Acta Cryst. (1992). B48, 499–511

Structure Determination of the Bacteriophage φX174

BY ROBERT MCKENNA, DI XIA, PETER WILLINGMANN,* LEODEVICO L. ILAG AND MICHAEL G. ROSSMANN†

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 26 *November* 1991; *accepted* 28 *January* 1992)

Abstract

The structure of the single-stranded DNA phage φX174 has been determined to 3.4 Å resolution. The crystal space group was $P2_1$ with one icosahedral particle per asymmetric unit, giving 60-fold noncrystallographic redundancy. Oscillation diffraction photographs were collected using synchrotron radiation at various wavelengths. The particle orientations in the unit cell were determined with a rotation function. Because cowpea mosaic virus has a similar external envelope to φX174, it was used as a search model to find the approximate positions of the φX174 particles in the unit cell relative to the crystallographic symmetry axes. An initial phase set to 12 Å resolution was then based on the cowpea mosaic virus atomic structure. These phases were improved by 20 cycles of real-space molecular replacement averaging. The phase information was gradually extended to 3.4 Å resolution by molecular replacement electron density averaging. One reciprocal lattice point was used for each extension followed by four cycles of averaging. The unusual particle capsid, with its 12 pentameric spikes, required the careful determination of a precise molecular envelope. This was redetermined at regular intervals, as was the particle center. The resultant electron density map was readily interpreted in terms of the F, G and J polypeptides in the capsid. A difference electron denisty map between full and partially empty particles showed some ordered DNA structure.

Introduction

The bacteriophage φX174 is an icosahedral ssDNA virus which used *E. coli* as its host. The circular ssDNA contains 5386 nucleotide bases (Sanger, Air, Barrell, Brown, Coulson, Fiddes, Hutchison, Slocombe & Smith, 1977). The genome encodes 11 genes [see Hayashi, Aoyama, Richardson & Hayashi (1988) for a review]. The capsids of the mature phages consist of four proteins J, F, G and H with molecular weights of 4200, 48400, 19050 and 34400 daltons, respectively. There are 60 copies of the proteins J, F and G as well as 12 copies of the protein H in the virion (Burgess, 1969; Edgell, Hutchison & Sinsheimer, 1969; Siden & Hayashi, 1974). The particle has a molecular weight of 6.2×10^6 and a sedimentation coefficient of 114S (Sinsheimer, 1959). Electron microscopy studies (Bayer & Starkey, 1972) suggested that the particles have capsid diameter of around 270 Å with 12 protruding spikes. The spikes consist of proteins G and H, project from the capsid surface (Edgell *et al.*, 1969) and have the ability to recognize the host receptor (Brown, MacKenzie & Bayer, 1971; Incardona & Selvidge, 1973; Jazwinski, Linberg & Kornberg, 1975; Feige & Stirm, 1976).

Monoclinic crystals of the 114S full particles and of the isomorphous 70S partially empty particles were reported earlier (Willingmann, Krishnaswamy, McKenna, Smith, Olson, Rossmann, Stow & Incardona, 1990). Here we report the structure determination of these crystals, the first using a newly developed molecular averaging program (Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi & Lynch, 1992), while the structure and its biological implications have been described earlier (McKenna, Xia, Willingmann, Ilag, Krishnaswamy, Rossmann, Olson, Baker & Incardona, 1992).

Crystal preparation and X-ray diffraction data collection

E. coli were infected at 310 K according to a standard protocol (Incardona, Tuech & Murti, 1985) and purified as described by Willingmann *et al.* (1990). The final purification step was centrifugation of the virus in a 5–30% (*w*/*v*) sucrose gradient for 4 h at 28 000 rev min^{-1} at 277 K in an SW28 rotor. The virus separated into two bands with an A_{260}/A_{280} ratio of 1.55 for the faster (114S particles) and 1.25 for the slower (70S particles). The 70S particles had a heterogeneous population of DNA fragments representing about 20% of the complete genomic sequence (Sinsheimer, 1959; Eigner, Stouthamer, van der Sluys & Cohen, 1963). The optimal crystallization conditions for both types of particles was 1.5–2.0% (*w*/*v*) PEG 8000, in 90 m*M* bis-tris methane buffer at pH 6.8 in the reservoir, over which was suspended a hanging drop of 5μ of virus

* Present address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, England.

† To whom reprint requests should be addressed.

0108-7681/92/040499-13$06.00

Table 1. *X-ray diffraction data collection*

Resolution range (Å)	Number of observations [$F^2 > 2\sigma(F^2)$] for each data set and percentage of possible total: Combined data[a] No. of unique reflections	% of theoretical	70S data[b] No. of unique reflections	% of theoretical	114S data[c] No. of unique reflections	% of theoretical	K_2PtCl_4 data[d] No. of unique reflections	% of theoretical	$KAu(CN)_2$ data[d] No. of unique reflections	% of theoretical	K_2OsCl_6 data[d] No. of unique reflections	% of theoretical
∞ -30.0	940	73	924	72	95	7	321	18	361	28	38	3
30.0-15.0	8133	91	7976	89	1044	12	2502	28	3564	39	447	5
15.0-10.0	22240	92	21794	90	3039	12	6769	28	9884	40	1450	6
10.0-7.5	40725	86	39467	83	5870	12	12786	27	17995	38	2841	6
7.5 5.0	148401	76	145712	75	24130	12	48816	25	68343	35	11715	6
5.0 4.0	193012	73	189722	72	31897	12	60812	23	84608	32	13220	5
4.0-3.5	177528	67	168282	64	28447	11	47694	18	66242	25	7949	3
3.5-3.2	148155	60	142153	58	22990	9	37038	15	54324	22	7407	3
3.2 3.0	115489	51	112079	49	17295	8	24909	11	38496	17	2264	1
3.0-2.9	53573	39	50945	37	4777	3	8242	6	15110	11	1373	1
2.9 2.7	92108	27	86295	25	8754	3	1364	4	27290	8	3411	1
Total observations	1701099		1575376		159289		286452		457717		54174	
Independent reflections	1000304		965349		148338		251253		386217		52115	
No. of films	299		279		20		39		48		9	
R factor[e]	12.9		12.0		12.5		10.8		11.5		9.4	

Notes: (*a*) Refers to data from both 70S and 114S particles. (*b*) Refers to data collected from crystals of 70S particles. (*c*) Refers to data collected from crystals of 114S particles. (*d*) K_2PtCl_4, $KAu(CN)_2$ and K_2OsCl_6 refer to data from crystals soaked in 10 m*M* solutions of these compounds. (*e*) $R = [\sum_h \sum_i |(F_h^2 \quad F_{hi}^2)| / \sum_h \sum_i F_h^2] \times 100$ where F_h^2 is the mean intensity of the i observations F_{hi}.

Table 2. *Variation of cell dimensions and R factors of data collected at various synchrotron sources*

Δa, Δb and Δc are the cell-dimension changes found during post-refinement. λ(used) is the wavelength used during processing and post-refinement. λ(corrected) is λ(used) after correction for the systematic change in cell dimensions found during post-refinement. $R_{sym} = [\sum_h \sum_i (\langle I_h \rangle \cdot I_{hi}) / \sum_h \sum_i \langle I_h \rangle] \times 100$ where $\langle I_h \rangle$ is the mean of the I_{hi} symmetry-related observations for reflection h on a given film.

Synchrotron	No. of films	Δa (Å)	Δb (Å)	Δc (Å)	$\Delta\beta$ ( )	λ(used)	λ(corrected)	R_{sym}
DESY, Sept. 1988	6	0.10	· 0.19	· 0.16	0.01	1.4607	1.4613	12.5
Brookhaven, Nov. 1988	20	+ 0.31	+ 0.23	+ 0.23	0.00	1.2174	1.2183	11.4
Brookhaven, Dec. 1988	3	+ 0.07	− 0.09	· 0.03	· 0.02	1.2176	1.2176	16.9
Daresbury, Mar. 1989	17	+ 0.16	0.01	· 0.01	0.00	0.9110	0.9112	8.4
Brookhaven, Apr. 1989	14	+ 0.43	+ 0.36	+ 0.47	· 0.01	1.2174	1.2190	14.0
Brookhaven, June 1989	111	0.06	− 0.23	0.18	0.00	1.1060	1.1054	10.4
CHESS (A1), Nov. 1989	28	0.31	− 0.40	0.47	0.00	1.5633	1.5614	5.8
CHESS (A1), Jan. 1990	20	− 0.19	0.51	0.36	0.00	1.5633	1.5616	8.9
CHESS (A1), Mar. 1990	12	− 0.23	0.33	− 0.21	0.01	1.5633	1.5620	10.8
CHESS (A1), May 1990	68	+ 0.57	+ 0.56	+ 0.46	0.00	1.5658	1.5683	8.8
	Total = 299	a = 305.58	b = 360.78	c = 299.46	β = 92.89			

solution (40 μg of virus) and 5 μl reservoir solution (Willingmann *et al*, 1990). Crystals up to 0.5 mm in width formed after 4 weeks and were shown to be isomorphous for the two particle types. The monoclinic $P2_1$ cell dimensions were found to be a = 305.6, b = 360.8, c = 299.5 Å and β = 92.89° (Willingmann *et al.*, 1990). The Matthews coefficient, V_M (Mathews, 1968), was 2.7 Å^3 per dalton for two particles per unit cell. Thus, there was one virion per asymmetric unit and the noncrystallographic redundancy was 60-fold.

