Acta Cryst. (1998). D**54**, 805–816

Structure of Bacteriophage T4 Fibritin M: a Troublesome Packing Arrangement

SERGEI V. STRELKOV,[a,b] YIZHI TAO,[a] MIKHAIL M. SHNEIDER,[b] VADIM V. MESYANZHINOV[b] AND MICHAEL G ROSSMANN[a]*

[a]*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA, and* [b]*Shemyakin-Ovchnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya Street, Moscow 117 871, Russia. E-mail: mgr@indiana.bio.purdue.edu*

(*Received 19 September 1997; accepted 8 December 1997*)

Abstract

Fibritin, a 52 kDa product of bacteriophage T4 gene *wac*, forms 530 Å long fibers, named whiskers, that attach to the phage neck and perform a helper function during phage assembly. Fibritin is a homotrimer, with its predominant central domain consisting of 12 consecutive α-helical coiled-coil segments linked together by loops. The central domain is flanked by small globular domains at both ends. Fibritin M is a genetically engineered fragment of the wild type and contains 74 amino-acid residues corresponding to the last coiled-coil segment and the complete carboxy-terminal domain. The crystals of fibritin M belong to the rare space group *P*3 with three crystallographically independent trimers in the unit cell. The structure has been established at 1.85 Å resolution by combining molecular and isomorphous replacement techniques. One of the two heavy-atom derivatives used was gaseous xenon. A substantial fraction of residues in each independent trimer is disordered to various extents in proportion to the lack of restraints on the molecules provided by the lattice contacts. Accurate modeling of the solvent present in the crystals was crucial for achieving good agreement with experimental data.

1. Introduction

More than 40 different structural proteins constitute one particle of bacteriophage T4. Fibritin, a product of gene *wac*, forms 530 Å long fibers, named whiskers. The whiskers attach to the phage neck and function during the later stages of the phage assembly. Interaction of the whiskers with the long tail fibers increases the rate of assembly and subsequent attachment of the long tail fibers onto the tail baseplate (Eiserling & Black, 1994).

Recombinant fibritin, overexpressed in *E. coli*, assembles into filamentous particles that have the same biochemical and immunogenic properties as whiskers that were purified from the phage. Recombinant fibritin complements *in vivo* 'whiskerless' T4 phage that lacks the gene *wac* (Efimov *et al.*, 1994). The fibritin molecule is formed by three identical 52 kDa polypeptide chains that run parallel along its length. Its predominant central part is a trimeric α-helical coiled-coil domain that is flanked by small globular domains at both ends. As in other coiled-coil structures, the amino-acid sequence of fibritin (486 residues) has a distinct heptad repeat $(abcdefg)_n$ which contains predominantly hydrophobic residues at the *a* and *d* positions, while the other positions are occupied mostly by polar residues. Since seven residues make approximately two α-helical turns, the hydrophobic residues of consecutive heptads are located on one side of the helix, and this provides for the formation of the coiled-coil core (Lupas, 1996). An unusual feature of fibritin is that the coiled coil is interrupted by insertion of loops, consisting of five to 20 residues, which create 12 separate segments (Efimov *et al.*, 1994).

The full-length fibritin could not be crystallized, probably because of its inherent flexibility. However, a series of successively smaller fibritin fragments (Fig. 1) was engineered and could be expressed in a soluble form in an *E. coli* system (Sobolev *et al.*, 1995). The linear repetitive structure of the segmented coiled coil suggests that such fragments can be studied with the possibility of subsequent interpolation onto the full-length fibritin. Crystals were obtained of fibritins E and M (Strelkov *et al.*, 1996), and also of fibritin B (unpublished results). The crystal structure of fibritin E has been determined and provided an insight into the mechanism of folding of this protein (Tao *et al.*, 1997). The key structural element of fibritin responsible for the initiation of the folding process was shown to be its C-terminal globular domain. This conclusion is supported by the studies of various deletion and single-point mutants of fibritin (Efimov *et*

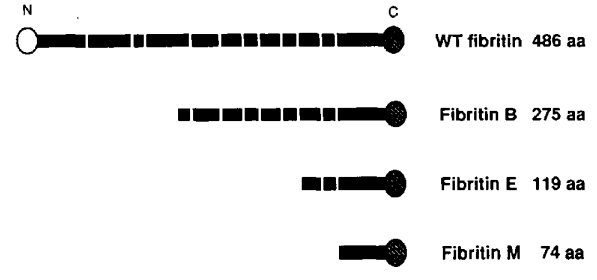

Fig. 1. Schematic diagrams of the wild-type fibritin and the engineered fragments. The black boxes represent the individual coiled-coil segments and the ovals show the N- and C- terminal domains.

Table 1. *Statistics for diffraction data used in structure determination*

Data set	Native 1	Xe derivative	Uranyl derivative
Precipitant		Lithium sulfate	
Temperature of data collection (K)	100	298	100
Oscillation range per frame (°)	0.95	1.7	1.5
Number of frames	103	101	125
Resolution limits (Å)	23-1.85	38-2.1	24-2.3
$\sigma(I)$ cut-off	None	None	None
No. of observations	39783	39649	45405
No. of unique reflections†	14861	11494	8423
Completeness (%)‡	89.8§	98.4	99.7
	(53.6)	(89.3)	(100.0)
Cell dimensions (Å) a	43.68	44.33	43.27
c	90.61	91.25	90.54
V_m (Matthews coefficient) with three molecules per asymmetric unit ($Å^3\ Da^{-1}$)		2.15	
Mosaic spread (°)	0.53	0.28	0.91
R_{merge}‡¶ in Laue class $\bar{3}$	0.036	0.077	0.081
	(0.182)	(0.373)	(0.235)
R_{ano}††	—	0.060	0.058
		(0.282)	(0.134)
R_{diff}‡‡	—	0.111	0.256

† For the derivatives, Friedel pairs were counted as one unique reflection. ‡ The data in parentheses are for the reflections in the outermost of ten equal bins in $(2\sin\theta/\lambda)^2$. § The completeness is 99.5% for resolution 40–2.3 Å. ¶ $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h\rangle| / \sum_h \sum_i \langle I_h\rangle$, where I_{hi} are all (both Bijvoet positive and negative) measurements of a reflection. †† $R_{ano} = \sum_h |\langle I_h\rangle^+ - \langle I_h\rangle^-| / 2 \sum_h \langle I_h\rangle$ where mean values $\langle I_h\rangle^+$ and $\langle I_h\rangle^-$ are calculated over all Bijvoet positive and negative instances of a reflection, respectively. ‡‡ $R_{diff} = \sum_h |F_{PH} - F_P| / \sum_h F_P$ R_{diff} for the uranyl derivative is against native data set 1; R_{diff} for the Xe derivative is against another native data set collected at room temperature.

al., 1994; M. M. Shneider, V. V. Mesyanzhinov, in preparation).

The subject of the present work is fibritin M that contains residues 413–486 of the wild type. These residues form the last coiled-coil segment ($5\frac{1}{2}$ heptad repeats) and the complete C-terminal domain. Fibritin M, but not fibritin E, has five point mutations relative to the wild-type sequence, all occurring in the coiled-coil part of the molecule (Fig. 2). These mutations were introduced to enhance the stability of the protein and, hence, the likelihood of a structure determination. In particular, the mutation Ser421→Lys was introduced to test the possible formation of an interchain salt bridge with Glu426. The mutations Asn428→Asp and Thr433→Arg were designed to create a similar interchain salt bridge between these two residues. Residue 425, an asparagine in a *d* position, was replaced by an isoleucine, which is generally favorable in this position for a trimeric coiled coil (Harbury *et al.*, 1994). Here, we present the crystal structure of fibritin M at 1.85 Å resolution and discuss the difficulties encountered during the structure determination.

