References

ALMENNINGEN, A., BASTIANSEN, O., FERNHOLT, L., CYVIN, B. N., CYVIN, S. J. & SAMDAL, S. (1985). *J. Mol. Struct.* **128**, 59–75.
BASTIANSEN, O. (1949). *Acta Chem. Scand.* **3**, 408–414.
BASTIANSEN, O. & SAMDAL, S. (1985). *J. Mol. Struct.* **128**, 115–125.
BAUDOUR, J. L., CAILLEAU, H. & YELON, W. B. (1977). *Acta Cryst.* B**33**, 1773–1780.
BAUDOUR, J. L., DÉLUGEARD, Y. & CAILLEAU, H. (1976). *Acta Cryst.* B**32**, 150–154.
BAUDOUR, J. L., DÉLUGEARD, Y. & RIVET, P. (1978). *Acta Cryst.* B**34**, 625–628.
BAUDOUR, J. L., DÉLUGEARD, Y. & SANQUER, M. (1974). *Acta Cryst.* B**30**, 691–696.
BAUDOUR, J. L. & SANQUER, M. (1983). *Acta Cryst.* B**39**, 75–84.
BAUDOUR, J. L., TOUPET, L., DÉLUGEARD, Y. & GHÉMID, S. (1986). *Acta Cryst.* C**42**, 1211–1217.
BENKERT, C. & HEINE, V. (1987). *Phys. Rev. Lett.* **58**, 2232–2234.
BENKERT, C., HEINE, V. & SIMMONS, E. H. (1987). *J. Phys. C*, **20**, 3337–3354.
BONADEO, H. & BURGOS, E. (1982). *Acta Cryst.* A**38**, 29–33.
BREE, A. & EDELSON, M. (1977). *Chem. Phys. Lett.* **46**, 500–504.
BRENNER, H. C., HUTCHISON, C. A. JR & KEMPLE, M. D. (1974). *J. Chem. Phys.* **60**, 2180–2181.
BUSING, W. R. (1983). *Acta Cryst.* A**39**, 340–347.
CAILLEAU, H., BAUDOUR, J. L., MEINNEL, J., DVORKIN, A., MOUSSA, F. & ZEYEN, C. M. E. (1980). *Faraday Discuss. Chem. Soc.* **69**, 7–18.
CAILLEAU, H., BAUDOUR, J. L. & ZEYEN, C. M. E. (1979). *Acta Cryst.* B**35**, 426–432.
CAILLEAU, H., GIRARD, A., MOUSSA, F. & ZEYEN, C. M. E. (1979). *Solid State Commun.*. **29**, 259–261.
CAILLEAU, H., MESSAGER, J. C., MOUSSA, F., BUGAUT, F., ZEYEN, C. M. E. & VETTIER, C. (1986). *Ferroelectrics*, **67**, 3.
CARREIRA, L. A. & TOWNS, T. G. (1977). *J. Mol. Struct.* **41**, 1–9.
CHARBONNEAU, G. P. & DÉLUGEARD, Y. (1976). *Acta Cryst.* B**32**, 1420–1423.
CHARBONNEAU, G. P. & DÉLUGEARD, Y. (1977). *Acta Cryst.* B**33**, 1586–1588.
CRUICKSHANK, D. W. J. (1965*a*). *Acta Cryst.* **9**, 754–756.
CRUICKSHANK, D. W. J. (1956*b*). *Acta Cryst.* **9**, 1005–1009.
CRUICKSHANK, D. W. J. (1956*c*). *Acta Cryst.* **9**, 1010–1011.
DEJACE, J. (1969). *Bull. Soc. Fr. Minéral. Cristallogr.* **92**, 141–159.
DUNITZ, J. D. (1979). *X-ray Analysis and the Structure of Organic Molecules*, pp. 244–261. Cornell Univ. Press.
FRIEDMAN, P. S., KOPELMAN, R. & PRASAD, P. N. (1974). *Chem. Phys. Lett.* **24**, 15–17.
HÄFELINGER, G. & REGELMANN, C. (1985). *J. Comput. Chem.* **6**, 368–376.
HARGREAVES, A. & RIZVI, S. H. (1962). *Acta Cryst.* **15**, 365–373.
HEINE, V. & PRICE, S. L. (1985). *J. Phys. C*, **18**, 5259–5278.
JOHNSON, C. K. (1970). *Thermal Neutron Diffraction*, edited by B. T. M. WILLIS, ch. 9, pp. 132–160. Oxford Univ. Press.
KATON, J. E. & LIPPINCOTT, E. R. (1959). *Spectrochim. Acta Part A*, **15**, 627–650.
NATKANIEC, I., BIELUSHKIN, A. V. & WASIUTYNSKI, T. (1981). *Phys. Status Solidi B*, **105**, 413–423.
PICKETT, L. W. (1936). *J. Am. Chem. Soc.* **58**, 2299–2303.
PLAKIDA, N. M., BIELUSHKIN, A. V., NATKANIEC, I. & WASIUTYNSKI, T. (1983). *Phys. Status Solidi B*, **118**, 129–133.
RAICH, J. C. & BERNSTEIN, E. R. (1984). *Mol. Phys.* **53**, 597–614.
REYNOLDS, P. A. & WHITE, J. W. (1972). *J. Chem. Soc. Faraday Trans.* **8**, 1434–1438.
RIETVELD, H. M., MASLEN, E. N. & CLEWS, C. J. B. (1970). *Acta Cryst.* B**26**, 693–706.
ROBERTSON, G. B. (1961). *Nature (London)*, **191**, 593–594.
SCHOMAKER, W. & TRUEBLOOD, L. N. (1968). *Acta Cryst.* B**24**, 63–76.
TAKEUCHI, H., SUZUKI, S., DIANOUX, A. J. & ALLEN, G. (1981). *Chem. Phys.* **55**, 153–162.
TOUDIC, B., GALLIER, J., RIVET, P. & CAILLEAU, H. (1983). *Solid State Commun.* **47**, 291–295.
TROTTER, J. (1961). *Acta Cryst.* **14**, 1135–1140.
TRUEBLOOD, K. N. & DUNITZ, J. D. (1983). *Acta Cryst.* B**39**, 120–133.
WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography*. Cambridge Univ. Press.

Acta Cryst. (1991). B**47**, 949–960

Structure Determination of the Bacteriophage MS2

BY KARIN VALEGÅRD, LARS LILJAS, KERSTIN FRIDBORG AND TORSTEN UNGE

Department of Molecular Biology, Uppsala University, Uppsala Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden

(*Received* 5 *September* 1990; *accepted* 11 *June* 1991)

Abstract

The structure of the bacterial virus MS2 has been solved at 3·3 Å resolution. Initial phases to 13 Å resolution were obtained from a model based on the known coordinates of the plant virus southern bean mosaic virus. These phases were extended in small steps to a resolution of 3·4 Å. The phases obtained represented essentially the Babinet opposite of the true structure and were not of a sufficiently good quality to allow an interpretation of the electron density contoured at negative levels. Difference Fourier maps of two heavy-atom derivatives based on these phases were interpretable, and these derivatives were used to calculate isomorphous replacement phases at 8·8 Å resolution. Phase extension to 3·3 Å resolution led to maps which could be easily interpreted.