Data were collected by oscillation photography primarily at the Cornell High-Energy Synchrotron Source (CHESS) at Cornell University and the National Synchrotron Light Source (NSLS) at Brookhaven, although some data were collected at the Deutches Elektronen Synchrotron (DESY) at Hamburg and the Science and Engineering Research Council (SERC) at Daresbury. The crystals were cooled to 277 K and diffracted to at least 2.6 Å resolution in exposures lasting approximately 2 min using a 0.4° oscillation angle. Between 4 and 10 photographs could be taken per crystal, dependent on crystal size and wavelength. No serious attempt was made to align the crystals optically and no 'setting photographs' were taken (Rossmann & Erickson, 1983) in order to avoid radiation damage prior to data collection. Among the 299 films used in the final native data set (Table 1), 32 were collected with a 300 mm crystal-to film distance (together with a helium path) in order to collect good-quality low-resolution data (150–8 Å resolution). For these it was found that a 3° oscillation angle could be used with no spot overlap. These data aided film scaling and provided a nearly complete low-resolution data set.

The films were initially indexed using an auto-indexing procedure (Kim, 1989). These were then processed (Rossmann, 1979), scaled together and post-refined (Winkler, Schutt & Harrison, 1979; Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979) to give 70% of all possible reflections to 3.0 Å resolution. Where there were partial reflections from sequentially abutting films of the same crystal, these were added to determine the full intensity. Reflections which had a partiality greater than 0.6 and for

which there were no abutting film data were also included, after adjustment by their calculated partiality (Table 1). Eventually, data from all X-ray sources, using crystals of both the full 114S and the partially empty 70S particles, were combined for the initial structure determination to maximize the number of observed structure factors (F_{obs}). Since post-refinement is independent of crystal-to-film distance, the relative X-ray wavelength used at each synchrotron could be adjusted to give the same cell dimensions from each source. The absolute cell dimensions were calibrated using human rhinovirus crystals whose cell dimensions had been determined originally from data collected with Cu $K\alpha$ radiation (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985). Table 2 shows the final variation of cell dimensions and R factors obtained in post-refinement of all films collected at different synchrotrons. In general, the sign and proportion of the cell dimensional changes for a particular synchrotron source are much the same for a, b and c, while β stays essentially constant. This is consistent with an assumption of a slightly erroneous wavelength during film processing and shows that the relative error in cell dimension is unlikely to be larger than 6 parts in 10000. The R factor shows that the best data were collected at CHESS and Daresbury.

Particle orientation and position

The orientations of the φX174 virus particles in the unit cell were determined with a self-rotation function (Rossmann & Blow, 1962). The initial calculations used 15–7 Å resolution data from a data set that contained only 20% of the observable structure factors. About 10% of the large terms were used to represent the second Patterson (Tollin & Rossmann, 1966). The radius of integration was set at 140 Å and the interpolation grid around each rotated non-integral reciprocal lattice point was $3 \times 3 \times 3$. The

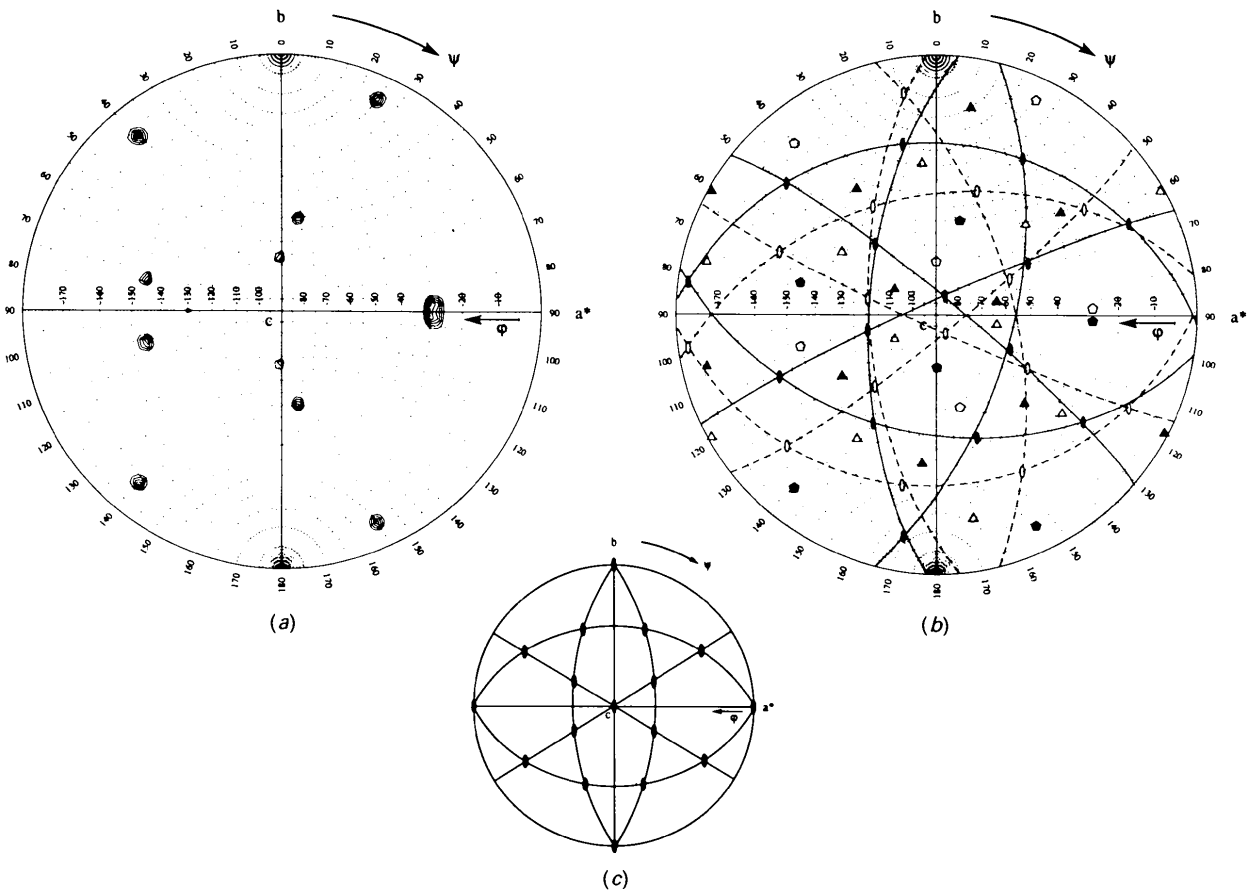

Fig. 1. Stereographic projections showing: (*a*) Rotation function for $\kappa = 72^\circ$ using the complete data set in the 15–7 Å resolution range. (*b*) Interpretation of the complete rotation function. Great circles pass through the twofold axes for each particle. One particle is represented by continuous lines, the other by dashed lines. (*c*) Definition of the standard icosahedral orientation. Only the twofold axes are shown. Three mutually perpendicular twofold axes are parallel to a^*, b and c.

function was explored in spherical polar coordinates in 2° intervals. The search for twofold ($\kappa = 180°$), threefold ($\kappa = 120°$) and fivefold ($\kappa = 72$ and 144°) axes established that the virion possesses icosahedral symmetry. The orientation was later refined using the complete data set (Fig. 1) and also by calculating a locked rotation function (Tong & Rossmann, 1990) at 3.5 Å resolution. The observed positions of the icosahedral symmetry axes had an angular r.m.s. deviation of 0.15° from that of an ideal icosahedron, corresponding to a distance of 0.3 Å at a radius of 130 Å (Rao & Rossmann, 1973).

