Reprinted from *J. Mol. Biol.* (1992) **226**, 1073–1084

New Triple-helical Model for the Shaft of the Adenovirus Fibre

Pieter F. W. Stouten, Chris Sander, Rob W. H. Ruigrok and Stephen Cusack

J. Mol. Biol. (1992) **226**, 1073–1084

New Triple-helical Model for the Shaft of the Adenovirus Fibre

Pieter F. W. Stouten and Chris Sander

European Molecular Biology Laboratory
Meyerhofstraße 1, D–6900 Heidelberg, Germany

Rob W. H. Ruigrok and Stephen Cusack

European Molecular Biology Laboratory, Grenoble Outstation
c/o I.L.L. 156X, F–38042 Grenoble, France

(Received 2 January 1992; accepted 17 April 1992)

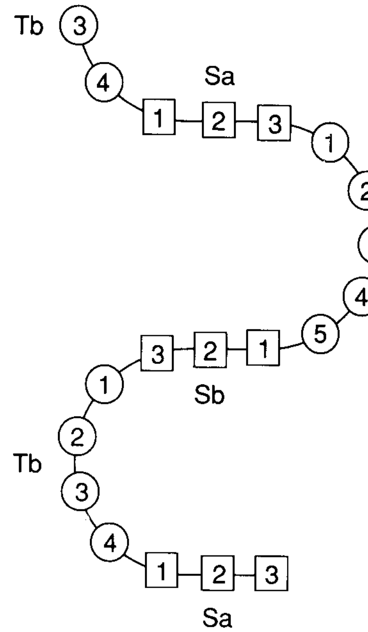
The adenovirus fibre is a trimeric protein with a globular head on a long thin shaft that projects from the twelve fivefold vertices of the virion. The shaft region of the fibre primary sequence has a unique pseudo-repeating motif of 15 residues. Using constraints derived from sequence analysis, the trimeric nature of the fibre, the experimental determination of the shaft length and general knowledge about protein structure, an atomic model of the fibre shaft has been constructed by computer modelling techniques. In the final model the three monomers form a left-handed triple-helical structure with threefold symmetry and with successive 15-residue repeats on the same chain related by an axial rise of 13·1 Å and a left-handed azimuthal rotation of close to 300°. Three threefold related β-sheets with short strands are formed by *inter*-monomer main-chain hydrogen bonds giving rise to superhelical ribbons covering the surface of the shaft. The model satisfies criteria of extensive hydrogen bonding, reasonable backbone torsion angles, burial of most hydrophobic residues and good packing of the hydrophobic core. Furthermore, the model is consistent with the observed shaft length of about 290 Å and its calculated X-ray fibre diffraction pattern shows the characteristic features found in the diffraction pattern of crystals of fibre, notably layer lines with a spacing of about 1/26 $Å^{-1}$ and strong meridional intensity at 1/4·4 $Å^{-1}$.

Keywords: adenovirus fibre; triple helix; computer model

1. Introduction

The adenoviruses constitute a family of icosahedral viruses containing a number of structural proteins whose position and role in the virion are beginning to be understood (Stewart *et al.*, 1991). The capsid consists of 252 capsomers including 240 hexons and 12 pentons which are situated at the twelve fivefold vertices. The pentons consist of a pentameric penton base embedded in the capsid and a protruding trimeric fibre, which is responsible for the attachment of the virus to the cell surface. Analysis of the primary sequences of six human adenovirus fibres, Ad2 (Hérissé *et al.*, 1981), Ad3 (Signäs *et al.*, 1985), Ad5 (Chroboczek & Jacrot, 1987), Ad7 (Hong *et al.*, 1988), Ad40 (Kidd & Erasmus, 1989) and Ad41 fibre (Pieniazek *et al.*, 1989) and of other experimental data suggests an almost identical overall structure for the fibre consisting of an N-terminal tail of about 40 residues, embedded non-covalently in the penton base (Devaux *et al.*, 1987; Weber *et al.*, 1989), an extended fibrous shaft of variable length, and a C-terminal knob of about 180 residues. In all six fibre sequences the shaft is made up of pseudo-repeating motifs of about 15 residues each (Green *et al.*, 1983). Ad3 and Ad7 fibre have 6, Ad40 has 21, and Ad2, Ad5 and Ad41 have 22 such motifs. The first 50 amino acid residues of Ad2 and Ad5 are fully conserved, which, based on the expectation of strong evolutionary constraints at the point of attachment, supports their assignment to the fibre tail (Chroboczek & Jacrot, 1987). The optimum alignment of the Ad2 and Ad5 fibre sequences (582 and 581 residues, respectively) requires only two gaps (of 1 and 2 amino acid residues) and one insertion (of 2 residues) in the Ad5 sequence. The

0022–2836/92/161073–12 $08.00/0



(a)

	Sa	Ta	Sb	Tb
47	**L** S **L**	R L S E P	**L** V **T**	S N G M
62	**L** A **L**	K M G N G	**L** S **L**	D E A G N
78	**L** T S	Q N V		
	T	T V S P P	**L** K K	T K S N
97	**I** N **L**	E I S A P	**L** T **V**	T S E A
112	**L** T **V**	A A A A P	**L** M **V**	A G N T
127	**L** T **M**	Q S Q A P	**L** T **V**	H D S K
142	**L** S **I**	A T Q G P	**L** T **V**	S E G K
157	**L** A **L**	Q T S G P	**L** T **T**	T D S S T
173	**L** T **I**	T A S P P	**L** T **T**	A T G S
188	**L** G **I**	D L K E P	**I** Y **T**	Q N G K
203	**L** G **L**	K Y G A P	**L** H **V**	T D D L N T
220	**L** T **V**	A T G P G	**V** T **I**	N N T S
235	**L** Q **T**	K V T G A	**L** G **F**	D S Q G N
251	**M** Q **L**	N V A G G	**L** R **I**	D S Q N R R
268	**L** I **L**	D V S Y P	**F** D **A**	Q N Q L N
284	**L** R **L**	G Q G P	**L** F **I**	N S A H N
299	**L** D **I**	N Y N K G	**L** Y **L**	F T A S N N S K K L E
321	**V** N **L**	S T A K G	**L** M **F**	D A T A
336	**I** A **I**	N A G D G	**L** E **F**	G S P N A P N T N P
357	**L** K **T**	K I G H G	**L** E **F**	D S N K A
373	**M** V **P**	K L G T G	**L** S **F**	D S T G A
389	**I** T **V**	G N		

(b)

Figure 1. (a) Secondary structure scheme of the shaft. Sa and Sb indicate short β-strands. Residues 1 and 3 in the strands are hydrophobic. Ta and Tb indicate the turns. Ta is the 5-residue turn, normally with a proline or glycine in position 5; Tb is the 4-residue turn. (b) The sequence of the shaft of the Adenovirus 5 fibre. The modelled part is enclosed by heavy lines. The assignment of strands and turns is given at the top of the figure. The numbers on the left are the residue numbers of the 1st residue in each segment in the sequence of the fibre. Boldface letters indicate hydrophobic residues in the 1st and last positions of the short strands. Hydrophobicity is according to the normalized consensus scale of Eisenberg *et al.* (1984) and residues with a hydrophobicity value of $-0{\cdot}05$ or higher are considered hydrophobic, i.e. ILVMFPAT.

overall identity is 65% with only five out of 103 hydrophobic residues in the Ad2 shaft replaced by hydrophilic ones in Ad5 (Chroboczek & Jacrot, 1987). The assumption that the global structure and properties of both fibres are essentially the same is, therefore, reasonable.