0108-7681/91/060949-12$03.00

Introduction

The symmetry of the protein shell promotes to a great extent the structure determination of icosahedral virus particles. The earliest crystal structure determinations of icosahedral viruses were based on isomorphous replacement combined with phase refinement using the non-crystallographic symmetry (Harrison, Olson, Schutt, Winkler & Bricogne, 1978; Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980; Liljas, Unge, Jones, Fridborg, Lövgren, Skoglund & Strandberg, 1982). With the structure determinations of rhinovirus (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Lou, Mosser, Rueckert, Sherry & Vriend, 1985) and poliovirus (Hogle, Chow & Filman, 1985) it was shown that isomorphous replacement could be performed at intermediate (5 Å) resolution followed by stepwise extension of the phases to a resolution where the polypeptide chain could be traced. It is often a problem to prepare heavy-atom derivatives that do not damage the crystals and are isomorphous to high resolution. More recently, several virus structures have been solved with the use of the known structure of a closely related virus as a low-resolution structural approximation followed by phase extension to higher resolution (Hogle, Maeda & Harrison 1986; Luo, Vriend, Kamer, Minor, Arnold, Rossmann, Boege, Scraba, Duke & Palmenberg, 1987; Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989; Chen, Stauffacher, Li, Schmidt, Bomu, Kamer, Shanks, Lomonosoff & Johnson, 1989). There have also been some attempts to solve virus structures by phase extension, starting *ab initio* with a spherical shell (Johnson, Akimoto, Suck, Rayment & Rossmann, 1976; Nordman, 1980), or starting with a very low-resolution model derived from electron micrographs (Rayment, Baker, Caspar & Murakami, 1982). These attempts were all done at low resolution and did not lead to useful high-resolution phases.

The MS2 virions contain 180 identical coat-protein subunits of 129 amino acids (M_r 13·7 kdaltons), one copy of a maturation protein of 393 amino acids (M_r 44·0 kdaltons) and a single-stranded RNA molecule of 3569 nucleotides (M_r 1·2 Mdaltons). The coat-protein subunits are arranged with icosahedral symmetry in a $T = 3$ (Caspar & Klug, 1962) surface lattice. This implies that the icosahedral asymmetric unit contains three protein subunits.

We have solved the structure of MS2 at 3·3 Å resolution by isomorphous replacement combined with molecular replacement using the known structure of southern bean mosaic virus (SBMV). A description of the structure has been presented (Valegård, Liljas, Fridborg & Unge, 1990). MS2 is the first bacteriophage for which a structure to atomic resolution has been determined. The coat protein has no structural similarity to the SBMV coat protein, which was used as the initial model, or to any other structurally determined virus. In this paper we describe in more detail the structure determination procedure. The partial success of the molecular replacement procedure described here indicates that in the presence of high non-crystallographic symmetry, a crystal structure can be obtained without the use of the isomorphous replacement technique. Originally the space group was assumed to be monoclinic ($C2$), with half a virus particle in the asymmetric unit and 30-fold non-crystallographic symmetry.

Crystallization

Crystals of MS2 were grown as reported (Valegård, Unge, Montelius, Strandberg & Fiers, 1986). They are rhombohedral, space group $R32$ with $a = 274·0$ Å and $\alpha = 63·43°$, and have one virus particle per unit cell. One twofold axis and one threefold axis of the virus particle coincide with crystallographic axes, giving a 10-fold non-crystallographic redudancy. The volume per mass value (V_m) is only 4·1 Å^3 dalton^{-1}. The crystals diffract to at least 2·6 Å resolution.

Data collection and all the calculations described in this paper were carried out as if the space group was $C2$, treating the crystal threefold axis as a non-crystallographic axis. However, since the cell parameters used were very close to those corresponding to the correct rhombohedral cell, we think it unlikely that the final phase-refined map is significantly influenced by this different cell choice.

Data collection and processing

The MS2 crystals are very sensitive, both to mechanical destruction and to radiation. Most of the crystals used in the data collection were therefore placed on a plateau in X-ray capillaries, surrounded by a solution containing 2% gelatin, 0·4 *M* sodium phosphate (pH 7·4) and 5% PEG 6000. The gelatin causes an increased background level on the films. To improve the quality of the data at high resolution, data were also collected from about 15 crystals mounted in capillaries without gelatin. Whilst several crystals cracked when mounted without gelatin, those which survived gave good-quality high-resolution data.

Three-dimensional X-ray diffraction data for native crystals were collected by oscillation photography. Initial 4 Å resolution data sets were recorded both on an Elliot rotating-anode generator in our laboratory and at the protein crystallographic station

7.2 at the SERC Synchrotron Radiation Source at Daresbury, England. The oscillation range was 1·4°. Most of the MS2 data to high resolution (2·8 Å) were collected on the Wiggler station 9.6 at the synchrotron at Daresbury. Data were collected on several occasions. Typical energies were 2 GeV with the ring current varying from 280 to 70 mA and a Wiggler magnetic field strength of 5 T. The wavelength varied from 0·88 to 0·92 Å and the crystal-to-film distance was 165–175 mm. The crystals were kept at a constant temperature. In different experiments this temperature varied between 275 and 279 K. Photographs were taken using an Enraf–Nonius Arndt–Wonacott camera. The crystal lifetime was dependent on the type of X-ray source. Using the rotating anode we could obtain two exposures of 30 h on each crystal compared with up to eight exposures of 4–16 min at the synchrotron source in Daresbury. The high-resolution films each covered an oscillation range of 0·5°. A typical oscillation photograph is shown in Fig. 1. It contains about 10 000 reflections of which 3000 are fully recorded.

The films were measured and digitized with an Optronics P-1000 microdensitometer using a 50 μm raster. Data were processed with the oscillation film evaluation program *OSC* by M. G. Rossman and M. Schmid (Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979; Schmid, Weaver, Holmes, Grütter, Ohlendorf, Reynolds, Remington & Matthews, 1981) and modified by S. J. Remington and T. A. Jones.

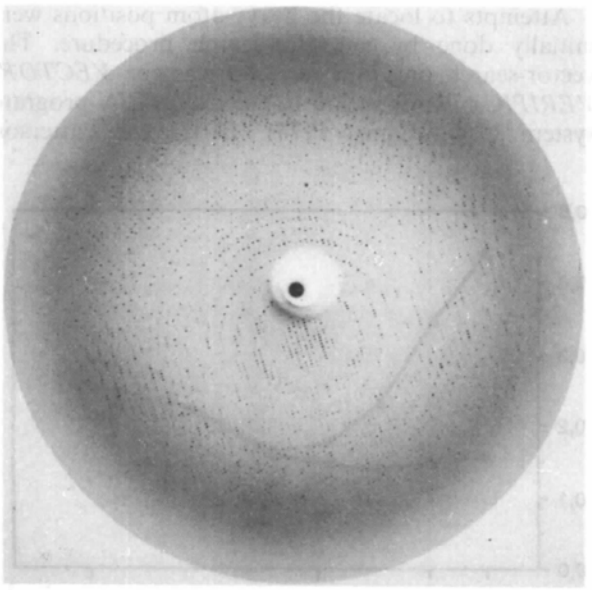

Fig. 1. A 0·5° oscillation photograph of an MS2 crystal collected at the synchrotron source at Daresbury (England). The diffraction pattern in this film extends to 2·8 Å. The exposure time was 12 min.

The crystals were difficult to align optically, so that the oscillation films were often obtained from crystals which had large setting errors or were randomly oriented. The crystal orientation was therefore determined by comparing an image of the film and a superimposed calculated pattern on a raster display (RAMTEK) using an option of the film processing program. Different possible orientations were systematically tested. If a calculated pattern of reflections to low resolution could be produced which was similar to the pattern on the film, the convolution technique of the film processing program was used to automatically refine the missetting angles, camera parameters and crystal-to-film distance. The correctness of the solution was checked by the criterion that most of the predicted reflections to low resolution should have significant intensities. If this was not the case another orientation was tried until the correct solution was found. The data were merged and scaled using the *PROTEIN* program system (Steigemann, 1974).

