Acta Cryst. (1995). D**51**, 792–804

Pf1 Filamentous Bacteriophage: Refinement of a Molecular Model by Simulated Annealing using 3.3 Å Resolution X-ray Fibre Diffraction Data

BY A. GONZALEZ AND C. NAVE

Daresbury Laboratory, Warrington WA4 4AD, England

AND D. A. MARVIN*

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, England

(*Received* 4 *July* 1994; *accepted* 1 *March* 1995)

Abstract

The filamentous bacteriophage Pf1 is structurally similar to the well known Ff (fd, f1, M13) strains, but it gives much better X-ray diffraction patterns, enabling a more detailed analysis of the molecular structure. The 46-residue protein subunit can be closely approximated by a single gently curved stretch of α-helix. The axes of the subunits are at a small angle to the virion axis, and several thousand subunits form an overlapping interdigitated helical array surrounding a DNA core. We have derived a detailed model of the virion based on X-ray data and stereochemical constraints. We have considered potential sources of error in the diffraction data, and used the improved data to study regions where the protein subunit of Pf1 may deviate from a continuous α-helix. We use simulated annealing to escape from local minima, and various kinds of electron-density maps to guide the model building. Refinement of the model shows that the first few residues at the N terminus are non-helical, and there is a slight discontinuity in the α-helix near the middle of the sequence. The model is consistent both with general structural principles derived from high-resolution analysis of other proteins, and with specific chemical and spectroscopic data about Pf1. We apply the same refinement techniques to an alternative model with a non-helical surface loop between residues 13 and 19. Comparative analysis of models with and without a loop shows that the loop model is not supported by 3.3 Å resolution X-ray diffraction data.

1. Introduction

The filamentous bacteriophage (genus *Inovirus*: Matthews, 1982) is a valuable model system in molecular biology. The virion is a flexible nucleoprotein rod about 1–2 μm by 60 Å diameter, comprising a helical tube of several thousand major coat protein subunits surrounding a core of single-stranded circular DNA. A few minor proteins cap the two ends. During the assembly, the virion extrudes through the bacterial membrane as the DNA passes from an intracellular DNA-binding protein to the viral coat protein, and *Inovirus* is a model to study protein–protein, DNA–protein and membrane–protein interactions. Foreign DNA can be spliced into the viral genome, either as a separately expressed gene or fused to an existing gene coding for a viral protein, and the longer genome can be encapsidated in a correspondingly longer virion and expressed in progeny viruses. This property makes *Inovirus* useful as a cloning vector and as a framework to display foreign peptides or proteins for ligand studies. Knowledge of the structure of the virion is essential to understanding these properties of the virus. [For general reviews, see Rasched & Oberer (1986); Model & Russel (1988); for a review of the virion structure, see Marvin (1990); for reviews of the assembly process, see Wickner (1988); Russel (1991); for reviews of peptide display, see Cesareni (1992); Smith (1993)].

X-ray fibre diffraction patterns of the virion in different *Inovirus* strains show that the major coat proteins are largely α-helix, with similar molecular packing for all strains. Slight differences between the diffraction patterns distinguish two symmetry classes, typified by strains fd and Pf1. The protein subunits of fd and Pf1 are similar in size (50 and 46 residues, respectively) and have the same 'trifunctional' primary structure: a collection of acidic residues in the N-terminal third, a 19-residue apolar domain in the middle, and a collection of basic residues near the C terminus; but despite this overall similarity, there is virtually no detailed homology between the two coat-protein sequences (Nakashima, Wiseman, Konigsberg & Marvin, 1975). The protein subunits can be closely approximated by gently curved α-helix rods about 7 nm long by 1 nm diameter in an overlapping interdigitated array. The symmetry difference between the two virion classes does not substantially affect either the conformation of the subunit or the interactions between neighbouring subunits, but it does alter the diffraction pattern enough so that significantly more data can be extracted

* Author to whom correspondence should be addressed.

ISSN 0907-4449

from Pf1 diffraction patterns than from fd diffraction patterns. Lowering the temperature below about 283 K slightly changes the Pf1 symmetry, enabling resolution of three-dimensional intensity data to 4 Å along the layer lines (Nave *et al.*, 1981). The unit twist and unit rise relating one helix subunit to the next are (66.667°, 2.90 Å) for the higher temperature symmetry ($Pf1^H$), and (65.915°, 3.05 Å) for the lower temperature symmetry ($Pf1^L$). These small differences in helix parameters imply only small differences in subunit structure. We therefore chose $Pf1^L$ as the representative of *Inovirus* for detailed X-ray structure analysis.

A model-independent maximum-entropy method of calculating an electron-density map from native data and single isomorphous derivative data (with iodine covalently bound to Tyr25) confirmed that the subunit of Pf1 is largely α-helix, and a molecular model of the protein subunit was built into this map (Bryan, Bansal, Folkhard, Nave & Marvin, 1983). This model is α-helix except for the first few residues at the N terminus, where the map is ill defined. The native electron-density map shows little detail that can be attributed to side chains, but the position of the iodinated Tyr25 was determined by calculation of a difference electron-density map (Bryan *et al.*, 1983; Marvin, Bryan & Nave, 1987), and this determines the azimuthal orientation of the α-helix subunit around its own axis. Independent chemical and spectroscopic data are consistent with this orientation of the α-helix. The side chains in the central apolar domain of each subunit interlock tightly with side chains on neighbouring subunits, but the N-terminal and C-terminal domains are relatively unconstrained by neighbouring subunits. Refinement of this model with respect to 3.3 Å resolution X-ray diffraction data confirmed the azimuthal orientation of the subunit, but gave no convincing unique non-helix solution for the N-terminal and C-terminal domains. Therefore, we reduced the number of independent variables in the model by constraining all main-chain (φ, ψ) and side chain χ^1 torsion angles to be within the range found for α-helix in high-resolution protein structures, to give a 100% α-helix approximation to the structure (Marvin, 1990; Marvin, Nave, Bansal, Hale & Salje, 1992).

A somewhat different model has been proposed on the basis of 8 Å resolution neutron-diffraction experiments on $Pf1^H$ (Nambudripad, Stark & Makowski, 1991). In this model, residues 13 to 19 are in a disordered surface loop. The overlapping interdigitated nature of the array of subunits implies that the electron density assigned to the N terminus in our model is assigned to the loop from a neighbouring subunit in the loop model.

Here we describe further refinement of the Pf1 model using simulated high-temperature annealing and calculation of various kinds of electron-density maps to explore systematically the accessible areas of model space, and we examine in particular whether the loop model is consistent with the 3.3 Å resolution diffraction data.

2. Theory and methods

2.1. *Fibre diffraction data*

The application of X-ray fibre diffraction to determine the structure of filamentous bacteriophage is discussed in detail by Nave *et al.* (1981); Marvin & Nave (1982); Marvin (1990), and Marvin *et al.* (1987, 1992); Marvin, Hale, Nave & Helmer Citterich (1994). Diffracted intensity from a helix is confined to a set of layer lines at reciprocal spacing $Z = n/P + m/p$, where n and m are integers, p is the axial rise between subunits in the helix, and P is the helix pitch. Within the limits of error, the $Pf1^L$ helix can be approximated by a helix with 71 units in 13 turns and a repeat $c = 216.5$ Å (Marvin *et al.*, 1987). Layer lines are assigned indices $l = Zc$, but the symmetry of the helix implies systematic absence of many layer lines or parts of layer lines. The amplitude distribution along the layer lines from independently diffracting helices corresponds to the continuous molecular transform of the virion, and this continuity constrains phases. Helix symmetry may also lead to superposition of different Bessel function orders on a layer line, but for our data this is not a problem to $R = 0.25$ Å^{-1} along the layer line, and we use the notation $F_o = |I_l^{1/2}(R)|$, $F_c = |G_{n,l}(R)|$.

We use the diffraction data reported by Marvin *et al.* (1987). The native diffraction data extend to 2.6 Å on some higher layer lines, but to between 3 and 4 Å for layer lines nearer the equator, and only to 4.2 Å on the equator ($l = 0$) itself. The number of independent data points is comparable to that expected for single-crystal data at 3.3 Å resolution from a 46-residue protein. The continuous transform was sampled at intervals of $R = 0.0025$ Å^{-1} along the layer line. To use standard crystallographic refinement programs, we converted the data to pseudo single-crystal data by defining an orthogonal unit cell with $a = b = 400$ and $c = 216.5$ Å, in space group $P1$. This unit cell contains 71 identical protein subunits in 13 turns of the helix. The DNA is only about 6% by weight of the virion, and it does not have the same symmetry as the protein, so we omit it from our calculations on non-equatorial layer lines. Each data point along a layer line l is assigned a Miller index $(h,0,l)$, so the data correspond to two-dimensional single-crystal data, and the non-crystallographic (helix) symmetry provides three-dimensional information.