Green *et al.* (1983) proposed a model (hereafter referred to as Green's model) of the Ad2 shaft (residues 45 to 401) in which each unit of 15 residues comprises two three-residue β-strands (denoted Sa and Sb in Fig. 1(a)) linked in anti-parallel fashion by two bends, one usually a four-residue β-bend (denoted by Ta) and a five-residue bend (denoted by Tb) which, with a single exception, always has a proline or glycine at position 5. This basic structural unit is shown in Fig. 1(a) and Fig. 1(b) shows the amino acid sequence of the shaft of Ad5 fibre arranged according to the pseudo-repeating motif. Green *et al.* (1983) suggested that the shaft region of each monomer could form a cross-β structure consisting of a long, short-stranded, amphipathic β-sheet which would most likely dimerize face-to-face to form a probably twisted sandwich-like structure running from the base of the fibre to its top. Although at the time there was some indication that the fibre was indeed a dimer, recent results unambiguously reveal that the fibre is a trimer (Van Oostrum & Burnett, 1985; Ruigrok *et al.*, 1990). Since the tertiary structure originally proposed by Green *et al.* (1983) for a dimer (face-to-face packing of the 2 monomers) cannot apply to a trimer, alternative structures must be sought. A possible trimeric structure, analogous to Green's original model, could be formed by edge-to-edge or edge-to-face packing of cross β-sheets, resulting in a fibre with triangular cross-section. A trimeric β-sheet structure with edge-to-face packing has been determined for tumour necrosis factor (Jones *et al.*, 1989). However this model, like Green's model, would result in a shaft length for the case of Ad2 and Ad5 of about 200 Å ($22 \times 2 \times d$, where the separation d of adjacent β-strands in a β-sheet is about 4·7 Å: 1 Å = 0·1 nm). This is in conflict with recent measurements using electron microscopy of the length of Ad2 fibre, as single particles and in crystals, which show that the shaft is about 290 Å long (Ruigrok *et al.*, 1990; Devaux *et al.*, 1990). It was our aim, with these observations in mind, to construct a new three-dimensional model of the adenovirus fibre shaft, which must be trimeric, have the observed length and predict the observed X-ray diffraction pattern (Devaux *et al.*, 1990).

2. Boundary Conditions and Constraints

In designing the atomic model of the fibre constraints arising from experimental data, information conveyed by the sequence and general knowledge of proteins must be taken into account.

The experimental data comprise the following.

(1) The length of the Ad2 fibre shaft is about 290 Å (Ruigrok *et al.*, 1990; Devaux *et al.*, 1990). The average axial repeat distance per 15-residue amino acid repeat is about 13·2 A (Ruigrok *et al.*, 1990).

(2) Low-resolution image reconstruction of adenovirus capsids (Van Oostrum *et al.*, 1987) suggest that the fibre has threefold symmetry. Although it is conceivable that the fibre is not threefold symmetric, despite being a trimer of identical polypeptide chains, we make the reasonable assumption that it is so.

(3) The X-ray diffraction pattern of Ad2 fibre crystals shows a very strong meridional reflection on the axis parallel to the fibre axis, corresponding to a Bragg spacing of 4·4 Å and weak layer lines corresponding to spacings of about 26·4 Å (Devaux *et al.*, 1990). Even orders (i.e. at spacings of 13·2 Å) tend to be meridional and odd orders off-meridional.

Analysis of the Ad2 fibre sequence reveals a repeating motif of about 15 residues in the shaft with sub-repeats of seven and eight residues. The six commonly used secondary structure prediction methods described by Lesk (1988) indicate that the shaft consists almost solely of β-structure, although this was only partially confirmed by circular dichroism experiments (Green *et al.*, 1983). The strands and loops can be assigned unambiguously in the region with the most regular 15-residue repeats (Thr85–Ala312). This assignment, which is the same as Green's, provides another important piece of information.

(4) The 15-residue repeats are each made up of two short (3-residue) strands connected by loops of 4 (Tb) and 5 (Ta) residues (Fig. 1(a) and (b)). The first and last residues of each strand are invariably apolar. Although the actual sequence indicates that deviations from the 15-residue motif are permissible, particularly in the length of loop Tb (Fig. 1(b)), we will concern ourselves only with the part of the structure containing exact 15-residue repeats.

General knowledge of protein structures that applies to this sequence include the following.

(5) β-sheets are formed by β-strands that are properly interconnected by main-chain hydrogen bonds.

(6) β-sheets generally have a right-hand twist when viewed along the polypeptide chain direction (Chothia, 1973).

(7) Hydrophobic residues tend to be removed from solvent. This applies specifically to the first and last β-strand residues (refer to constraint (4)). This strongly suggests that the core of the fibre is basically hydrophobic and therefore it is unlikely that there is a water-filled central channel in the fibre.

Any model that is based on a trimer whose monomers form individual anti-parallel β-sheets that independently extend along the entire shaft will result in a fibre that is too short. The simplest shaft model that can overcome this length problem is an intertwined trimer (a rope-like structure), where the monomers are related by strict threefold symmetry and where each β-sheet contains successive β-strands coming from all three chains in turn. One can picture this abstractly as three chains that run parallel to one another along the surface of a hypothetical cylinder at some angle relative to the fibre axis. As a consequence, the β-sheet surfaces also run parallel to each other along this surface, roughly perpendicular to the individual chains. The main feature of these sheets is that main-chain hydrogen bonding occurs between strands in different monomers rather than within a monomer as originally envisaged in Green's cross-β model. *A priori* it is not evident whether the monomeric chains will wrap around the cylinder in a left- or right-handed fashion.

Consider three completely regular, extended chains of 330 amino acid residues each (22 repeats of 15 residues), wrapped around a hypothetical cylinder in such a way that every repeat of 15 is related to the preceding one in the same chain by a screw operation about the cylinder axis (z-axis). The chains are also related by threefold symmetry about the z-axis. The relevant parameters for such an idealized model are shown in Figure 2, where P is the helical pitch, α is the angle between the chains and the xy-plane, T is the axial rise per 15 residues, N ($= 1/3\ P \cos \alpha$) is the perpendicular interstrand distance and R ($= N \cos \alpha$) is this distance projected onto the fibre axis. Rough estimates of these parameters can be derived as follows. Since a polypeptide chain of 15 residues has a maximum length of about 50 Å (in extended conformation), and the axial rise T is 13·2 Å (point (1) above), the average tilt angle α is not less than 15·4° If we take N as 4·8 Å, approximately the normal interstrand distance for parallel β-strands, then $P \approx 14{\cdot}9$ Å and $R \approx 4{\cdot}6$ Å. We note that this value of R is quite close to the Bragg spacing of the observed strong meridional reflection (1/4·4 Å^{-1} : Devaux *et al.*, 1990) which could arise from the projection of the tilted β-sheet onto the fibre axis. In reality, however, the model would not be so regular since there is no reason to suppose that the loops and strands make the same angle α with the xy-plane. Thus, if the 1/4·4 Å^{-1} meridional reflection does arise from the perpendicular distance between adjacent strands, the strands may make an angle α of as much as 24° which would lead to the correct interstrand distance of 4·8 Å. This larger inclination of the strands could be compensated for by the loops making a smaller angle.

Constraint (5) is invaluable in the model building procedure, because the requirement of properly hydrogen-bonded β-sheets severely limits the number of possible structures. The simplest *inter*-chain β-sheet structure would involve the hydrogen bonding of a three-residue strand from one chain with a three-residue strand from an adjacent chain. Residue i in one chain would then be aligned with residue $i+7$ or $i+8$ (depending on the length of the intervening loop) in the next. Such a hydrogen-bonding pattern can be described by the sequence

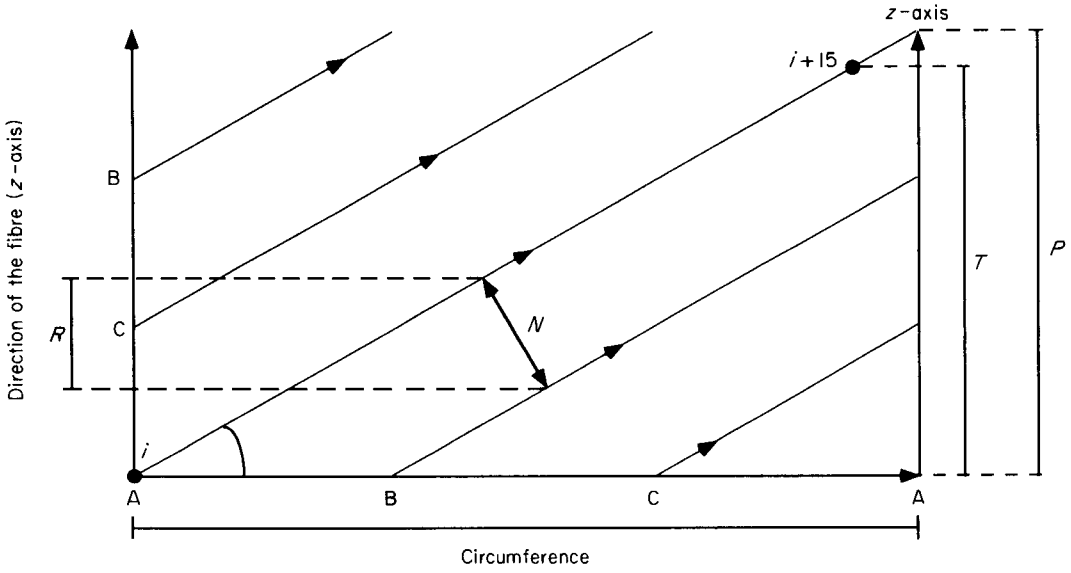


Figure 2. Schematic drawing of the unfolded ideal cylinder composed of 3 chains, A, B and C. In reality the left and right vertical axes (z-axes) coincide. Characteristic parameters are: α, the out-of-plane tilt of a chain relative to the plane perpendicular to the fibre axis; N, the average distance between 2 different chains; T, the rise along the z-axis corresponding to 1 repeat of 15 residues; and P, the pitch (the rise after the chain has gone around the cylinder one complete 360° turn). i and $i+15$ indicate 2 residues in chain A that are separated by exactly 1 repeat.