Refinements of crystal cell parameters, missetting angles and mosaic spread were done on an Alliant FX/90 computer using a version of the *CCP*4 program *POSTREF*, originally written by C. E. Schutt and F. Winkler (Winkler, Schutt & Harrison, 1979) and modified by A. Bloomer and P. Evans. The measured intensities of full and partial reflections were transformed to the format used by the *POSTREF* program. Missetting angles were converted from those obtained in the *OSC* program to the corresponding ones for the totally different convention of axis definitions and order of application used in *POSTREF*. Low-resolution and high-resolution native films were refined together. After a first scaling, cycles of post-refinement and scaling were done. Cell dimensions, missetting angles and mosaicity were refined individually for each film. The missetting angles were refined in one cycle, and mosaicity and cell parameters in another. The cell parameters were originally determined from 'still' photographs using the program *IDXREF* (Nyborg & Wonacott, 1977), and these parameters were post-refined using 24 low-resolution films collected with Cu $K\alpha$ radiation. The changes were of the order of 0·5%. For films collected with synchrotron radiation either wavelength or absolute scale of the cell parameters could be refined, and we chose to refine the cell parameters. To obtain the correct scale of the cell axes we adjusted the wavelength for all films collected on one occasion so that the mean of the refined cell axes was the same as that of the first 24 films.

The refinement of crystal mosaicity was critical. A cosine profile for the crystal mosaicity was used (Winkler, Schutt & Harrison, 1979). Values for the vertical and horizontal beam divergence and

wavelength dispersion of the synchrotron beam were obtained from Daresbury Laboratory. For about 30 films the refinement increased the mosaicity so that all reflections became classified as partial and the films had to be discarded. Some of these films were the last exposures of a crystal, and the others were from crystals of low quality or from crystals which had moved signficantly during data collection. The final mosaicity value varied considerably. For each crystal it normally increased for the last few films. For some crystals the value was high for all films, due to movement of the crystal during exposure. This movement was probably caused by temperature-dependent shrinking or swelling of the gelatin surrounding the crystal.

The intensities of partially recorded reflections measured on two adjacent films were added for all crystals where slippage during exposure was negligible, as judged from the refined missetting angles. The data collection was done in space group $C2$. In total 1 436 330 reflections were used, and of these 836 548 were fully recorded. The final data set consisted of 530 500 independent reflections, which in space group $R32$ corresponds to 230 000 independent reflections. Those partial reflections which were recorded to more than 80% and not used for addition were instead scaled to their fully recorded equivalent. The final data set consisted of 424 film packs taken from 103 crystals and the R value was 0·182 for all reflections.

Table 1. *Data collection*

	Native	K_2PtCl_4	$KAu(CN)_2$
6·0 Å resolution			
No. of unique reflections	71 231	19 892	13 179
Percent of possible unique total	96·9	27·1	17·9
3·3 Å resolution			
No. of unique reflections	408 773	·	
Percent of possible unique total	89·2	·	·
2·8 Å resolution			
No. of unique reflections	530 522	·	
Percent of possible unique total	74·6		
No. of film packs	424	32	19
R value	0·182	0·125	0·106

Notes: $R = \sum_h \sum_i |(F_h^2 - F_{hi}^2 G_i)| / \sum_h \sum_i F_h^2$, F_h = mean for independent reflection h, F_{hi} = the individual measurement of reflection h on film i, G_i = the scale factor factor on film i. The data collection and all the calculations were carried out as if the space group was $C2$.

Packing of the virus particle

The cell dimensions indicate that the packing of the virus particle in the rhombohedral cell is such that three fivefold axes align along the axes of the cell. This packing is identical to that of one crystal form of sourthern bean mosaic virus (SBMV) (Akimoto, Wagner, Johnson & Rossmann, 1975). An oscillation photograph from a crystal aligned with the X-ray beam parallel to one of the rhombohedral axes showed the spikes typically produced by a fivefold axis. A self-rotation function calculation (Rossmann & Blow, 1962; Tollin & Rossmann, 1966) confirmed this orientation. Post-refinement of the cell dimensions indicated that the fivefold axes are aligned along the rhombohedral axes within experimental error, and the rhombohedral angle is fixed by the geometry of the icosahedron.

Analysis of heavy-atom derivatives

A number of heavy-atom derivatives were prepared by the soaking technique. Partial data sets were collected and analysed. Statistics for two useful derivatives are shown in Table 1. A platinum derivative was prepared by soaking MS2 crystals for three days at room temperature in hanging drops containing 5 mM K_2PtCl_4 in 0·4 M sodium phosphate buffer pH 7·4 and 5% PEG 6000. The crystal was washed once a day in freshly prepared soaking liquid to prevent precipitate in the soaking solution from sticking to the crystals and making mounting more difficult. A gold derivative was prepared by soaking crystals for 15 h at room temperature in small plastic wells containing 10 mM $KAu(CN)_2$ in 0·4 M sodium phosphate buffer pH 7·4 and 5% PEG 6000. Analysis of the differences as a function of resolution shows that the platinum derivative displays very large differences at low resolution (Fig. 2). This is possibly due to binding of the metal complex to the disordered RNA molecule.

Attempts to locate the heavy-atom positions were initially done by a vector-search procedure. The vector-search program we used was the *VECTOR-VERIFICATION* option of the *PROTEIN* program system (Steigemann, 1974). Difference Patterson

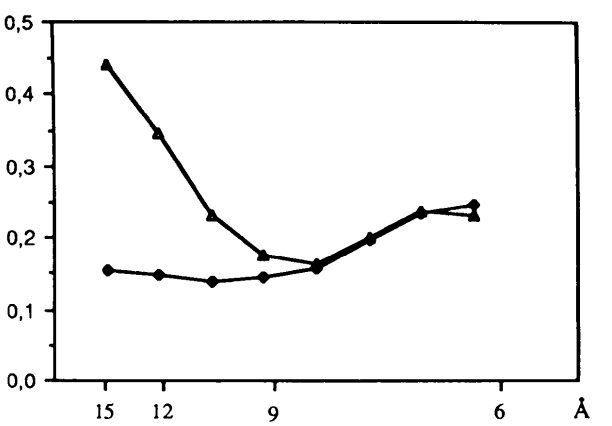

Fig. 2. Relative differences in amplitude as a function of resolution between heavy-atom derivative and native data sets: (△) 5 mM K_2PtCl_4 and (◆) 10 mM $KAu(CN)_2$.

maps for the two derivatives were calculated using data from 15 to 6 Å resolution at a grid spacing of 1 Å. The maps were searched within one icosahedral asymmetric unit for vectors between positions related by non-crystallographic and crystallographic symmetry. In this way the sum of the Patterson function values at 1740 positions was calculated for each position in the icosahedral asymmetric unit. Peaks were only found at special positions in the icosahedral asymmetric unit, on the threefold and fivefold axes and on the planes containing these axes. All these peaks turned out to be false peaks.

