-related β

hairpins. This β sheet is in a plane perpendicular to the coiled-coil axis.

Coiled-coil domain

Unlike most other coiled-coil structures [11], each α helix in fibritin E is interrupted by two insertion loops, resulting in three segments each containing 2, 1.5 and 5.5 heptad repeats, respectively. The downstream helices almost exactly follow the continuation of the preceding ones (Fig. 3a). Two mainchain hydrogen bonds are formed between consecutive helical segments which bring the helices into register after omitting approximately one turn (Fig. 3b,c). In a coiled coil, residues at the *a* or *d* positions are arranged in alternate layers along the coiled-coil axis. As each helical turn contains about 3.5 residues, the residues at *a* and *d* have to be positioned at different sides of the supercoil axis. In this arrangement, every internal sidechain, termed a 'knob', is fitted into the empty space, a 'hole', provided by other peptide chains in the coiled coil [12,13]. Helical segments α10 and α11 end at position *a* and helical segments α11 and α12 start at position *d*. This makes the abutting helical segments consistent with the requirements for forming a continuous coiled coil. Four mainchain hydrogen bonds are omitted at the site of every insertion loop, which creates greater flexibility in comparison with an uninterrupted α helix (Fig. 3d). This

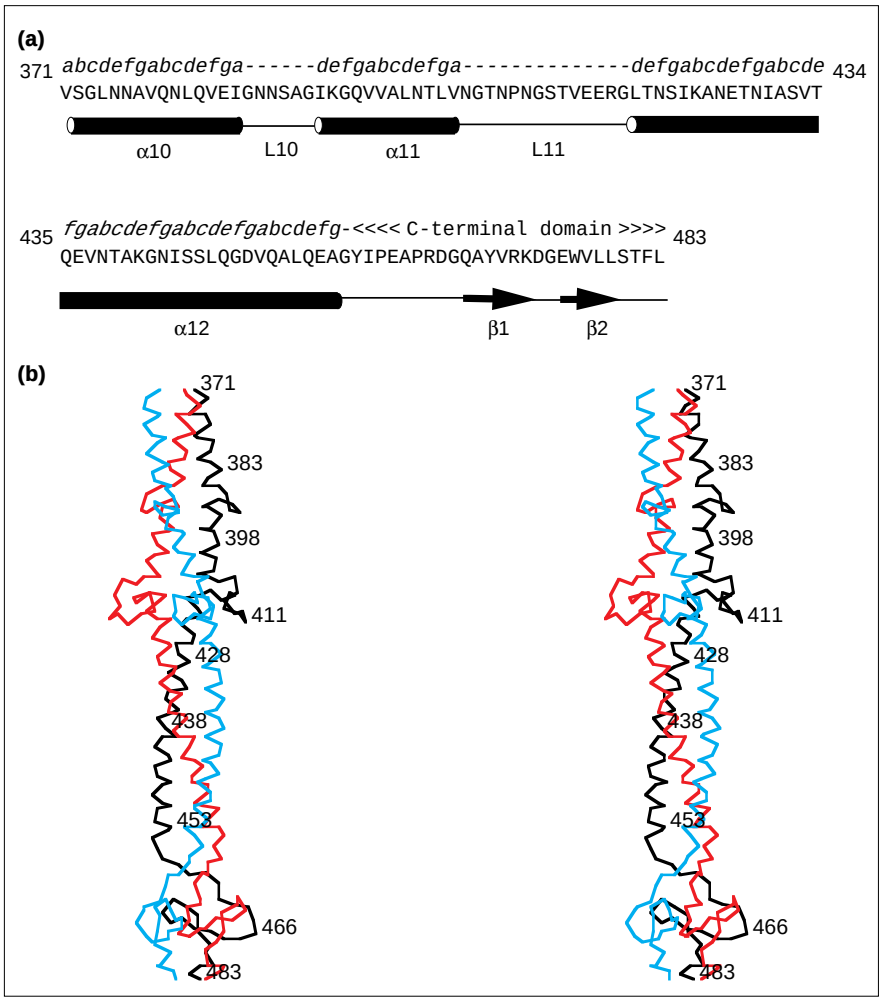
Table 1

Phasing statistics.

Derivative	R_{cullis}*	Phasing power†	Site	Ligand
$UO_2(CH_3COO)_2$	0.64	1.87	a	D449,E475
			b	E455
$Pb(CH_3COO)_2$	0.72	1.56	a	D449,E475
K_2PtCl_4	0.90	0.77	c	Q467