2. Experimental

2.1. *X-ray data collection*

The crystal form of fibritin M used for the structure determination reported here was obtained with Li_2SO_4 as the precipitant (Strelkov *et al.*, 1996). In addition, two other crystal forms were examined. The first one was obtained with 22%(*v*/*v*) polyethylene glycol (average M_r 400), 0.05 *M* Zn acetate in 0.1 *M* MES/Na buffer, pH 6.0, as the precipitant solution, while the other crystal form was grown from 0.55 *M* Na/K tartrate in 0.07 *M* HEPES/Na buffer, pH 7.5. In both cases, hanging drops were made by mixing equal volumes of 20 mg ml^{-1} protein solution with the precipitant solution that was placed in the well.

Diffraction data (Tables 1, 2) were collected either at room temperature from a crystal mounted in a capillary or at 100 K from a flash-frozen crystal in a loop. Gradual addition of glycerol as a cryoprotectant to a final concentration of 25%(*v*/*v*) was used for the crystals grown with Li_2SO_4. The crystals grown with PEG were flash-frozen in their mother liquor. Cu $K\alpha$ radiation (λ =

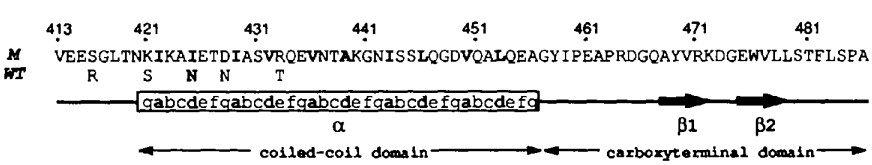

Fig. 2. Amino-acid sequence and observed secondary-structure elements of fibritin M. Residue numbering of the wild-type fibritin is preserved. The wild-type residues that were mutated are given below the complete fibritin M sequence. The heptad repeat is shown as $(\mathbf{abc}\mathbf{d}\mathbf{efg})_n$, with the hydrophobic core residues in bold.

Table 2. *Comparison of three native crystal forms*

Crystal form (precipitant)	Litihium sulfate		PEG 400	Na/K tartrate
Temperature of data collection (K)	298	100	100	298
Highest resolution (Å)	2.1	1.85	2.2	2.1
Trigonal cell dimensions (Å) a	44.26	43.68	43.12	44.38
c	91.23	90.61	90.39	91.26
R_{merge}† in Laue class $\bar{3}$	0.067	0.036	0.048	0.065
R_{merge}† assuming Laue class $\bar{3}1m$	0.456	0.454	0.061	0.158
Twinning fraction α	0	0	0.49	0.41

† $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$.

1.5418 Å) and an R-axis IIc image-plate detector were used for the data collection. The data were processed with the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Many low-resolution (>6 Å) reflections on the diffraction images were 'smeared' in radial reciprocal lattice directions. This phenomenon was noticeable at room temperature, but more pronounced after freezing. Smeared reflections extended into the designated region used for measuring background and, therefore, could be distinguished by an increased variability of the background. All reflections with large random variation of optical density in the designated background pixels were rejected. Nevertheless, the remaining data within 6 Å resolution was 86.5% complete and had a surprisingly good R_{merge} of 0.020 with an average redundancy of 2.5 (native data set 1, Table 1).

The crystals grown with Li_2SO_4 belong to space group $P3$, which was confirmed by observing significant $(00l)$ reflections with $l \neq 3n$ (Table 3). The subsequent computations were performed with programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994), unless otherwise noted.

Table 3. *List of the (00l) reflections for the native data set 1*

Reflections with $l = 3n$ (n = 1, 2, ...) are given in bold type.

l	F	$\sigma(F)$
4	394.9	9.7
5	448.7	10.4
6	**331.9**	**7.4**
7	406.2	5.9
8	156.9	2.6
9	**67.1**	**1.6**
10	285.2	6.3
11	168.7	3.8
12	**244.8**	**5.3**
13	511.8	10.8
14	245.7	6.4
15	**323.5**	**12.0**
16	793.5	16.9
17	Not observed	–
18	**464.9**	**10.8**
19	708.9	14.7
20	411.7	9.4

2.2. *Molecular replacement*

The model for molecular replacement (MR) was based on the structure of fibritin E (Tao *et al.*, 1997). The fibritin M sequence is completely contained within the longer fibritin E sequence, but differs from it by the five point mutations described above. The model was derived by taking the residues 418–483 of the fibritin E structure (the complete coiled-coil segment 12 and the C-terminal domain) and replacing the mutated residues by alanines (Fig. 2). Residues 413–417 were not included since they are part of a loop structure in fibritin E and could have a different conformation in fibritin M.

Cross-rotation and translation searches were performed with the program *AMoRe* (Navaza, 1994). A trimeric model was placed on the c axis of a triclinic cell with the shape of a cube and edges of length 120 Å. The cross-rotation searches gave consistent results for various combinations of low (8–15 Å) and high (4–2.5 Å) resolution limits of the diffraction data (Table 4). The two highest peaks corresponded to two trimers (A and B) oriented in opposite directions along the c axis of the $P3$ cell. Furthermore, a self-rotation function, calculated with the program *GLRF* (Tong & Rossmann, 1997), showed a prominent peak in the $\kappa = 180°$ section corresponding to a direction perpendicular to the crystallographic c axis (data not shown). This peak was consistent with the rotational relationship between molecules A and B. The third highest peak of the cross-rotation function corresponded to molecule C that was parallel to molecule B, but their azimuthal angles about the threefold axis differed by 14°. A translation-function search using *AMoRe* established the positions of molecules A, B and C on the three different threefold axes of the $P3$ unit cell (Table 5). The superiority of the true translational solutions over false ones was minimal but consistent with searches using various resolution limits imposed on the data. These solutions were subsequently refined in *AMoRe* as rigid bodies. Furthermore, a model

Table 4. *Cross-rotation function search results for data between 10 to 3 Å resolution*

Peak No.	Correlation coefficient†	Molecule	Eulerian angles (°) α	β	γ
1	0.198	*A*	3.5	0.0	0.0
2	0.175	*B*	0.0	180.0	9.1
3	0.108	*C*	0.0	180.0	114.8
First false peak for trimer parallel to *c* axis	0.074	–	93.9	0.0	0.0
First false peak for general orientation	0.065	–	84.2	123.1	96.9

† Corr = $[\sum(F_{obs} - \langle F_{obs}\rangle)(F_{calc} - \langle F_{calc}\rangle)]/\{[\sum(F_{obs} - \langle F_{obs}\rangle)^2]^{1/2}[\sum(F_{calc} - \langle F_{calc}\rangle)^2]^{1/2}\}$ where $\langle F\rangle$ is an average over all reflections.

Table 5. *Translation search results for data between 10 to 3 Å resolution*

Molecule†	Correlation coefficient	*R* factor‡	Fractional coordinates *x*	*y*	*z*
A	0.272	0.526	0.0	0.0	0.0
B	0.394§	0.493§	1/3	2/3	0.4811
C	0.404¶	0.488¶	2/3	1/3	0.1214

† Positioning molecule *A* on one of the three threefold axes, as well as assigning its *z* coordinate, is arbitrary. Molecules *B* and *C* are placed relative to the assumed position of molecule *A*. ‡ $R = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$. § With molecule *A* in place. ¶ With molecules *A* and *B* in place.

consisting only of the trimeric C-terminal domain (residues 458–483) was found to give the best rotation-function signal (results not shown), despite representing only 35% of the protein. Using this model avoided the many false solutions that resulted from alternative superpositions of the coiled coil onto itself after screw translations along its axis. Confirmation of these results was subsequently obtained from the isomorphous replacement technique.