Molecular replacement with an SBMV model

Construction of the model

Owing to the failure to locate any true heavy-atom positions in the tested heavy-atom derivatives, we decided to try to solve the MS2 structure using the molecular replacement technique. Structures of several icosahedral RNA viruses of plant or animal origin have been determined and they share a common structural motif in that the capsid proteins form an eight-stranded antiparallel β-sandwich of identical topology (Liljas, 1986). Assuming also that MS2 would have the same topology, we constructed a low-resolution model of the MS2 protein shell based on the known SBMV structure (Abad-Zapatero *et al.*, 1980; Silva & Rossmann, 1987). There are two problems in the construction of such a model. Firstly, the size of the MS2 coat protein (129 amino acids) is much smaller than that of SBMV, which has 196 amino-acid residues in its shell domain with conserved topology. Secondly, the icosahedral protein shells of SBMV and MS2 differ in size, having outer diameters of 290 and 275 Å, respectively. To produce a model of the subunit with a resonable size, all the helices and most of the loops were removed. The model used consisted of an eight-stranded antiparallel β-sandwich of 90 amino acids (Fig. 3). We then tried to position the subunit model at the correct radius by making an R-value search. Since SBMV is also a $T = 3$ virus, an icosahedral asymmetric unit could be built by placing this model at the positions of the three independent (A, B and C) subunits of SBMV. The change in radius was achieved by moving pentamers of A subunits and hexamers of B and C subunits radially along the fivefold and the threefold axis, making the assumption that the pentameric and hexameric contacts of the subunits would be conserved. The crystal asymmetric unit was produced using the icosahedral symmetry. Electron density was generated and structure factors to 10 Å were calculated for models of the protein shell where the subunits were moved inwards radially in steps of 5 Å. When the calculated structure factors were compared with the observed amplitudes, no clear minimum in R value was obtained, and we therefore decided to move the subunits 18 Å, which was consistent with the difference in radius between the two viruses, and which did not lead to any collisions of the subunits at the dimeric and trimeric contacts (Fig. 3*b*).

Definition of envelopes

Envelopes were defined as described earlier (Liljas *et al.*, 1982) using the averaging program system written by Bricogne (1976) and locally written programs to define an envelope by tracing the outer contours of electron density on sections of the map on the tablet of an Evans & Sutherland PS330 graphics system. The first particle envelope was defined as a shell, 46 Å thick, with an outer radius of 137 Å. At later stages of the calculations, the envelope was redefined several times. The envelope was drawn as tightly as possible without cutting off any electron density that belonged to the capsid. The inner limit of the envelope was always a sphere of radius 92 Å. The density outside the envelope was set to zero, both in the interior of the particle and between particles.

Phase refinement

Phase refinement utilizing the non-crystallographic symmetry was performed using a slightly modified version of the program system written by Bricogne

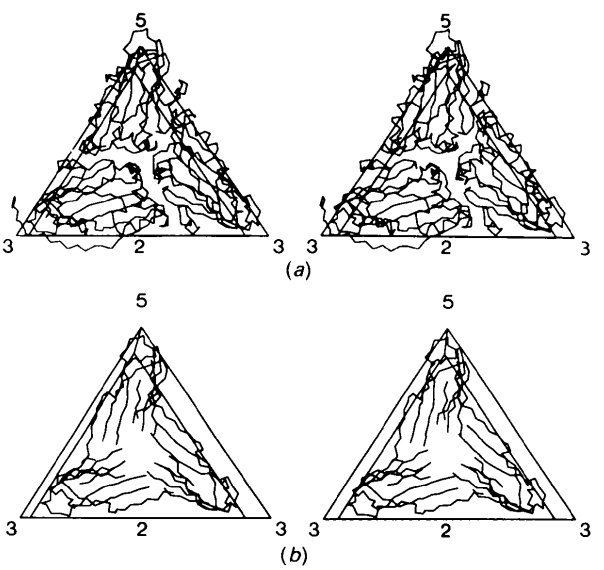

Fig. 3. Stereo drawings of the α-carbon skeleton of the three subunits in the asymmetric unit for (*a*) southern bean mosaic virus (SBMV) and (*b*) the final SBMV-MS2 model. The lines are connecting points at 110 and 128 Å radius on the threefold and fivefold axes.

(1974, 1976). All these calculations were carried out on a MicroVAX 3600. To save computer time, an averaged map for the icosahedral asymmetric unit (1/60th of the virion) was first computed and the crystallographic asymmetric unit (half a virus particle) was then rebuilt from this averaged map (Bricogne, 1976). The grid spacings used for the unaveraged Fourier map and the averaged subunit were always smaller than or equal to a fifth of the actual resolution of the data used. For the rebuilt averaged map the grid spacing was twice as large. After local scaling in several resolution ranges, the calculated structure factors were combined with the observed structure factors in a version of the program *COMBINE* to give weighted ($2F_{obs} - F_{calc}$, α_{calc}) structure factors. A combination of Sim weights and weighting based on the similarity of F_{obs} and F_{calc} was used (Rayment, 1983).

An initial density map was calculated from the model phases using all reflections between 300 and 13 Å. The low-resolution limit of the observed data was about 35 Å and for all missing reflections the calculated structure factors were used, weighted with the average weight of the observed reflections in the same resolution range. Phase refinement was done for several cycles (Fig. 4).

Phase extension

Phase extension from 13 Å to higher resolution was done in steps of less than 1·5 reciprocal lattice points per extension, which corresponds to 0·8 Å at 13 Å and 0·05 Å at 3·5 Å resolution (Fig. 4). During each phase extension, structure factors for the new reflections were calculated based on the current map. The new reflections were scaled and weighted in the same way as the other reflections and included in the structure-factor list. At each extension step, all phases were refined for several cycles using the 30-fold non-crystallographic symmetry. Normally the phases of the new reflections had converged after five cycles of averaging.

Occasionally during the phase extension the data set was extended, the cell dimensions were corrected by post-refinement of the film data, and the envelope was modified. To further improve the phases using the new information provided by the added reflections or improved parameters, the phase extension was on several occasions repeated from a lower resolution, but starting with the current phases. As judged from the statistics, this was the only way to improve the phases in a certain resolution range once convergence had occurred.

The final R factor and correlation coefficient to 3·4 Å resolution were 0·204 and 0·890, respectively (Fig. 4). By comparison with the statistics obtained from earlier phase-extension calculations (Hogle *et al.*, 1986; *Luo et al.*, 1987; Acharya *et al.*, 1989; Chen *et al.*, 1989), the statistics for the MS2 phase extension seemed reasonable. Nevertheless, the 3·4 Å averaged electron density map was not interpretable, and the expected approximate threefold symmetry due to the quasi-threefold axis was not obvious. Maps were calculated using reflections in various resolution ranges, but none of them showed any clear threefold symmetry.

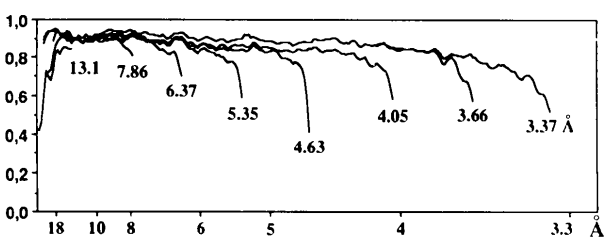

Fig. 4. Phase refinement and extension based on molecular replacement using the SBMV-MS2 model. Plot of the correlation coefficients $\{(\langle F_o F_c\rangle - \langle F_o\rangle\langle F_c\rangle)/[(\langle F_o^2\rangle - \langle F_o\rangle^2)(\langle F_c^2\rangle - \langle F_c\rangle^2)]^{1/2}\}$ as a function of resolution for a selection of cycles.

