FEJES, P. L. (1977). *Acta Cryst.* A**33**, 109-113.
FUJIMOTO, F. (1978). *Phys. Status Solidi A*, **45**, 99-106.
HOWIE, A. & BASINSKI, Z. S. (1969). *Philos. Mag.* **149**, 1039-1063.
ISHIZUKA, K. (1980). *Ultramicroscopy*, **5**, 55-65.
KUWANO, N., HAMADA, Y., MANABE, T. & EGUCHI, T. (1984). *J. Electron Microsc.* **33**, 335-343.
MARKS, L. D. (1986). *Ultramicroscopy*, **18**, 33-38.
OHSHIMA, K., MATSUI, M., HARADA, J. & ADACHI, K. (1986). *J. Magn. Mater.* **54-57**, 157-158.
O'KEEFE, N. A. (1979). *37th Ann. Proc. EMSA*, edited by G. W. BAILEY, pp. 556-567. Baton Rouge: Claitor.
RUEDL, E., DELAVIGNETTE, P. & AMELINCKX, S. (1968). *Phys. Status Solidi*, **28**, 305-328.
RÜHLE, M. & WILKENS, M. (1974). *J. Appl. Cryst.* **8**, 223-224.
SHINDO, D. (1982). *Acta Cryst.* A**38**, 310-317.
SPENCE, J. C. H. (1978). *Acta Cryst.* A**34**, 112-116.
TANAKA, N. & COWLEY, J. M. (1986). *Ultramicroscopy*, **17**, 365-378.
TANAKA, N., COWLEY, J. M. & OHSHIMA, K. (1987). *Acta Cryst.* B**43**, 41-48.
TANAKA, N. & OHSHIMA, K. (1984). *Phys. Status Solidi A*, **81**, 129-138.
TANAKA, N., OHSHIMA, K., HARADA, J. & MIHAMA, K. (1979). Proc. Mod. Str. Conf., Hawaii, pp. 292-294.
TERASAKI, O., WOOD, G. J. & WATANABE, D. (1984). *Proc. Mater. Res. Soc. Symp.* **21**, 253-257.
VAN DYCK, D., VAN TENDELOO, G. & AMELINCKX, S. (1982). *Ultramicroscopy*, **10**, 263-280.
VAN TENDELOO, G. & AMELINCKX, S. (1983). *Phys. Status Solidi A*, **77**, K9-K11.
VAN TENDELOO, G., DE RIDDER, R. & AMELINCKX, S. (1978). *Phys. Status Solidi A*, **49**, 655-666.
WARREN, B. E. (1969). *X-ray Diffraction*. Reading, MA: Addison-Wesley.
YAMAGUCHI, S., WATANABE, D. & OGAWA, S. (1961). *J. Phys. Soc. Jpn*, **17**, 1030-1041.

Acta Cryst. (1987). A**43**, 346-361

The Structure Determination of a Common Cold Virus, Human Rhinovirus 14

BY EDWARD ARNOLD, GERRIT VRIEND,* MING LUO, JAMES P. GRIFFITH, GREG KAMER, JOHN W. ERICKSON,† JOHN E. JOHNSON AND MICHAEL G. ROSSMANN

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

(*Received* 21 *May* 1986; *accepted* 12 *November* 1986)

Abstract

The methods used to solve the structure of human rhinovirus 14 at 3·0 Å resolution are described in detail. The crystals are cubic, space group $P2_13$, a = 445·1 Å with 20-fold non-crystallographic redundancy and with approximately 55% solvent and RNA content. The data used to solve the structure were collected at the Cornell High Energy Synchrotron Source (CHESS) using oscillation photography. Most of the computations were performed on Purdue University's Cyber 205 supercomputer. Two heavy-atom derivative data sets from crystals soaked in 1 and 5 m*M* $KAu(CN)_2$ were used to provide isomorphous phasing to 4 Å resolution, although it was subsequently shown that phases beyond 5 Å resolution were random. The phases were refined at 5 Å resolution by five cycles of real-space molecular replacement. Phase extension from 5 to 3 Å was then performed using 60 cycles of real-space molecular replacement while extending the resolution in steps of three reciprocal-lattice points at a time once every three cycles. The 3·5 Å skew-averaged map was easily interpreted and showed 811 of the 855 amino acids in the four distinct viral polypeptide chains. A complete atomic model has been built using *FRODO* on an Evans & Sutherland PS300 graphics system with respect to the 3·08 Å resolution electron density map. The roles of the non-crystallographic symmetry, solvent content, errors in amplitudes, orientation and translation in the molecular replacement process are discussed.