2.3. *Multiple isomorphous replacement (MIR)*

Diffraction data were collected for many potential heavy-atom derivatives. The space group *P*3 has four possible settings resulting from various combinations of 60° rotations of the coordinate system about the three-fold axis and 180° rotations about the [110] direction. The four settings correspond to transformations (h,k,l), $(k,h,-l)$, $(-h,-k,l)$ and $(-k,-h,-l)$ of the reciprocal space. For each derivative data set, $R_{diff} = \sum|F_{PH} - F_P|/\sum F_P$ was calculated in the four possible ways relative to an arbitrarily chosen setting for the native data set. One of the resulting *R* values was always considerably lower, while the others were greater than 0.35. The former was, therefore, the true value obtained with both data sets indexed in the same way.

Fibritin M contains no cysteine, methionine or histidine residues that are useful as specific heavy-atom substitution sites in proteins (Fig. 2). Nevertheless, a few heavy-atom compounds yielded significant R_{diff} values, and yet did not cause crystal cracking or changes in cell dimensions. However, none of the derivatives gave difference Patterson maps that could be interpreted. After the MR solution was obtained, the binding sites of two heavy-atom compounds could be determined from difference Fourier maps calculated with the model phases. A Xe derivative data set was collected at room temperature from a crystal mounted in a capillary. Xe gas was applied with a pressure of 10 bars for 2 h prior to and also during the room temperature data collection. The data were first scaled to another native data set that had been collected at room temperature, resulting in $R_{diff} = 0.11$ for data between 10 and 2.1 Å resolution. An isomorphous difference Fourier map and a Bijvoet difference Fourier map (Rossmann, 1960) had their major peaks in the same positions and suggested that all Xe binding sites were located on the crystallographic threefold axes, predominantly in the coiled-coil domain (Table 6). Two binding sites, inserted between the adjacent triplets of hydrophobic side chains from the residues in the *a* and *d* positions, were the same in each of the three independent molecules *A*, *B* and *C* (Fig. 3). In addition, molecules *A* and *C* had a third minor binding site. Although there were significant changes in cell dimensions upon freezing (see Table 2), the Xe derivative data set was also scaled to the native data set collected under cryogenic conditions, yielding R_{diff} of 0.18 for data between 10 and 3 Å resolution. The corresponding difference Fourier maps confirmed all major peaks present in the map calculated using the room temperature native data. This justified the subsequent use of this derivative data set for phasing the 'cryo' native data. Furthermore, Xe difference Fourier maps contained clear peaks corresponding to the binding sites on molecule *C* even if this molecule was excluded from the MR phasing calculation. This

Table 6. *Heavy-atom positions*

HA	Site†	Occupancy‡	Ligand(s)	Distance(s) to HA (Å)
Xe	1*A*	0.45	Ile429 $C_{\gamma 2}$/Val432 $C_{\gamma 1}$	4.1/3.8
	2*A*	0.93	Val436 $C_{\gamma 2}$/Ala439 C_{β}	4.1/3.7
	3*A*	0.22	Val470 C_{α}/Trp476 $C_{\zeta 3}$	3.9/3.7
	1*B*	0.51	Ile429/Val432 $C_{\gamma 2}$	Residue disordered/3.8
	2*B*	0.72	Val436 $C_{\gamma 2}$/Ala439 C_{β}	3.9/4.0
	1*C*	0.30	Ile429 $C_{\gamma 1}$/Val432 $C_{\gamma 1}$	4.2/4.4
	2*C*	0.51	Val436 C_{α}/Ala439 C_{β}	4.6/3.6
	3*C*	0.36	Val450 $C_{\gamma 2}$/Leu453 $C_{\delta 1}$	3.2/3.2
Uranyl	1*A*	0.55	Glu414, Glu415?	Residue disordered
	2*A*	0.48	Glu414, Glu415?	Residue disordered
	3*A*	0.42	Glu414, Glu415?	Residue disordered
	4*B*	0.34	Gln451($O_{\varepsilon 1}$)	3.8
	5*B*	0.29	Gln451($O_{\varepsilon 1}$)	4.2

† A letter in a site name designates the molecule in which this site is located. ‡ Occupancies, based on placing the native data onto a roughly absolute scale, after refinement with *MLPHARE*.

confirmed the presence of molecule *C* in the structure for which the MR indication had been weak (Table 4). In addition, a $(2F_{obs}-F_{calc})$ map based on MR phases from molecules *A* and *B* combined with isomorphous Xe phases showed density for a part of molecule C so that it could be unambiguously positioned, this was the same position as found by the MR search procedure mentioned above.

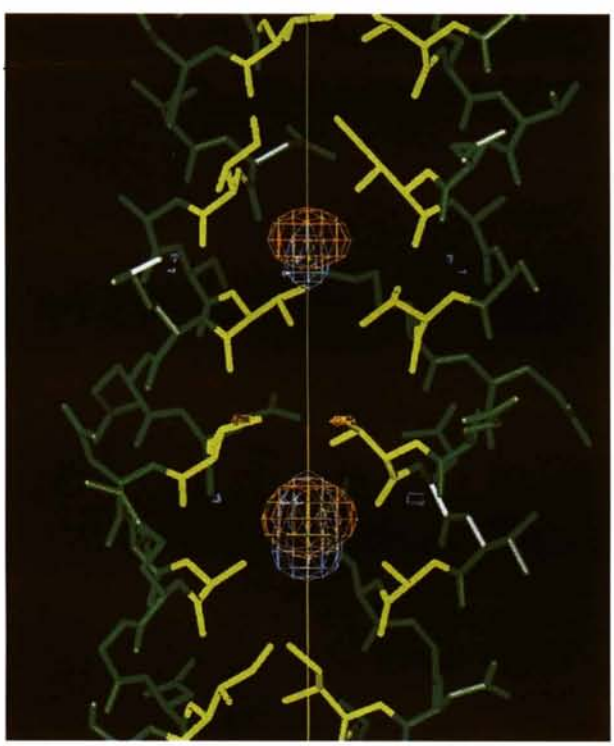

Fig. 3. The two major xenon binding sites in the coiled-coil domain of molecule *A*. Shown are isomorphous (coral) and anomalous (cyan) difference Fourier maps with phases from the final model. Both maps were calculated from data between 23 and 3 Å resolution and contoured at 4σ. Only two chains of trimer *A* are shown, with the hydrophobic core residues in yellow. The figure was drawn with program *O* (Jones *et al.*, 1991).

Gaseous xenon had been successfully used to prepare an isomorphous derivative of another coiled-coil protein, cartilage oligomeric matrix protein (Malashkevich *et al.*, 1996), where Xe atoms were found to insert into the hydrophobic core of a pentameric coiled coil in a fashion similar to that now observed in fibritin M. Thus, Xe derivatization may be especially useful for proteins containing coiled-coil motifs. Furthermore, (Soltis *et al.*, 1997) reported a technique for freezing crystals that had been under gas pressure. This technique should further enhance the use of noble gases for preparing isomorphous derivatives.

A uranyl acetate derivative was obtained by soaking the crystals in 4 m*M* solution for 4 d. An isomorphous difference Fourier map and a Bijvoet difference Fourier map calculated with data between 8 and 2.5 Å resolution suggested a consistent set of five heavy-atom binding sites. The sites are clustered together between molecules *A* and *B*, forming a 'network' with 3.2 to 4.0 Å separation between the uranyls. Two of the binding sites can be directly attributed to interactions with the side chain of Gln451 in molecule *B* (Table 6). The other three sites can be potentially associated with Glu414 and Glu415 of molecule *A*. The side chains of these residues are disordered in the electron-density maps calculated from the native data, but might become ordered upon heavy-atom binding. Uranyl acetate derivatization was also used in the structure determination of fibritin E, where there was one major uranyl binding site due to interactions with Asp449 and Glu475 located in two neighboring molecules. Such favorable location of these two residues does not exist in fibritin M crystals.