Isomorphous replacement and phase extension

Difference electron density maps of the two heavy-atom derivatives, 5 mM K_2PtCl_4 and 10 mM $KAu(CN)_2$, were calculated using the phases obtained by phase refinement and extension from the MS2-SBMV model (the 'SBMV' phases), assuming that these phases were correct to some extent. Fig. 5 shows those sections of these maps which contain the major peaks. These peaks were all at negative electron density, but they were found at positions agreeing with $T = 3$ symmetry. We therefore assumed that the peaks obtained represented

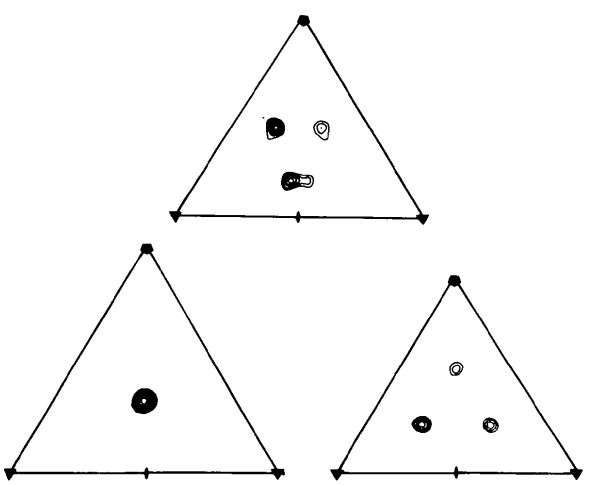

Fig. 5. Sections of difference Fourier maps for 10 mM $KAu(CN)_2$ (top) and 5 mM K_2PtCl_4 (bottom). Top: sites *A*, *B* and *C*; bottom right: sites *D*, *E* and *F*; bottom left: site *G*.

Table 2. *Heavy-atom parameters used for double isomorphous replacement phasing*

Derivative	Site	Fractional cell coordinates x	y	z	Relative occupancy	Binding site
$KAu(CN)_2$	*A*	0·0568	0·4015	0·1376	16·0	Cys *A*46
	B	0·0221	0·4042	0·0551	16·0	Cys *C*46
	C	−0·0295	0·3935	−0·1343	7·0	Cys *B*46
K_2PtCl_4	*D*	0·0466	0·3894	0·0774	11·0	Arg *A*38
	E	−0·0053	0·3897	−0·0984	10·5	Arg *C*38
	F	−0·0519	0·3897	−0·1586	11·8	Arg *B*38
	G	0·0356	0·4518	0·1116	23·3	N-terminus of *A* subunit

difference electron density from normal heavy-atom substitution but that systematic errors in the phases had caused the negative sign (see below). Table 2 shows the positions and relative occupancy of these peaks. The occupancies, but not the positions were refined, constraining the relative occupancies between the sites to the values calculated from the relative heights of the difference electron density peaks. We also checked if the peaks obtained were represented in the Patterson search maps from the *VECTORVERIFICATION* program. Only site *G* gave a noticeable peak in these calculations, but it was still not the highest of all the peaks found at general positions.

Isomorphous replacement phases were calculated from 20·0 to 6·0 Å resolution and a native electron density map was calculated based on the partial data set which was phased in the isomorphous replacement. The phases of all reflections were refined for 13 cycles using the 30-fold non-crystallographic symmetry. The final map was very distinct and showed clearly the quasi-threefold nature of the icosahedral asymmetric unit. The statistics were also considerably improved compared to the molecular replacement phase refinement, especially at a resolution lower than 8·8 Å. A new map was therefore calculated at this resolution, and phase extension from 8·8 to 3·3 Å resolution was performed in 101 steps (Fig. 6*a*). The phases were extended as described above, but in steps of less than 0·45 reciprocal lattice points. For the total set of 416 185 independent reflections in the range of 20·0 to 3·3 Å a final overall R factor of 0·157 and a correlation coefficient of 0·921 were obtained. Fig. 6(*b*) shows the improvement in correlation coefficient after phase extension from the isomorphous replacement model, compared to the values obtained from the SBMV-MS2 model.

Comparison of phases obtained by molecular replacement and isomorphous replacement

A comparison of the final phase sets obtained by phase extension from the SBMV-MS2 model and by phase extension from the isomorphous replacement map clearly shows that the majority of the phases out to 5 Å resolution differ by roughly 180° ($\alpha_{SBMV} = 180° + \alpha_{MIR}$, Fig. 7). The final 'MIR' phases are probably mostly correct, as judged by the high correlation coefficients and the quality of the map. Thus, to this resolution the electron density obtained after phase extension from the SBMV-MS2 model is mostly the Babinet opposite of the true structure. For some reflections, the relation of the two phases corresponds to other solutions, namely the opposite hand ($\alpha_{SBMV} = -\alpha_{MIR}$) or the Babinet opposite with opposite hand ($\alpha_{SBMV} = 180 - \alpha_{MIR}$). Curiously, only the real solution is not represented in the SBMV-MS2 model phase set. This is true at all resolution ranges to 5 Å, and these different solutions have thus co-existed during the phase extension from the original model. In contrast, the 'SBMV' phases at higher resolution are more or less random compared to the final 'MIR' phases. Most of the phases between 35 and 13 Å did not converge to values corresponding to the Babient opposite of the true structure at the initial refinement; only the phases of reflections at higher resolution converged to these values.

To test the influence of the envelope on the convergence we have repeated the initial phase refinement. In these calculations we initially used a similar spherical envelope, but slightly improved cell dimensions and the final larger data set after post-

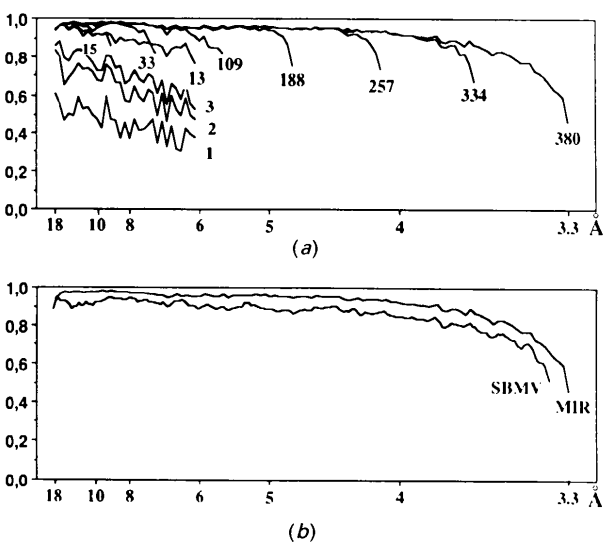

Fig. 6. (*a*) Phase refinement and extension based on isomorphous replacement. Correlation coefficients as a function of resolution for a selection of cycles. The curves are denoted by the cycle number. The first 13 cycles were at 6 Å resolution. The phase extension started at 8·8 Å resolution (cycle 15). (*b*) Comparison of the final correlation coefficients as a function of resolution for structure factors from phase extension based on molecular replacement (SBMV) and from isomorphous replacement (MIR).

refinement. However, after phase refinement and extension to 5 Å the phases had converged to a phase set which corresponded to the true structure rather than the Babinet opposite, although again some reflections in this case obtained phases corresponding to other phase sets (Fig. 7*d*). The electron density map obtained from this phase extension (Fig. 8*d*) is clearly similar to the map obtained using heavy-atom derivatives (Fig. 8*c*).