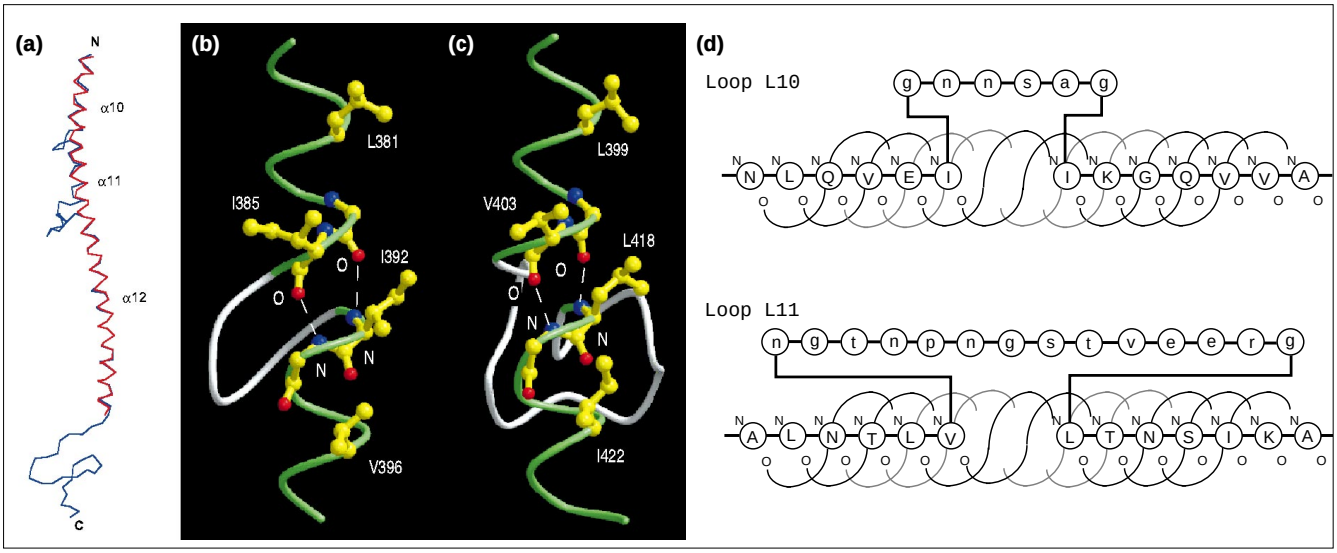
*$R_{cullis} = \Sigma||F_{PH}-F_P|-|F_H(calc)|| / \Sigma|F_{PH}-F_P|$, where the F_P, F_{PH} and F_H are the structure amplitudes of the native, heavy atom derivative and heavy atom compounds, respectively. †Phasing power = rms (F_H/E), where E is the lack of closure error.

Figure 2



Fibritin E secondary and tertiary structure. (**a**) Assignment of secondary structure elements to the fibritin E sequence. Rods represent α helices and arrows represent β strands. The periodic pattern $(abcdefg)_n$ denotes heptad repeats. Loops are shown by thin lines. The numbering of the α-helical segments and insertion loops is derived from the predicted structure of the wild-type fibritin [8]. Fibritin E contains α-helical segments α10 (partially), α11 and α12, as well as loops L10 and L11. (**b**) Stereo diagram of the fibritin E Cα backbone, drawn with the program MolView [42]. Each monomer is shown in a different color.

Figure 3



Insertion loops in the coiled-coil region of fibritin E. **(a)** Superposition of the fibritin E Cα backbone (blue) onto an ideal structure (red), which is calculated from parameters defining helix α12. **(b,c)** The arrangement of hydrophobic sidechains and hydrogen bonds at the insertion loops L10 and L11 (parts (b) and (c) are rotated by 50° and 100° about the coiled-coil axis relative to (a), respectively). α helices are in green and the insertion loops are in white. **(d)** The pattern of hydrogen bonds across each insertion loop. Hydrogen bonds are represented by black solid lines, whereas missing hydrogen bonds are in grey. Part (a) was drawn with the program MolView [42]; parts (b) and (c) were drawn with MOLSCRIPT [43] and RASTER3D [44].

is consistent with the differences of curvature of fibritin molecules viewed by electron microscopy [8] and with the increased temperature factors of atoms in the insertion loops of fibritin E. This pliability may permit the whiskers to bind the long tail fibers.

There is a glycine residue near the beginning and end of each insertion loop. The mutation of these glycines to non-glycine residues, especially to residues with large sidechains, would result in steric hindrance because the mainchain dihedral angles of these four glycines are in the regions of the Ramachandran plot that are unfavored for other residues. These glycines might be important for helix initiation and termination. The hydrogen bonds between sidechains within the loop structures might provide some stability to the loop conformation. Hence, the glycines at the loop termini may provide some flexibility between the rigid α-helical segments and the somewhat rigid loop structure itself.

Table 2

Comparison of various coiled-coil structures.

		Supercoil	
Oligomerization	Protein	Radius(Å)	Pitch(Å)
Dimer	Average*	5.0	128.5
Trimer	Average†	6.6	175.5
	Hemagglutinin‡	6.7	335.1
	Fibritin E§	6.4	137.8
Tetramer	GCN4 pLI#	7.6	200.5
Pentamer	COMP**	8.6	204.0