The Xe and uranyl positions and occupancies were refined first independently and then simultaneously using the program *MLPHARE* (Table 7). The Xe derivative data were used only to 3 Å resolution. Anomalous scattering data from both derivatives were also included in the phasing.

Table 7. *MIR phasing and phase-combination statistics*

Resolution† (Å)		10.13	5.98	4.65	3.48	2.90	2.46	2.09	Overall
Xe	PhP‡	1.49	2.13	1.79	0.92	0.89	-	-	1.07
	R_{Cullis}§	0.76	0.58	0.69	0.89	0.90	—	—	0.83
UO_2	PhP‡	7.39	2.36	1.53	1.12	1.40	1.44	1.17	1.40
	R_{Cullis}§	0.17	0.57	0.78	0.88	0.76	0.77	0.83	0.77
$\langle m \rangle_{\text{MIR}}$¶		0.757	0.672	0.580	0.444	0.441	0.352	0.302	0.412
$\langle \cos(\alpha_{\text{comb}} - \alpha_{\text{MIR}}) \rangle$		0.968	0.841	0.603	0.428	0.433	0.346	0.097	0.242
$\langle \cos(\alpha_{\text{comb}} - \alpha_{\text{MR}}) \rangle$		0.223	0.656	0.906	0.957	0.925	0.921	0.981	0.949

† High-resolution limit of a bin. ‡ PhP(phasing power) = $\langle |F_H^{\text{calc}}| / |F_{\text{PH}} - |F_{\text{PH}}^{\text{calc}}| | \rangle$. § $R_{\text{Cullis}} = \sum |F_{\text{PH}} - |F_{\text{PH}}^{\text{calc}}| | / \sum |F_{\text{PH}} - F_{\text{P}}|$. ¶ Mean figure of merit.

2.4. *Refinement of the model*

The MIR phases, in the form of Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970), were combined with the MR phases in the program *SIGMAA* using the default σ_A weighting (Read, 1986), and further improved by solvent flattening and histogram matching with the program *DM* (Cowtan, 1994). The $(2F_{\text{obs}} - F_{\text{calc}})$ and $(F_{\text{obs}} - F_{\text{calc}})$ difference maps based on the resulting phases were used for manual correction of the MR model using the program *O* (Jones *et al.*, 1991). The electron density for most parts of all three molecules *A*, *B* and *C*, especially for the C-terminal domains, was reliable and only needed minor corrections and improvements, such as building the side chain of Arg433. However, the quality of the electron density gradually decreased towards the N-terminal end of the coiled-coil domain in all three molecules, especially in molecule *B*. At this stage, the MR phases were recalculated after residues *A*418–*A*423, *B*418–*B*437 and *C*418–*C*428 had been removed from the model. These phases were again combined with the MIR phases and used to calculate $(2F_{\text{obs}} - F_{\text{calc}})$ and $(F_{\text{obs}} - F_{\text{calc}})$ maps. The last few residues in each of the omitted regions were observable in both maps and appeared correct. However, no reliable tracing of the remaining omitted residues was possible; these were, therefore, excluded from the model at the initial stage of refinement and gradually built back during the subsequent refinement rounds, as the quality of the maps improved.

Rigid-body, least-squares positional, simulated-annealing and individual temperature-factor refinement protocols were carried out using the program *X-PLOR* (Brünger, 1992), version 3.8. F_{calc} scaled well to F_{obs} only for data outside 3.8 Å resolution. This was largely mitigated by introducing a bulk solvent correction which involved calculation of a solvent mask and refinement of a bulk solvent electron density level and temperature factor. This correction allowed the use of data between 23 and 1.85 Å resolution and was updated after each significant modification of the model. A 'test' set of 740 reflections (5% of the total data set) was excluded from all stages of refinement and used for evaluating the free *R* factor. Refinement runs with different values of the weight of the crystallographic residual with respect to the empirical chemical energy were executed. Thereupon, the optimal value of this weight was estimated as about half of the ratio of the empirical energy gradient and the crystallographic residual gradient, while higher values resulted in overfitting judged by increase of R_{free}. The use of non-crystallographic symmetry restraints or constraints resulted in higher R_{free} values. During the final stage of refinement, 'omit' $(2F_{\text{obs}}-F_{\text{calc}})$ maps were calculated with model phases after additional simulated-annealing refinement and used to verify the tracing of small stretches of the polypeptide and side chains that caused concern.

When the refinement of the model containing only the protein appeared to have converged, the first solvent 'shell' was added to the model. Water molecules were accepted at all peaks greater than 1.2σ in a $(2F_{\text{obs}} - F_{\text{calc}})$ map that corresponded to peaks greater than 2.5σ in $(F_{\text{obs}} - F_{\text{calc}})$ map and were located at distances of 2.4–3.6 Å to possible donors or acceptors, with reasonable angular positions for hydrogen-bond formation. Upon the subsequent refinement of individual temperature factors and update of the bulk solvent correction, the addition of water molecules yielded a drop of 0.046 in R_{free}. The electron-density maps also improved, which allowed tracing more of the poorly ordered N-terminal protein residues. After positional refinement, two more passes were made to select additional water molecules. The largest temperature factor among the accepted water molecules was 83.6 Å^2.

The final coordinates and experimental structure factors have been deposited with the Brookhaven Protein Data Bank.†

2.5. *Hemihedral twinning*

Twinning by hemihedry (Yeates, 1997) is a potential possibility for crystals in space group *P*3. Two domains of a twin may be related by one of three twinning operations: a 60° rotation about the *c* axis, 180° rotation about the *a* axis or a 180° rotation about the axis

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1AVY).

perpendicular to a and c. In each case, the diffraction patterns from the two domains overlap completely, so that the following expression for the observed intensities is true,

$$I_{\text{twin}}(hkl) = (1 - \alpha)I(hkl) + \alpha I(h'k'l'),$$

where $\alpha(0 < \alpha \leq 0.5)$ is the ratio, referred to as the twinning fraction, of the volume of the less abundant twin domain to that of the entire crystal specimen; $I(hkl)$ and $I(h'k'l')$ are the unknown intensities of two reflections from a single crystal that are related by the twinning operation.

Two other crystal forms of fibritin M (grown with PEG 400 or Na/K tartrate as precipitants, respectively) had trigonal lattices with cell parameters remarkably close to those of the form obtained with Li sulfate (Table 2). Several data sets were collected from crystals grown with PEG 400, and in each case they could be indexed in Laue group $\bar{3}1m$ (implying space group $P312$) with a reasonable R_{merge}. However, in this space group the length of a trimer residing on a crystallographic threefold axis is limited to $c/2$ (45 Å) due to the presence of the additional twofold crystallographic symmetry. This is not compatible with the fibritin M model which has a length of about 85 Å. However, these crystals might be nearly perfectly twinned ($\alpha \simeq 0.5$) with true $P3$ symmetry with the twinning operation being a twofold rotation about an axis perpendicular to a and c. Indeed, after indexing the data set in $P3$, it was possible to find six, rather than three, MR solutions. The solutions could be divided into two sets, and each solution in one set was related to a solution in another by the twinning operation. A single domain of the twinned crystal had an arrangement of three trimers in the unit cell that was