Interpretation of electron density map and model building

The polypeptide chain of the MS2 coat protein was first traced in a 3·8 Å resolution mini-map. The quality of the electron density map was such that it was easy to follow the chain in all three independent coat-protein subunits. A portion of the electron density is shown in Fig. 9. The complete atomic model

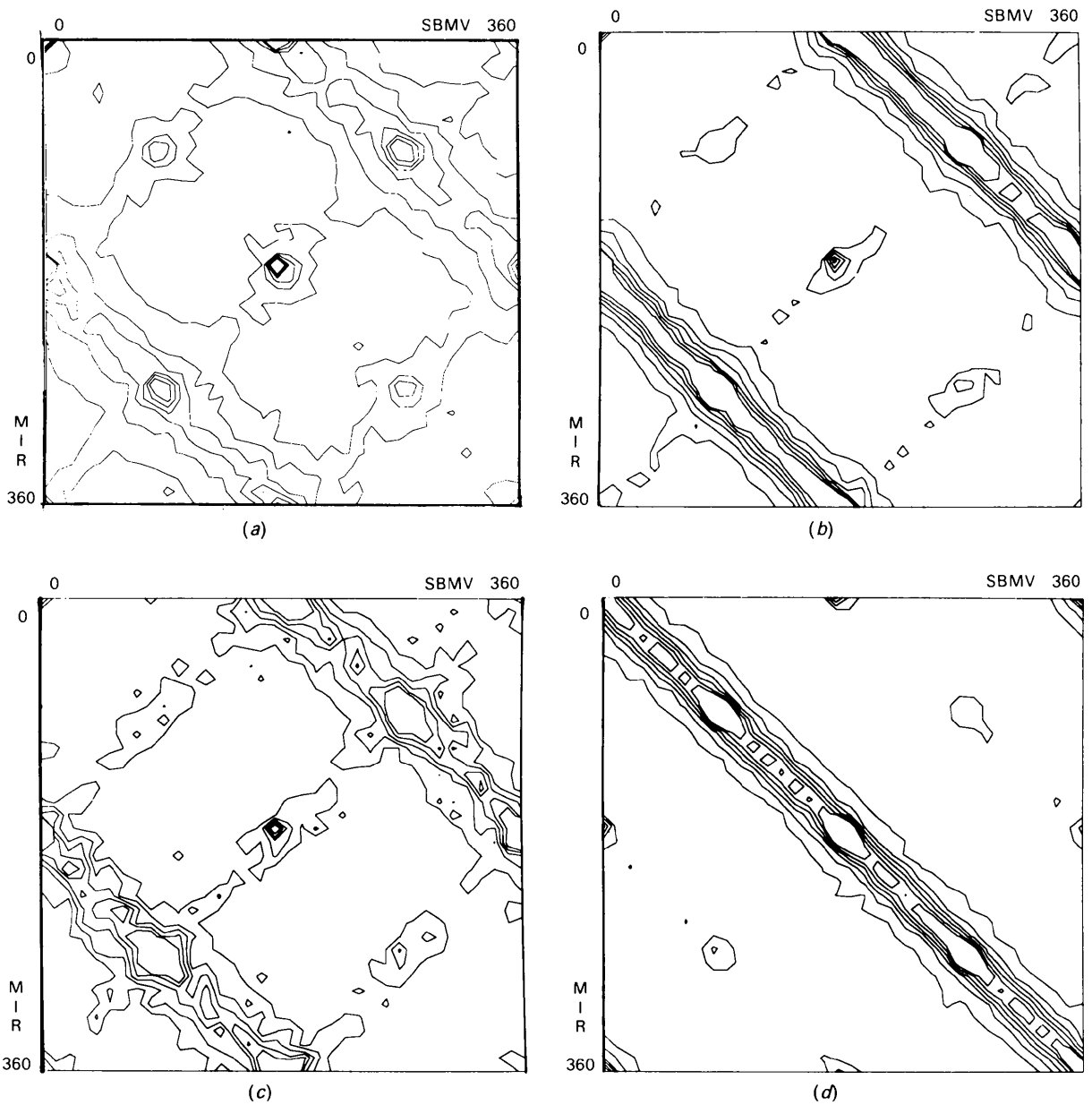

Fig. 7. Comparison of phase angles between the phase set after phase extension from the SBMV-MS2 model and the final phase set based on isomorphous replacement. The diagram shows a plot contoured at the number of reflections with a certain combination of phase angles α_{SBMV} and α_{MIR}, contoured at suitable levels. (*a*) 15–9 Å; (*b*) 9–6·4 Å; (*c*) 6·4–5·2 Å. (*d*) Shows the phase-angle comparison after the repeated phase refinement using slightly improved data and parameters (6·4–5·2 Å).

was built in the final 3·3 Å resolution electron density map using an Evans and Sutherland PS330/PS390 graphics system and the program *O*, developed by Alwyn Jones and coworkers (Jones, Bergdoll & Kjeldgaard, 1990). Fig. 10 shows an α-carbon tracing of one MS2 coat-protein subunit. All residues in the three subunits were visible except the N terminus of the *B* subunit, but the four N-terminal residues and the loop between the first two strands (residues Gly13 to Thr15) were poorly defined in all subunits. The RNA core has no interpretable density, indicating that the nucleic acid of the virion lacks the icosahedral symmetry necessary for detection by the present method.

Even if most of the phases of the SBMV-MS2 model phase extension converged to values roughly corresponding to the Babinet opposite of the true structure, a comparison of this map, contoured at negative density, shows that is is clearly inferior to the isomorphous replacement map (Fig. 8). This 'SBMV' map, contoured at positive density, shows density in the protein region, but has slightly less continuous density than the same map, contoured at negative density.

Discussion

Our results indicate that the structure of a crystal having high non-crystallographic symmetry can be solved by phase extension from a very crude low-resolution model. In our case the initial model was based on an unrelated virus, and the phases were refined and extended using the non-crystallographic symmetry. However, the procedure was only partly