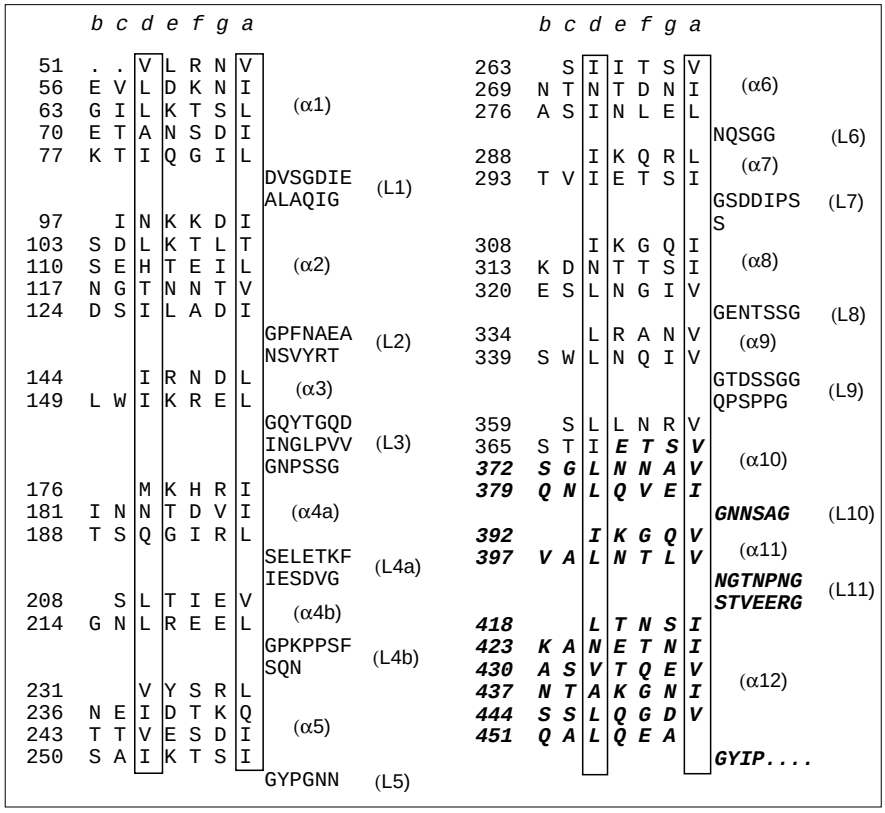
*Average of GCN4 leucine zipper p1 [15], catabolite gene activator protein [45] and GAL4 yeast transcription factor [46]. †Average of GCN4 leucine zipper pII [13], mannose-binding protein (rat) [47] and mannose-binding protein (human) [20]. ‡Hemagglutinin [14] has occasional additional inserted residues that interrupt the regular repeat pattern. §Average of fibritin E segments α10, α11 and α12. #GCN4 leucine zipper pLI [16]. **Cartilage oligomeric matrix protein (rat) [48].

No aromatic residues have been found at the *a* or *d* positions in parallel coiled-coil structures with strict heptad repeats and which have been solved to atomic resolution (for examples see Table 2). Although there are two phenylalanines at *a* positions in the coiled-coil region of hemagglutinin [14], the irregularity of the heptad repeats shifts these two *a* positions further out of the interior than in a regular coiled-coil structure. For a coiled-coil structure with exact heptad repeats, like fibritin, aromatic residues at either the *a* or *d* position are unlikely because of their bulky sidechains, which would result in steric hindrance.

The above considerations lead to some guidelines for the sequence requirements in the formation of coiled-coil structures with insertion loops, such as those that occur in fibritin. Firstly, a helix prior to an insertion loop ends at position *a*, and the next helix starts at position *c* or *d* in order to facilitate the helix continuation. Secondly, glycines or small polar residues are generally found at the ends of insertion loops. Thirdly, there are no aromatic residues at the *a*

Figure 4

Predicted secondary structure of the coiled-coil domain of complete wild-type fibritin. Residues listed in each column headed '*abcdefg*' are predicted to form α-helices (α1–α12) and residues at the right side of each column form insertion loops (L1–L11) between the two neighboring α-helical segments. Residues in the *a* or *d* positions are highlighted in boxes. Residues corresponding to the known fibritin E structure are shown in bold italics.



and *d* positions in regular parallel coiled-coil structures. These guidelines have now been applied to the complete wild-type fibritin sequence, leading to small revisions in the secondary structure predictions of Efimov *et al*. [8] (Fig. 4). The differences to the earlier prediction are in the positioning of the N termini of α-helical segments and the introduction of new loops where there are aromatic residues.

Simple rules have been proposed that predict whether the spontaneous oligomerization of coiled coils forms dimers, trimers or tetramers. These are based on the mutant studies of the GCN4 leucine zipper and are the consequence of efficient 'knobs-into-holes' packing of sidechains at the *a* and *d* positions (Table 3) [13,15,16]. However, analysis of a large variety of coiled-coil structures shows less clear preference of polypeptide sequence for different oligomerization states [17] (Table 3), although it is not clear whether the state of oligomerization of these structures has been affected by adjacent globular domains. Furthermore, a particular isolated heptad-repeat sequence can form a mixture of dimers and trimers if the residues at the *a* and *d* positions do not fit easily into available holes [18,19]. Out of 46 residues at *a* positions in the wild-type fibritin, there are 16 isoleucines, 14 valines and 10 leucines, and, out of 43 residues at *d* positions, there are 15 leucines, 13 isoleucines and 14

Table 3

Preferred residues in parallel coiled-coil oligomers.