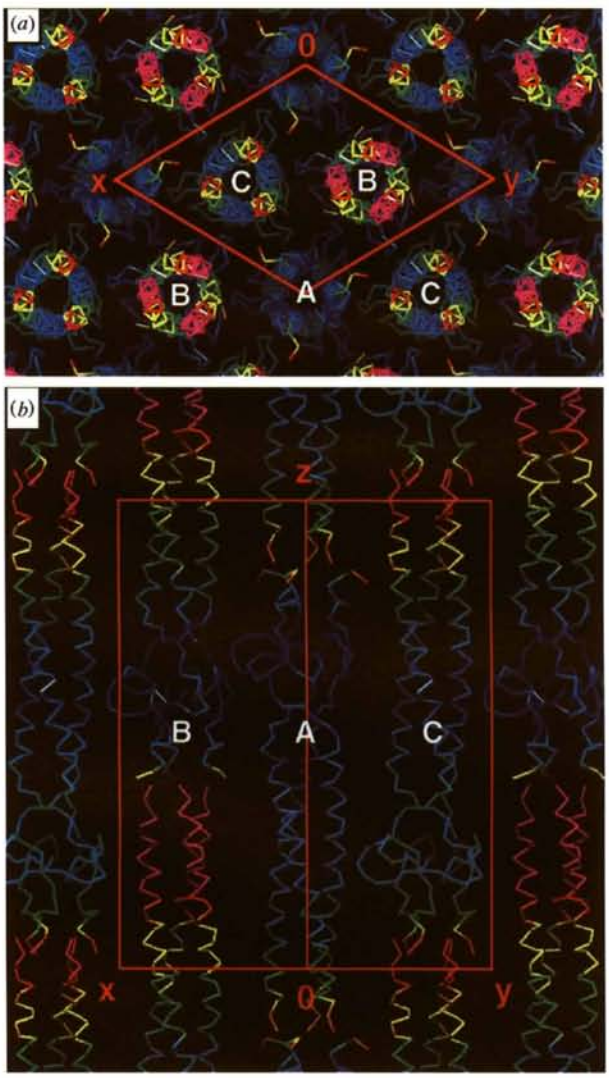

Fig. 4. Crystal packing arrangement in two orthogonal views. Shown are Cα traces colored by a rainbow sequence of colors representing the atomic temperature factors, from 5 Å^2 (dark blue) to 82 Å^2 (red). In addition, the expected conformation of residues 418–431 of molecule B (that were not ordered in the crystal structure) was obtained from the coordinates of trimer A after superimposing it onto the known part of B. These residues are colored magenta. A unit cell is shown with the x, y and z axes radiating from the origin '0'.

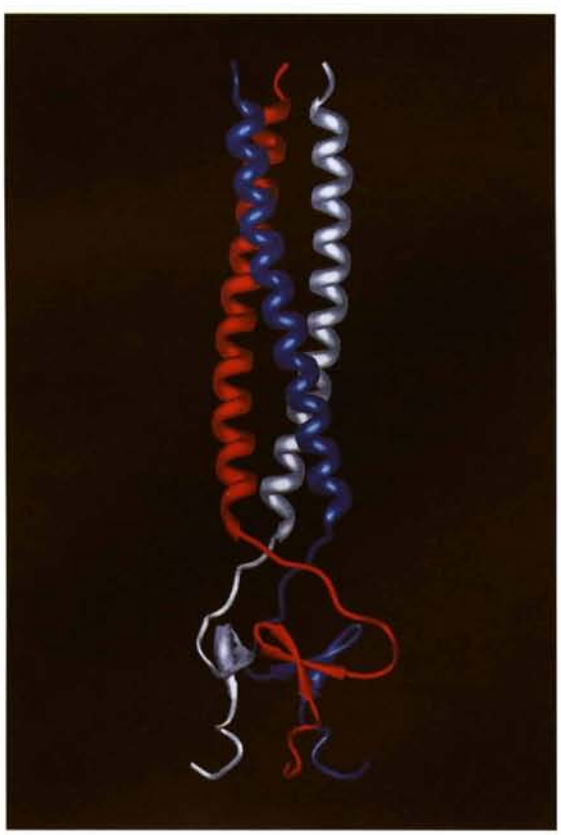

Fig. 5. Ribbon diagram of fibritin M. Drawn with *SETOR* (Evans, 1993).

Table 8. *Properties of the final model*

Range of ordered residues	
Molecule *A*	419–486
Molecule *B*	432–485
Molecule *C*	418–485
Total number of protein non-H atoms	1441
Number of water molecules	
First solvent shell	183
Secondary shells	94
Bulk solvent density (e Å^{-3})	0.432
R_{free}†	0.253 (0.416)
R_{work}†	0.220 (0.317)
R.m.s.d. bonds‡ (Å)	0.006
R.m.s.d. angles‡ (°)	1.3
R.m.s.d. fixed dihedrals‡ (°)	24.9
Average atomic mean temperature factors§ (Å^2)	
Molecule *A*	24.8
Molecule *B*	30.2
Molecule *C*	38.0
First solvent shell	38.9
Secondary shells	48.3

† $R = \sum_h |F_{obs} - F_{calc}| / \sum_h F_{obs}$ for 23–1.85 Å resolution (data in parentheses are for the highest resolution bin, 1.93–1.85 Å). ‡ Compared to ideal values from (Engh & Huber, 1991). § The average temperature-factor estimate from Wilson plot was 20.9 Å^2 (calculated with the program *TRUNCATE*).

very similar to that in the crystal grown with Li sulfate, despite very different crystallization conditions.

An estimate of the twinning fraction α can be obtained solely from the cumulative distribution of observed intensities (Yeates, 1997). However, if a model for a single crystal is available, α can be more accurately determined by minimizing the function,

$$E(\kappa, \alpha) = \sum_{hkl} [I_{twin}(hkl) - \kappa(1-\alpha)F^2_{calc}(hkl) - \kappa\alpha F^2_{calc}(h'k'l')]^2,$$

where $F_{calc}(hkl)$ and $F_{calc}(h'k'l')$ are the calculated structure amplitudes of two single crystal reflections related by the twinning operation and κ is the scale factor between the observed and calculated intensities. Such minimization for the data set collected from the PEG-grown crystal yielded $\alpha = 0.49$. The crystal form obtained with Na/K tartrate also exhibited the same type of hemihedral twinning, with the true space group being $P3$ and the single domain packing arrangement closely resembling the Li sulfate form; however, the twinning was not perfect ($\alpha = 0.41$) (Table 2). The above expression for function $E(\kappa,\alpha)$ can be used to refine the fibritin M structure using the data from the twinned crystals.

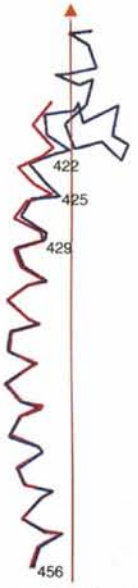

Fig. 6. Superposition of the coiled-coil domains of fibritins M (magenta) and E (cyan). Cα traces of one chain in a trimer are shown. Vertical line shows the trimer axis. Deviation between the Cα positions of residue 422 is 1.4 Å.

3. Results and discussion

3.1. *Overall structure*

The structure was determined by MR with additional help from MIR. Crystals of fibritin M contain three symmetry-independent trimers, *A*, *B* and *C*, each located on a different crystallographic threefold axis (Fig. 4). Phase information from two isomorphous derivatives was helpful for verification of a poor MR solution of molecule *C* and also during the tracing of the poorly ordered N-terminal parts of all three molecules.