It is useful to define two types of unit cells in order to give a description of the determination and refinement of the particle position and the determination of a molecular envelope (Rossman *et al.*, 1992). The '*p*-cell' is the unit cell of the unknown crystal structure with the oblique coordinate system **y**, which is orthogonalized to the Cartesian coordinate system **Y** (Rossmann & Blow, 1962); in this case it is the monoclinic $P2_1$ cell of φX174. The '*h*-cell' is the unit cell that contains a single particle in the standard orientation (Fig. 1*c*) and has a Cartesian coordinate system **X**. The matrix [*P*] defines the rotational relationship between the structures within the two cells (Rossman *et al.*, 1992) such that

$$\mathbf{X} = [P]\,\mathbf{Y}.$$

The rotation function showed that

$$[P] = \begin{pmatrix} 0.9899 & -0.1362 & -0.0403 \\ 0.1290 & 0.9809 & -0.1454 \\ 0.0593 & 0.1388 & 0.9885 \end{pmatrix}$$

which corresponds to a rotation of 11.56° about an axis whose direction cosines are (0.709, 0.243, 0.658).

Packing considerations provided a rough position of the virions in the crystal unit cell. The maximum particle separation within the $P2_1$ unit cell occurs when they are placed at $(\frac{1}{4}, y, \frac{1}{4})$ and $(-\frac{1}{4}, \frac{1}{2}+y, -\frac{1}{4})$. This implies that the virus would have a diameter of about 285 Å.

The initial goal had been to phase the crystal data of φX174 using phases obtained from a frozen hydrated electron microscopy reconstructed image (Olson, Baker, Willingmann & Incardona, 1992) (Fig. 2). The low-resolution EM reconstructed image was placed into the *h*-cell and used to calculate structure factors to 25 Å resolution. These were used to calculate a cross-rotation function between the *p*-cell and the electron microscopy image in the *h*-cell. The function showed excellent correspondence between the X-ray and EM data (Fig. 3). The *h*-cell structure factors were then also used to calculate *R*-factor translation searches based on a knowledge of the particle orientations in the *p*-cell (Rossmann, 1972). Several minima, with *R* factors around 50%, could be observed in a ring around $(\frac{1}{4}, y, \frac{1}{4})$ at a distance of 3 to 5 Å (Fig. 4). Structure factors were derived from the EM image with respect to various minima, but they failed to give successful phase extension on using real-space molecular averaging (Rossmann, 1990).

We are grateful for the observation by John E. Johnson that the overall external shape, appearance and size of the EM reconstruction of φX174 were rather similar to those of cowpea mosaic virus (CpMV) (Fig. 2) (Olson *et al.*, 1992). CpMV was oriented in the *h*-cell into the standard orientation, permitting the calculation of structure factors from the known atomic coordinates (Stauffacher, Usha, Harrington, Schmidt, Hosur & Johnson, 1987; Chen, Stauffacher & Johnson, 1990) to 12 Å resolution. In contrast, structure factors derived from φX174 EM data had essentially zero amplitudes beyond 25 Å

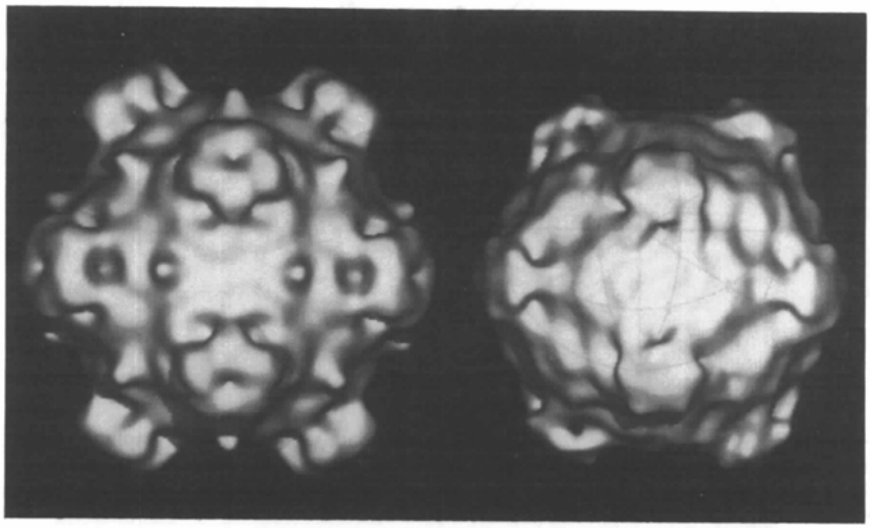

Fig. 2. Shaded surface representations of three-dimensional reconstructions of φX174 (left) and CpMV (right) calculated from particle images recorded in the electron microscope. Unstained virus samples were frozen in a layer of vitreous ice and images were recorded under minimal electron dose conditions ($\sim 20\ e\ Å^{-2}$). 25 (φX174) and 17 (CpMV) particle images were selected, their centers of density (origins) and orientation positions in the layer of ice were determined, and reconstructions were calculated to 21 and 33 Å, respectively [see Olson *et al.* (1992) and Wang, Porta, Chen, Baker & Johnson (1992) for details on procedures]. Surface representations are viewed down the icosahedral twofold axis.

resolution. Using the CpMV structure factors, a search at 0.5 Å intervals with a radius of integration of 180 Å (Argos & Rossmann, 1980) and an interpolation grid of 3 × 3 × 3 gave a slight minimum at (0.2440, y, 0.2480) with an R factor of 57.3% (Fig. 5*a*). This differed by at least 2 Å from any of the previous particle positions which had been tried with the EM data and led to a successful structure solution.

It was subsequently found that the atomic structures of φX174 and CpMV differed greatly. The small protein of CpMV corresponds very roughly to the G protein spikes of φX174, giving both CpMV and φX174 an appearance of having 12 spikes (Fig. 2). However, there is an approximate 11 Å radial shift inwards of the G protein for it to correspond roughly to CpMV when the icosahedral axes of the two viruses are superimposed. The structure of the large protein, consisting of two β-barrels in CpMV (Stauffacher *et al.*, 1987), is very different to the structure of the F protein in φX174 which has only one β-barrel with two large insertions (McKenna *et al.*, 1992).

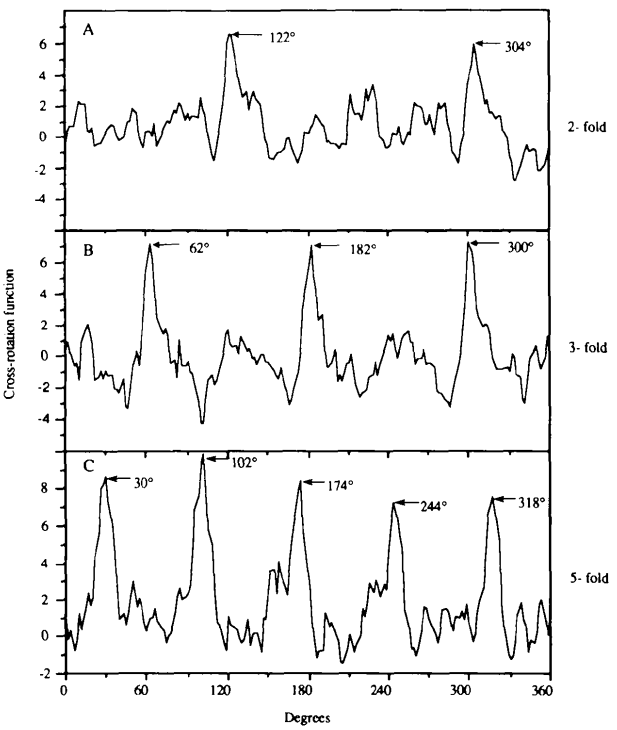

Fig. 3. The cross-rotation function between the structure amplitudes corresponding to the EM image reconstruction and single-crystal data of φX174. A selected two-, three- and fivefold axis of the EM image was superimposed on a corresponding two-, three- and fivefold rotation axis of a particle in the crystal unit cell. The EM image was then rotated in 2° intervals about the common superimposed axes. Data were selected between 50 and 25 Å resolution. The rotation function was calculated from the Fourier amplitudes representing the EM image and the observed X-ray data.

Phase extension by molecular replacement to 5.6 Å resolution

Phases were generated in the 30–12 Å resolution range using the CpMV model given the known orientations and particle centers in the φX174 unit cell. Phase improvement, using 20 cycles of icosahedral averaging and solvent flattening at 12 Å resolution, followed the procedure described by Rossmann (1990) and Rossmann *et al.* (1992) (Fig. 6). Calculated structure factors (F_{calc}) were not used to supplement the 10% missing data of the observed structure factors (F_{obs}), nor were structure amplitudes weighted in calculating electron density maps because of the poor quality of the correlation coefficients. A spherical envelope was defined by an outer radius of 165.0 Å and an inner radius of 80.0 Å. Where there was overlap between particles, a tangential plane was used to separate them. Any density within a radius of 80 Å was set to the mean nucleic acid value; any density outside 165 Å was set to the mean solvent value. The mean correlation coefficient improved from an initial value of 0.06 to 0.54 (Fig. 7).