	Rules based on GCN4 leucine zipper mutants*		Analysis of various coiled-coil structures†	
Oligomer	Position *a*	Position *d*	Position *a*	Position *d*
Dimer	β-branched residues	Leucines	Few valines	Few isoleucines, few valines
Trimer	Predominantly β-branched		No preference	
Tetramer	Leucines	β-branched residues	Not studied	

*See references [13,15,16]. †See reference [17].

Table 4

Parameters of the three fibritin E α-helical segments.

Parameters*	α10	α11	α12
Supercoil radius (Å)	6.4	6.6	6.2
Number of residues per supercoil turn	112.4	76.8	95.1
Supercoil pitch (Å)	165.2	109.7	138.6
Supercoil cross angle (°)	27.2	41.3	31.5
Radius of curvature (Å)	114.9	53.0	84.4
Radius of α-helix (Å)	2.27	2.24	2.30
Number of residues per α-helix turn	3.64	3.66	3.63
Rise per residue for α-helix (Å)	1.51	1.53	1.51
α-helix pairwise cross angle (°)	23.4	34.8	26.9
Pairwise inter-helical distance (Å)	11.0	11.4	10.8
Number of heptad repeats	2	1.5	5.5
Rms deviation (Å)†	0.32	0.49	0.33

*See references [13] and [41]. †Rms deviation is between calculated and observed positions of Cα atoms.

valines. This distribution is roughly consistent with an expectation for a typical trimer, rather than a dimer or a tetramer. On the other hand, in fibritin E, of the 10 residues at the *a* position, 6 are valines and 4 are isoleucines (all β branched) and of the 10 residues at the d positions, there are 6 leucines, 1 isoleucine, 1 alanine, 1 valine and 1 asparagine (i.e. leucines are predominant). This suggests that fibritin E would be more stable as a dimeric, rather than a trimeric, coiled-coil structure. A similar situation exists in the trimeric coiled-coil domain of human mannose-binding protein [20]. The lack of suitable residues at the *a* and *d* positions in fibritin E is also apparent in the large distances between closest contacts of symmetry-related sidechain atoms. In fibritin E, the average distance is 4.42 Å, whereas in an isoleucine zipper trimer it is 3.95 Å [13]. The larger distance in fibritin E is because of the dominance of valines ('knobs'), which do not extend far enough to completely fill the holes in the presence of some isoleucines that keep the α helices separated. This accounts for the larger distance between valines than there would be between isoleucines. Thus, the coiled coil in fibritin appears to be less energetically favorable than a perfect trimer.

A least-squares procedure was used to parameterize coiled-coil segments, as described in Materials and methods (Table 4). These parameters were compared to those for other coiled-coil structures (Table 2). The relatively short pitch of fibritin E compared to other trimeric coiled-coil structures can be explained by the efficient packing of valines at the *a* position [18].