Ca–Bb–Aa–Cb–Ba–Ab, where in each case the capitalized letter is a chain identifier and the small letter indicates the strand type (i.e. whether it precedes a short loop (Tb) or long loop (Ta); see Fig. 1). Figure 3 shows four strands of a β-sheet with this non-staggered basic hydrogen-bonding pattern. Residues C112 to C114 (strand type Sa in chain C) are connected via strands B120 to B122 (strand type Sb in chain B) and A127 to A129 (strand type Sa in chain A) to residues C135 to C137 (strand type Sb) in the same chain. The intervening two strands in chain C (residues C120 to C122 and C127 to C129, not shown in Fig. 3) are part of two other, independent β-sheets that are related to the first one by threefold symmetry.

In the model just described all three-residue strands are connected to adjacent ones by at least three hydrogen bonds. Alternative hydrogen-bonding patterns are obtained if one chain is shifted by one or more residues in the direction of the chain relative to the adjacent chains. However, this implies that the strands are no longer perfectly aligned and that strand residues no longer make hydrogen bonds to other strand residues only. The outward appearance of such a shifted arrangement is that successive strands in the sheets are staggered, which is typically observed in known proteins. Odd staggers are prohibited by constraints (4), (5) and (7), because they would lead to exposure of the hydrophobic strand residues.

3. Model Building

In order to construct an atomic model of the Ad5 fibre shaft we have focused on the stretch of 62 residues from Lys94 to Gly155 because here the repeat is exactly 15 residues, whereas it is 15 to 17 over the next 158 residues. The ranges Pro44–Thr93 and Asn314–Asn393 are even more irregular (see Fig. 1(b)).

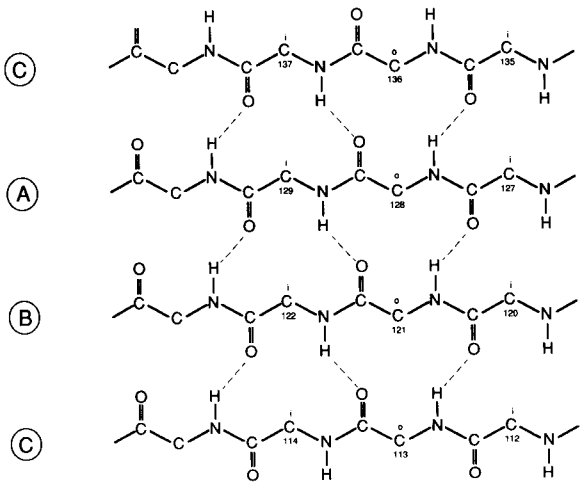


Figure 3. Basic non-staggered hydrogen-bonding pattern for the model wrapped around the cylinder in a left-handed fashion. The chain identifiers are given in circles on the left. i and o indicate side chains that point towards the fibre axis (inwards) and into the solvent (outwards), respectively. i corresponds to the hydrophobic residues 1 and 3 in the strands depicted in Fig. 1. The numbers below the C^{α} atoms are the residue numbers according to the sequence of the Ad5 fibre. The pattern used for the final model (see Fig. 4) differs from this basic pattern by a relative shift of 2 residues between adjacent strands.

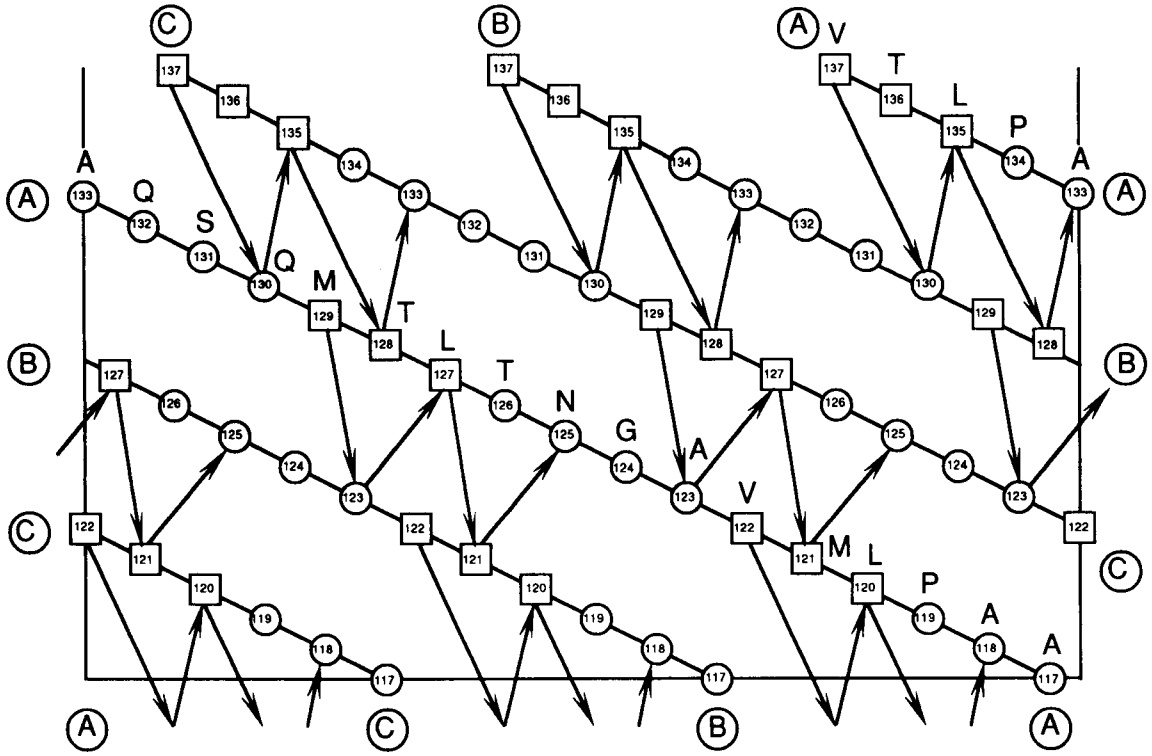


Figure 4. Schematic drawing of part of the unfolded cylinder of the left-handed model with a β-sheet stagger of 2 residues. The left and right vertical axes coincide. The encircled letters indicate the 3 separate monomeric chains. The 1-letter amino acid codes and residue numbers are also given. As in Fig. 1(a), the boxed numbers represent strand residues, the encircled ones loop residues. The arrows represent the main-chain hydrogen-bonding pattern.