A further check was then conducted on the particle position. The p-cell electron density was averaged and skewed to generate a single particle in the h-cell. This density was back-transformed to give structure factors for calculating a new R-factor translation search. This yielded a very well defined minimum at 0.2520, y, 0.2500, with an R factor of 42.3% (Fig. 5*b*) and represented a shift of 2.5 Å from the initial starting center found by using the CpMV

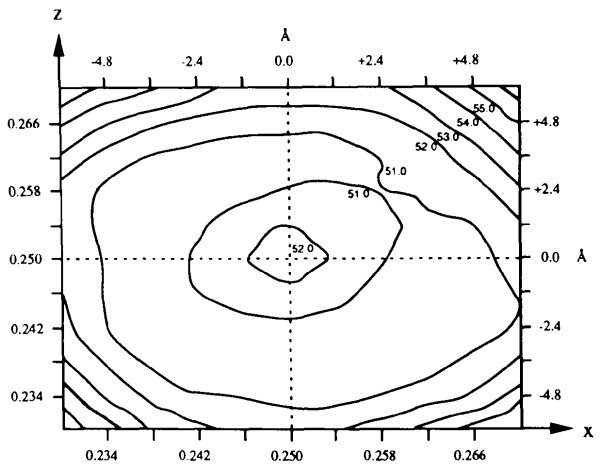

Fig. 4. The R-factor (%) search, around ($\frac{1}{4}$, y, $\frac{1}{4}$), for the particle position using the correctly oriented EM model of φX174 with 50–25 Å resolution data.

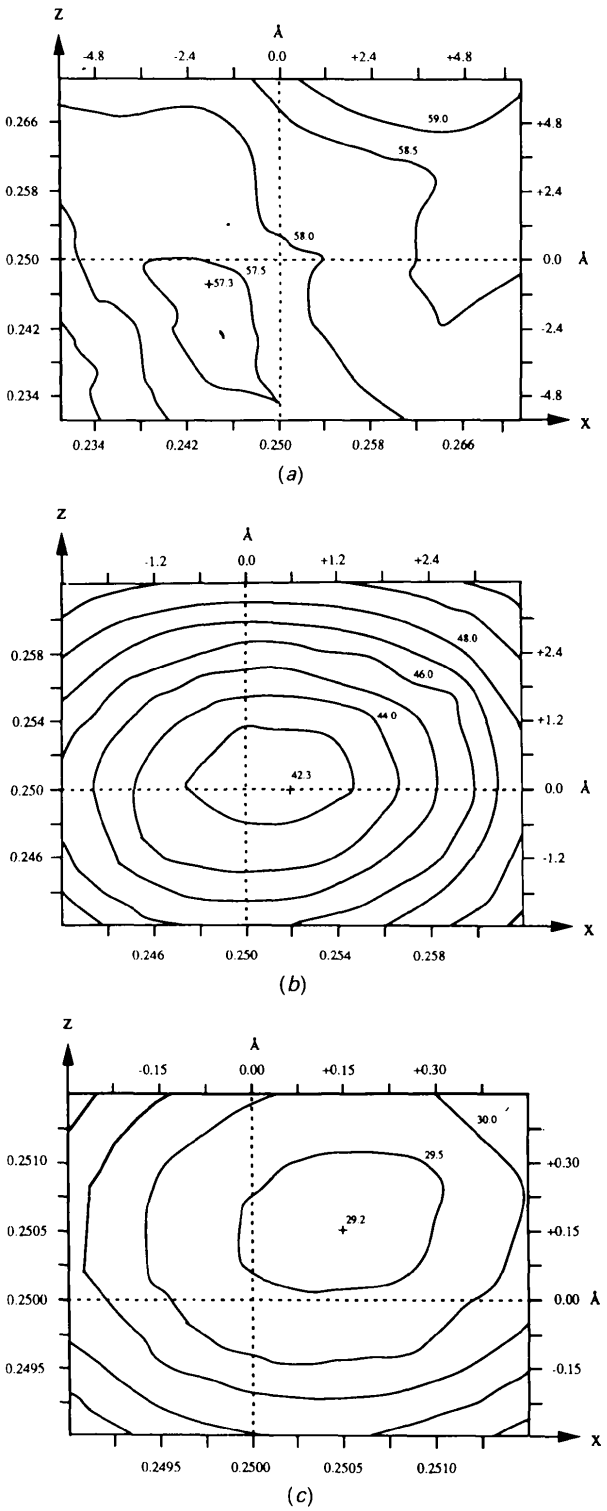

Fig. 5. (*a*) *R*-factor translation search using the CpMV model with 30–12 Å resolution data. (*b*) *R*-factor search using a model obtained after 20 cycles of real-space averaging with 30–12 Å resolution data. (*c*) *R*-factor search using a model obtained after 40 cycles of real-space averaging and phase extension to 11 Å resolution.

model (Fig. 5*a*). After 20 cycles using this new center, the mean correlation coefficient had increased from 0.59 to 0.69. Phase extension now proceeded by one reciprocal lattice unit at a time to 11 Å resolution, with four cycles at each extension step. At this point a further attempt was made to refine the particle position using a newly derived averaged *h*-cell density as a search model. A slight positional shift of 0.5 Å to 0.2505, *y*, 0.2505 was obtained and the overall *R* factor was 29.2% (Fig. 5*c*). No further change in particle position was detected as phase extension proceeded to higher resolution (Fig. 7).

Confirmation of molecular replacement phase determination using isomorphous replacement

Partial heavy-atom data sets had been collected (Table 1) because it had not been possible to obtain a successful starting position or phasing model from the EM data. Although the eventual structural solution was independent of any heavy-atom information, the data were used to confirm the correctness of phasing when the extension had attained 8.8 Å resolution.

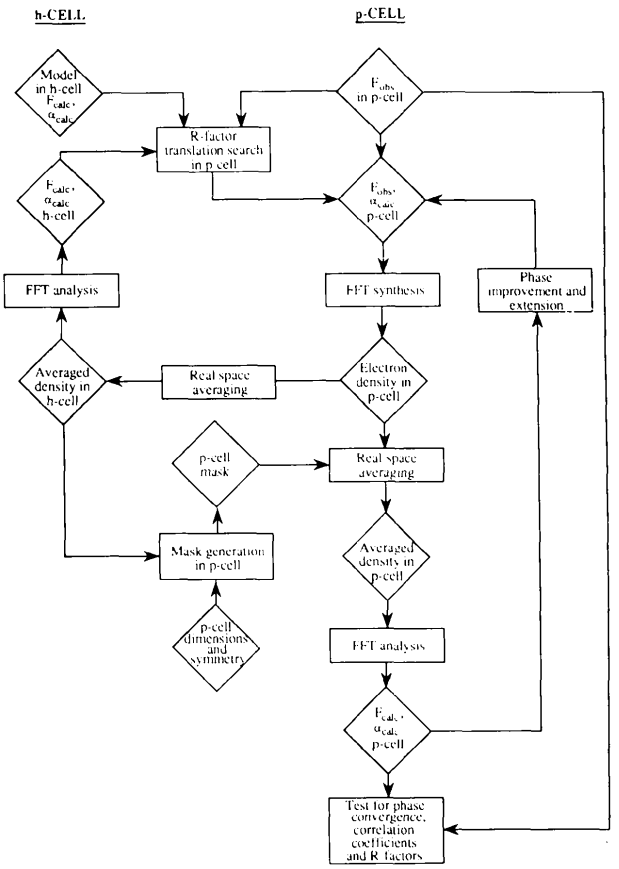

Fig. 6. Scheme for molecular replacement real-space averaging for phase improvement and extension.

Heavy-atom derivatives were prepared by soaking native crystals of 70S particles in a solution that consisted of 10 m*M* heavy-atom compound, 100 m*M* bis-tris buffer at pH 6.8 and 4% (*w*/*v*) of PEG 8000. All soaking experiments were carried out at 277 K for periods ranging from 12 to 36 h. Over ten compounds inhibited crystal diffraction. The chosen compounds were selected on the basis of only mild crystal damage after more than 24 h. Partial data sets were collected of K_2PtCl_4, $KAu(CN)_2$ and K_2OsCl_6. Native and heavy-atom data were scaled locally in 45 resolution ranges. The difference *R* factors are shown in Fig. 8 and these gave reasonable correspondence to similar results for other viruses (Arnold & Rossmann, 1986). Mean differences increase beyond about 4.5 Å resolution, suggesting lack of isomorphism at higher resolution. The K_2PtCl_4 derivative data showed the largest differences. There was only a small amount of K_2OsCl_6 derivative data that had been collected and it showed roughly the same average differences as the more extensive $KAu(CN)_2$ derivative data.