Introduction

The earlier successful determinations of the small RNA plant viruses tomato bushy stunt virus (TBSV; Harrison, Olson, Schutt, Winkler & Bricogne, 1978), southern bean mosaic virus (SBMV; Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980) and satellite tobacco necrosis virus (STNV; Liljas, Unge, Jones, Fridborg, Lövgren, Skoglund & Strandberg, 1982) represented the largest molecular analyses by crystallographic techniques prior to the determination of two animal virus structures (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985; Hogle, Chow & Filman, 1985). The plant virus structures not only laid the technological foundations but also provided a structural framework on which the subsequent work on animal viruses could be based. The major

* Present address: Department of Structural Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

† Present address: Department of Physical Biochemistry, AP-9A D-47E, Abbott Laboratories, Abbott Park, North Chicago, Illinois 60064, USA.

0108-7673/87/030346-16$01.50

additional problems presented in the work on human rhinovirus related to the small amounts of virus that could be produced weekly (2 mg instead of 300 mg), the far greater rate of radiation damage in the X-ray beam symptomatic of an inherently less-stable protein capsid, and the absence of quasi-symmetry. The greater viral capsid fragility also reflected itself in the difficulty of preparing heavy-atom derivatives that did not damage the crystals and were isomorphous at high resolution. These problems were tackled by using synchrotron radiation and a supercomputer. Improved techniques for the X-ray diffraction analysis using oscillation photography at a synchrotron (Vriend, Rossmann, Arnold, Luo, Griffith & Moffat, 1986) and the optimal use of non-crystallographic symmetry for phase determination (Rossmann & Blow, 1963; Main & Rossmann, 1966; Main, 1967; Crowther, 1969; Nordman, 1980; Gaykema, Hol, Vereijken, Soeter, Bak & Beintema, 1984) were developed in conjunction with a Cyber 205 supercomputer.

Human rhinoviruses are classified as picornaviruses (small animal RNA viruses). They are icosahedral, have a diameter of about 300 Å, a molecular weight of about $8{\cdot}5 \times 10^6$ and contain about 30% by weight positive-stranded RNA. The capsid consists of 60 copies of each of four proteins, VP1, VP2, VP3 and VP4, with relative molecular masses of 32 000, 29 000, 26 000 and 7000 daltons respectively. Picornaviruses cause disease in man, animals and insects and, apart from the rhinoviruses, include polio (Hogle *et al.*, 1985), foot-and-mouth disease, Mengo (Luo, Vriend, Kamer, Minor, Arnold, Rossmann, Boege, Scraba, Duke & Palmenberg, 1987), hepatitis A and cricket paralysis virus. The RNA sequence of human rhinovirus 14 (HRV14) has been determined by Stanway, Hughes, Mountford, Minor & Almond (1984) and Callahan, Mizutani & Colonno (1985) from which the protein sequences can be inferred.

The first crystals of HRV14 that were examined diffracted X-rays to high resolution and were in the orthorhombic crystal system. These crystals were shown to be isomorphous with poliovirus crystals at least to 30 Å resolution (Erickson, Frankenberger, Rossmann, Fout, Medappa & Rueckert, 1983; Finch & Klug, 1959). However, closer inspection of the diffraction pattern showed multiple lattices. The second high-resolution HRV14 crystal form was grown from tris/PEG/Ca^{2+} solutions instead of ammonium sulfate (Arnold, Erickson, Fout, Frankenberger, Hecht, Luo, Rossmann & Rueckert, 1984). These cubic HRV14 crystals diffract X-rays to at least 2·5 Å resolution and were used in the structure determination (Rossmann *et al.*, 1985).

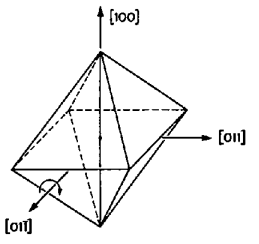


Fig. 1. HRV14 cubic crystal morphology.

Space group and crystal packing

The morphology of cubic HRV14 crystals is shown in Fig. 1. Their space group is $P2_13$ [originally reported as $P23$ by Arnold *et al.* (1984)] with $a = 445{\cdot}1$ Å. Each of the four icosahedral virions is situated on a threefold axis, giving a non-crystallographic redundancy of 20. Each non-crystallographic asymmetric unit contains approximately 90 000 daltons of protein and the approximate fractional volume of solvent and disordered RNA is 55%. The positional parameters for the reference virion are, therefore, constrained to be at (x, x, x). $P2_13$ is the cubic special case of the common orthorhombic space group $P2_12_12_1$ which would have one particle per asymmetric unit. A diagram showing the two most efficient packing arrangements for the four nearly spherical particles in a $P2_13$ cell is shown in Fig. 2. These differ in that the reference particle is positioned near (0, 0, 0) or $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ relative to the conventionally chosen origin. The observed low-resolution face-centering condition dictates that the reference particle center must be close to $x = 0$ or $\frac{1}{4}$.