The properties of the final model are presented in Table 8. 91.4% of all non-glycine and non-proline protein residues map onto the most favored regions of the Ramachandran plot [as defined by the program *PROCHECK* (Laskowski *et al.*, 1993)], and no residues map onto disallowed or generously allowed regions. Residues 421–456 form an α-helical coiled-coil domain. The C-terminal part of each polypeptide chain contains two small β-strands (residues 468–472 and 475–479, respectively) forming a 'hairpin' (Fig. 2). Three symmetry-related β-hairpins are located roughly in the plane perpendicular to the trimer axis and form a propeller-like structure. Together with the loops that connect it to the α-helical part, this structure constitutes the C-terminal domain which has a highly hydrophobic interior (Fig. 5). Asp473 is in an unusual conformation with $\varphi = 45°$, $\psi = 50°$ and is located in a turn connecting the two β-strands.

3.2. *Comparison with fibritin E and effect of the point mutations*

The three fibritin M trimers *A*, *B* and *C* were each superimposed on the structure of fibritin E (Tao *et al.*, 1997), as well as on each other. The r.m.s. deviations of the main-chain atomic positions between pairs of the fibritin M molecules are 0.6 Å. The r.m.s. deviations of

the main-chain atomic positions of the *A*, *B* and *C* molecules of fibritin M and the fibritin E molecule are 0.75, 0.67 and 1.07 Å, respectively, somewhat larger than, but comparable to, those between the individual fibritin M trimers. The Cα positions of the residues 429–456 in both fibritins are almost identical, *i.e.* the parameters (Crick, 1953) of the predominant part of the coiled coil in fibritin M are the same as in the last segment of fibritin E. The only significant deviation is in the N-terminal part of the coiled coil (residues 418–428): in fibritin M there is a gradual increase of distances between the three helices towards the N terminus and a slight unwinding of the supercoil (Fig. 6). As a result, the first residue involved in the hydrophobic core of fibritin M is Ile422, while it is Leu418 in fibritin E and most probably in the wild type. The distortion of the N-terminal end of the coiled coil in fibritin M may, therefore, be the result of the point mutations and the 'head-to-tail' interactions of the molecules related to the **c** lattice translation, as well as by the absence of the preceding coiled-coil segments.

The mutation Ser421→Lys creates a salt bridge between residues Lys421 and Glu426 in agreement with predictions. These residues occupy the heptad positions *g* and *e* in different chains within one fibritin M trimer, respectively. Such an interchain salt bridge may have a stabilizing effect on the coiled coil (Lumb & Kim, 1995*b*). Mutations of residue 428 (heptad position *g*) to aspartic acid and residue 433 (heptad position *e*) to arginine were expected to create an extra interchain salt bridge between these residues. However, Arg433 is equally capable of forming an interchain salt bridge with Glu435. Neither of the two potential salt bridges can be identified with confidence from the crystal structure, as the side chain of Arg433 resides at roughly equal distances of about 4.5 Å to both Asp428 and Glu435. Another mutation, Asn425→Ile, disables an unusual interaction between the asparagines in a *d* position that is mediated in fibritin E crystals by a chloride ion located on the threefold axis. This interaction, also found in other coiled-coil proteins, is considered to be important for correct alignment of polypeptide chains upon formation of a coiled coil (Lumb & Kim, 1995*a*). In fibritin, such an alignment is governed by the formation of the C-terminal domain (Tao *et al.*, 1997). Furthermore, Ile425 is well accommodated at its *d* position in the trimeric coiled coil as anticipated (Harbury *et al.*, 1994). Thus, this mutation is unlikely to have a serious effect on the folding of fibritin M.

3.3. *Crystal packing*

As of August 1997, the Brookhaven Protein Data Bank contained only two entries with space group *P*3, among about 5200 X-ray crystal structures of proteins. This space group, like other space groups without screw rotational elements, is infrequent among crystals of globular proteins because it would normally have a non-compact crystal packing arrangement (Padmaja *et al.*, 1990). For elongated trimers such as fibritin molecules, the packing in this space group can be much better, especially if the trimers are positioned on crystal-

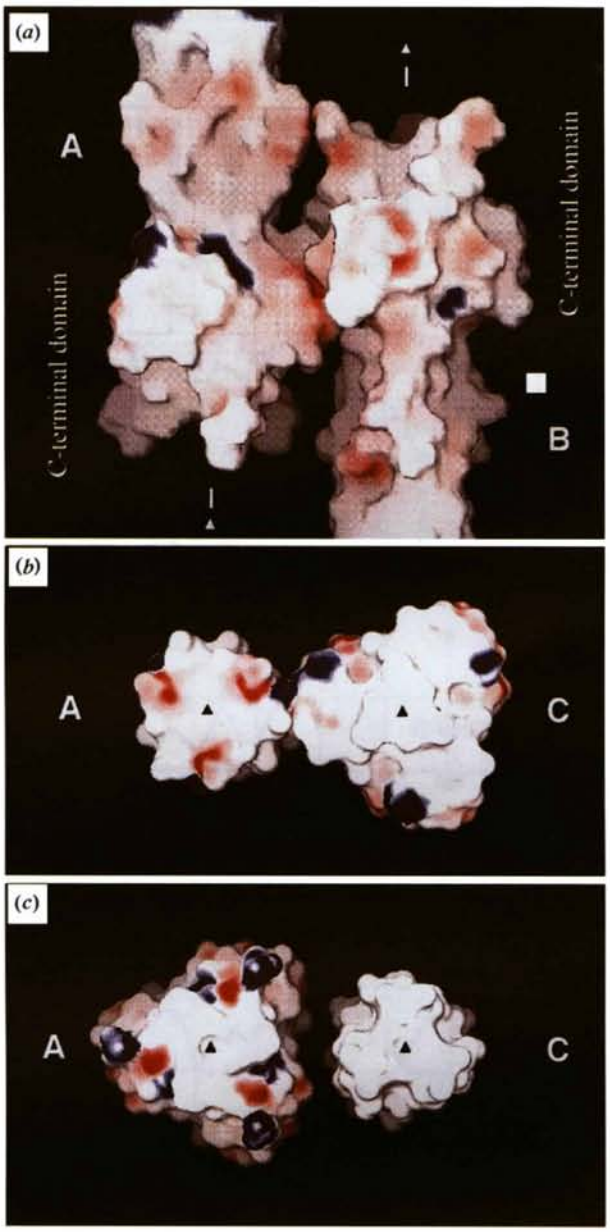

Fig. 7. Crystal contacts. Surfaces of trimeric molecules were calculated with program *GRASP* (Nicholls *et al.*, 1993) and colored by approximate electrostatic potential revealing the location of acidic (red) and basic (blue) side chains. (*a*) Contact between the C-terminal domains of molecules *A* and *B*. (*b*) Contact between the coiled-coil domain of *A* and the C-terminal domain of *C*. There is a similar contact between the coiled-coil domain of *C* and the C-terminal domain of *B*. (*c*) The C-terminal domain of *A* and the coiled-coil domain of *C* make no contacts. The same is true for the C-terminal domain of *C* and the coiled-coil domain of *B*.

lographic threefold axes. Since the C-terminal domain of fibritin M has a larger cross section than the coiled-coil domain, the C-terminal domains of the three trimers *A*, *B* and *C* should have different *z* coordinates for compact packing, which is indeed the case (Fig. 4*b*). Besides the head-to-tail interaction of the molecules residing on the same threefold axes, the most critical interaction occurs between the C-terminal domains of molecules *A* and *B* which are antiparallel. These contacts appear to be sufficient for the formation of the crystal lattice even if the *C* molecules were absent, consistent with our preliminary crystallographic analysis (Fig. 2 in Strelkov *et al.*, 1996). However, the *C* molecule reduces the allowed flexibility of molecules *A* and *B* within the lattice. The lattice formed by molecules *A* and *B* would potentially allow insertion of molecule *C* in either an 'up' or 'down' orientation with many feasible rotations and translations along the *c* axis. The possibility was explored that there might be alternative positions of molecule *C*, in addition to the major position found by MR, but no convincing evidence could be found. Furthermore, refinement of the occupancy of molecule *C* in the position found by MR in the presence of fully occupied molecules *A* and *B* resulted in a value of 0.95 and no change in R_{free} and R_{work}; therefore, molecule *C* can be considered fully occupied within experimental accuracy. With three molecules in the asymmetric unit, the V_m (Matthews coefficient) is 2.15 Å^3 Da^{-1}, which gives fairly tight packing and may be the reason for diffraction extending to 1.85 Å resolution.