Electron density difference maps for all three compounds, with respect to the native data, using the phases extended to 8.8 Å resolution showed significant negative peaks for the K_2PtCl_4 and $KAu(CN)_2$ compounds (Fig. 9 and Table 3), but not for the K_2OsCl_6 compound. The negative peaks are reminiscent of those seen in the structure determination of the phage MS2 (Valegård, Liljas, Fridborg & Unge, 1990) and canine parvovirus (Tsao, Chapman, Agbandje, Keller, Smith, Wu, Luo, Smith, Rossmann, Compans & Parrish, 1991). Furthermore, difference maps calculated with various resolution data all showed negative peaks consistent with a single Babinet solution. Clearly, the opposite Babinet solution to the true structure had been obtained. In addition, the particle position could be verified by an analysis of the 60 equivalent heavy-atom sites in the difference electron density maps. Two different criteria were used, both dependent on averaging the difference electron density maps. The first criterion determined the particle position by maximizing the 'height' of the averaged icosahedrally equivalent negative 'peaks' (Tsao, Chapman, Wu, Agbandje,

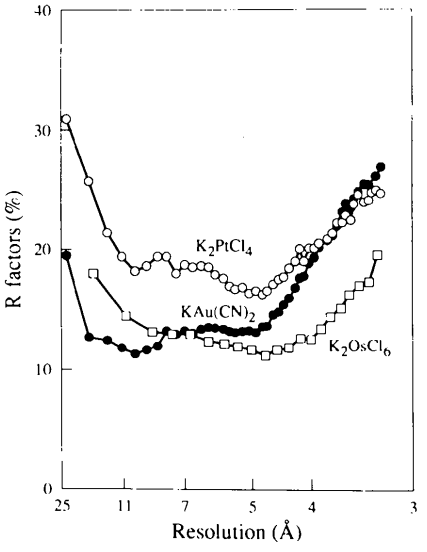

Fig. 8. Variation between heavy-atom derivative and native structure amplitudes as a function of resolution. The difference *R* factors are defined as $[\Sigma (F_{PH} - F_P)/\Sigma F_P] \times 100$ where F_{PH} and F_P are the derivatized and native structure amplitudes.

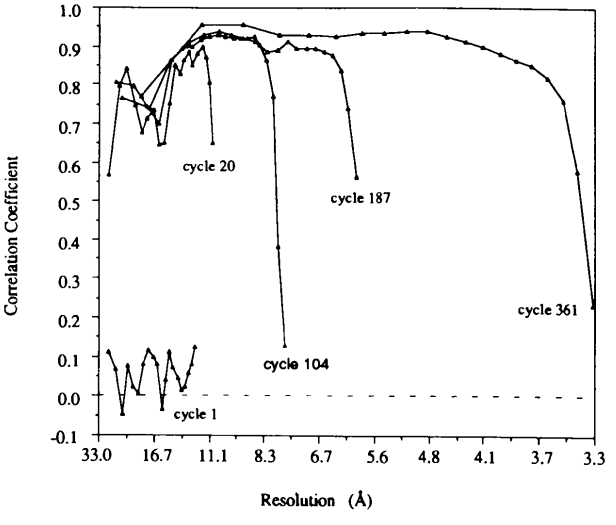

Fig. 7. Correlation coefficient for φX174 for phase extensions from 12 to 3.4 Å resolution. Correlation coefficients are defined as $\{[\Sigma(\langle F_{obs}\rangle - F_{obs})(\langle F_{calc}\rangle - F_{calc})]/[\Sigma(\langle F_{obs}\rangle - F_{obs})^2\Sigma(\langle F_{calc}\rangle - F_{calc})^2]^{1/2}\}$.

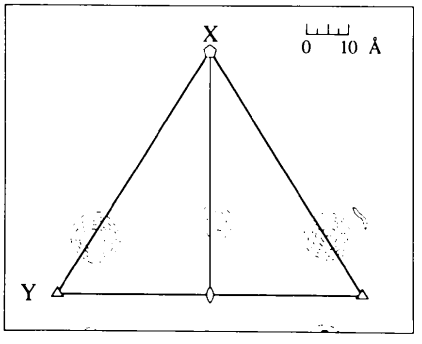

Fig. 9. Section $Z = 95$ Å of the K_2PtCl_4 difference electron density map based on 8.8 Å phases extended from a 12 Å resolution CpMV model. The triangle represents one icosahedral asymmetric unit. Three peaks are shown. The peaks on the line between the five- and threefold axes are related by the icosahedral fivefold axes. The resultant two independent negative peaks indicate that the phase extension had produced phases with the wrong Babinet solution. Contours are at equal but arbitrary intervals. Symmetry axes of the icosahedron are shown symbolically.

Table 3. *Heavy-atom sites in the structure of φX174*

Heavy-atom positions refer to an arbitrary icosahedral asymmetric unit in the *h*-cell, with Cartesian coordinates *X*, *Y*, *Z*.

Compounds	Sites	*X*	*Y*	*Z*	Height in difference map	Liganding residues	Protein subunits	Comment
K_2PtCl_4	1	14.5	−28.9	95.0	−279	Met10	F	On interior
	2	17.2	−1.8	97.0	−154	Met238	F	On interior
Largest noise peak					+55			
$KAu(CN)_2$	3	0.0	−48.5	127.0	153	Thr187	F	On threefold axis
	4	83.1	0.0	134.5	−143	Asp117	G	On fivefold axis in ion channel at putative Ca^{2+} binding site
Largest noise peak					+40			

Keller & Rossmann, 1992). This corresponds to the center of gravity of the 60 independent equivalent difference peaks. The second criterion searched for the position which gave the least r.m.s. scatter of all the 60 noncrystallographically related grid points within the peak position. These tests were conducted with respect to maps computed with 8.8 Å and, later, with 5.8 Å resolution data. The results were scattered within a box of about 0.25 × 0.25 Å, roughly centered around the site that had been found from phases extended to 11 Å resolution (Fig. 5*c*). No further adjustment of the particle center was deemed appropriate. The heavy-atom data, thus, had been used to establish the Babinet solution and to confirm the particle position.

Phase extension from 5.6 to 3.4 Å resolution

At 5.6 Å resolution the mean correlation coefficient was 0.86, but further phase extension started to show deterioration in the higher-resolution ranges. It was reasoned that this might be attributed to the simple spherical mask which was causing solvent flattening of the interdigitated spikes in the crystal structure. Fig. 10 shows a typical section of φX174 electron density in the *p*-cell at 3.4 Å resolution. Superimposed on that density are the spherical and molecular masks used in averaging. It clearly shows that the simple spherical mask with tangential planes was inadequate in accounting for the convoluted shape of φX174. The spherical mask used covered most of the φX174 particle but had to have such a large radius to account for the spikes that most of the solvent region was within the mask and, therefore, could not be used for solvent flattening. Therefore, a true molecular envelope was generated (Rossmann *et al.*, 1992). This had the further advantage of reducing the number of grid points to be averaged by up to a factor of two (mask 6 in Table 4). The first mask was generated at 5.6 Å resolution by averaging and skewing the *p*-cell into the *h*-cell. Such averaging, given 60-fold redundancy, essentially cancels out all other particles that do not obey the local noncrystallographic symmetry in the *p*-cell with respect to a chosen particle center. This density was then used to establish a mask employing a series of criteria (Table 4), including a 165 Å radial limit to the particle size.* The most sensitive of these criteria was the absolute height of density (the density 'cutoff') in the *h*-cell that determined the best molecular envelope. This value was established by plotting the number of grid points (proportional to volume) within the molecular mask expressed as a percentage of the number of grid points within one asymmetric unit of the *p*-cell (Fig. 11). As the density cutoff decreased, the volume of the molecular mask increased roughly linearly. When the density cutoff had been reduced to below the mean smeared density encountered within the protein density, then the volume of the molecular mask increased precipitously. At that point the volume of the mask was about 45% of the volume of the *p*-cell asymmetric unit, consistent with expectation. On using the new mask, the correlation coefficients improved dramatically, particularly in the higher-resolution ranges. Table 4 shows that the percentage of grid points that had been wrongly assigned by using the spherical mask was only 2%. On the other hand, the percentage of grid points that were solvent flattened was increased by about 30% in the unit cell when the molecular mask was used. Thus, the improvement of correlation coefficients was seen to be (once the structure had been solved) not due to the problem of interdigitating spikes, but due to the larger volume available for flattening.

Beyond 5.6 Å resolution, starting from the time an accurate mask (mask 6 in Table 4) was employed, the observed structure-factor (F_{obs}) data were supplemented with calculated structure factors (F_{calc}). In addition, the geometric mean of Sim and exponential weights (Arnold & Rossmann, 1986) were applied to the structure amplitudes in calculating new electron density maps. Phase extension then continued from

* With hindsight, 165 Å was an insufficient radial cutoff value. When interpreting the 3.5 Å minimap, it was seen that amino acids 45–50 of the βB–βC corner of the G protein had been solvent flattened. On the regeneration of a new mask at 3.4 Å resolution (mask 9 in Table 4), using a 175 Å radius cutoff, the density for these amino acids was recovered after four cycles of averaging.