In our refinement we used only the data in the resolution range 12–2.6 Å, which is largely beyond spacings where it is necessary to correct for bulk solvent, and we calculated the structure factors without correcting for solvent or including the DNA core. The complete data were put on an approximate absolute scale, allowing for solvent, by Marvin *et al.* (1987). If this scale is correct, the scale factor $\sum F_o F_c / \sum F_c^2$ calculated for our models using F between 12 and 2.6 Å resolution should be 1/71, to allow for the 71 asymmetric units used

to calculate F_c, and this is approximately correct: using this scale factor gives $G_{0,0} = 620\,e$, compared with $G_{0,0} = 490\,e$ estimated by Marvin *et al.* (1987) for the protein shell alone. We used the correlation coefficient (CC) as well as the R factor to measure the fit of calculated to observed continuous transform. The CC is relatively insensitive to errors in scaling, and does not presuppose that all errors are in the observed data: it allows for intrinsic errors in the method of calculation, such as imperfect treatment of solvent and disordered regions of the molecule. With only one Bessel function term contributing to each data point, the expected R factors are similar to those for single-crystal data. We used all the data which could successfully be deconvoluted into individual Bessel function terms. We tested the success of the deconvolution using the error analysis described below. Using this approach, we could refine the structure and follow the refinement using standard crystallographic methods, for instance simulated annealing to obtain unbiased omit maps.

By the sampling theorem (*e.g.* Bracewell, 1965), the Fourier transform of a structure of maximum dimension 70 Å is determined by sampling at 0.007 $Å^{-1}$, so our data are oversampled by about a factor of 3. We used this fact to estimate the errors in our data by iteratively refining a map phased on a model (Nave *et al.*, 1981), also called Fourier–Bessel smoothing. The smoothed-calculated transform (F_s) fits the observed data (F_o) with CC = 0.979 in the resolution range 12–2.6 Å. There is no point in attempting to fit the data better than this. Similar estimates of the quality of the data have been obtained by a maximum-entropy method (Marvin *et al.*, 1987). The Fourier–Bessel smoothing gave an estimate of the standard deviation σ in the observed data and enabled outliers in the data to be identified. Data points with $|F_o - F_s| > 3\sigma$ were examined and discarded if there was a reasonable explanation for the unreliable measurement. A few data points showing residual crystallinity or spill-over from adjacent very strong layer lines were identified by this method. The lower resolution data that we used (Nave *et al.*, 1981) contains some zero intensities, since the maximum-entropy procedure (Bryan, 1987) was not used to process these data. In these cases, data points measured as zero were set to $\sigma = 0.8\,e$ (the observed standard deviation for weak measured reflections) in regions where diffracted amplitude is consistent with the helix symmetry. Inclusion of these weak reflections increased the number of data points used in the refinement from 3252 to 3548.

2.2. *Refinement by energy minimization*

One common method to refine a macromolecular structure with respect to X-ray data is to use a computer program developed to minimize the potential energy within a molecular model, and to add a weighted term to minimize $|F_o - F_c|$. To derive model A (Marvin *et al.*, 1992), we used the Jack–Levitt method (Jack & Levitt, 1978; Deisenhofer, Remington & Steigemann, 1985) based on the energy-minimization program *EREF*; to refine other models we used the *X-PLOR* package (Brünger, 1992), based on the energy-minimization program *CHARMm*. We used the energy parameters of Levitt (1983) with *EREF*, but the energy parameters of Engh & Huber (1991) with *X-PLOR*. Charges were set to zero on the charged side chains. With the Jack–Levitt method, we used a difference electron-density map calculated in helix symmetry, with constant atomic temperature factors $B = 10\,Å^2$ on all atoms. With *X-PLOR* we refined 'group' temperature factors, where all non-H atoms of the main chain (C^α, C, N and O atoms, taken as one group) and side chain (another group) of each residue have the same atomic temperature factor B. It was not necessary to modify *X-PLOR* to calculate the cylindrically averaged transform: at the resolution of the data, no more than one Bessel function term contributes to any layer line, so the transform is already cylindrically symmetrical.

To identify possible errors in the structure, we compared values of torsion angles in the refined structure with values found in surveys of high-resolution protein structures, as described for Ff by Marvin *et al.* (1994). For the α-helix, we used idealized values $\varphi = -65°$, $\psi = -39°$. We use the notation of Ponder & Richards (1987) for side-chain rotamers: − is in the −60° rotamer; + is in the +60° rotamer; t is in the 180° rotamer; 90 is near 90°; and 0 is near 0°. For most side chains in proteins, and especially for those in an α-helix, the values of χ^1 are in the − rotamer, and for most side chains we use the idealized value $\chi^1 = -67°$ (Morris *et al.*, 1992). For Val, the t rotamer is more frequent, but by the IUPAC–IUB conventions, this implies that the second C^γ is in the − rotamer, so this conformation is equivalent to Ile with $\chi^1 = -60°$. Therefore, for Val we used the idealized value $\chi^1 = 173° = -67° + 240°$. To refine the α-helix regions we constrained (φ, ψ) angles of all residues including Gly, and all χ^1 except the χ^1 of the bulky side chains Met, Tyr, Lys and Arg. The constraint function that we used gives no energy penalty within ±15° of the idealized values, but an increasing penalty outside this range.

Non-bonded interchain contacts between symmetry-related subunits in Pf1 are as close as the intrachain contacts found in many globular proteins, and provide important constraints on the model. From the cylindrical-polar coordinates (r_j, φ_j, z_j) of the atoms in an arbitrary subunit of the virion, which we give index $k = 0$, we generate the coordinates of subunit k by adding $13k/71$ turn to φ_j and $3.05k$ Å to z_j. The neighbours in non-bonded contact with subunit $k = 0$ are $k = \pm1$, ±5, ±6, ±11 and ±17, with ±6 and ±11 the nearest neighbours. The contacts between units 0 and k are identical to the contacts between units $-k$ and 0. We modified *EREF* to include these non-bonded contacts in the calculation of

van der Waals energy; with *X-PLOR* these contacts are included within the set of subunits in one pseudo unit cell of $c = 216.5$ Å ($0 \leq k < 71$).

2.3. *Refinement by simulated annealing*

Computer programs to simulate macromolecular dynamics can be used to escape from local minima when refining a molecular structure (Brünger, 1991). We used the simulated-annealing option of the *X-PLOR* package (Brünger, 1992). Our annealing schedule included cycles of molecular dynamics at a high 'temperature', interspersed with cycles of energy minimization to optimize the stereochemistry and the fit to the X-ray data. We tested various annealing schedules, each preceded and followed by energy minimization, including quenching, quick cooling and slow cooling. We chose a slow-cooling schedule, starting at 4000 K, decreasing by intervals of 500 to 500 K, and finishing at 300 K.

2.4. *Calculation of electron-density maps*

We calculated electron density for a single helix asymmetric unit in cylindrical polar coordinates using the program *DIFF* (Nave *et al.*, 1981). In the refinement with *EREF*, we used the complete solvent-corrected transform of the model, including cylindrically symmetrical density corresponding to the DNA core. For our refinement using only data beyond 12 Å resolution, we did not include solvent or DNA when calculating F_c.

Refinement that uses models to supply phases is biased towards the model, and although the refinement may converge, it is possible that the bias may remain. We tried various methods to calculate electron density from observed data and model phases, including σ-weighted maps (Read, 1986), but we found the $2F_o - F_c$ map to be just as useful. This is the appropriate map for fibre diffraction data with no overlapping Bessel function terms (Namba & Stubbs, 1987).

There can even be bias in omit maps, which use observed amplitudes with phases from a partial model to calculate the map. These phases differ from those of the full model, but might be close enough to regenerate the original map even if the omitted portion of the original model were incorrect. To calculate omit maps, we used molecular dynamics and energy minimization to refine the partial model with respect to the X-ray data. The phases thus take up a new set of values that should not be directly related to those of the original model. To prevent distortions in the stereochemistry during refinement of the partial model, we used the full original model but set the occupancy of the omitted atoms to zero. The 'omitted' atoms thus contribute to calculations of potential energy, but not to calculations of diffracted amplitudes. Brünger (1992) describes a slightly different method to do this. The r.m.s. coordinate deviation between the original and the refined partial models is typically about 0.7 Å for models with four residues deleted.

We also used the maximum-entropy method of Bryan (1987) to calculate electron-density maps, as described by Marvin *et al.* (1987). In this method, a starting map (which may be flat or may represent a model) is iteratively refined until its transform fits the observed data to within predetermined limits. The map may be constrained to be as close as possible to a 'prior' map, which may be identical to the starting map. We used this method to calculate omit maps, where the starting and prior maps were calculated from a model with a few residues deleted. As the maximum-entropy algorithm generates the full map, the omitted portion and the retained portion of the map are both free to change, which should reduce the bias due to the retained portion of the map.