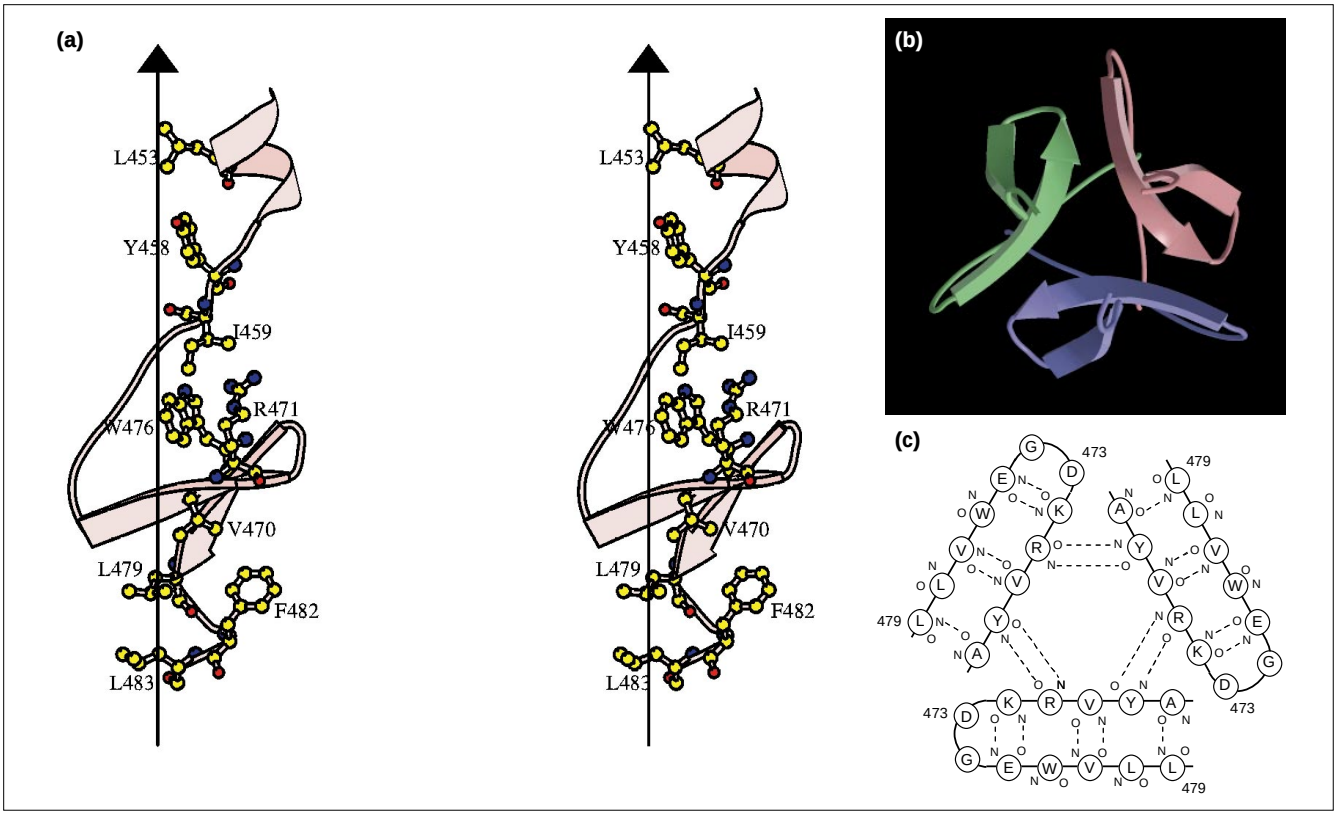
There was unassigned electron density on the coiled-coil axis, at the center between the symmetry-related Asn425 residues at the *d* position. This density was modeled as a chloride ion, which was present in the protein buffer. If this site had been a Zn^{2+} ion, it would probably have been substituted by Pb^{2+} in the heavy-atom derivative. In wild-type fibritin, the predicted structure has five asparagines at *d* positions. It has been suggested that asparagines at the *d* positions, which are stabilized by a chloride ion, increase the specificity of alignment [21,22]. In addition, there are salt bridges between residues from adjacent subunits — Glu384–Lys393, Glu414–Lys423 and Arg416–Glu426, which help to stabilize the trimeric coiled coil. However, the predicted salt bridge between residues Glu435 and Lys440 [8], at positions *g* and *e*, respectively, is not formed in the crystal structure.

C-terminal domain and its role in folding and trimerization

Each subunit in the fibritin E trimer contains a C-terminal region that has a β-hairpin motif and a loop connecting the β hairpin to the end of the last α-helical segment (Figs 1,2). Within a subunit, the connecting loop is stabilized by hydrophobic interactions and hydrogen bonds. Residues Tyr458, Ile459, Trp476, Val470, Leu479, Phe482 and Leu483 form a hydrophobic interior within the C-terminal domain of the trimer (Fig. 5a). In particular, Trp476 interacts with the hydrocarbon chain of Arg471, which in turn interacts with a symmetry-related Trp476 in another subunit. The three symmetry-related β hairpins, each stabilized by five mainchain hydrogen bonds, form a propeller-like structure (Fig. 5b). There are two mainchain hydrogen bonds between each pair of neighboring β hairpins (Fig. 5c). In addition, there is a salt bridge between Glu461 and Arg471 from adjacent subunits.

In order to investigate the folding and oligomerization pathway of fibritin, a number of truncated fibritin molecules were expressed in *E. coli* [23]. These mutant molecules lacked the N-terminal domain and a progressively larger amount of the amino end of the α-helical domain. The shortest mutant was composed of only the last 58 residues. All of the mutants formed soluble expression products, except a mutant without the last 18 residues of the C-terminal domain which produced an insoluble product. Similarly, substitution of Trp476 by leucine or serine in the C-terminal domain also produced an insoluble product (MM Shneider & VVM, unpublished results). On the other hand, substitution of Trp476 by tyrosine or phenylalanine had no effect on protein solubility. These observations are consistent with intersubunit interactions seen in the fibritin E structure. Substitution of Trp476 by leucine or serine disrupts this interaction, as the relatively small sidechains do not reach to the hydrocarbon chain of Arg471. However, mutations of Trp476 to aromatic residues preserve this interaction. Therefore, the residues which form the hydrophobic core of the C-terminal domain are essential for trimerization, so that mutations which disrupt the hydrophobic core apparently result in improper subunit association, leading to the formation of inclusion bodies.

Figure 5



The C-terminal domain of fibritin E. (**a**) Stereo diagram of the C-terminal domain of a fibritin E subunit. The sidechains shown are those located in the hydrophobic interior formed at the interface between three symmetry-related subunits. The vertical line shows the trimer axis. Atoms are shown in standard colors. (**b**) Ribbon diagram of the C-terminal domain looking along the trimer axis, each subunit is shown in a different color. (**c**) Mainchain hydrogen bonds formed within the C-terminal domain of a fibritin E trimer. Parts (a) and (b) were drawn with MOLSCRIPT [43] and RASTER3D [44].