Table 4. *Mask generation*

	Spherical mask[a]					Molecular mask			
Mask	1	2	3	4	5	6	7	8	9
Resolution of data used for mask generation (Å)	12.0	11.0	8.2	7.6	6.1	5.6	4.3	3.9	3.4
No. of grid steps along *a*, *b*, *c*	90–90–90	108–108–108	120–120–120	140–160–140	150–180–150	180–220–180	220–260–220	220–260–220	220–260–220
External, internal and core radii[b] (Å)	165–80–80	165–80–80	165–80–80	165–80–80	165–80–80	165–120–80	165–120–80	165–120–80	175–120–80
Total No. of points in asymmetric unit	364500	629856	864000	1568000	2025000	3564000	6292000	6292000	6292000
Grid points in mask based on *h*-cell map or defined by spherical limits	304990	527189	723168	1312416	1694925	1570108	2430512	2674031	2721716
Grid points added by hole filling[c]	–	–	–	–	–	10659	12568	15784	21596
Grid points added by symmetrization[d]	–	–	–	–	–	333101	117764	53497	–
% of points within mask[e]	83.7	83.7	83.7	83.7	83.7	53.7	40.7	43.6	43.6
% of points within solvent	3.3	3.3	3.3	3.3	3.3	24.3	32.9	31.0	31.0
% of points within nucleic acid	13.0	13.0	13.0	13.0	13.0	22.1	26.4	25.4	25.4
% of grid points common to spherical mask and molecular mask	–	–	–	–	–	51.7	38.6	41.9	41.9
% of grid points in spherical mask that were in wrong particle	–	–	–	–	–	1.9	1.3	1.6	1.6
% of grid points in spherical mask that were actually solvent	–	–	–	–	–	20.9	30.3	27.8	27.8
% of grid points in spherical mask that were actually in nucleic acid region	–	–	–	–	–	9.1	13.5	12.4	12.4

Notes: (*a*) The initial spherical mask had tangent planes separating overlapped particles. (*b*) All grid points outside the external radius were assumed to be in solvent. Grid points inside the external radius and outside the internal radius were assumed to be solvent if the corresponding *h*-cell density was less than a selected value. Grid points inside the internal radius but outside the core radius were assumed to be in the nucleic acid region if the corresponding density in the *h*-cell was less than a certain value. All grid points inside the core radius were assumed to be in the nucleic acid region. (*c*) Grid points that were surrounded by at least four grid points that had been assigned to a specific molecular mask were assigned to that mask. (*d*) Grid points were added to the mask to give it icosahedral symmetry. This operation, together with (*b*) and (*c*), is described in more detail by Rossmann *et al.* (1992). (*e*) All percentages are given with respect to the total number of grid points in the asymmetric unit.

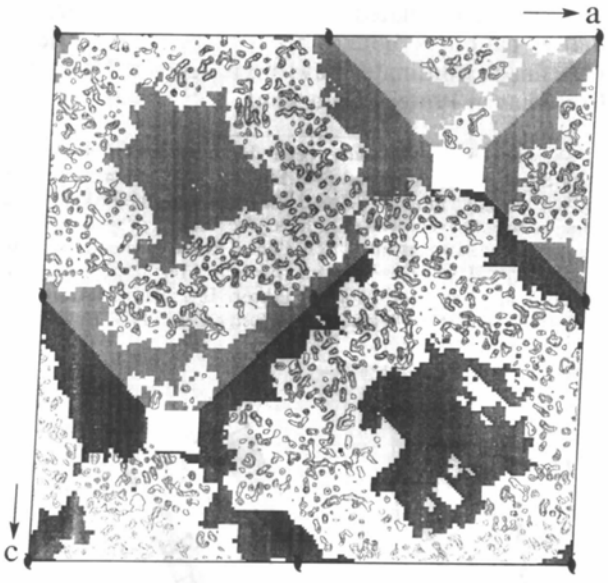

Fig. 10. Section $y = 0$ showing the final averaged electron density superimposed onto the spherical mask and the molecular mask. Spherical masks with respect to different molecular centers are differentiated by various degrees of shading outside the molecular mask. The molecular mask has the lightest shading. Volumes outside both the spherical and molecular masks are left clear.

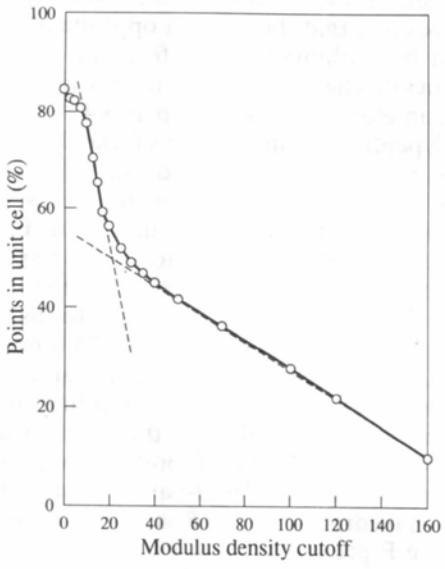

Fig. 11. The volume of the molecular mask, expressed as a percentage of the volume of the *p*-cell asymmetric unit, as determined by the density cutoff in the *h*-cell. When the modulus of the density cutoff is decreased to less than the mean smeared electron density within the protein, the mask volume increases rapidly. Intersection of the tangents suggests the most appropriate density cutoff value for mask generation.

5.6 to 3.4 Å resolution without problems. The molecular envelope was redefined when the electron density map sampling became too coarse at 4.3 Å resolution (mask 7 in Table 4) and again at 3.9 Å resolution (mask 8 in Table 4). Linear interpolation to determine density at non-integral grid positions was used at all times. The number of cycles required before the mean correlation coefficient in the outermost current resolution range increased by less than 0.01 was about the same for intervals ranging from $\frac{1}{4}$ to $\frac{1}{2.4}$ of the resolution. It had previously been suggested that sampling using linear interpolation of $\frac{1}{6}$ of the resolution would be essential for successful phase extension (Bricogne, 1976). A total of 74 steps of phase extension were used between 12 and 3.4 Å resolution (Fig. 7).

The number of points that had to be symmetrized after mask generation was fewer as phase extension progressed (see successive masks 6, 7 and 8 in Table 4). Although the *h*-cell density, used for generating a new mask, was itself icosahedrally averaged, the *p*-cell mask symmetry is destroyed when there is conflict between neighboring particles as to the assignment of individual grid points. Thus, the number of grid points that needed to be eliminated reflects both a more accurate knowledge of the molecular envelope and the use of a higher density cutoff in mask generation.

Map interpretation

It had been shown (see above), using the heavy-atom derivative data, that the Babinet opposite phase solution had been obtained. Therefore, all phases were systematically changed from α to $\pi + \alpha$ before calculating an electron density map at 3.4 Å resolution. The polypeptide chain tracing of the F, J and G proteins and their amino-acid identification was completed with ease on a 'mini-map' representation. An atomic model was built using an Evans & Sutherland PS390 graphics system with respect to a 3.4 Å resolution map, with the use of the program *FRODO* (Jones, 1978). The F protein forms a $T = 1$ (Caspar & Klug, 1962) capsid. All 175 amino acids of the G spike protein and the 426 amino acids, apart from the first and last two residues, of the F capsid protein were built into the electron density. The major folding motif of both proteins was an eight-stranded antiparallel β-barrel, found in many other icosahedral viruses (Rossmann & Johnson, 1989). The F protein has two large inserted loops of 170 and 110 amino acids at the βE–βF and βH–βI turns, respectively [see McKenna *et al.* (1992) for the standard nomenclature of the secondary structural elements βB to βI]. The small 37 amino-acid internal DNA packaging J protein has a high content of basic residues. Only the less-basic carboxy-terminal residues 28–37 were ordered. A typical piece of the high-quality electron density map and its interpretation is given in Fig. 12. After the polypeptide chain had been traced, it was found that the heavy-atom binding sites were at chemically reasonable positions (Table 3). There is no compelling evidence in the electron density for the presence of the H protein. Nevertheless, there is mutational evidence (Ilag, Tuech, Beisner & Incardona, 1992) that diffuse density observed on each of the fivefold axes in a cavity formed by the pentamers of G proteins may be part of the H protein. A fuller description of the structure is given by McKenna *et al.* (1992).