We used the maximum-entropy method to determine the positions of I atoms with respect to the model as described by Marvin *et al.* (1987). We calculated a maximum-entropy map by refinement against native data and then used this as the starting and prior maps in refinement against derivative data, to give a new map that differed from the start map by only enough to fit the differences between native and derivative data. The differences between these two maps give the electron density attributable to iodine.

3. Results

3.1. *The simulated-annealing model*

The starting point of our structure analysis was model A in Table 1. This model was derived by building a fully α-helix model into the model-independent electron-density map of Marvin *et al.* (1987), and is described in detail by Marvin *et al.* (1992). A first round of simulated annealing led to a model which differs from model A by a deviation of C^α positions of 0.3 to 1.0 Å for residues 7 to 43, but with much larger deviations (1.0 to 2.5 Å) at the two ends. The (φ, ψ) torsion angles remained within the broadly defined α-helix range for all residues except 2 and 41. Using this simulated-annealing model to supply phases, we calculated a $2F_o - F_c$ map. The N-terminal region of the map was poorly defined, with no evidence for a helix, and we rebuilt residues 1 to 4 to fit the density as well as possible and to bury the Ile3 side chain. We repeated the cycle of simulated annealing followed by rebuilding based on the $2F_o - F_c$ and omit maps. The group temperature factors were refined for the final model. Some properties of this final simulated-annealing model (model SA) are given in Table 1.

The stereochemical properties of model SA are generally better than expected for single-crystal protein structures solved at 3.3 Å resolution, and indeed are as good as structures solved at 2.0 to 2.5 Å resolution, using the criteria of Morris *et al.* (1992), but the intrasubunit

Table 1. *Comparison of model properties*

Model	A	AR	SA	SAR
Mean φ (°)*	−63±10	−63±11	−61±17	−65±15
Mean ψ (°)*	−41±11	−41±13	−44±18	−41±17
Mean ω (°)	179.4±5.9	179.6±1.6	179.2±1.4	179.5±1.2
Bond length r.m.s. (Å)	0.018	0.019	0.012	0.017
Bond angle r.m.s. (°)	2.2	2.2	2.4	2.3
Torsion angle r.m.s. (°)	16	18	22	19
Potential energy [kcal(mol asymmetric unit)$^{-1}$]†				
Bond length	178	123	57	82
Bond angle	415	219	261	176
Torsion angle	388	161	194	159
van der Waals				
Intraunit	−105	−26	78	16
Interunit	−40	−30	−29	−23
Electrostatic				
Intraunit	−1103	−1114	−1114	−1087
Interunit	−7	−12	−1	−5
Total	−274	−679	−554	−682

* The mean α-helix torsion angles φ, ψ were calculated for all residues of models A and AR, but only for residues 8–38 of model SA and residues 6–45 of model SAR.

† Calculated according to Brünger (1992) (1 kcal mol^{-1} = 4.184 kJ mol^{-1}.

van der Waals energy is considerably higher than for model A, suggesting some close non-bonded contacts (Table 1). To improve the stereochemistry we rebuilt distorted regions of model SA to be more like the corresponding regions of model A, and refined this model while constraining the (φ, ψ) torsion angles and the χ^1 torsion angles of residues 12 to 45 to be near the idealized α-helix values. In this simulated-annealing-refined model (model SAR), no backbone bond or angle is more than $3\frac{1}{2}$ standard deviations from the ideal, and only two (in the disordered residues 1 and 2) are more than 2 standard deviations from the ideal. There are no close non-bonded contacts. All main-chain torsion angles are in allowed regions of the Ramachandran diagram, and residues 6 to 45 are in the α-helix region.

The $2F_o - F_c$ map based on this model is shown in Fig. 1. Continuous helix density fits the main chain from residues 6 to 45 at a contour level of 0.65 e Å^{-3}, and some side-chain density is visible. The map is less well defined near the two ends. The first few residues at the N

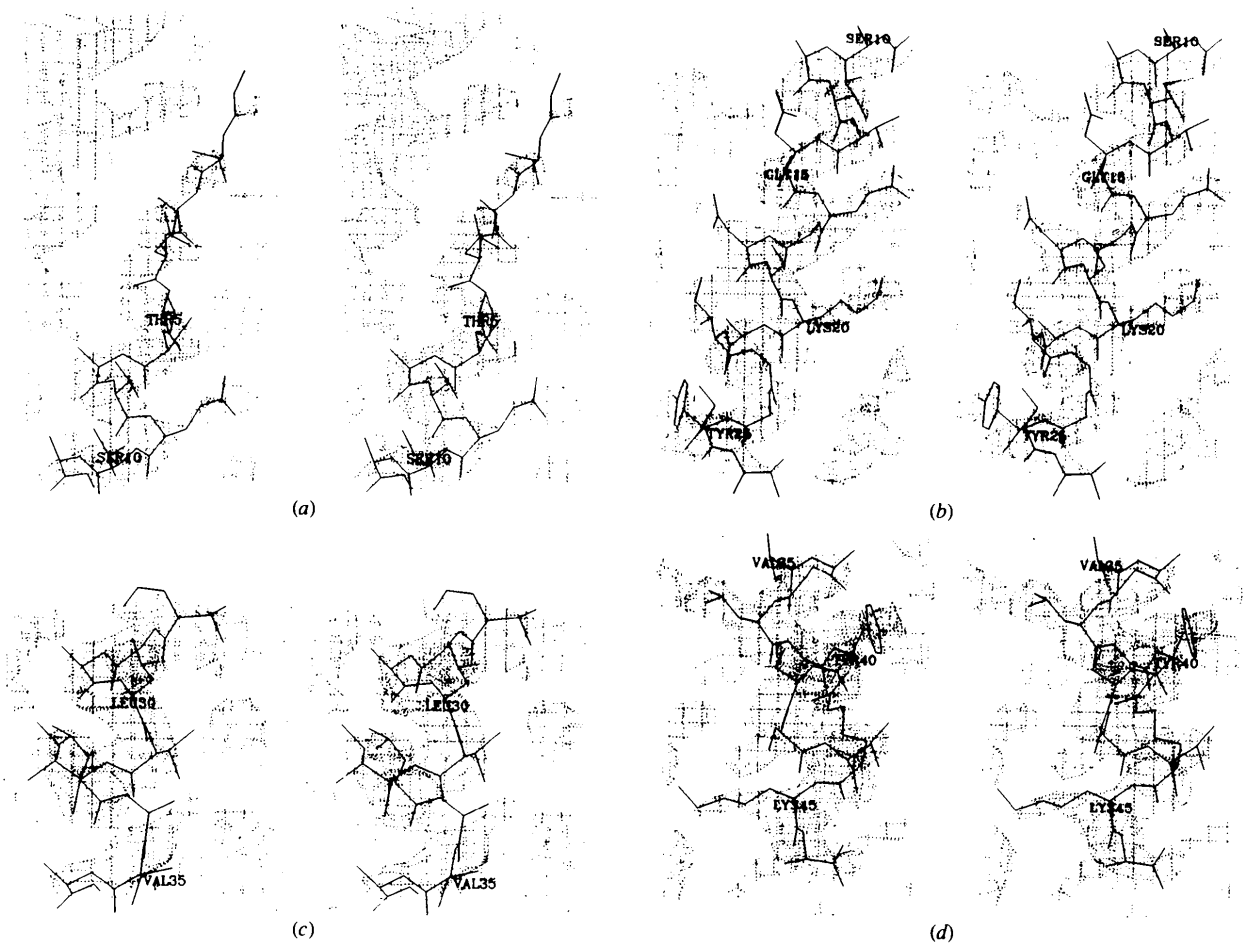

Fig. 1. $2F_o - F_c$ map of model SAR. In all views the virion axis is vertical, and the view is parallel to the radius connecting the C^α of Tyr25 to the virion axis. The four views show successive overlapping segments of the map corresponding to a single subunit of the virion, starting from the top (N terminus), with the corresponding portion of model SAR superimposed on the map. (*a*) to (*c*) are viewed from outside the virion towards the centre. (*d*) is viewed from inside the virion towards the outside. Electron density is contoured at 0.65 e Å^{-3}. (Stereo pairs.)

terminus are not α-helix, and may be flexible. Neither DNA nor possible bound water molecules were included in the model used for phasing, and this may affect the density distribution at the C terminus. The structure of the DNA in the virion is not known, but the DNA cannot have the same symmetry as the protein, because the circular topology of the DNA means that it must have an 'up' strand and a 'down' strand, whereas the protein does not. Two types of model have been proposed for Pf1 DNA. In one, the bases are near the outer DNA radius and the phosphates are near the centre (Day, Marzec, Reisberg & Casadevall, 1988). In the other, bases are near the centre, intercalated between bases on the opposite strand, and the phosphates are near the outer radius of the DNA (Marvin *et al.*, 1992). We have not defined a specific model for the DNA, but there is electron density near positively charged side chains in the N-terminal region of the protein, consistent with the DNA phosphates at the outer radius of the DNA.