A monomeric α-helical structure, with its amphipathic character, would be unstable. There is also some evidence that, in a coiled coil, the α helices are not formed until the individual chains come together [24]. The formation of inclusion bodies implies random association of the polypeptide chains due to non-specific hydrophobic interactions between the helices. Therefore, the likely function of the C-terminal domain of fibritin is perhaps to provide correct alignment of the subunits to each other. Thus, after expression of the polypeptide, the correctly aligned trimer would only be able to form once the complete polypeptide has been synthesized. By that time, however, the rest of the molecule may already have been associated randomly. Presumably, the instability of the random associations would permit their re-association into correctly aligned trimers upon the formation of α helices and coiled coils. Apparently, the interactions within the stable trimeric C-terminal domain appear to override the tendency of the last three segments of the coiled-coil domain to form dimers.

Examples of proteins that contain coiled-coils associated with non-coiled-coil C-terminal domains are the type I macrophage receptor [25] and the large class of 'collectin' proteins [26]. The assembly pathway of fibritin proposed here might apply especially to α-fibrous proteins, such as tropomyosin [27], intermediate filament protein [28] and lamin [29]. A C-terminal domain is also required for the assembly of procollagen, in which folding of the collagen helix proceeds from the carboxy to the amino end [30]. The structure of fibritin E should, therefore, be valuable for the study of mechanisms that determine folding and oligomerization and to provide insights into the function of coiled-coil proteins with stabilizing non-coiled-coil domains.

Biological implications

Fibritin is a fibrous structural protein of bacteriophage T4. During phage assembly, six fibritin molecules attach to each virion neck through their N-terminal domains, to form a collar with six fibers ('whiskers'). The whiskers

interact with the long tail fibers and accelerate the assembly and attachment of the long tail fibers to the baseplate.

The structure of fibritin E, a deletion mutant consisting of the last 119 residues of the 486-residue wild-type protein, has been determined. The structure of fibritin E consists of a long coiled coil and a small globular C-terminal domain. The coiled coil is composed of three segments linked together by exposed loop structures. These insertion loops introduce flexibility into the otherwise rigid coiled coil, increasing the rotational freedom of the whiskers, which may be necessary to accelerate the attachment of the long tail fibers to the tail baseplate.

The sequence of the coiled-coil domain of fibritin E is not optimal for the formation of either a stable dimer or trimer. The small, globular, C-terminal domain contains a β-hairpin structure and its residues are involved in extensive hydrophobic and some polar interactions between the subunits of the trimer. There is mutational and structural evidence that the C-terminal domain provides correct alignment of the subunits and defines the oligomerization state. The same strategy for attaining the correct alignment between monomers might be used by the large class of α-fibrous proteins, many of which have a globular domain associated with a coiled-coil structure. The structure of fibritin, therefore, provides valuable details of the mechanisms of protein folding and oligomerization. It also provides insights into the functions of coiled-coil proteins.

Using fibritin as a template, a surface-display T4 vector was recently developed. This is the first example where fused chimeric molecules can be designed for a phage display system with knowledge of the atomic structure of the template.

Materials and methods

Crystallization

Recombinant fibritin E was expressed in *E. coli* strain BL21(DE3) [31] and purified as described by Efimov *et al.* [8]. Crystals of fibritin E were obtained using the hanging drop vapor diffusion method [10]. The best crystals grew at 22°C when 1 μl droplets of the recombinant protein (29 mg ml^{-1} in 10 mM Tris-HCl, pH 7.5) were mixed with 1 μl well solution containing 34% PEG400, 0.1 M $Zn(CH_3COO)_2$, 0.1 M MES, pH 6.0. Rhombohedral crystals appeared after one day. Their space group was R32 with hexagonal cell dimensions a = 41.2, c = 358.7 Å.

Data collection

Native and derivative data sets were recorded at room temperature on an R-axis IIC image plate detector, which was mounted on a Rigaku rotating anode X-ray source operated at 50 kV and 100 mA. Each data set was from a single crystal, and all the data were processed using the programs DENZO and SCALEPACK [32] (Table 5).

Structure determination

Difference Patterson maps were used to identify the heavy-atom sites of the uranyl acetate derivative. Thereafter, the heavy-atom sites of the Pb and Pt derivatives were located by cross Fourier maps combined with difference Patterson maps. The position of the Pb site was the same as the major site of the UO_2 derivative. Heavy-atom parameters were refined using MLPHARE [33] (Table 1). The anomalous signals from the UO_2 and Pb derivatives were included in the refinement and the phasing. The overall figure of merit for the calculated phases was 0.54 for data between 15 and 3 Å resolution. The initial MIR map calculated to 3.0 Å resolution showed good density for most of the α helices. The program DM [34] (employing solvent flattening and histogram matching) was used for further phase improvement. The CCP4 suite of programs [35] was used for all crystallographic calculations.