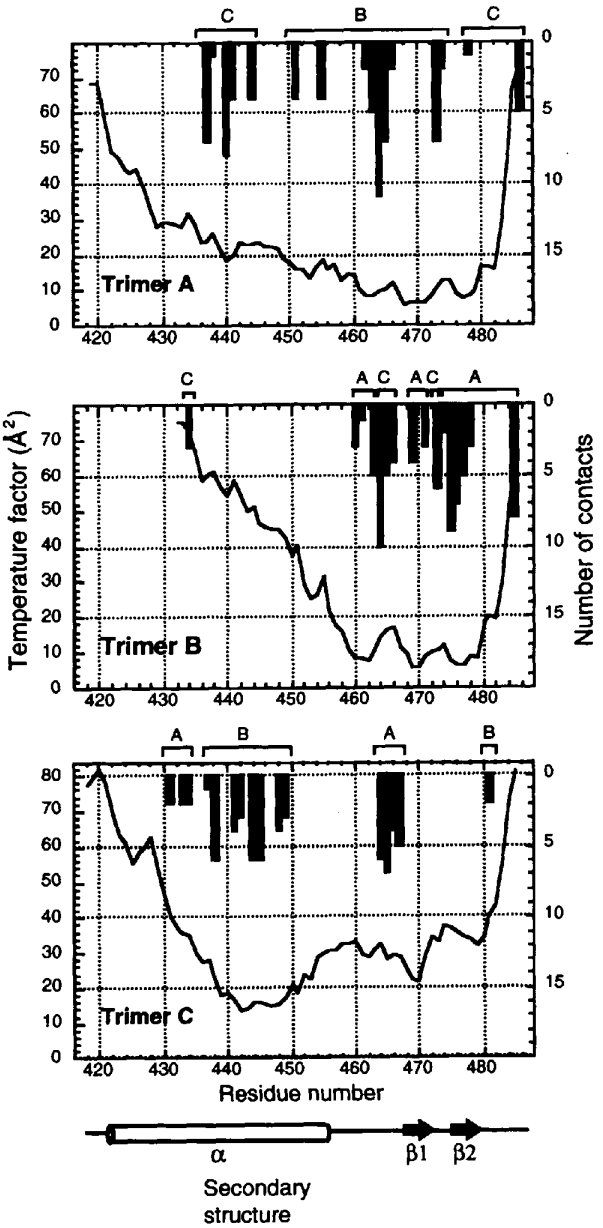

Fig. 8. Plots of average main-chain temperature factors in chains *A*, *B* and *C* *versus* residue number. In addition, the parts of each molecule making crystal lattice contacts with other molecules are marked above the corresponding plot. The number of atoms in each residue that are involved in the contacts are shown as vertical bars. An atom was assumed to make a lattice contact if this atom was closer than 5.0 Å to any of the atoms in neighboring trimers.

The three molecules *A*, *B* and *C* have different parts involved in their contacts with neighbors in the crystal lattice (Figs. 7 and 8). Coiled-coil domains of adjacent trimers do not make contacts with each other anywhere. The head-to-tail interactions of molecules on the same threefold axis are not completely clear because the five N-terminal residues are not ordered in any of the three trimers.

3.4. *Partial disorder*

Low and poor electron density in the ($2F_{obs}-F_{calc}$) maps (Fig. 9), as well as large temperature factors (Fig. 8), show that the coiled-coil domains of molecules *A*, *B* and *C* are disordered to different extents. The disorder increases gradually towards the N terminus of each molecule, particularly in molecule *B*, and significant numbers of residues at the N termini could not be traced (Table 8). In addition, even though the C-terminal domain of molecule *C* could be traced reliably in its entirety, the temperature factors for this region are considerably higher than those for *A* and *B*. The lack of order in the C-terminal domain of molecule C, as well as its larger deviation from the model used for MR compared to molecules *A* and *B*, probably explains why the MR solution for molecule *C* was only slightly above the noise level, while the solutions for *A* and *B* were unambiguous (Table 4).

The individual patterns of disorder (or increased mobility) along the three trimers can be explained by the lattice contacts. The N-terminal part of each molecule is progressively more disordered the greater the distance from that molecule's amino terminus to the closest point of contact of its coiled coil with a neigh-

Table 9. *Properties of solvent in various protein structures*

Protein	Type	Resolution (Å)	V_m (Å^3 Da^{-1})	N_{res}†	α_{solv}‡	N_{solv}§	N_{ord}¶	N_{ord}/N_{res}	N_{ord}/N_{solv}
Fibritin M	Coiled-coil	1.85	2.15	190	0.45	770	277	1.46	0.36
Collagen††	Collagen triple helix	1.85	1.95	90	0.38	195	162††	1.80	0.83
NDP kinase‡‡	Globular	1.9	2.50	287	0.49	1362	219	0.76	0.16
Thrombin§§	Globular	2.6	2.33	307	0.48	1639	95	0.31	0.06

† Number of protein residues in the model, per asymmetric unit. ‡ Fraction of volume occupied by the solvent. $\alpha_{solv} = V_{solv}/V_{au}$; $V_{solv} = V_{au} - V_{prot}$; $V_{prot} = m_{prot}\bar{v}_{prot}$, where m_{prot} is the total mass of protein in the asymmetric unit. A typical value $\bar{v}_{prot} = 0.73$ cm^3 g^{-1} (Harpaz *et al.*, 1994) was used. § Estimate of the total number of water molecules in asymmetric unit. $N_{solv} = V_{solv}/v_{wat}$, where $v_{wat} = M_{wat}/\rho_{wat}N_A$ is the volume of one water molecule (~29.9 Å^3), M_{wat} is molecular weight, $\rho_{wat} = 1$ g cm^{-3} is the density of liquid water, and N_A is Avogadro constant. ¶ Number of ordered water molecules in the model, per asymmetric unit. †† Trimeric collagen peptide (Bella *et al.*, 1995). One asymmetric unit contains 141 water molecules and seven acetic acid molecules. One acetic acid molecule was assumed here to be equivalent to three water molecules. ‡‡ Nucleoside diphosphate kinase/cAMP complex (Strelkov *et al.*, 1995). §§ Human α-thrombin/inhibitor complex (Tabanero *et al.*, 1995).

boring molecule (Fig. 9). In addition, there is a distinct correlation in the atomic temperature factors of adjacent trimers in the lateral direction (Fig. 4*a*).