We calculated a set of simulated-annealing omit maps having successive groups of four residues deleted along the subunit. There was no indication of alternative chain tracing. As an example, the omit map with residues 16–19 deleted is shown in Fig. 2. This can be compared with Fig. 1(*b*), which is contoured at a higher level. Density of the omitted sequence was regenerated for both the main-chain helix and the side chains of Gln16 and Asp18. Analogous full and omit maps were also calculated using the maximum-entropy method, and showed similar features.

The *R* factor and CC of the models are shown in Table 2. After group temperature-factor refinement, CC = 0.92 for model SAR. The model is illustrated in Figs. 3, 4 and 5. Accessibility of some side-chain atoms is shown in Table 3, to compare with chemical and spectroscopic data. Side chains found experimentally to be accessible to non-disruptive reagents are accessible in the model,

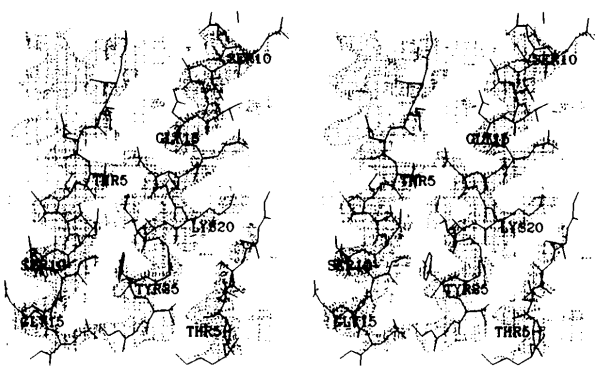

Fig. 2. $2F_o - F_c$ omit map of model SAR. Residues 16–19 were omitted from the model used to supply phases for calculating the map. The omit model was refined to reduce bias as described in §2.4. View approximately as Fig. 1(*b*), but broader to include not only the $k = 0$ unit (centre) but also the $k = -6$ unit (left) and the $k = -11$ unit (right). Electron density is contoured at 0.5 e Å^{-3}. (Stereo pair.)

Table 2. *R factor and correlation coefficient*

Calculated for data points between 12 and 2.6 Å resolution, first with $B = 15$ Å^2 on all atoms, and then after the group temperature factors were refined with fixed atomic positions. Each value was calculated for data with an amplitude $\geq 1\sigma$ (the threshold), and therefore includes the data points measured as zero at *R* values for which diffracted intensity is consistent with the helix symmetry (see §2.1). The calculation was repeated for data points $\geq 3\sigma$ (about 3/4 of the number of data points $\geq 1\sigma$).

	Model			
	A	AR	SA	SAR
	R factor			
$B = 15$ Å^2				
$\geq 1\sigma$	0.45	0.37	0.35	0.37
$\geq 3\sigma$	0.36	0.30	0.27	0.29
Group *B*				
$\geq 1\sigma$	0.38	0.32	0.29	0.33
$\geq 3\sigma$	0.32	0.26	0.23	0.26
	Correlation coefficient			
$B = 15$ Å^2				
$\geq 1\sigma$	0.84	0.90	0.91	0.89
$\geq 3\sigma$	0.85	0.90	0.92	0.90
Group *B*				
$\geq 1\sigma$	0.89	0.92	0.94	0.92
$\geq 3\sigma$	0.90	0.93	0.94	0.93

Table 3. *Accessibility in some side chains of model SAR*

			Accessible area (Å^2)‡	
Residue	Atom*	Ref. area (Å^2)†	Pf1^L	Pf1^H
Gly1	N terminus	44	48	46
Asp4	$O^{\delta 2}$	41	25	25
Glu9	$O^{\epsilon 2}$	42	25	23
Asp14	$O^{\delta 2}$	41	27	26
Asp18	$O^{\delta 2}$	41	21	18
Met19	S^{δ}	43	0	0
Lys20	N^{ζ}	41	8	7
Tyr25	O^{η}	40	24	23
Tyr25	$C^{\epsilon 1}$	35	8	8
Tyr25	$C^{\epsilon 2}$	35	31	30
Tyr40	O^{η}	40	21	20
Met42	S^{δ}	43	0	1
Arg44	$N^{\eta 1}$	46	35	34
Arg44	$N^{\eta 2}$	52	59	59
Lys45	N^{ζ}	41	20	16
Ala46	C-terminal O	42	13	11

* The potentially interactive side-chain atom in the residue. Residues from Tyr40 onwards are accessible only to the inner DNA core.

† Defined after Eisenberg & McLachlan (1986) as the accessible area of the atom in a completely unfolded polypeptide chain.

‡ Pf1^L is the low-temperature form, model SAR. Pf1^H is the high-temperature form, model SAR slewed into the room-temperature symmetry as described by Marvin *et al.* (1992).

and potentially reactive side chains which are not accessible to reagents are generally not accessible in the model (Nakashima & Konigsberg, 1980). Zimmermann, Hagedorn, Heuck, Hinrichsen & Ludwig (1986) used charged small-molecule probes with radius about 1.4 Å to measure accessible charges on the surface of Pf1 as a function of pH. The p*K* of charged side chains can be used to calculate a corresponding theoretical curve, which is fitted by the data if the N terminus, Asp4, Glu9, Asp14, Asp18 and Lys20 are assumed to be accessible, as found for model SAR (Table 3). The Lys20 side chain

is exposed on the outer surface of the model, but the calculated accessibility of its N^{ζ} is relatively low (Table 3) because the N^{ζ} is tucked behind the N terminus of a neighbouring subunit (Fig. 5). If explicit water molecules had been included in the refinement, the N^{ζ} site in our model could be replaced by a water molecule, and the side chain moved to a more accessible position, without changing the ability of Lys20 to crosslink to the N terminus of a neighbouring subunit (see below). The model is consistent with accessibility of the Lys20 N^{ζ} to solvent.

The planes of both Tyr25 and Tyr40 are roughly parallel to the virion axis, consistent with polarized Fourier-transform infrared spectroscopy of oriented films of the virion (Azpiazu, Gomez-Fernandez & Chapman, 1993). Tyr25 is accessible but Tyr40 faces the inner core of the model, consistent with iodination of Tyr25 but not Tyr40 (Nave *et al.*, 1981). The positions of iodines determined from a maximum-entropy difference map are shown in Fig. 3. Considering that the derivative data are less accurate than the native data, and the derivative structure is not perfectly isomorphous with the native structure, these iodine positions are remarkably close to the expected sites on $C^{\varepsilon 1}$ and $C^{\varepsilon 2}$ of Tyr25.

Figs. 4 and 5 illustrate the close packing between the subunit and its nearest neighbours. The N terminus is 3.6 Å from the N^{ζ} of Lys20 on the $k = 11$ subunit, consistent with the ability of cross-linking reagents to join the Lys20 of one subunit with the N terminus of a neighbour (Nave *et al.*, 1981). Lys45 is close enough to the C terminus of the $k = 1$ subunit so that it could contribute to neutralizing the negative charge on the α-

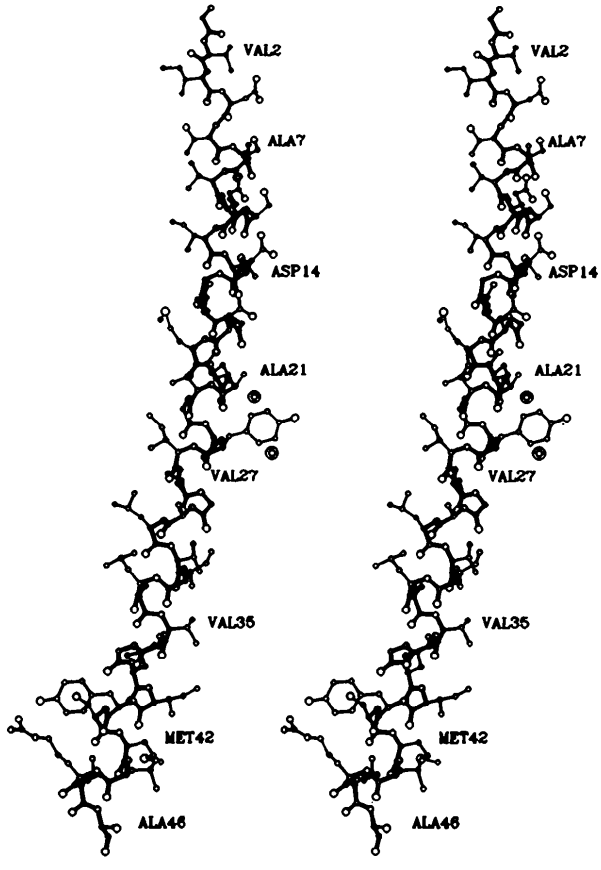

Fig. 3. Subunit of model SAR. Virion axis is vertical, with the N terminus of the subunit towards the top. Heavy lines join main-chain atoms, and light lines join side-chain atoms. The concentric circles near Tyr25 indicate the positions of the iodine atoms selected from the difference map calculated using the iodine derivative data. Atom types are coded by circles whose size increases with atomic number: C < N < O < S. Tangential view, perpendicular to the view of Fig. 1, looking from smaller to larger φ away from the viewer, so the inside of the virion is to the left and the outside to the right. (Stereo pair.)