A difference electron density map was calculated between the partial diffraction data (Table 1) of the full 114S particles and the mostly empty 70S particles, using the phases, α_{combined}, based on the 3.4 Å resolution determination. This shows regions where there is some partially ordered DNA structure. A total of 11 nucleotides per icosahedral asymmetric unit could be built into the difference electron density. An electron density map (McKenna *et al.*, 1992) using $(F_{114S} - kF_{70S})e^{i\alpha_{\text{combined}}}$, where $k = 0.9$, showed the DNA density to have roughly the same height as the protein. Full substitution of DNA would be expected for $k = 0.5$ (Chapman, Tsao & Rossmann, 1992). Thus, only about one in five sites on the virus interior contains the icosahedrally ordered DNA structure. A large negative peak occurs in the difference map within the G protein pentamer channel. This density is liganded by the five symmetry-related Asp117 residues, suggesting that a cation is bound here in the 70S particles. Unrefined coordinates have been deposited with the Brookhaven Protein Data Bank.*

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1BPA), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37061 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

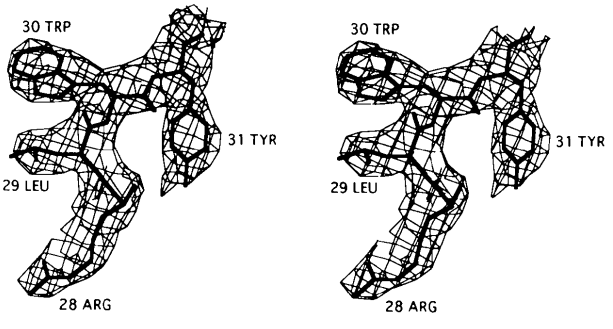

Fig. 12. A piece of the J protein (residues 28–31) in its electron density. The J protein is disordered prior to residue 28.

Postmortem on the method of phase determination

The initial phases derived from a model of CpMV eventually led to a correctly phased 3.4 Å electron density map of φX174, although the initial mean correlation coefficient was only 0.056 at 12 Å resolution, corresponding to an almost random set of phases (Fig. 13*a*; Table 5). An analysis (Valegård *et al.*, 1990) of the early phase solutions was made by plotting (Fig. 13) the final 'true' phases, α, against those derived for the three successive particle positions used in the initial stages of the phase determination (Table 5). Both the hand (phase change α to $-\alpha$) and Babinet inversion (phase change α to $\pi + \alpha$) changed as the particle position was refined within the unit cell. The phases show an initial, apparently random, starting set (Fig. 13*a*), which evolves into a Babinet opposite phase solution after 20 cycles of averaging at the first position (Table 5; Fig. 13*b*). After a further 20 cycles of averaging at the second particle position (Table 5) the phases contained two competing solutions: the enantiomorph and the Babinet inversion of the enantiomorph (Fig. 13*c*). The third and final particle position caused the phases to change solution again and converge on the Babinet opposite phase set which led to the final structure solution (Fig. 13*d*). It would seem that various phase sets were competing with each other, of which one eventually became dominant. A similar example of differing phase solutions that competed with each other occurred in the initial attempt of a structure determination of MS2 using molecular replacement averaging (Valegård *et al.*, 1990). These results suggest that it might be possible to initiate phasing from an entirely random phase set. Such experiments are in progress.

It is remarkable that on finding the correct particle position the phases changed to a unique solution other than the phase solutions found at the intermediate positions. This was possible perhaps primarily because of the availability of a nearly complete data set and the high noncrystallographic

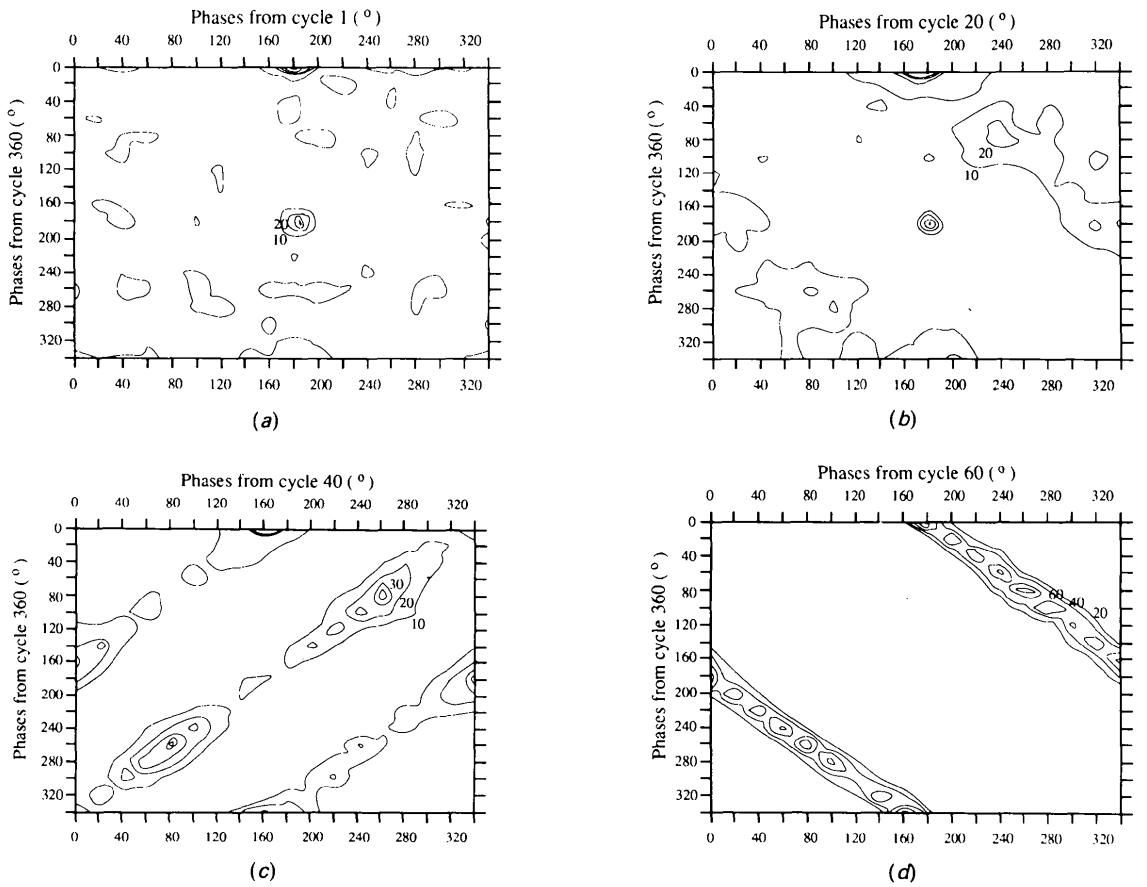

Fig. 13. Phase comparison plots for data in the 15–12 Å resolution range, comparing early phase determinations (Table 5) with final phases at cycle 360. Contours represent the frequency of phase correspondence. (*a*) Initial phases to 12 Å resolution derived from the CpMV model placed at position 1. (*b*) After 20 cycles of averaging at position 1 at 12 Å resolution. (*c*) After another 20 cycles of averaging at position 2 and extension to 11 Å resolution. (*d*) After a further 20 cycles of averaging at position 3 at 11 Å resolution.

Table 5. *Phase solutions at various particle positions in the resolution range* 15–12 Å

No.	Position[a] x	y	z	Cycle number[b]	Deviation (Å) from ($\frac{1}{4}$,	y	$\frac{1}{4}$)	Phase solution	Phases[c]
1	0.2440	y	0.2480	1	−1.83	y	0.60	No solution	
1	0.2440	y	0.2480	20	−1.83	y	0.60	Babinet solution	$\alpha \rightarrow \pi + \alpha$
2	0.2520	y	0.2500	40	+0.61	y	0.00	Mirror and mirror-Babinet (competing solutions)	$\alpha \rightarrow -\alpha$ $\alpha \rightarrow \pi - \alpha$
3	0.2505	y	0.2505	60	+0.15	y	+0.15	Babinet solution (final solution)	$\alpha \rightarrow \pi + \alpha$

Notes: (*a*) Center used in determining phases for given cycle number. (*b*) Cycle number after which phases were compared with the final phase solution. (*c*) α are the final phases at 3.4 Å resolution.

symmetry. Inclusion of an excessive number of F_{calc}'s might have restrained the phases to the current mix of solutions. On the other hand, it is clear that when opposing structural solutions have been incorporated into the phasing set, then their convergence onto a single solution is probably impossible at higher resolution where there is less overall interaction between structure factors due to the decay of the roughly spherical *G* function (Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). Even at low resolution it is often difficult to extract the correct solution from a mixture of solutions (Chapman *et al.*, 1992). This could be to some extent alleviated by detailed information about the molecular envelope, which will have the effect of locally linking structure factors by variations in the decay of the *G* function.