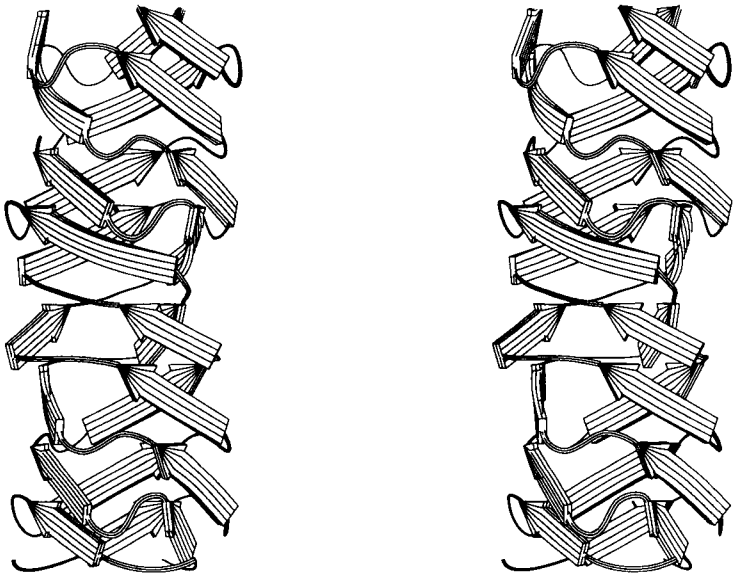
Based on the ideas developed above, a 3×62 residue left-handed trimer was constructed (alternative, rejected models are detailed below). It consisted entirely of alanine in order to avoid the problem of packing the side chains in the interior in the initial stages of the modelling process. This cylindrical model was improved in a controlled fashion, using restrained energy minimization (EM†) and molecular dynamics (MD) techniques, for which the GROMOS package was employed (Van Gunsteren & Berendsen, 1987). First, energy minimizations were carried out with distance restraints on selected strand–strand hydrogen bonds, thus imposing the desired hydrogen-bond pattern with a β-sheet stagger of two residues (Fig. 4). Threefold symmetry was enforced by using the main-chain atoms of the threefold rotationally averaged structure as position restraints. All ω angles were restrained to 180°, while the ϕ and ψ angles of the parallel β-strand segments were restrained to $-119°$ and $+113°$, respectively (Sherman & Andrianov, 1987). No non-bonded interactions were included until the β-strand torsion angles had approximately the right values (average deviation $\approx 10°$). After some energy minimization, isothermal MD simulations *in vacuo* were carried out at several temperatures between 300 and 450 K. Starting from several points in the trajectory, the temperature of the system was decreased and the resulting structures were energy minimized while gradually diminishing the restraining force constants.

At this stage the side chains were put onto the polyalanine structure that had the lowest energy, using the WHAT IF modelling package (Vriend, 1990). Some hydrophilic residues that were pointing inwards without being engaged in favourable interactions were manually turned to point outwards. A reasonable (initial) packing of side chains in the interior was established by an iterative process of rotation about χ-angles.

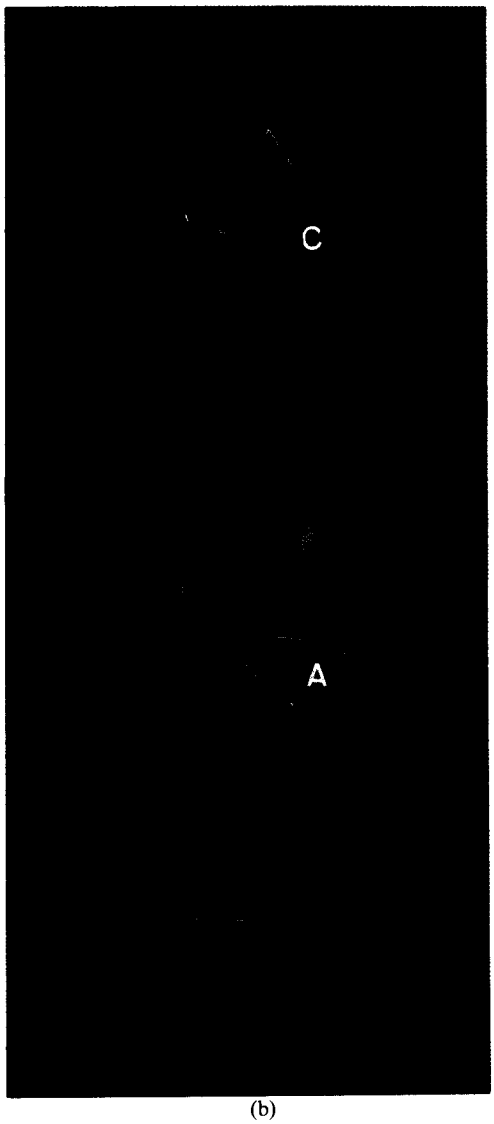
Subsequently, new EM calculations were carried out. Initially, the co-ordinates of the optimized polyalanine model were used for position restraints. Later, they were replaced by position restraints that ensure threefold symmetry. Hereafter, MD simulations were carried out for the full model, followed again by EM calculations, in the same fashion as in the case of the polyalanine model. This left-handed model with a hydrogen bond stagger of two (Fig. 4) proved most satisfactory. It is shown in its final form in Figs 5 and 6 and is discussed in detail below.

Several alternative models were investigated. Two such models are the right-handed and left-handed models with a zero stagger. Starting with an idealized cylinder, these two models were optimized in the same fashion as the left-handed model with a stagger of two. The resulting optimum structures were only ~223 Å long, much shorter than 290 Å.

†Abbreviations used: EM, energy minimization; MD, molecular dynamics.



(a)



(b)

Figure 5. Two representations of the final model. Strands are represented by broad bands, loops by thin ribbons. The assignment of strands and loops is according to the DSSP program (Kabsch & Sander, 1983). The strands do not have equal lengths, indicating that different repeats have different conformations which depend on the actual amino acid sequence. On the outside of the fibre the surfaces of 3 curved β-sheets make left-handed super β-helices. The short strands that make up the sheets are clearly staggered. The stereo picture (a) was generated using the RIBBON program (Priestle, 1988). The coloured mono picture (b) was generated using the UCSF MIDAS program (Ferrin *et al.*, 1988). Different colours indicate different monomeric units. Marks A and C represent strands in chains A and C; see the text for explanation.

Moreover, visual inspection revealed an empty channel running along the axis of the fibre, a highly unlikely feature since the interior of the fibre is hydrophobic. In addition, the right-handed model had the incorrect (left-handed) β-sheet twist. A third alternative was modelled starting with the ideal cylinder, but without any restraints on hydrogen bonds. This attempt led to a structure whose hydrophobic strand residues alternatingly pointed inwards and outwards in every subsequent 15-residue repeat, which is not consistent with the assumption that the repeats have the same three-dimensional structure. The fourth alternative model was arrived at by taking fragments from the Protein Data Bank (Bernstein *et al.*, 1977) and building the model with these fragments, using WHAT IF (Vriend, 1990). A 24-residue sandwich (6 strands) was taken from the triosephosphate isomerase (TIM) barrel (Banner *et al.*, 1975; Lasters *et al.*, 1988). Its front was rotated slightly with respect to the back in order to ensure the proper angle α with the xy-plane and the new sandwich was superimposed on the ideal cylinder several times. In order to connect these pieces, the best fitting loops with the correct length were taken from the Protein Data Bank and the resulting structure was regularized (Dodson *et al.*, 1976). The resulting 166-residue model did not possess threefold symmetry at all and the sandwich structure had been disrupted by the regularization procedure necessary to accommodate the loops. None of these four alternative models led to a plausible structure and they were therefore rejected.

4. Results

The model shown in Figs 5 and 6, which is left-handed (i.e. the chains are wrapped around a hypothetical cylinder making a left-handed screw) and has a hydrogen-bonding pattern with a stagger of two (Fig. 4), appears to be the most satisfactory. This judgement is based on two key observations.

First, the model is consistent with all experimental data as enumerated above. In particular it has the correct shaft length of about 290 Å, it has the correct β-sheet twist and it explains the observed fibre diffraction pattern. Second, it satisfies the criteria for a plausible protein structure. These points will now be discussed in detail.

As we have modelled only part of the shaft, the ends of the model lack the correct interactions with the rest of the shaft, and may result in quite unrealistic atomic arrangements. In order to assess the quality of our model, we consequently restrict ourselves to the two central repeats (Glu110 to Asp139) as a representation of the shaft. Superposition of the main-chain segment Glu110 to Gly124 onto segment Asn125 to Asp139 (root-mean-square deviation of C^{α} positions is 1·6 Å) gives the rotational part of the screw operation relating two consecutive 15-residue repeats as +60·63°. Since the model is left-handed, this implies that one repeat corresponds to a left-handed rotation of about 300°. The rise per 15-residue repeat was found to be 13·12 Å (T in Fig. 2), leading to a shaft length of 288·8 Å for 22 repeats. Although the path followed by the polypeptide chain is not a regular helix (due to the different conformations of the loop and strand regions), the effective pitch of a single chain is, due to the +60·63° azimuthal repeat, $-360/-299{\cdot}37 \times 13{\cdot}12 = 15{\cdot}8$ Å (P in Fig. 2). The average distance of the main-chain centres of mass (N, C, C^{α} atoms) to the fibre axis in the central two repeats is 8·03 Å. This can be regarded as the radius of the "ideal cylinder" shown in Figure 2, which leads to an average tilt angle α of 17·4° and an average chain length l_c (the distance between residue i and $i+15$ along chain A in Fig. 2) for a 15-residue repeat of 44·0 Å. These values should be compared with the estimates $\alpha = 15{\cdot}4°$ and $l_c =$ 50 Å for a completely extended polypeptide chain. The average tilt angle of the strands (calculated in the same fashion) is 26·8°, which is slightly more than the initially estimated value of 24°. This means that the strands and loops make a zig-zag pattern: the strands have a tilt angle almost double the average value and the loops consequently have a tilt angle of almost 0°.