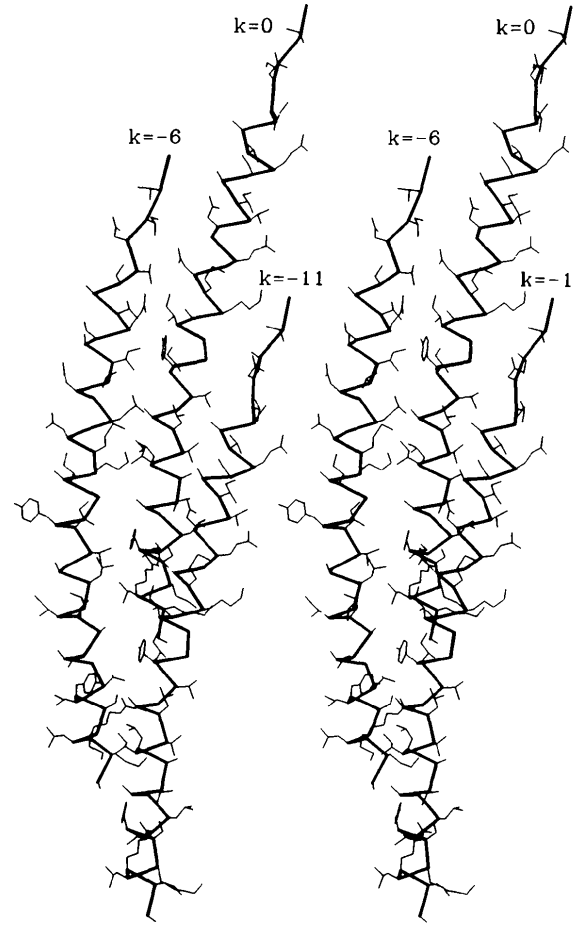

Fig. 4. Nearest-neighbour interactions between subunits of model SAR. The virion axis is vertical, and the view is from outside the virion, parallel to the radius connecting the C^{α} of Tyr25 on the $k = 0$ unit to the virion axis. Heavy lines join C^{α} atoms and light lines join side-chain atoms. Three subunits are shown. The central (highest) subunit has index $k = 0$; to the left is the neighbour with $k = -6$; to the right is the neighbour with $k = -11$. (Stereo pair.)

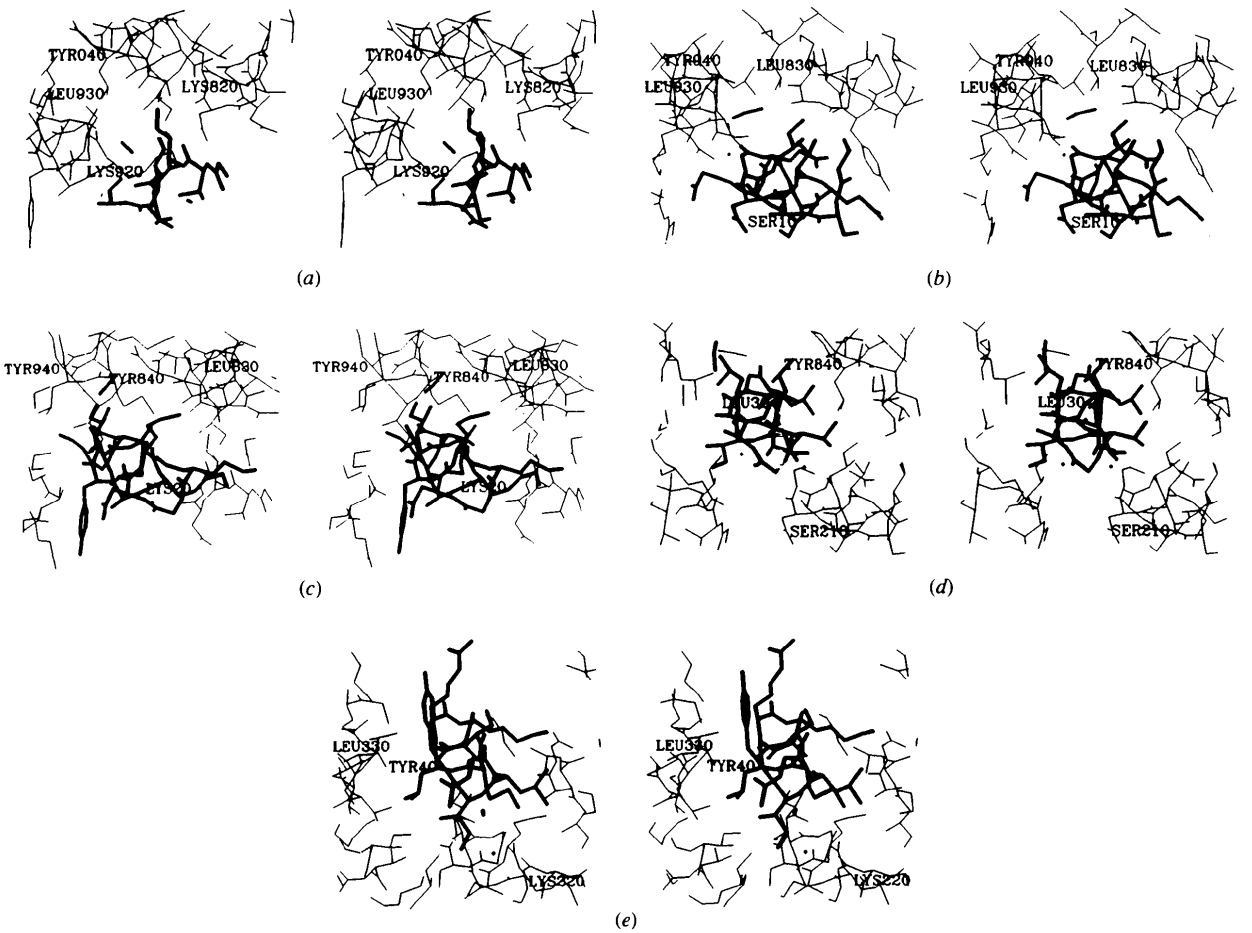

Fig. 5. Subunit of model SAR (heavy lines), shown together with parts of symmetry-related neighbouring subunits (light lines). Series of diagrams with the view direction parallel to the virion axis. Each view direction is from larger towards smaller z and from smaller towards larger residue number away from the viewer; the series, (*a*) to (*e*), moves towards smaller z. The inside of the virion is towards the top of the figure and the outside towards the bottom. Amino acids are coded as follows. RESn is near the C^{α} of residue n in the reference asymmetric unit. RESmn is near the C^{α} of residue n in symmetry-related neighbour with index m, where m is 1 for the $k = -17$ subunit; 2 for -11; 3 for -6; 4 for -5; 5 for -1; 6 for +1; 7 for +5; 8 for +6; 9 for +11; and 0 for +17. (Stereo pairs.)

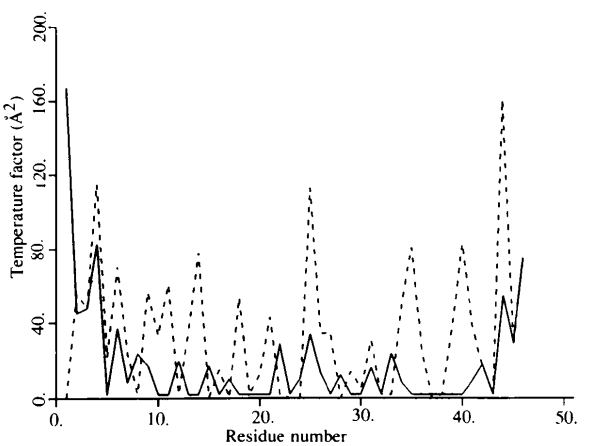

Fig. 6. Group temperature factors B (Å^2) as a function of residue number in model SAR. Solid curve, main-chain B; broken curve, side-chain B.

helix dipole through a water bridge. The O^{ϵ} of Glu9 on the $k = -6$ subunit is 3.4 Å from the O^{η} of Tyr25; with small adjustments of the model, Gln16 on the $k = -6$ subunit might hydrogen-bond to the O^{η} of Tyr25; the N^{η} of Arg44 is 3.3 Å from the O^{η} of Tyr40. These interactions are consistent with spectroscopic evidence (Thomas, Prescott & Day, 1983).