The structure determination succeeded even with the 'irrelevant' structural model of CpMV, possibly because significantly sized structure amplitudes were available to 12 Å resolution while the EM model had essentially zero amplitudes beyond 25 Å resolution. Nevertheless, it is of some interest whether the EM image of φX174 could have led to a correct phase solution given the correct particle position. Phase extension was attempted using the correct position and this did eventually succeed (unpublished results). It can be concluded, therefore, that, given a correct particle position and an EM image, it is then possible to correctly initiate a high-resolution structure determination given fairly complete and accurate structure amplitudes with sufficient noncrystallographic redundancy. Hence, assuming the availability of reasonably good very-low-resolution data, the most critical component in obtaining a phasing start is to determine an accurate particle position.

We greatly appreciate Nino Incardona introducing us to this fascinating virus, S. Krishnaswamy's initial work on the structure determination and many useful discussions with Thomas J. Smith concerning purification and crystallization. We thank Norman Olson and Tim Baker for their electron microscopy reconstruction work of φX174. We also thank the many helpers in data-collection trips (including Mavis Agbandje, Jodi Bibler, Michael Chapman, Hok-Kin Choi, Vincent Giranda, Andrea Hadfield, Walter Keller, Marcos Oliveira, Andrew Prongay, Thomas Smith, Liang Tong and Hao Wu) and the outstanding support at the various synchrotron facilities that we have visited (NSLS, CHESS, Daresbury and DESY). We thank Helene Prongay and Sharon Wilder for help in the preparation of the manuscript. The work was supported by a National Science Foundation grant and a Lucille P. Markey Charitable Trust grant for the development of structural studies at Purdue to MGR and fellowship Wi873/1-1 from the Deutsche Forschungsgemeinschaft to PW.

References

ARGOS, P. & ROSSMANN, M. G. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 361–417. New York: Plenum.

ARNOLD, E. & ROSSMANN, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489–5493.

ARNOLD, E., VRIEND, G., LUO, M., GRIFFITH, J. P., KAMER, G., ERICKSON, J. W., JOHNSON, J. E. & ROSSMANN, M. G. (1987). *Acta Cryst.* A**43**, 346–361.

BAYER, M. E. & STARKEY, T. W. (1972). *Virology*, **49**, 236–256.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

BROWN, D. T., MACKENZIE, J. M. & BAYER, M. E. (1971). *J. Virol.* **7**, 836–846.

BURGESS, A. B. (1969). *Proc. Natl Acad. Sci. USA*, **64**, 613–617.

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

CHAPMAN, M. S., TSAO, J. & ROSSMANN, M. G. (1992). *Acta Cryst.* A**48**, 301–312.

CHEN, Z., STAUFFACHER, C. V. & JOHNSON, J. E. (1990). *Sem. Virol.* **1**, 453–466.

EDGELL, M. H., HUTCHISON, C. A. III & SINSHEIMER, R. L. (1969). *J. Mol. Biol.* **42**, 547–557.

EIGNER, J., STOUTHAMER, A. H., VAN DER SLUYS, I. & COHEN, J. A. (1963). *J. Mol. Biol.* **6**, 61–84.

FEIGE, U. & STIRM, S. (1976). *Biochem. Biophys. Res. Commun.* **71**, 566–573.

HAYASHI, M., AOYAMA, A., RICHARDSON, D. L. JR & HAYASHI, M. N. (1988). *The Bacteriophages* (*The Viruses*), edited by R. CALENDAR, Vol. 2, pp. 1–71. New York: Plenum Press.

ILAG, L. L., TUECH, J. K., BEISNER, L. & INCARDONA, N. L. (1992). *J. Mol. Biol.* In the press.

INCARDONA, N. L. & SELVIDGE, L. (1973). *J. Virol.* **11**, 775–782.

INCARDONA, N. L., TUECH, J. K. & MURTI, G. (1985). *Biochemistry*, **24**, 6439–6446.

JAZWINSKI, S. M., LINDBERG, A. A. & KORNBERG, A. (1975). *Virology*, **66**, 268–282.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KIM, S. (1989). *J. Appl. Cryst.* **22**, 53–60.
MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491 497.
McKENNA, R., XIA, D., WILLINGMANN, P., ILAG, L. L., KRISHNASWAMY, S., ROSSMANN, M. G., OLSON, N. H., BAKER, T. S. & INCARDONA, N. L. (1992). *Nature (London)*, **355**, 137–143.
OLSON, N. H., BAKER, T. S., WILLINGMANN, P. & INCARDONA, N. L. (1992). *J. Struct. Biol.* **108**, 168–175.
RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.
ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H. J., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER, A. G., RUECKERT, R. R., SHERRY, B. & VRIEND, G. (1985). *Nature (London)*, **317**, 145–153.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
ROSSMANN, M. G. & ERICKSON, J. W. (1983). *J. Appl. Cryst.* **16**, 629–636.
ROSSMANN, M. G. & JOHNSON, J. E. (1989). *Annu. Rev. Biochem.* **58**, 533–573.
ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.
ROSSMANN, M. G., McKENNA, R., TONG, L., XIA, D., DAI, J., WU, H., CHOI, H. K. & LYNCH, R. E. (1992). *J. Appl. Cryst.* In the press.
SANGER, F., AIR, G. M., BARRELL, B. G., BROWN, N. L., COULSON, A. R., FIDDES, J. C., HUTCHISON, C. A. III, SLOCOMBE, P. M. & SMITH, M. (1977). *Nature (London)*, **265**, 687–695.
SIDEN, E. J. & HAYASHI, M. (1974). *J. Mol. Biol.* **89**, 1–16.
SINSHEIMER, R. L. (1959). *J. Mol. Biol.* **1**, 37–42.
STAUFFACHER, C. V., USHA, R., HARRINGTON, M., SCHMIDT, T., HOSUR, M. V. & JOHNSON, J. E. (1987). *Crystallography in Molecular Biology*, edited by D. MORAS, J. DRENTH, B. STRANDBERG, D. SUCK & K. WILSON, pp. 293–308. London: Plenum.
TOLLIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 872–876.
TONG, L. & ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
TSAO, J., CHAPMAN, M. S., AGBANDJE, M., KELLER, W., SMITH, K., WU, H., LUO, M., SMITH, T. J., ROSSMANN, M. G., COMPANS, R. W. & PARRISH, C. R. (1991). *Science*, **251**, 1456–1464.
TSAO, J., CHAPMAN, M. S., WU, H., AGBANDJE, M., KELLER, W. & ROSSMANN, M. G. (1992). *Acta Cryst.* B**48**, 75–88.
VALEGÅRD, K., LILJAS, L., FRIDBORG, K. & UNGE, T. (1990). *Nature (London)*, **345**, 36–41.
WANG, G., PORTA, C., CHEN, Z., BAKER, T. S. & JOHNSON, J. E. (1992). *Nature (London)*, **355**, 275–278.
WILLINGMANN, P., KRISHNASWAMY, S., McKENNA, R., SMITH, T. J., OLSON, N. H., ROSSMANN, M. G., STOW, P. L. & INCARDONA, N. L. (1990). *J. Mol. Biol.* **212**, 345–350.
WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.

Acta Cryst. (1992). B**48**, 511–514

A Note on the Conformational Flexibility of the Antiestrogenic Drug Tamoxifen: Preferred Conformations in the Free State and Bound to the Protein Calmodulin

BY KAREN J. EDWARDS, CHARLES A. LAUGHTON AND STEPHEN NEIDLE*

Cancer Research Campaign Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, England

(*Received* 13 *May* 1991; *accepted* 13 *December* 1991)

Abstract

The conformational properties of the antiestrogenic drug tamoxifen, a triphenylbut-1-ene derivative, have been studied using molecular mechanics. Four distinct conformers have been identified, and the energy barriers between them have been established. The orientation of the ethyl group substitutent has been examined in particular, since the lowest-energy conformers have this group orientated 180° away from its position in the crystal structures of tamoxifen and its derivatives. These differences have implications for the interactions of tamoxifen with the calcium-binding protein calmodulin; relevant results from a molecular-modelling study of this protein–drug complex are presented.

* To whom correspondence should be addressed.

0108-7681/92/040511-04$06.00

Introduction

The *trans*-triphenylbut-1-ene compound tamoxifen (Fig. 1) has established clinically useful anticancer activity (Jordan, Fritz & Gottardis, 1987) with its binding to the estrogen receptor believed responsible for its action against hormone-positive human breast cancer. There is, however, increasing evidence that the drug acts on other macromolecular targets as well. In the course of molecular-modelling studies in this laboratory on structure–activity relationships of tamoxifen and its derivatives and their interactions with non-estrogenic receptors, especially the calcium-binding protein calmodulin (Rowlands, Parr, McCague, Jarman & Goddard, 1990), it has become necessary to establish the conformational flexibility and energetics of tamoxifen itself. Several crystallographic studies in the tamoxifen series have been