Although the MIR electron density map allowed most of the amino acids to be fitted unambiguously using the program O [36] and the known sequence, the bigger insertion loop (L11) and the N-terminal helical segment had been deleted by the solvent flattening procedure. Another MIR map, modified using DM without solvent flattening, gave good density for these regions. The initial model included 111 amino acid residues out of a total of 119. Crystallographic refinement was performed using the program X-PLOR [37] with idealized amino acid parameters as defined by Engh and Huber [38]. Four percent of the reflections were omitted during the refinement and used as a test set to monitor the refinement process by calculating R_{free} [39]. The F_{obs} amplitudes were re-scaled to F_{calc} using an anisotropic scale factor given by the expression $k\exp(B_{11}h^2a^{*2} + B_{22}k^2b^{*2} + B_{33}k^2c^{*2} + 2B_{12}hka^*b^* + 2B_{23}klb^*c^* + 2B_{13}hla^*c^*)$, where $B_{11} = B_{22} = -7.36$, $B_{33} = 14.71$, $B_{12} = -9.83$ and $B_{23} = B_{13} = 0.0$ Å^2.

Individual atomic temperature factors were refined after several rounds of positional refinement and manual correction. A total of 80 water molecules and a bulk solvent model (B = 160 Å^2, k = 0.43) were added. The final R_{free} was 0.267 and R_{work} was 0.217 for data between 30.0 and 2.2 Å resolution. The stereochemistry of the model had a root mean square (rms) deviation of 0.010 Å for bond lengths from the idealized values and of 1.3° for bond angles. The overwhelming majority of amino acids could be associated with well-defined electron density. The exceptions are the sidechains of Pro408, Asn409 and Ser411,

Table 5

Data collection statistics.

Data set	Soaking time (days)	Heavy atom concentration (mM)	Highest resolution (Å)	No. of measured reflections	No. of unique reflections (% of total)	R_{merge}*	R_{diff}†
Native	–	–	2.2	57901	6490 (92.6%)	0.053	–
$UO_2(CH_3COO)_2$	1	20	3.0	20545	2666 (99.6%)	0.057	0.142
$Pb(CH_3COO)_2$	3	1	2.5	54054	4527 (99.5%)	0.046	0.109
K_2PtCl_4	3	1	2.5	30181	4521 (95.2%)	0.031	0.090

*$R_{merge} = \sum_h\sum_i |I_{hi} - \langle I_h\rangle| / \sum_h\sum_i I_{hi}$. †$R_{diff} = \sum_h |F_{P_h} - F_{PH_h}| / \sum_h F_{P_h}$,
where the F_P and F_{PH} are the structure amplitudes of the native and heavy atom derivative compounds, respectively.

which are located in an insertion loop (L11) and have high temperature factors. The final protein model was analyzed with the program PROCHECK [40]. The Ramachandran plot has 91.7% of the residues in the most favored regions. The only residues in the generously allowed region is Asp473, which makes a salt bridge with Lys472 within the same subunit.

Coiled-coil parameterization

A least-squares refinement procedure was used to refine 13 parameters which defined the Cα positions of the coiled coil in the reference strand. Six parameters described the orientation and position of the coiled coil. Another seven parameters [41] described the coiled coil itself. All the Cα atom positions in the coiled coils were used for refinement. The positions of the other strands in the coiled coils were generated from the reference strand by applying the rotational symmetries (two, three, four or fivefold). The computer program is available from the authors.

Accession numbers

Coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank (entry codes 1aa0 and r1aa0sf, respectively).

Acknowledgements

We would like to thank Guoguang Lu, Jordi Bella, Alan Friedman, Prasanna Kolatkar, Sukyeong Lee, Lidiya Kurochkina, Yuri Londer and Mikhail Shneider for their helpful advice and discussion. We also thank Thomas J Smith for the use of the program MolView in creating Figures 2b and 3a. The work was supported by a grant from the National Science Foundation (MCB-9102855) to MGR, and also by grants from the Protein Engineering Council of Russia and from the Russian Foundation of Basic Research (96-04-48035) to VVM. VVM is an International Howard Hughes Medical Institute Scholar.