The CC is larger (and the R factor smaller) after refinement of the group temperature factor B. The group temperature factors are large for both side chain and main chain of the first and last few residues of the subunit. Along the rest of the subunit, B is less than 40 Å^2 for main chain, but larger for the longer side chains, in particular for side chains on the outer or inner surface of the virion, such as Tyr25 (Fig. 6). These are physically reasonable results: the two ends of the subunit and the side chains on the surface of the virion are expected to have relatively undefined positions because they are

relatively unconstrained by packing with neighbouring subunits. Regions of the electron-density map (Fig. 1) corresponding to the two ends and to side chains on the surface are relatively undefined, correlating with the regions of the model that have large temperature factors. The large temperature factors may indicate time-averaged mobility (groups of atoms are in motion) or space-averaged disorder (groups of atoms on different subunits are in different positions) or both. The temperature factors for buried regions are lower than those commonly found in single-crystal protein structures, and are comparable to those observed in small-molecule crystallography. A plausible explanation is the absence of rigid-body displacements of the protein in the lattice. Such displacements occur in protein crystals, but are not expected in the tightly packed array of subunits in the virion.

3.2. *Comparison of models*

We illustrate the course of the refinement by comparing several properties of our intermediate models, rather than simply following a single parameter such as the *R* factor. Properties of the models are shown in Table 1. The energy terms of model *A* shown in Table 1 are different from the energy terms of the same model calculated using *EREF* (Marvin *et al.*, 1992), because the functional form and energy parameters used in *EREF* are slightly different from those used in *X-PLOR*. To provide a benchmark for comparison with models SA and SAR, we refined model *A* with respect to the data between 12 and 2.6 Å by molecular dynamics and energy minimization using *X-PLOR*, while maintaining the α-helix constraints on all (φ, ψ) and on χ^1. For this model A-refined (model AR), main-chain torsion angles and hydrogen bonds remained within the α-helix range as found for model A. The covalent and torsion potential energy decreased slightly at the expense of the non-bonded energy, indicating different relative weights of the force constants, and the fit to the X-ray data improved considerably relative to model A (Tables 1 and 2).

Model SA is better than model AR as regards CC, *R* factor and the quality of the $2F_o - F_c$ maps. Refining model SA within the torsion constraints on all (φ, ψ) and χ^1 within the α-helix region improved the stereochemistry of this region (model SAR). The models are compared with respect to the r.m.s. coordinate deviation in Table 4.

Many of the larger side chains are in the same conformation for all models, and most are in favoured rotamers (Table 5). Exceptions are the two Tyr in models A and AR, and Tyr40 in model SAR, which are in the rare t 0 conformation; Tyr40 in model SA, which is in the rare + 90 conformation; and Met42 in model SA, which is in the rare t − conformation.

The subunit in the virion assembly is stabilized with respect to the isolated subunit by a solvation free energy of 33.8 kcal mol^{-1} (141.4 kJ mol^{-1}) for model AR and 37.6 kcal mol^{-1} (157.3 kJ mol^{-1}) for model SAR (Table 6). This is about 0.8 kcal (mol-residue)$^{-1}$ [3.3 kJ (mol-residue)$^{-1}$], which compares with the approximately 1 kcal(mol-residue)$^{-1}$ [4.2 kJ (mol-residue)$^{-1}$] found for two globular proteins with respect to a hypothetical completely unfolded strand (Eisenberg & McLachlan, 1986). The reduction in solvation free energy results from masking apolar C; masking other atoms which favour a polar environment reduces the value slightly. The structure is stabilized by apolar interactions.

We used the method of three-dimensional profile analysis (Bowie, Lüthy & Eisenberg, 1991; Lüthy, Bowie & Eisenberg, 1992) to study residue-by-residue side-chain environment. In this method, the probability of finding a given type of side chain in a given environment (whether exposed or buried; and, if buried, whether in contact with polar or apolar groups) is defined by analysis of known three-dimensional structures. In our models, most side chains in the central apolar domain of the sequence are surrounded by apolar groups in the virion assembly, and apolar side chains nearer the two ends of the subunit are also generally buried. The profile scores are 21 for model AR and 27 for model SAR,

Table 4. *Comparison of model coordinates*

The r.m.s. coordinate deviation (Å) between non-H atoms was calculated as described by McLachlan (1982). Models listed on the left were compared with models listed at the top. The upper number in each triplet is the r.m.s. coordinate deviation over all atoms; the middle number is the deviation over main-chain atoms only; the lower number is the deviation over side-chain atoms only. Models A and AR are compared with each other over all residues; the other comparisons involve residues 6–45 only.

	A	AR	SA
	0.55		
AR	0.40		
	0.70		
	1.56	1.55	
SA	0.82	0.85	
	2.18	2.14	
	1.13	1.08	1.33
SAR	0.53	0.53	0.70
	1.61	1.52	1.85

Table 5. *Conformation of larger side chains*

Side-chain rotamers χ^1 and χ^2 are tabulated using the notation of Ponder & Richards (1987), see §2.2.

	A	AR	SA	SAR
Glu9	− t	− t	− t	− t
Gln16	− t	− t	− t	− t
Met19	− t	− t	− t	− t
Lys20	t t	t t	t t	− −
Tyr25	t 0	t 0	− 90	− 90
Tyr40	t 0	t 0	+ 90	t 0
Met42	t +	− −	t −	t +
Arg44	− t	− t	− −	+ t
Lys45	t t	t t	t t	t t

Table 6. *Solvation free energy*

Calculated as described by Eisenberg & McLachlan (1986), with the modified atomic solvation parameters of Eisenberg, Wesson & Yamashita (1989). The solvation free energy is calculated using accessibility of an isolated asymmetric unit (Single), and using accessibility of an asymmetric unit surrounded by nearest neighbours with index ±1, ±5, ±6, ±11 and ±17 (Virion).

	Net energy ΔG_s (kcal mol^{-1})*			
	AR		SAR	
Type	Single	Virion	Single	Virion
Non-polar (C)	−33.0	−74.6	−33.3	−77.1
Polar (neutral N/O)	11.1	13.6	10.6	12.6
Sulfur	0.2	0.4	0.2	0.4
Charged O^-	1.9	4.0	2.1	3.6
Charged N^+	−0.2	2.7	−0.6	1.8
$\sum \Delta G_s$	−20.1	−53.9	−21.1	−58.7

* 1 kcal mol^{-1} = 4.184 kJ mol^{-1}.

which are in the range expected for proteins of 46 residues (Lüthy *et al.*, 1992).

Profile window plots of the models are illustrated in Fig. 7. The profile score of model SAR in the virion assembly is substantially higher than the score of the isolated subunit along most of its length. This is consistent with the calculation shown in Table 6, that the protein subunit has a lower solvation free energy in the virion than in the monomer. The profile score of the virion assembly along the length of the sequence is generally as high as for the correct model of various small proteins (Lüthy *et al.*, 1992). At the N terminus the profile score is low for models A (not shown) and AR. Rebuilding the N terminus has buried some apolar side chains in model SAR that are accessible in the other models.

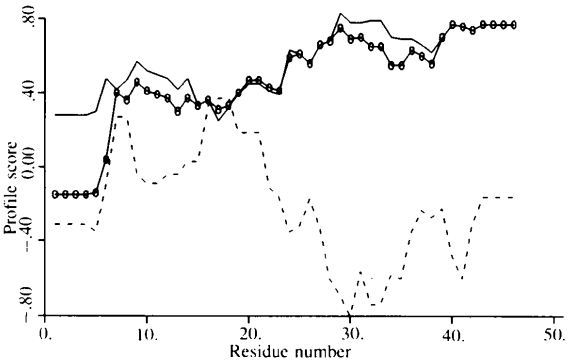

Fig. 7. Profile window plots for models AR and SAR. The vertical axis gives the average score for residues in a seven-residue sliding window, the centre of which is at the residue number indicated by the horizontal axis. Scores for the first three and the final three residues therefore have no meaning. Profile for the isolated protein subunit of model SAR (broken curve); for the subunit of model SAR surrounded by its neighbours with indices $k = \pm1$, ±5, ±6, ±11, and ±17 (solid curve) and for the subunit of model AR surrounded by its nearest neighbours (solid curve with symbol ○). Secondary-structure class was defined as α-helix throughout.