3.5. *Ordered solvent and implications towards an optimal refinement protocol*

Because of the elongated shape and the small molecular weight of fibritin M, its polypeptide chain, especially the coiled-coil domain, is exposed to solvent to a larger extent than in a typical globular protein. All of the 277 ordered solvent molecules in one crystallographic asymmetric unit were interpreted as water molecules, while no sulfate ions were found. There are 183 water molecules that have direct interactions (the first solvation 'shell') with polar groups on the protein, while the remaining molecules interact only with other water molecules. There are 1.46 ordered solvent positions per one protein residue in fibritin M, which is close to the value in the trimeric collagen peptide structure (Bella *et al.*, 1995), but significantly larger than in typical structures of globular proteins (Table 9). The ratio of the number of observable water sites to the total number of water molecules presumably present in the asymmetric unit is about $\frac{1}{3}$ for fibritin M. In globular proteins, this ratio is typically much lower and decreases further if high-resolution data are not available (Table 9). In the course of refinement of the fibritin M structure, R_{free} and

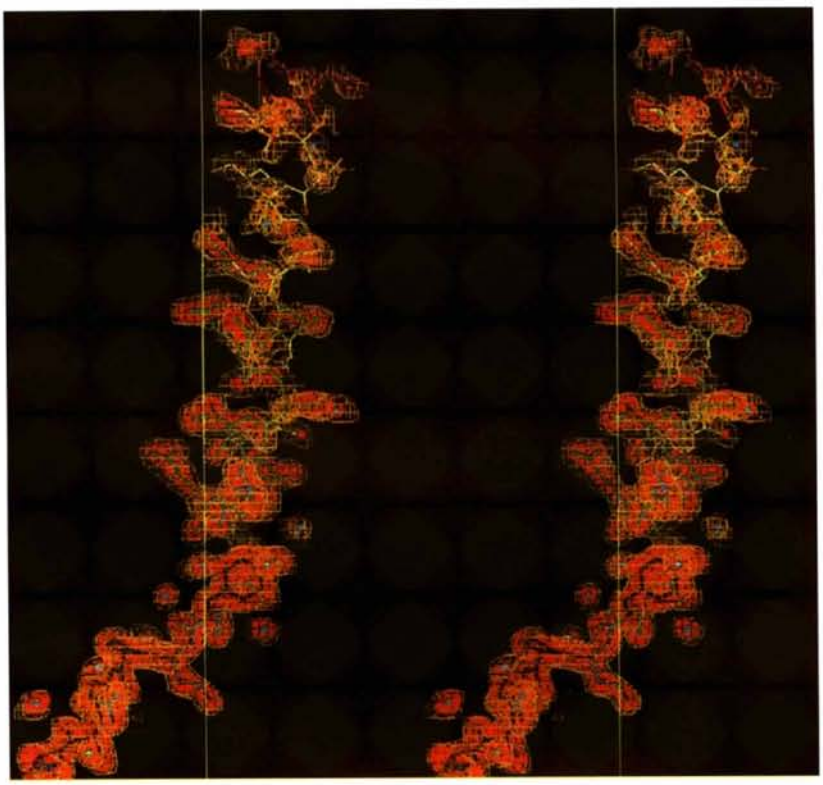

Fig. 9. Stereo diagram of the structure of residues 432 (top) and 463 (bottom) of molecule *B* superimposed on a ($2F_{obs} - F_{calc}$) electron-density map. Only one chain of the trimer is shown, with 'rainbow' coloring for the atomic temperature factors, from 5 Å^2 (dark blue) to 78 Å^2 (red). Water molecules directly associated with the chain are represented by light blue spheres. The map is contoured at two levels, 0.65σ (conventional 'net', yellow) and 1.5σ (semitransparent surface, orange), to show the gradual loss of the map contrast and quality towards the N terminus.

Table 10. *R factors of the model at various stages of refinement (23–1.85 Å resolution)*

	Model after rigid-body refinement†	Before building in solvent molecules‡	Final model, including solvent	Final model, protein only§
R_{free}	0.375	0.332	0.253	0.321
R_{work}	0.370	0.276	0.220	0.286

† After MR, bulk solvent correction, rigid-body and atomic temperature-factor refinement. ‡ After manual rebuilding and automated refinement of the model containing protein atoms only. § After removal of solvent molecules, re-scaling F_{calc} to F_{obs} and re-calculating the bulk solvent correction.

R_{work} could be lowered to acceptable values only after individual water molecules were included (Table 10). Moreover, after the structure determination had been completed, it was shown that many water molecules could have been positioned immediately after the molecular replacement. Indeed, despite R_{free} being 0.375 before any positional refinement of the protein model, 132 (more than $\frac{2}{3}$) of all molecules in the first solvent shell were present as significant peaks in $(2F_{obs} - F_{calc})$ and $(F_{obs} - F_{calc})$ maps. The above experience shows that incorporation of solvent molecules at an intermediate step of refinement may significantly improve the agreement between the model and the data and allow improved refinement of the protein model, although such a procedure should be carefully monitored by R_{free}.

We thank Dino Moras, University of Strasbourg, France, for providing us with the device for Xe derivatization, as well as Jordi Bella and many other colleagues at Purdue University for helpful advice. The work was supported by a Howard Hughes Medical Institute grant to VVM and MGR, a National Science Foundation grant to MGR, and in part by grants from the Protein Engineering Council of Russia and the Russian Foundation for Basic Research (grant No. 96-04-48035) to VVM. We are also grateful to Purdue University for a reinvestment grant and to the Lucille P. Markey Foundation, both to support the development of structural biology at Purdue.

References

Bella, J., Brodsky, B. & Berman, H. M. (1995). *Structure*, **3**, 893–906.

Brünger, A. T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*. New Haven and London: Yale University Press.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. (1994). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.

Crick, F. H. C. (1953). *Acta Cryst.* **6**, 689–697.

Efimov, V. P., Nepluev, I. V., Sobolev, B. N., Zurabishvili, T. G., Schulthess, T., Lustig, A., Engel, J., Haener, M., Aebi, U., Venyaminov, S. Y., Potekhin, S. A. & Mesyanzhinov, V. V. (1994). *J. Mol. Biol.* **242**, 470–486.

Eiserling, F. A. & Black, L. W. (1994). *Pathways in T4 Morphogenesis. Molecular Biology of Bacteriophage T4*, edited by J. D. Karam, pp. 209–212. Washington DC: American Society for Microbiology.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Evans, S. V. (1993). *J. Mol. Graphics*, **11**, 134–138.

Harbury, P. B., Kim, P. S. & Alber, T. (1994). *Nature (London)*, **371**, 80–83.

Harpaz, Y., Gerstein, M. & Chothia, C. (1994). *Structure*, **2**, 641–649.

Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lumb, K. J. & Kim, P. S. (1995*a*). *Biochemistry*, **34**, 8642–8648.

Lumb, K. J. & Kim, P. S. (1995*b*). *Science*, **268**, 436–439.

Lupas, A. (1996). *Trends Biochem. Sci.* **21**, 375–382.

Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T. & Engel, J. (1996). *Science*, **274**, 761–765.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, 166–170.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst.* A**46**, 725–730.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rossmann, M. G. (1960). *Acta Cryst.* **13**, 221–226.

Sobolev, B. N., Shneider, M. M., Marusich, E. I. & Mesyanzhinov, V. V. (1995). *Folding and Oligomerization of Fibrous Proteins. Evolutionary Biochemistry and Related Areas of Physicochemical Biology*, edited by B. Poglazov, pp. 351–374. Moscow: Bach Institute of Biochemistry and ANKO.

Soltis, S. M., Stowell, M. H. B., Wiener, M. C., Phillips, G. N. Jr & Rees, D. C. (1997). *J. Appl. Cryst.* **30**, 190–194.

Strelkov, S. V., Perisic, O., Webb, P. A. & Williams, R. L. (1995). *J. Mol. Biol.* **249**, 665–674.

Strelkov, S. V., Tao, Y., Rossmann, M. G., Kurochkina, L. P., Shneider, M. M. & Mesyanzhinov, V. V. (1996). *Virology*, **219**, 190–194.

Tabanero, L., Chang, C. Y. Y., Ohringer, S. L., Lau, W. F., Iwanowicz, E. J., Han, W. C., Wang, T. C., Seiler, S. M., Roberts, D. G. M. & Sack, J. S. (1995). *J. Mol. Biol.* **246**, 14–20.

Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). *Structure*, **5**. 789–798.

Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**. 594–611.

Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.