Table 1 gives the main chain and C^{β} co-ordinates of the two central repeats, from which a model of the entire shaft can be derived by application of the screw operation and threefold symmetry.

(i) *Comparison with X-ray diffraction data*

A striking feature of the diffraction pattern of three-dimensional crystals of Ad2 fibre is the occurrence of diffuse layer lines at a spacing of 26·4 Å which presumably arise from a rotational disordered component of the crystals (Devaux *et al.*, 1990). That this scattering arises from the fibre itself is confirmed by the observed sampling of the diffuse scattering by the crystal Bragg reflections (Devaux *et al.*, 1990). Since the only significant pseudo-periodic structure in the fibre is the shaft, we further assume that this fibre-like diffraction pattern arises from the fibre shaft and thus our model of the shaft should explain the observed pattern, i.e. the occurrence of $1/26{\cdot}4$ Å^{-1} layer lines with meridional intensity on even orders, particularly strong on the sixth order ($1/4{\cdot}4$ Å^{-1}).

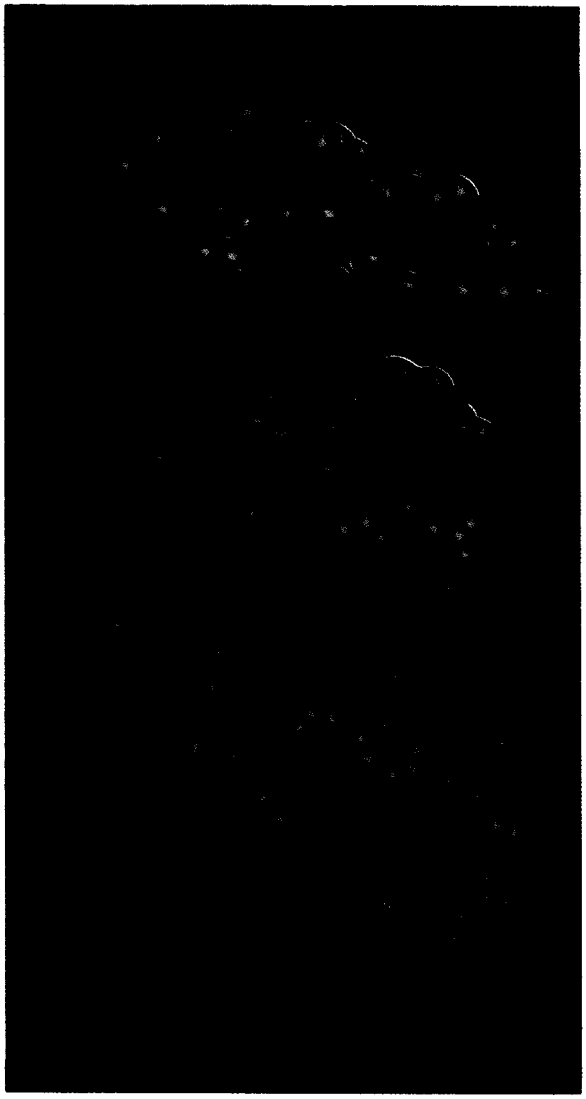


Figure 6. Space-filling representation of the final model. Different colours indicate different chains. The model is roughly 62 Å long and 28 Å wide when all atoms, including their van der Waals' radii are taken into account. Based on the positions of the C^{α} atoms only, the dimensions are 56 and 21 Å, respectively. The actual shaft is 5·3 times as long as this model. This picture was generated using WHAT IF (Vriend, 1990).

The atomic model presented here clearly explains the main features of the fibre diffraction pattern. Firstly, the axial rise per 15-residue repeat is 13·1 Å and this corresponds to the occurrence of strong meridional reflections on the 13·1 Å layer lines (even orders of the 26·2 Å layer lines). Secondly, the left-handed azimuthal rotation per repeat of nearly 300° (5/6 turn), combined with the threefold symmetry of the fibre, shows that the true axial repeat is 26·2 Å, which explains the observed layer lines at this