Fig. 8 shows the complete Fourier–Bessel transform of model SAR, with correction for solvent and inclusion of DNA. The good fit of calculated to observed over all data confirms the validity of using only high-resolution data in the refinement.

An alternative explanation for the small temperature factors discussed in §3.1 is that they are an artifact of solvent. Since solvent is not included in the refinement calculation, low-resolution calculated structure factors are too large relative to high-resolution structure factors. This effect might be counterbalanced during refinement by generating artificially small temperature factors. However, most of the region where solvent contribution is large (*i.e.* below 12 Å resolution) was not included in the refinement. Since the calculated transform of the final model with solvent included (Fig. 8) fits the observed data, any distortion of temperature factors due to lack of solvent correction is probably small.

Increasing the temperature above 283 K changes the symmetry of the Pf1 virion and, therefore, the shape of the subunit from $Pf1^L$ to $Pf1^H$. Any valid model of $Pf1^L$ must be consistent with the small structural changes that take place on conversion to $Pf1^H$. We slewed model SAR into the $Pf1^H$ symmetry and corrected local distortions in the stereochemistry by a few cycles of *EREF* energy minimization, from +159 to −261 kcal mol^{-1} (+665 to −1092 kJ mol^{-1}), as described for model A by Marvin *et al.* (1992). The energy-minimized model differs from the initial slewed model by 0.1 Å r.m.s. deviation over all non-H atoms. No distortions in model SAR are introduced by the temperature-induced structural transition, and accessibility of side chains is virtually the same for $Pf1^L$ and $Pf1^H$ (Table 3). The transform of the $Pf1^H$ model fits the room-temperature X-ray data of Marvin *et al.* (1992) with a non-equatorial CC = 0.81.

3.3. *The loop model*

We tested loop models of the type proposed by Nambudripad *et al.* (1991), making no assumptions about the structure in the loop region. An omit map calculated from model SAR with residues 16–19 deleted (Fig. 2) shows density in the deleted α-helix region, and not in the putative loop region. We also examined a model in which residues 1–12 were built to occupy the positions of residues 8–19 of model A. Residues 13–19 were omitted during refinement and calculation of the electron-density map (Fig. 9). This omit map shows density extending beyond the N terminus of the model used to supply phases, and this density is consistent with model SAR (see Fig. 2) but not with a loop model. Density ascribed to the putative loop of unit $k = 0$ could extend either towards the N terminus of unit $k = -11$ (as shown in Fig. 11 of Nambudripad *et al.*, 1991) or towards the N terminus of unit $k = -6$. There is no suggestion of density extending in either direction in Fig. 9.

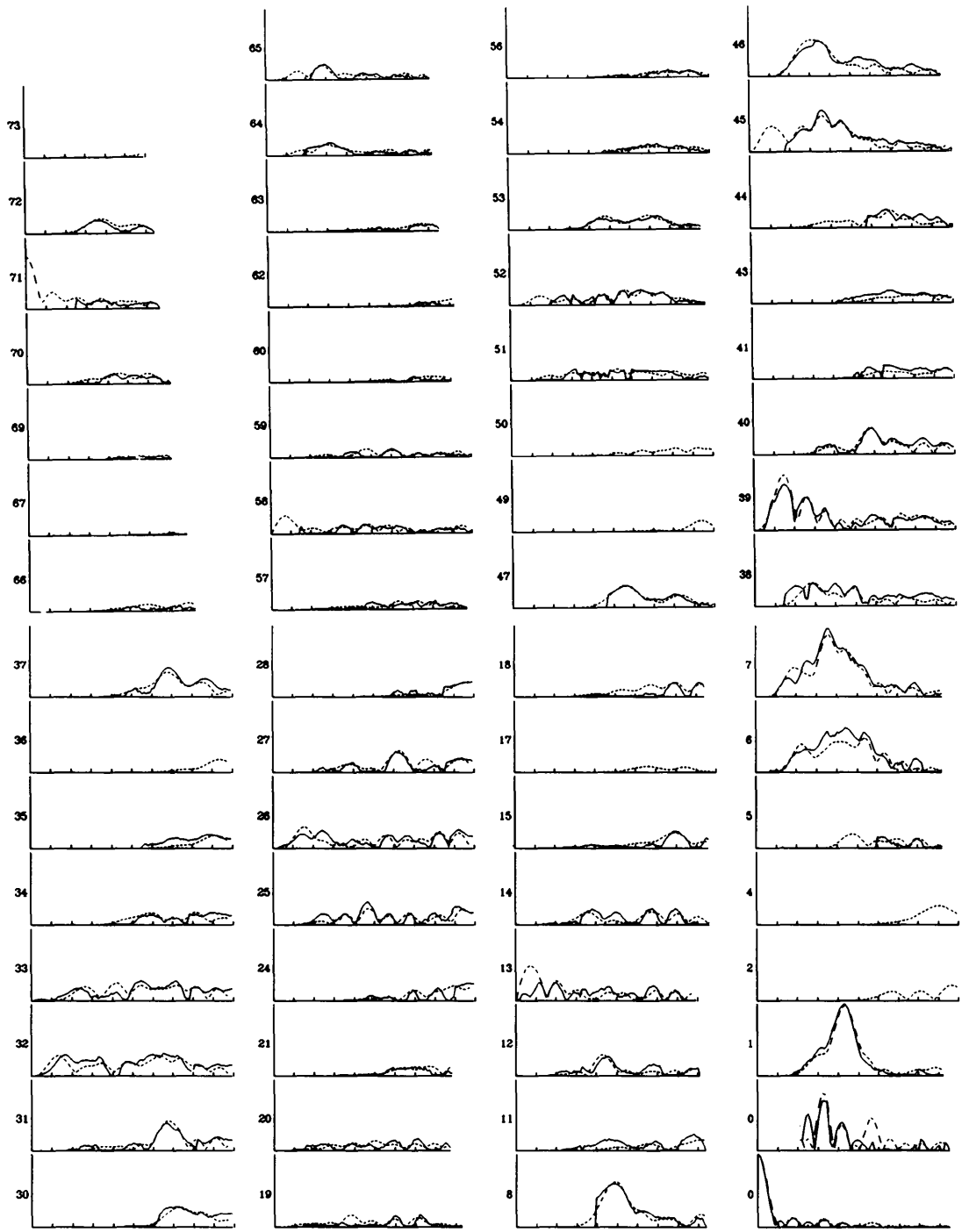

Fig. 8. Fourier–Bessel transform of model SAR relative to solvent, compared with the X-ray fibre diffraction data. Data correspond to one quadrant of the fibre-diffraction pattern, with the meridian (fibre axis direction) vertical and the equator horizontal. Each curve represents one layer line of index l, shown to the left of the curve. Intensity is not observed for Bessel function orders $n > 30$, so the corresponding layer lines are not shown. Scale divisions along R are at intervals of 0.025 Å^{-1} from $R = 0$ Å^{-1} at the left-hand side of each curve. Amplitudes $|I_l(R)|^{\frac{1}{2}}$ (continuous curves) for observed data are compared with the calculated amplitudes $|G_{n,l}(R)|$ (broken curves). The two curves for $l = 0$ show the same data on two different scales to enable the origin peak to be plotted; top, at the same scale as the rest of the data; bottom, at a scale 0.15 times that of the rest of the data. The transform of the model relative to solvent is the sum of three transforms: the transform relative to vacuum of all atoms in the protein subunit, including H atoms; the negative transform of uniform solvent density filling the molecular envelope; and the cylindrically symmetrical transform (that is, $l = 0$ and $n = 0$ only) of the DNA model described by Marvin *et al.* (1992). The non-equatorial CC for the full diffraction pattern is 0.86.

4. Discussion

The major coat proteins of Ff and Pf1 are simple models for monotonic membrane proteins. They reside in the membrane during virion assembly, but they then come out of the membrane into a form more easily studied, the virion. The model of the Ff subunit described by Marvin, Hale, Nave & Helmer Citterich (1994) is essentially a continuous α-helix, like our Pf1 model, although the lower resolution and overlapping layer lines of that study precluded identification of possible non-helical or disordered regions at the ends of the subunit. Some experimental and theoretical considerations suggest that Pf1 and Ff coat proteins in lipid-membrane environments are single membrane-spanning α-helices (Azpiazu *et al.*, 1993; Sanders, Haris, Chapman, Otto & Hemminga, 1993; Turner & Weiner, 1993); other studies suggest that a short N-terminal α-helix segment lies in the plane of the membrane (Shon, Kim, Colnago & Opella, 1991; McDonnell, Shon, Kim & Opella, 1993). This second model of the protein in the membrane is related to the virion model of Nambudripad *et al.* (1991). Accurate knowledge of the subunit structure is necessary for accurate models of virion assembly through membranes.