References

1. Eiserling, F.A. & Black, L.W. (1994). Pathways in T4 morphogenesis. In *Molecular Biology of Bacteriophage T4*. (Karam, E., ed), pp. 209–212, American Society for Microbiology, Washington, DC, USA.
2. Wood, W.B. & Conley, M.P. (1979). Attachment of tail fibers in bacteriophage T4 assembly. *J. Mol. Biol.* **127**, 15–29.
3. Bloomfield, V.A. (1983). Physical studies of morphogenetic reactions. In *Bacteriophage T4*. (Mathews, C.K., Kutter, E.M., Mosig, G. & Gerget, P.B., ed), pp. 270–276, American Society for Microbiology, Washington, DC, USA.
4. Wood, W.B. & Henninger, M. (1969). Attachment of tail fibers in bacteriophage T4 assembly: some properties of the reaction *in vitro* and its genetic control. *J. Mol. Biol.* **39**, 608–618.
5. Conley, M.P. & Wood, W.B. (1975). Bacteriophage T4 whiskers: a rudimentary environment-sensing device. *Proc. Natl. Acad. Sci. USA* **72**, 3701–3705.
6. Sobolev, B.N. & Mesyanzhinov, V.V. (1991). The *wac* gene product of bacteriophage T4 contains coiled-coil structure patterns. *J. Biomol. Struct. Dynam.* **8**, 953–965.
7. Cohen, C. & Parry, D.A.D. (1990). α-helical coiled coils and bundles: how to design an α-helical protein. *Proteins* **7**, 1–15.
8. Efimov, V.P., *et al.,* & Mesyanzhinov, V.V. (1994). Fibritin encoded by bacteriophage T4 gene *wac* has a parallel triple-stranded α-helical coiled-coil structure. *J. Mol. Biol.* **242**, 470–486.
9. Efimov, V.P., Nepluev, I.V. & Mesyanzhinov, V.V. (1995). Bacteriophage T4 as a surface expression vector. *Virus Genes* **10**, 173–177.
10. Strelkov, S.V., Tao, Y., Rossmann, M.G., Kurochkina, L.P., Shneider, M.M. & Mesyanzhinov, V.V. (1996). Preliminary crystallographic studies of bacteriophage T4 fibritin confirm a trimeric coiled-coil structure. *Virology* **219**, 190–194.
11. Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.
12. Crick, F.H.C. (1953). The packing of α-helices: simple coiled-coils. *Acta Cryst.* **6**, 689–697.
13. Harbury, P.B., Kim, P.S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80–83.
14. 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.
15. 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.
16. 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.
17. Woolfson, D.N. & Alber, T. (1995). Predicting oligomerization states of coiled coils. *Protein Sci.* **4**, 1596–1607.
18. Gonzalez Jr., L., 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.
19. Ogihara, N.L., Weiss, M.S., DeGrado, W.F. & Eisenberg, D. (1997). The crystal structure of the designed trimeric coiled coil coil-V_aL_d: implications for engineering crystals and supramolecular assemblies. *Protein Sci.* **6**, 80–88.
20. Sheriff, S., Chang, C.Y. & Ezekowitz, R.A. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α-helical coiled coil. *Struct. Biol.* **1**, 789–794.
21. Fass, D., Harrison, S.C. & Kim, P.S. (1996). Retrovirus envelope domain at 1.7 Å resolution. *Nat. Struct. Biol.* **3**, 465–469.
22. Lumb, K.J. & Kim, P.S. (1995). A buried polar interaction imparts structural uniqueness in a designated heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648.
23. Sobolev, B.N., Shneider, M.M., Marusich, E.I. & Mesyanzhinov, V.V. (1995). Fibritin and adhesin of bacteriophage T4 tell us how fibrous protein folds and assembles. In *Evolutionary Biochemistry and Related Areas of Physicochemical Biology*. (Poglazov, B., Kurganov, B.I., Kritsky, M.S. *&* Gladilin, K.L., eds), pp. 351–374, Bach Institute of Biochemistry and ANKO, Moscow, Russia.
24. Sosnick, T.R., Jackson, S., Wilk, R.R., Englander, S.W. & DeGrado, W.F. (1996). The role of helix formation in the folding of a fully α-helical coiled coil. *Proteins* **24**, 427–432.
25. Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P. & Krieger, M. (1990). Type I macrophage scavenger receptor contains α-helical and collagen-like coiled coils. *Nature* **343**, 531–535.
26. Hoppe, H. & Reid, K.B.M. (1994). Collectins — soluble proteins containing collagenous regions and lectin domains — and their roles in innate immunity. *Protein Sci.* **3**, 1143–1158.
27. Phillips Jr., G.N., Fillers, J.P. & Cohen, C. (1986). Tropomyosin crystal structure and muscle regulation. *J. Mol. Biol.* **192**, 111–131.
28. Parry, D.A.D. & Fraser, R.D.B. (1985). Intermediate filament structure. 1. Analysis of IF protein sequence data. *Internatl. J. Biol. Macromol.* **7**, 203–213.
29. McKeon, F.D., Kirschner, M.W. & Caput, D. (1986). Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. *Nature* **319**, 463–468.
30. Prockop, D.J. & Kivirikko, K.I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Ann. Rev. Biochem.* **64**, 403–434.
31. Studier, F.W., Rosenberg, J.J., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.
32. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276A**, 307–326.
33. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering. Proceedings of the CCP4 Study Weekend.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, UK.
34. Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34–38.
35. Collaborative computational project number 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
36. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
37. Brünger, A.T. (1992). *X-PLOR. Version 3.1. A system for X-ray Crystallography and NMR*. Yale university press, New Haven & London.

38. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392–400.
39. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
40. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemistry quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
41. Crick, F.H.C. (1953). The Fourier transform of a coiled-coil. *Acta Cryst.* **6**, 685–689.
42. Smith, T.J. (1995). MolView: a program for analyzing and displaying atomic structures on the Macintosh personal computer. *J. Mol. Graphics* **13**, 122–125.
43. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
44. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst. D* **50**, 869–873.
45. Weber, I.T. & Steitz, T.A. (1987). Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J. Mol. Biol.* **198**, 311–326.
46. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356**, 408–414.
47. Weis, W.I. & Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. *Structure* **2**, 1227–1240.
48. Malashkevich, V.N., Kammerer, R.A., Efimov, V.P., Schulthess, T. & Engel, J. (1996). The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel? *Science* **274**, 761–765.