Table 1

Backbone and C^{β} co-ordinates of the two central repeats of the fibre model

Residue	Atom	x	y	z
Glu110	N	5·830	−6·214	−14·688
Glu110	C^{α}	5·322	−7·212	−15·656
Glu110	C	4·003	−7·855	−15·210
Glu110	O	2·954	−7·386	−15·644
Glu110	C^{β}	6·377	−8·227	−16·122
Ala111	N	4·068	−8·869	−14·345
Ala111	C^{α}	2·877	−9·509	−13·751
Ala111	C	2·260	−8·602	−12·670
Ala111	O	2·720	−8·535	−11·529
Ala111	C^{β}	3·259	−10·886	−13·197
Leu112	N	1·299	−7·807	−13·130
Leu112	C^{α}	0·707	−6·714	−12·338
Leu112	C	−0·831	−6·800	−12·286
Leu112	O	−1·558	−6·171	−13·058
Leu112	C^{β}	1·216	−5·400	−12·954
Thr113	N	−1·314	−7·576	−11·322
Thr113	C^{α}	−2·764	−7·765	−11·104
Thr113	C	−3·200	−6·887	−9·923
Thr113	O	−3·294	−7·322	−8·776
Thr113	C^{β}	−3·080	−9·260	−10·918
Val114	N	−3·570	−5·666	−10·286
Val114	C^{α}	−3·817	−4·576	−9·320
Val114	C	−5·285	−4·140	−9·227
Val114	O	−5·780	−3·284	−9·966
Val114	C^{β}	−2·860	−3·377	−9·491
Ala115	N	−5·945	−4·768	−8·260
Ala115	C^{α}	−7·344	−4·483	−7·896
Ala115	C	−7·374	−3·262	−6·960
Ala115	O	−7·061	−3·342	−5·770
Ala115	C^{β}	−7·961	−5·721	−7·235
Ala116	N	−7·510	−2·115	−7·616
Ala116	C^{α}	−7·466	−0·793	−6·967
Ala116	C	−8·754	0·012	−7·174
Ala116	O	−9·325	0·041	−8·265
Ala116	C^{β}	−6·275	−0·016	−7·535
Ala117	N	−9·177	0·673	−6·097
Ala117	C^{α}	−10·325	1·602	−6·126
Ala117	C	−9·878	2·967	−6·685
Ala117	O	−9·497	3·871	−5·944
Ala117	C^{β}	−10·923	1·711	−4·716
Ala118	N	−9·770	2·980	−8·015
Ala118	C^{α}	−9·262	4·068	−8·884
Ala118	C	−8·377	5·194	−8·291
Ala118	O	−7·173	5·058	−8·516
Ala118	C^{β}	−10·325	4·526	−9·897
Pro119	N	−8·853	6·257	−7·599
Pro119	C^{α}	−7·973	7·334	−7·090
Pro119	C	−7·086	6·894	−5·910
Pro119	O	−7·421	7·039	−4·734
Pro119	C^{β}	−8·921	8·481	−6·733
Leu120	N	−5·905	6·417	−6·280
Leu120	C^{α}	−4·904	5·870	−5·350
Leu120	C	−3·680	6·781	−5·225
Leu120	O	−2·893	6·941	−6·160
Leu120	C^{β}	−4·460	4·488	−5·840
Met121	N	−3·603	7·440	−4·076
Met121	C^{α}	−2·420	8·228	−3·686
Met121	C	−1·503	7·288	−2·889
Met121	O	−1·427	7·305	−1·658
Met121	C^{β}	−2·828	9·478	−2·889
Val122	N	−0·815	6·467	−3·676
Val122	C^{α}	−0·112	5·272	−3·185
Val122	C	1·400	5·314	−3·463
Val122	O	1·893	4·954	−4·530
Val122	C^{β}	−0·819	3·987	−3·673
Ala123	N	2·125	5·720	−2·431
Ala123	C^{α}	3·597	5·664	−2·452
Ala123	C	4·100	4·656	−1·405
Ala123	O	4·596	4·994	−0·329
Ala123	C^{β}	4·201	7·064	−2·303
Gly124	N	3·766	3·399	−1·690
Gly124	C^{α}	4·183	2·229	−0·889
Gly124	C	5·613	1·814	−1·260
Gly124	O	5·867	0·678	−1·660
Asn125	N	6·529	2·760	−1·065
Asn125	C^{α}	7·923	2·664	−1·541
Asn125	C	8·693	1·426	−1·066
Asn125	O	9·371	0·792	−1·871
Asn125	C^{β}	8·677	3·964	−1·250
Thr126	N	8·598	1·127	0·227
Thr126	C^{α}	9·084	−0·150	0·795
Thr126	C	7·980	−0·791	1·652
Thr126	O	7·998	−0·750	2·885
Thr126	C^{β}	10·379	0·045	1·602
Leu127	N	6·987	−1·334	0·950
Leu127	C^{α}	5·749	−1·837	1·574
Leu127	C	5·248	−3·150	0·945
Leu127	O	4·518	−3·179	−0·047
Leu127	C^{β}	4·690	−0·727	1·492
Thr128	N	5·626	−4·251	1·581
Thr128	C^{α}	5·111	−5·586	1·214
Thr128	C	4·231	−6·072	2·376
Thr128	O	4·702	−6·637	3·368
Thr128	C^{β}	6·229	−6·582	0·863
Met129	N	2·945	−5·781	2·219
Met129	C^{α}	1·991	−5·745	3·341
Met129	C	0·813	−6·711	3·154
Met129	O	0·102	−6·700	2·146
Met129	C^{β}	1·491	−4·307	3·499
Gln130	N	0·632	−7·509	4·199
Gln130	C^{α}	−0·477	−8·467	4·326
Gln130	C	−1·361	−8·019	5·500
Gln130	O	−0·911	−7·891	6·640
Gln130	C^{β}	0·098	−9·870	4·566
Ser131	N	−2·554	−7·572	5·130
Ser131	C^{α}	−3·553	−7·090	6·103
Ser131	C	−4·911	−7·783	5·918
Ser131	O	−5·126	−8·483	4·927
Ser131	C^{β}	−3·688	−5·575	5·934
Gln132	N	−5·838	−7·531	6·840
Gln132	C^{α}	−7·228	−8·045	6·779
Gln132	C	−8·115	−7·342	5·720
Gln132	O	−9·313	−7·599	5·654
Gln132	C^{β}	−7·860	−7·975	8·188
Ala133	N	−7·449	−6·721	4·743
Ala133	C^{α}	−7·968	−5·788	3·719
Ala133	C	−8·446	−4·406	4·227
Ala133	O	−7·581	−3·527	4·195
Ala133	C^{β}	−8·816	−6·467	2·628
Pro134	N	−9·675	−4·148	4·731
Pro134	C^{α}	−10·074	−2·788	5·159
Pro134	C	−9·336	−2·338	6·426
Pro134	O	−9·401	−3·021	7·449
Pro134	C^{β}	−11·588	−2·851	5·364
Leu135	N	−8·453	−1·361	6·237
Leu135	C^{α}	−7·696	−0·751	7·348
Leu135	C	−7·715	0·784	7·299
Leu135	O	−7·395	1·401	6·283
Leu135	C^{β}	−6·249	−1·269	7·358
Thr136	N	−8·160	1·360	8·409
Thr136	C^{α}	−8·119	2·814	8·662
Thr136	C	−7·069	3·120	9·739
Thr136	O	−7·266	2·838	10·924
Thr136	C^{β}	−9·493	3·353	9·085
Val137	N	−5·926	3·605	9·259
Val137	C^{α}	−4·727	3·839	10·093
Val137	C	−4·304	5·317	10·146
Val137	O	−3·837	5·892	9·162
Val137	C^{β}	−3·587	2·869	9·711
His138	N	−4·431	5·884	11·340
His138	C^{α}	−4·168	7·318	11·573
His138	C	−3·135	7·537	12·686

Table 1 *(continued)*

Table 1 *(continued)*

Residue	Atom	x	y	z
His138	O	−3·439	7·341	13·862
His138	C^{β}	−5·462	8·077	11·902
Asp139	N	−1·890	7·678	12·237
Asp139	C^{α}	−0·695	8·070	13·025
Asp139	C	0·533	7·984	12·107
Asp139	O	1·027	6·897	11·818
Asp139	C^{β}	−0·449	7·193	14·271

The complete structure can be reconstructed from these coordinates by application of threefold symmetry about the z-axis and a screw operation with $\Delta z = 13{\cdot}12$ Å and $\Delta\phi = +60{\cdot}63°$.

spacing. In other words, going along the fibre parallel to its axis, starting at a given point in chain A (e.g. the green arrow indicated with A in Fig. 5(b)) upon arrival at the blue arrow in chain C (marked C in Fig. 5(b)), one encounters an identical situation both geometrically and in terms of amino acid sequence: both these arrows have equivalent amino acid sequence, they are related by a 26·2 Å translation and, because of the azimuthal rotation of almost 300° per repeat, by very nearly a 0° rotation ($2 \times 300° + 120°$).

To demonstrate the occurrence of the 26·2 Å repeat in the model, a fibre diffraction pattern was simulated for a full length Ad5 fibre shaft (i.e. 22 × 15-residue repeats) and compared to the experimental pattern (Fig. 7). Since the real adenovirus fibre shaft is non-periodic in sequence, a full-length model was constructed in the following way to avoid, as much as possible, artificial repeats. As mentioned above, only the central two 15-residue repeats (denoted by A, residues 110 to 124, and B, residues 125 to 139) of the model are considered reliable. A pseudo-random sequence of these two repeats was devised and the shaft constructed by applying the required multiple of the screw transformation described above to successive repeats. The actual sequence used was ABABBAABABAABBBABAABAB. The Fourier transform was calculated by direct summation on a cylindrical grid (pixel spacing 1/384 $Å^{-1}$) followed by cylindrical averaging, using the threefold symmetry. The calculated Ad5 fibre diffraction pattern shows a striking resemblance to the observed Ad2 diffraction pattern in the distribution of meridional and non-meridional intensity (Fig. 7). The predicted position of the subsidiary maximum on the equator, which is a measure of the fibre shaft radius, is at 1/14·2 $Å^{-1}$ corresponding to that observed at about 1/15 $Å^{-1}$. These results suggest that there is no special significance to the observed very strong 4·4 $Å^{-1}$ meridional reflection. It simply arises from the sampling of the molecular transform of one repeat by the 26·2 Å layer line spacing, of which it is the sixth order. In this connection we note that there are minima on the meridion of the molecular transform of one repeat at about 6·5 $Å^{-1}$ (close to the fourth layer line) and 4·9 $Å^{-1}$.