Model SAR fits the diffraction data as well as model AR (Table 2), but this alone might not be convincing. Also important is the environmental profile analysis (Fig. 7), which clearly favours model SAR in preference to model AR. The N terminus of the subunit is not helical. The N terminus might be a specific conformation that is required for the enzyme that cleaves the leader sequence off the protein subunit when it is processed in the bacterial membrane, or it may instead be relatively unspecific and flexible.

There is an interesting effect around residues 16–18: the environment profile is higher for the isolated monomer than for the subunit in the assembly (Fig. 7). That is, not only are there no exposed apolar side chains in this region of the monomer, but also the exposed polar side chains are not buried by apolar groups after surrounding the monomer with its neighbours. A similar effect in this region is also seen for model A, and for our model of Ff (Marvin *et al.*, 1994). Inspection of the model (Figs. 4 and 5) shows that in this region, side chains are either absent (Gly15 and Gly17) or are polar side chains oriented towards the outside surface of the virion (Glu16 and Asp18). There is a second interesting effect around residues 21–25, where there is a slight discontinuity in the α-helix hydrogen-bonding pattern (Fig. 3). This discontinuity is more marked for model SA, which was refined without constraints on the torsion angles. These effects occur in a region of low α-helix propensity (Turner & Weiner, 1993) that may have biological significance, for instance as the site of a bend in the subunit when it is first inserted into the membrane, before the leader sequence is removed (Model & Russel, 1988; Wickner, 1988).

Coordinates have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977) under entries 1IFM (model A, which is the same as model A of Marvin *et al.*, 1992); 2IFM (model AR); 3IFM (model SA); 4IFM (model SAR); and 2IFN (the $Pf1^H$ model corresponding to model SAR). The $Pf1^L$ diffraction data have also been deposited in the Protein Data Bank.*

We are indebted to the staff of the Medical Research Council Laboratory of Molecular Biology, Cambridge for advice and discussion and for generous access to computing and computer graphics facilities; to Professor R. N. Perham for support; and to the Leverhulme Trust and the Science and Engineering Research Council for research funding.

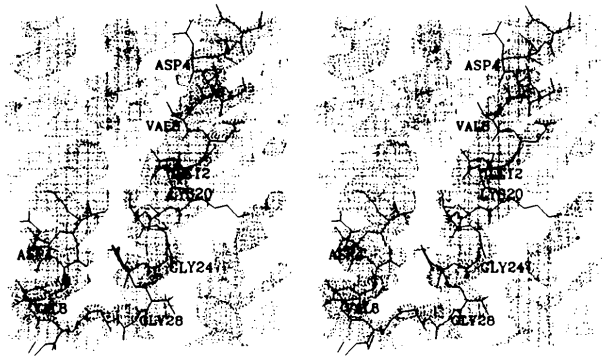

Fig. 9. $2F_o - F_c$ omit loop map. Residues 13–19 were omitted from the model used to supply phases for calculating the map; their positions in the α-helix are occupied by residues 1 to 12, as described in §3.3. View as Fig. 2, with the omit loop model superimposed on the map. Electron density is contoured at 0.45 e $Å^{-3}$. Stereo pair.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1IFM, 2IFM, 3IFM, 4IFM, 2IFN & 4IFMSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AD011). At the request of the authors, the atomic coordinates will remain privileged until 1 January 1996 and the structure factors will remain privileged until 1 January 1996.

References

AZPIAZU, I., GOMEZ-FERNANDEZ, J. C. & CHAPMAN, D. (1993). *Biochemistry*, **32**, 10720–10726.

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). *J. Mol. Biol.* **112**, 535–542.

BOWIE, J. U., LÜTHY, R. & EISENBERG, D. (1991). *Science*, **253**, 164–170.

BRACEWELL, R. (1965). *The Fourier Transform and its Applications.* New York: McGraw-Hill.

BRÜNGER, A. T. (1991). *Annu. Rev. Phys. Chem.* **42**, 197–223.

BRÜNGER, A. T. (1992). *X-PLOR. Version* 3.1. New Haven: Yale Univ. Press.

BRYAN, R. K. (1987). *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, edited by C. R. SMITH & G. J. ERICKSON, pp. 207–228. Dordrecht: Reidel.

BRYAN, R. K., BANSAL, M., FOLKHARD, W., NAVE, C. & MARVIN, D. A. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4728–4731.
CESARENI, G. (1992). *FEBS Lett.* **307**, 66–70.
DAY, L. A., MARZEC, C. J., REISBERG, S. A. & CASADEVALL, A. (1988). *Annu. Rev. Biophys. Biophys. Chem.* **17**, 509–539.
DEISENHOFER, J., REMINGTON, S. J. & STEIGEMANN, W. (1985). *Methods Enzymol.* **115**, 303–323.
EISENBERG, D. & MCLACHLAN, A. D. (1986). *Nature (London)*, **319**, 199–203.
EISENBERG, D., WESSON, M. & YAMASHITA, M. (1989). *Chem. Scr. A*, **29**, 217–221.
ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**47**, 392–400.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.
LEVITT, M. (1983). *J. Mol. Biol.* **168**, 595–620.
LÜTHY, R., BOWIE, J. U. & EISENBERG, D. (1992). *Nature (London)*, **356**, 83–85.
MCDONNELL, P. A., SHON, K., KIM, Y. & OPELLA, S. J. (1993). *J. Mol. Biol.* **233**, 447–463.
MCLACHLAN, A. D. (1982). *Acta Cryst.* A**38**, 871–873.
MARVIN, D. A. (1990). *Int. J. Biol. Macromol.* **12**, 125–138; 335.
MARVIN, D. A., BRYAN, R. K. & NAVE, C. (1987). *J. Mol. Biol.* **193**, 315–343.
MARVIN, D. A., HALE, R. D., NAVE, C. & HELMER CITTERICH, M. (1994). *J. Mol. Biol.* **235**, 260–286.
MARVIN, D. A. & NAVE, C. (1982). *Structural Molecular Biology*, edited by D. B. DAVIES, W. SAENGER & S. S. DANYLUK, pp. 3–44. New York: Plenum Press.
MARVIN, D. A., NAVE, C., BANSAL, M., HALE, R. D. & SALJE, E. K. H. (1992). *Phase Transit.* **39**, 45–80.
MATTHEWS, R. E. F. (1982). *Intervirology*, **17**, 1–200.
MODEL, P. & RUSSEL, M. (1988). *The Bacteriophages*, edited by R. CALENDAR, Vol. 2, pp. 375–456. New York: Plenum. Press.
MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992). *Proteins Struct. Funct. Genet.* **12**, 345–364.
NAKASHIMA, Y. & KONIGSBERG, W. H. (1980). *J. Mol. Biol.* **138**, 493–501.
NAKASHIMA, Y., WISEMAN, R. L., KONIGSBERG, W. & MARVIN, D. A. (1975). *Nature (London)*, **253**, 68–71.
NAMBA, K. & STUBBS, G. (1987). *Acta Cryst.* A**43**, 533–539.
NAMBUDRIPAD, R., STARK, W. & MAKOWSKI, L. (1991). *J. Mol. Biol.* **220**, 359–379.
NAVE, C., BROWN, R. S., FOWLER, A. G., LADNER, J. E., MARVIN, D. A., PROVENCHER, S. W., TSUGITA, A., ARMSTRONG, J. & PERHAM, R. N. (1981). *J. Mol. Biol.* **149**, 675–707.
PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.
RASCHED, I. & OBERER, E. (1986). *Microbiol. Rev.* **50**, 401–427.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
RUSSEL, M. (1991). *Mol. Microbiol.* **5**, 1607–1613.
SANDERS, J. C., HARIS, P. I., CHAPMAN, D., OTTO, C. & HEMMINGA, M. A. (1993). *Biochemistry*, **32**, 12446–12454.
SHON, K.-J., KIM, Y., COLNAGO, L. A. & OPELLA, S. J. (1991). *Science*, **252**, 1303–1305.
SMITH, G. P. (1993). *Gene*, **128**, 1–144.
THOMAS, G. J. JR, PRESCOTT, B. & DAY, L. A. (1983). *J. Mol. Biol.* **165**, 321–356.
TURNER, R. J. & WEINER, J. H. (1993). *Biochim. Biophys. Acta*, **1202**, 161–168.
WICKNER, W. (1988). *Biochemistry*, **27**, 1081–1086.
ZIMMERMANN, K., HAGEDORN, H., HEUCK, C. C., HINRICHSEN, M. & LUDWIG, H. (1986). *J. Biol. Chem.* **261**, 1653–1655.