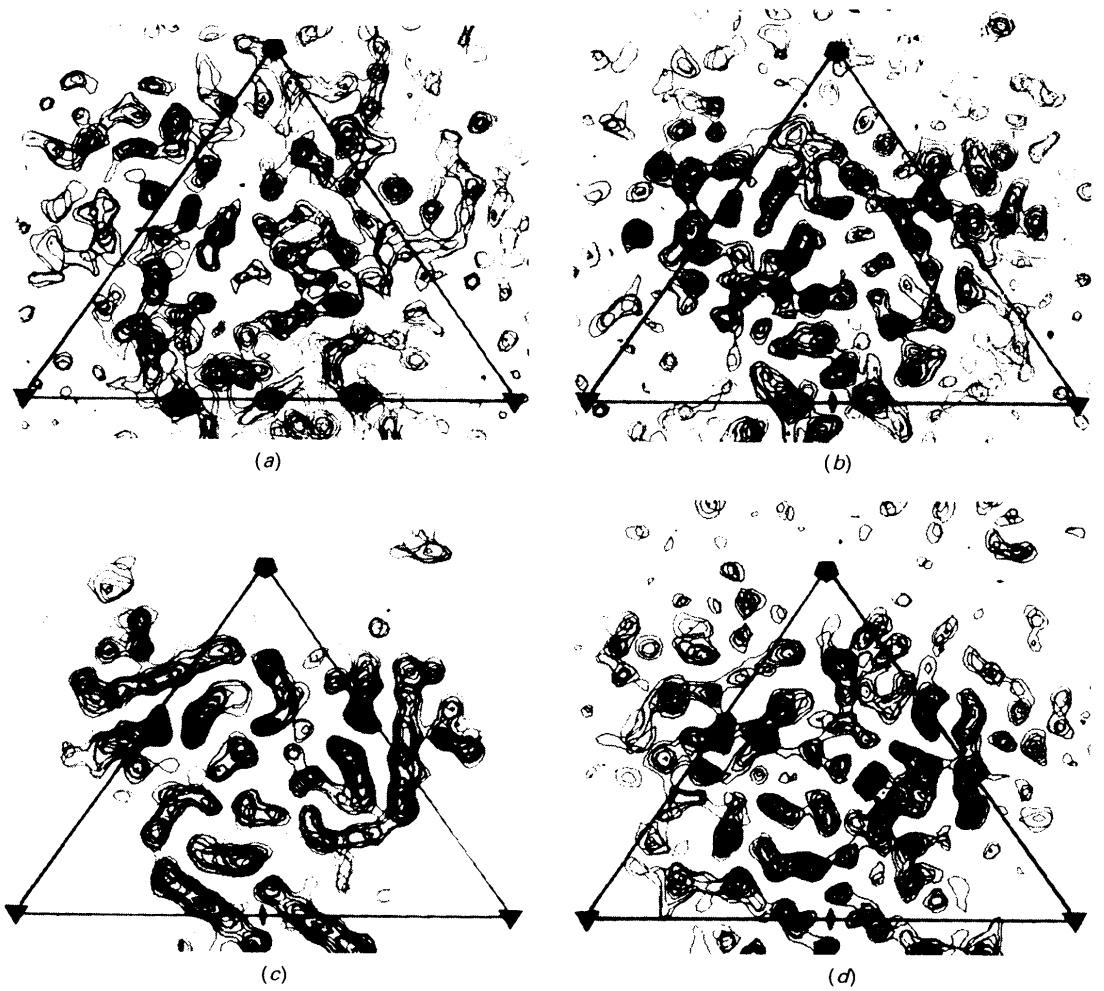

Fig. 8. Electron density sections of 5 Å resolution averaged maps, based on molecular replacement: (*a*) positively contoured and (*b*) negatively contoured; (*c*) from isomorphous replacement; (*d*) positively contoured map obtained at the repeated phase refinement using slightly improved data and parameters. The sections are cut perpendicular to an icosahedral twofold axis between 119 and 124 Å from the particle centre. A triangle covering one icosahedral asymmetric unit is included.

successful and most of the phases converged to values corresponding to the Babinet opposite of the true structure. Since the true structure, its Babinet opposite and the mirror images of these structures have the same icosahedral symmetry, the four different phase sets $\alpha_1 = \alpha_{true}$, $\alpha_2 = -\alpha_{true}$, $\alpha_3 = 180 + \alpha_{true}$ and $\alpha_4 = 180 - \alpha_{true}$ will fulfil the phase constraints imposed by the non-crystallographic symmetry equally well. When phases are extended from a low-resolution model and refined using the icosahedral symmetry, the phases might therefore converge to any of these sets depending on the initial model. The choice of phase set will also depend on how those very low-resolution terms are treated, as these terms contain information about the contrast between the protein shell, the particle interior and the surrounding buffer. If the initial model is sufficiently good, only the first phase set should be obtained.

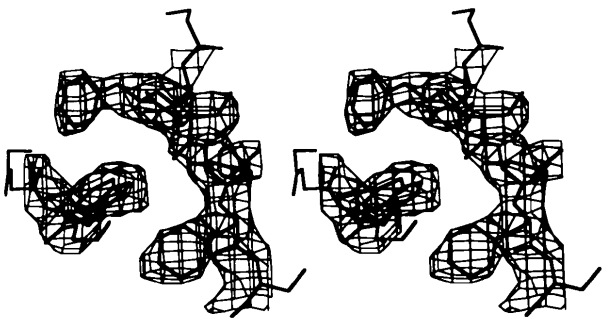

Fig. 9. Portion of the *C* subunit showing the electron density for residues Phe4, Thr5, Gln6 and Phe7 (left), and Trp32 (right).

The initial model was in our case clearly different from the true structure. The treatment of the low-resolution terms did not greatly favour the true structure over its Babinet opposite, because we had no observations of the reflections below 35 Å resolution. In addition, the F_{000} term was not included in the Fourier calculation, which gives an average value of zero for the map, and we used a zero value for the electron density outside the envelope.

If a uniform shell of density had been used as the initial model, the initial phases would have been centric. We might then have expected problems in separating the true structure from its mirror image,

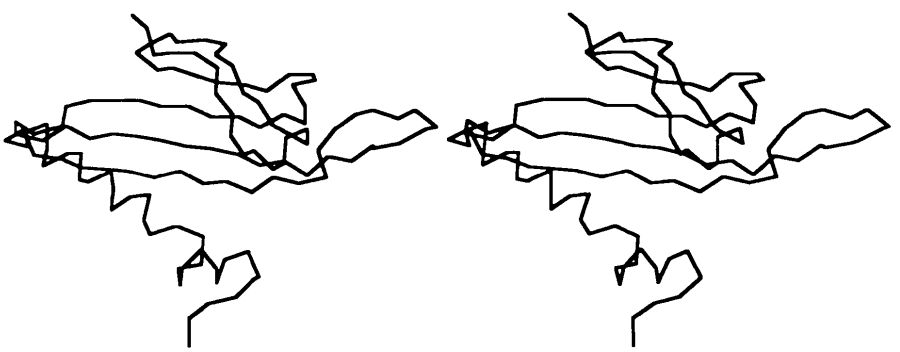

Fig. 10. Stereo drawings of the main-chain conformation of an MS2 *C* subunit as viewed along the twofold axis.

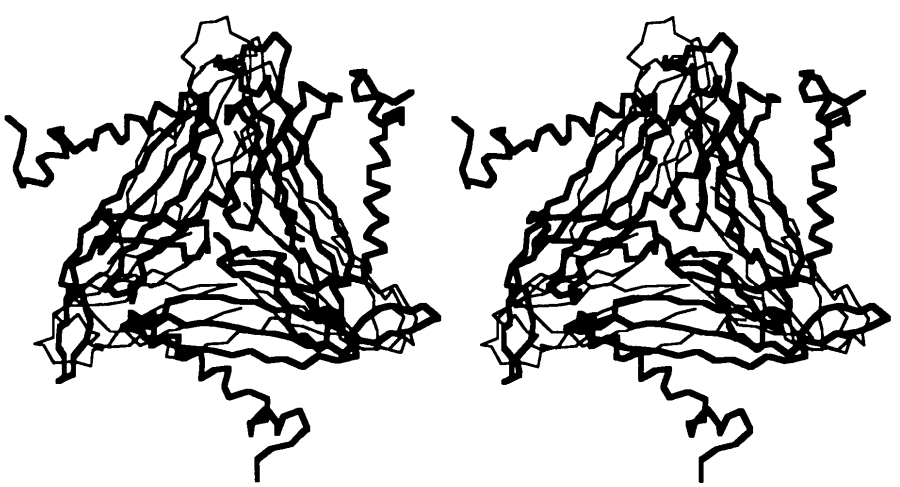

Fig. 11. Stereo diagrams of α-carbon models of the three independent MS2 coat-protein subunits (thick line) superimposed with the corresponding SBMV-MS2 model (thin line).

especially so because the packing of the icosahedra is also centric in this case (*cf* Nordman, 1980).

Our results show that phase sets corresponding to different solutions can co-exist during the phase refinement. This seems to indicate that the phase refinement does not constrain the phases globally, but that some reflections can form a subgroup corresponding to another phase set than that of the majority of the reflections, and that this subgroup includes new reflections at the next phase extension.