Diffraction patterns were also calculated for models $(A)_{22}$ and $(B)_{22}$ which differ in detail from the pseudo-random sequence, but again show strong layer lines of spacing 26·2 Å with meridional inten-

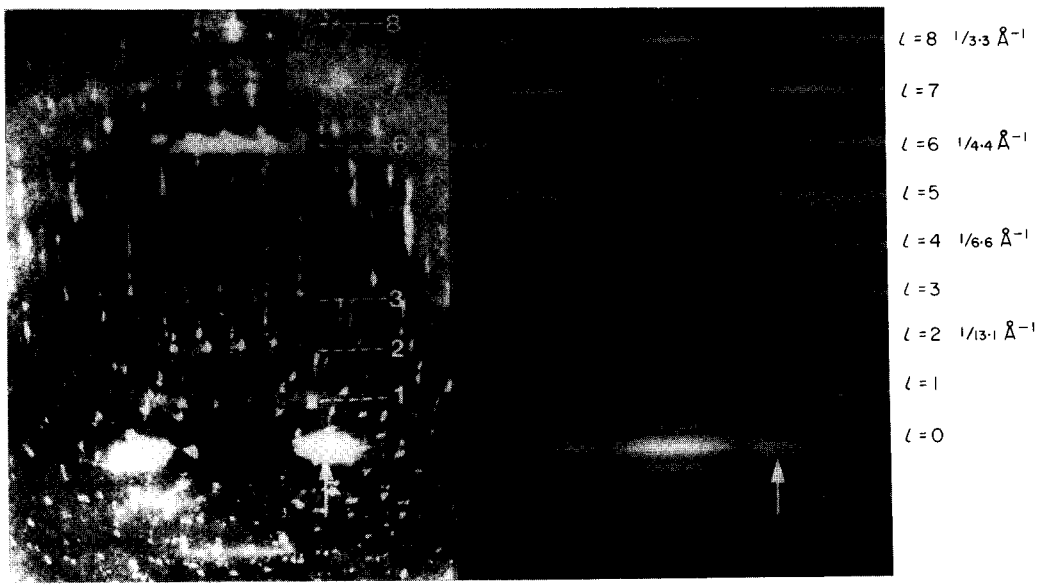


Figure 7. Experimental and simulated diffraction patterns. The left panel shows the experimental X-ray diffraction pattern of cleaved Ad2 fibre crystals with a weak fibre pattern superimposed on crystalline diffraction (taken with permission from Fig. 7 in Devaux *et al.*, 1990). The right panel shows the simulated fibre diffraction pattern of the full-length atomic model of the Ad5 fibre shaft with random alternation of A and B units (see text). The top half planes of the diffraction patterns are shown, the fibre axis being vertical. Orders of the 1/26·2 $Å^{-1}$ layer lines are marked. The equatorial subsidiary maxima (at 1/15 $Å^{-1}$ in the experimental pattern and at 1/14·2 $Å^{-1}$ in the simulated pattern) are indicated with white vertical arrows. The 2 patterns have been put as closely as possible on the same scale, although in the experimental pattern reciprocal space is distorted slightly due to the diffraction geometry.

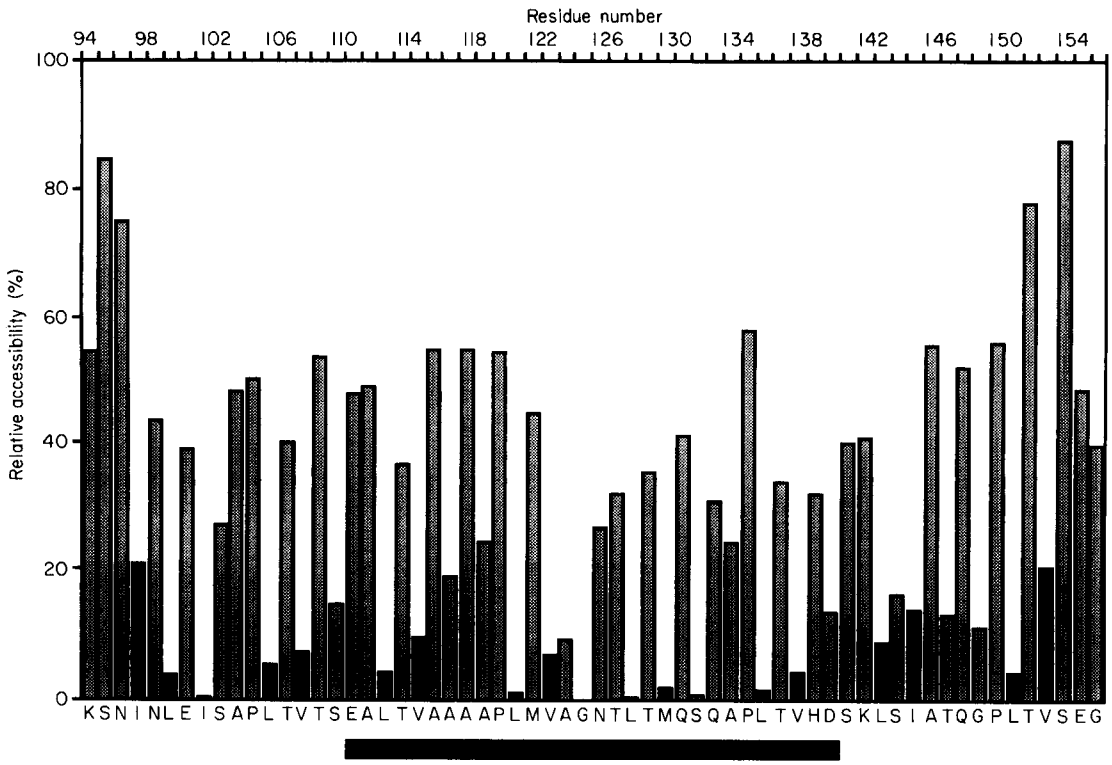


Figure 8. Relative accessibility as a function of residue. The relative accessibility of a residue is equal to its absolute accessible surface area divided by the maximum accessible surface area for that residue type in completely extended model peptides. Accessible surface areas are according to the DSSP program (Kabsch & Sander, 1983). The residue numbers are given on the upper ordinate, corresponding 1-letter amino acid types on the lower ordinate. The filled bar below residues Glu110 to Asp139 indicates the 2 central repeats on which the evaluation of the model is based. Black columns represent the hydrophobic 1st and 3rd residues of the short β-strands; grey columns stand for all other residues. It is clear that apart from the outer edges (1st and last 6 residues), where end effects play a role, the hydrophobic strand residues and other hydrophobics are buried to a great extent.

sity on the even orders. However, the model $(AB)_{11}$ would introduce an artificial exact axially projected repeat of 26·2 Å and would thus give meridional reflections with this spacing, which is not observed.

(ii) *Is this a plausible protein structure?*

The final model is a plausible protein structure based on the following criteria: (1) it consists of the strands and loops as suggested by secondary structure prediction methods; (2) it contains properly main-chain hydrogen-bonded β-sheets; (3) it generally has normal backbone torsion angles, although five ω-angles in the two central repeats have deviations of more than 10° and four ϕ–ψ-combinations in the loop regions of the two central repeats fall outside the "allowed" region; (4) its hydrophobic side chains are predominantly buried, as illustrated in Fig. 8, which shows the calculated relative accessible surface areas as a function of residue; and (5) its interior is well-packed as judged by a cavity search using WHAT IF (Vriend, 1990) with a probe of 1·4 Å radius, which only revealed one small isolated hole in the vicinity of Ile120 $C^{\delta}2$ and Val122 $C^{\gamma}2$. In addition, the model also passed a test for proper packing based on the contact quality index which is a measure for the agreement between the distribution of atoms around all residues in the model and the equivalent distributions observed in known high-resolution structures (Vriend & Sander, 1992).

5. Discussion

The assumption that the fibre shaft has predominantly β-strand features would lead to too short a fibre if *intra*-chain hydrogen bonding is expected to stabilize the structure (as in Green's model). Since the three monomeric units are identical and run from the tail all the way to the knob, it is plausible to assume that they are related by threefold symmetry. Any model of the shaft that features inter-chain main-chain hydrogen bonding and threefold symmetry almost naturally leads to a structure with parallel chains that are intertwined like a rope. For such a model the only basic degrees of freedom lie in the handedness and the exact nature of the hydrogen-bonded network.

The choice of hydrogen-bonding pattern has a dramatic effect on the predicted shaft length. A shift of two residues in the pattern of *inter*-chain hydrogen bonds (going from a stagger of 2 to a stagger of 0) brings about a length decrease of 70 Å (from 290 to 220 Å) and a corresponding increase in radius with an empty channel in the middle of the

hydrophobic core. A stagger of +4 need not be considered since it would lead to a shaft much longer than 290 Å and to a radius so small that the side chains could not be accommodated in the interior. Going to negative staggers would lead to shafts that are too short and thick.

We did not build a right-handed model with a hydrogen-bond stagger of 2, which would lead to a shaft of approximately correct length. Such a model would have a left-handed (incorrect) β-sheet twist as already seen explicitly in the right-handed model with a stagger of 0, which we did build. This left-handed twist is expected to be more pronounced for a more tightly wound model (i.e. a model with a larger hydrogen-bond stagger) and it is therefore unlikely that any right-handed model is correct.