It is not clear why the Babinet opposite of the true structure became the dominating solution in our case. A model where a major feature is at a slightly different radius than a similar feature in the true structure might lead to this solution. However, a comparison of the SBMV-MS2 model and the final MS2 model (Fig. 11) shows that though they are quite different, which is quite obvious from their completely different topology, the MS2 β-sheet is more or less at the same radius as the inner sheet of the SBMV-MS2 model. The fact that we obtained a completely different phase set starting from the same model and using a very similar procedure shows that the choice of phase set at the phase extension is quite sensitive to variations in envelope definition and possibly to other factors.

Are the errors which made the map based on the 'SBMV' phases uninterpretable only due to the incorrect starting model, or are they partly or completely due to errors in the envelope definitions? The initial envelope was crude and also had some errors. In contrast, the final envelope closely followed the outer contours of the protein. We also noted that an exact definition was critical for the convergence of the phase extension from the isomorphous replacement map. The vast amount of computing necessary to extend the phases from 13 Å to a resolution where the polypeptide chain can be traced has prevented us from making any thorough analysis of the influence of various factors on the convergence. However, the results presented in this paper indicate that in certain cases, phase extension using non-crystallographic symmetry might lead to an interpretable electron density map even with a crude initial low-resolution model.

We thank Professor Bror Strandberg for support and interest. Pär Nordlund kindly helped us with setting up the programs to transfer film data to the *CCP*4 programs. We thank D. Caspar and J. Johnson for pointing out that the MS2 crystals might be rhombohedral rather than monoclinic. We thank Alwyn Jones for introducing us to *O*. Comments on similar studies made by David Stuart and Michael Rossmann have been of importance in this work. We thank Seved Lövgren for drawing several of the illustrations and Janet Newman for linguistic corrections. The coordinates of the MS2 coat protein have been depostied with the Protein Data Bank* from which copies are available. This work has been supported by the Swedish Natural Science Research Council and the Swedish Council for Planning and Coordination of Research.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1MS2, R1MS2SF), 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 37049 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 1 January 1993.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33–39.

ACHARYA, R., FRY, E., STUART, D., FOX, G., ROWLANDS, D. & BROWN, F. (1989). *Nature (London)*, **337**, 709–716.

AKIMOTO, T., WAGNER, M. A., JOHNSON, J. E. & ROSSMANN, M. G. (1975). *J. Ultrastruct. Res.* **53**, 306–318.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

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

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

CHEN, Z., STAUFFACHER, C., LI, Y., SCHMIDT, T., BOMU, W., KAMER, G., SHANKS, M., LOMONOSOFF, G. & JOHNSON, J. E. (1989). *Science*, **245**, 154–159.

HARRISON, S. C., OLSON, A. J. SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368–373.

HOGLE, J. M., CHOW, M. & FILMAN, D. J. (1985). *Science*, **229**, 1358–1365.

HOGLE, J. M., MAEDA, A. & HARRISON, S. C. (1986). *J. Mol. Biol.* **191**, 625–638.

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394–400.

JONES, T. A., BERGDOLL, M. & KJELDGAARD, M. (1990). *Crystallographic and Modeling Methods in Molecular Design*, edited by C. BUGG & S. EALICK, pp. 189–199. New York: Springer-Verlag.

LILJAS, L. (1986). *Prog. Biophys. Mol. Biol.* **48**, 1–36.

LILJAS, L., UNGE, T., JONES, T. A., FRIDBORG, K., LÖVGREN, S., SKOGLUND, U. & STRANDBERG, B. (1982). *J. Mol. Biol.* **159**, 93–108.

LUO, M., VRIEND, G., KAMER, K., MINOR, I., ARNOLD, E., ROSSMANN, M. G., BOEGE, U., SCRABA, D. G., DUKE, G. M. & PALMENBERG, A. C. (1987). *Science*, **235**, 182–191.

NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747–754.

NYBORG, J. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 139–152. Amsterdam: North-Holland.

RAYMENT, I. (1983). *Acta Cryst.* A**39**, 102–116.

RAYMENT, I., BAKER, T. S., CASPAR, D. L. D. & MURAKAMI, W. T. (1982). *Nature (London)*, **295**, 110–115.

ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H.-J., JOHNSON, J. E., KAMER, G., LOU, 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., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

SCHMID, M. F., WEAVER, L. H., HOLMES, M. A., GRÜTTER, M. G., OHLENDORF, D. H., REYNOLDS, R. A., REMINGTON, S. J. & MATTHEWS, B. W. (1981). *Acta Cryst.* A37, 701–710.
SILVA, A. M. & ROSSMANN, M. G. (1987). *J. Mol. Biol.* **197**, 69–87.
STEIGEMANN, W. (1974). PhD Thesis, Technische Univ., München, Germany.
TOLLIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 872–876.
VALEGÅRD, K., LILJAS, L., FRIDBORG, K. & UNGE, T. (1990). *Nature (London)*, **345**, 36–41.
VALEGÅRD, K., UNGE, T., MONTELIUS, I., STRANDBERG, B. & FIERS, W. (1986). *J. Mol. Biol.* **190**, 587–591.
WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A35, 901–911.

Acta Cryst. (1991). B47, 960–968

The Variety of X-ray Diffuse Scattering from Macromolecular Crystals and its Respective Components

BY I. D. GLOVER

Department of Physics, University of Keele, Keele, Staffordshire ST5 5BG, England, and SERC Daresbury Laboratory, Daresbury, Cheshire WA4 4AD, England

G. W. HARRIS*

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

J. R. HELLIWELL

Department of Chemistry, University of Manchester, Manchester M13 9PL, England, and SERC Daresbury Laboratory, Daresbury, Cheshire WA4 4AD, England

AND D. S. MOSS

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

(*Received* 30 *November* 1989; *accepted* 16 *April* 1991)

Abstract

A range of protein single-crystal diffraction patterns, recorded with intense collimated synchrotron X-radiation are presented. These patterns illustrate the rich and varied nature of the diffuse scattering from this kind of crystal and suggest that the continuous background diffraction will be a source of specific information on the molecular dynamics and flexibility of particular proteins. The relative contributions of the background X-ray diffuse scattering from the solvent, the glass capillary and the sample have been quantified for two of the samples; the so-called solvent ring is shown to be due principally to protein disorder in the crystal both because of the intensity and the marked anisotropy of the ring in specific cases. The form of the acoustic scattering associated with the Bragg peaks was studied in ribonuclease, and is shown, at low resolution at least, to be explained in terms of single-phonon interactions.

* Present address: AFRC Institute of Food Research, Shinfield, Reading, Berkshire RG2 9AT, England.

1. Introduction

A characteristic feature of protein crystals is that their diffraction patterns often extend to limited resolutions. This feature is particularly apparent in medium to large proteins where diffraction data often do not reach 2 Å resolution. This restricted quantity of Bragg data becomes a limiting factor when studying conformational flexibility and it is precisely those proteins where flexibility is likely to be most important which yield fewest Bragg data and exhibit the strongest diffuse scattering.

The diffuse background on an X-ray diffraction pattern may arise from several sources including thermal diffuse scattering, static disorder, solvent disorder, Compton scattering, fluorescence, scattering from mounting tubes, air scattering and the intrinsic film fog. The static or dynamic displacement of atoms in a crystal may be due to a variety of causes, such as the optic and acoustic modes of vibration within the crystal, diffusional motion on the surface of macromolecules (dynamic), or positional and orientational disorder (static) within mol-

0108-7681/91/060960-09$03.00