The model presented cannot be expected to be correct in all details. It has generally acceptable backbone torsion angles with four exceptions in loop regions. Glu110 and Ala133 fall slightly outside the "allowed" range, while Ala118 ($\phi = +19°$, $\psi = -80°$ and Asp139 ($\phi = +175°$, $\psi = +106°$) deviate more. As there are no experimental guiding constraints on the conformations of the loops, further refinement of these residues was not attempted. For the same reason, the loops may not have the correct overall conformation, so at present there is no basis for rationalizing the fairly regular alternation of four- and five-residue turns or the role of the conserved proline (or glycine) residue in the five-residue turn. The real hydrogen-bond pattern might also be less regular than that presented here because of the possible presence of β-bulges (Richardson *et al.*, 1978). Not taking into account protein–water interactions in the simulations causes hydrophilic side chains to fold back onto the protein where they normally would protrude into the solvent. The lack of protein–water interactions is also known to have a slight compressive effect (Levitt & Sharon, 1988).

In principle, any final model of the fibre should be stable without restraints. However, because we model a small part of the shaft which may be unstable by itself, and also because we do not take protein–water interactions into account, we did not remove all restraints completely. Specifically, the restraints on the hydrogen bonds that stabilize the β-sheets were retained with a weak force constant of 0·023% of a normal O–H covalent bond force. This led to acceptable hydrogen bonds: the minimum and maximum H...O distances of the restrained hydrogen bonds were 1·76 Å and 2·27 Å and their average and root-mean-square difference values were 1·88 Å and 0·26 Å. All threefold symmetry and torsion angle restraints were removed completely. We observed that, for the two central repeats, five out of 31 peptide units deviate more than 10° from planarity (but nowhere more than 21°). This is acceptable, however, since present-day force fields are known to allow ω to deviate considerably from planarity (Avbelj *et al.*, 1990) while deviations of more than 10° are also frequently observed in real linear unstrained peptides (Ashida *et al.*, 1987).

6. Conclusion

We have demonstrated that it is possible to build, from first principles, a plausible model of a fragment of a protein whose unique sequence precludes use of a known homologous structure. Unlike the extremely difficult general case of a globular protein without any symmetry, we have chosen the more tractable case of a fibrous protein in which the known rotational symmetry and repeat distance are constraints that severely limit the possible folding patterns. In this situation the inherently blind methods of MD (which can sample many conformations) and EM (which can find local conformational energy minima), corrected and guided by human intervention, are very powerful tools for arriving at a model that satisfies experimental constraints and conforms to general principles of protein structure. The final model we present of the adenovirus fibre shaft is consistent with all experimental data and features a novel triple helix of β-strands.

We thank Bernard Jacrot for valuable discussions, Wilfred van Gunsteren and Herman Berendsen for supplying us with the GROMOS package, Gerrit Vriend for placing the WHAT IF package at our disposal, Jolanta Stouten for critically reading the manuscript and Ulrike Göbel, Clemens Hladek and Alan Summerfield for the preparation of illustrations.

References

Ashida, T., Tsunogae, Y., Tanaka, I. & Yamane, T. (1987). Peptide chain structure parameters, bond angles and conformational angles from the Cambridge structural database. *Acta Crystallogr. sect. B*, **43**, 212–218.

Avbelj, F., Moult, J., Kitson, D. H., James, M. N. G. & Hagler, A. T. (1990). Molecular dynamics study of the structure and dynamics of a protein molecule in a crystalline ionic environment, *Streptomyces griseus* protease A. *Biochemistry*, **29**, 8658–8676.

Banner, D. W., Bloomer, H. C., Petsko, G. A., Phillips, D. C., Pogson, C. I. & Wilson, A. I. (1975). Structure of chicken muscle triosephosphate isomerase determined crystallographically at 2·5 Å resolution using amino acid sequence data. *Nature (London)*, **255**, 609–614.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Chothia, C. (1973). Conformation of twisted β-pleated sheets in proteins. *J. Mol. Biol.* **75**, 295–302.

Chroboczek, J. & Jacrot, B. (1987). The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. *Virology*, **161**, 549–554.

Devaux, C., Caillet-Boudin, M. L., Jacrot, B. & Boulanger, P. (1987). Crystallization, enzymatic cleavage, and the polarity of the adenovirus type 2 fiber. *Virology*, **161**, 121–128.

Devaux, C., Adrian, M., Berthet-Colominas, C., Cusack, S. & Jacrot, B. (1990). Structure of adenovirus fibre. I. Analysis of crystals of fibre from adenovirus serotypes 2 and 5 by electron microscopy and X-ray crystallography. *J. Mol. Biol.* **215**, 567–588.

Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976). A method for fitting satisfactory models to sets of atomic positions in protein structure refinements. *Acta Crystallogr. sect. A*, **32**, 311–315.

Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125–142.

Ferrin, T. E., Huang, C. C., Jarvis, L. E., Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13–27.

Green, N. M., Wrigley, N. G., Russell, W. C., Martin, S. R. & McLachlan, A. D. (1983). Evidence for a repeating cross-β sheet structure in the adenovirus fibre. *EMBO J.* **2**, 1357–1365.

Hérissé, J., Rigolet, M., Dupont de Dinechin, S. & Galibert, F. (1981). Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic region of the fiber protein and the entire E4 region. *Nucl. Acids Res.* **9**, 4023–4042.

Hong, J. S., Mullis, K. G. & Engler, J. A. (1988). Characterization of the early region 3 and fiber genes of Ad7. *Virology*, **167**, 545–553.

Jones, E. Y., Stuart, D. I. & Walker, N. P. C. (1989). Structure of tumour necrosis factor. *Nature (London)*, **338**, 225–228.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kidd, A. H. & Erasmus, M. J. (1989). Sequence characterization of the adenovirus 40 fiber gene. *Virology*, **172**, 134–144.

Lasters, I., Wodak, S. J., Alard, P. & Van Cutsem, E. (1988). Structural principles of parallel β-barrels in proteins. *Proc. Nat. Acad. Sci., U.S.A.* **85,** 3338–3342.

Lesk, A. M. (1988). Prediction of secondary structure of proteins. In *Computational Molecular Biology: Sources and Methods for Sequence Analysis* (Lesk, A. M., ed.), pp. 192–197, Oxford University Press, Oxford.

Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 7557–7561.

Pieniazek, N. J., Slemenda, S. B., Pieniazek, D., Velarde, J. Jr. & Luftig, R. B. (1989). Sequence of human enteric adenovirus type 41 tak fiber protein gene. *Nucl. Acids Res.* **17**, 9474.

Priestle, J. P. (1988). RIBBON: a stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* **21**, 527–576.

Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). The β bulge: a common small unit of nonrepetitive protein structure. *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2574–2578.

Ruigrok, R. W. H., Barge, A., Albiges-Rizo, C. & Dayan, S. (1990). Structure of adenovirus fibre. II. Morphology of single fibres. *J. Mol. Biol.* **215**, 589–596.

Sherman, S. A. & Andrianov, A. M. (1987). Relationship of NMR spectral parameters with conformational characteristics of amino acid backbone. Analysis of elements of protein secondary structure. *Mol. Biol.* **21**, 1284–1291.

Signäs, C., Akusjärvi, G. & Pettersson, U. (1985). Adenovirus 3 fiber polypeptide gene: implications for the structure of the fiber protein. *J. Virol.* **53**, 672–678.

Stewart, P. L., Burnett, R. M., Cyrklaff, M. & Fuller, S. D. (1991). Image reconstruction reveals the complex molecular organization of adenovirus. *Cell*, **67**, 145–154.

Van Gunsteren, W. F. & Berendsen, H. J. C. (1987). GROMOS, Groningen molecular simulation computer program package. University of Groningen, The Netherlands.

Van Oostrum, J. & Burnett, R. M. (1985). Molecular composition of the adenovirus type 2 virion. *J. Virol.* **56**, 439–448.

Van Oostrum, J., Smith, P. R., Mohraz, M. & Burnett, R. M. (1987). The structure of the adenovirus capsid. III. Hexon packing determined from electron micrographs of capsid fragments. *J. Mol. Biol.* **198**, 73–89.

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–55.

Vriend, G. & Sander, C. (1992). Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.* accepted for publication.

Weber, J. M., Talbot, B. G. & Delorme, L. (1989). The orientation of the adenovirus fiber and its anchor domain identified through molecular mimicry. *Virology*, **168**, 180–182.

Edited by M. F. Moody