Initial 5 Å resolution data, collected on rotating-anode generators (Table 1), were used to determine the virion orientation (Fig. 3) as $\kappa = -6{\cdot}0°$ with a locked rotation function (Arnold *et al.*, 1984). This value of the particle orientation parameter was retained throughout the structure determination in all computations involving heavy-atom search and phasing as well as all operations involving non-crystallographic symmetry averaging.

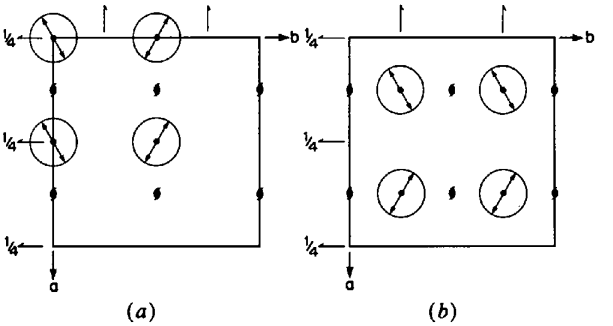


Fig. 2. Packing arrangements shown in *c*-axis projection with reference particle near (*a*) 0, 0, 0 and (*b*) $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ for $P2_13$ cell with conventionally chosen origin. The relative orientations of the icosahedral virions are schematically indicated by the arrows contained in the circles.

Table 1. *Summary of HRV14 cubic crystal data sets*

Derivative conditions	Source	Date	λ	Osc. range (°)	Resolution processed (Å)	Number of films	Number of data	Number of unique data	Percent of possible unique total	*R* factor*
(Native)	Elliott GX20	Apr.–July 1983	1·542	0·8	5·0	29	170 601	82 926	67·3	11·9 (1σ)
Native	CHESS	May 1984	1·57	0·3	2·6	83	2 706 020	505 896	57·7	10·9 (2σ)
$KAu(CN)_2$ 5 m*M* 12 h	CHESS	May, June 1984	1·57	0·8	5·0	11	103 370	55 320	45·3	10·7 (2σ)
	DESY	Aug. 1984	1·40			8				
$KAu(CN)_2$ 1 m*M* 36 h	CHESS	June 1984	1·57	0·6	5·0	8	50 285	35 137	28·5	10·4 (1σ)
$KAu(CN)_2$ 1 m*M* 24 h	CHESS	Nov. 1984	1·57	0·3	2·6	46	814 535	336 319	38·4	10·1 (1σ)

* $R = \left[\sum_h \sum_i |(F_h^2 - F_{hi}^2)| / \sum_h \sum_i F_h^2\right] \times 100$ for reflections with $F_{hi}^2 > n\sigma$ where n is defined in parentheses.

The rotation-function results established the relationship of the icosahedral axes to the crystal axes. For the special value of $\kappa = 0°$, in which some of the icosahedral twofold axes are parallel to the crystal twofold axes, the packing arrangements with the reference virion centered at (0, 0, 0) and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ are indistinguishable. However, the two packing arrangements become distinct when $\kappa \neq 0°$, since the particle contacts are asymmetric, relative to the crystal and icosahedral symmetry elements.

The 20 subunits in the crystallographic asymmetric unit are generated by successive application of the specific five-, two- and twofold axes of non-crystallographic symmetry shown in Fig. 3. Particle building is completed by application of the crystallographic threefold axis of symmetry. The orientation of the non-crystallographic axes used for symmetry generation is also described in Table 2, both before and after application of the κ rotation of $-6·0°$ around the threefold axis.

Data collection

The majority of the cubic HRV14 crystal data sets were collected at the synchrotron sources CHESS (Cornell High Energy Synchrotron Source, in Ithaca, New York) and DESY (Deutsches Electronen Synchrotronen, in Hamburg, West Germany). Table 1 summarizes some essential characteristics of the data sets which were used for the eventual structure determination.

Special consideration was given to transporting HRV14 crystals to the synchrotron sources (see supplemental item No. 1*). A Supper OSCCAM oscillation-precession camera (kindly donated to CHESS by Merck, Sharp & Dohme) was used to record all of the oscillation photographs collected at CHESS on Kodak DEF X-ray film. Typical operating energies were 5·5 GeV with the ring current varying from 40 to 20 mA. A bent triangular monochromator is used for horizontal focusing and wavelength selection and currently a platinum-coated mirror is used for vertical gathering (Schildkamp & Bilderback, 1986). Movable lead slits were used to truncate each edge of the synchrotron beam to enable spot resolution on the photographs ($a = 445$ Å, $F = 100$ mm, $\lambda \simeq 1·57$ Å). The slits are located at the beam entrance to the camera hutch, roughly 80 cm from the camera. Proportional intensities were measured by a removable ionization chamber which was placed in the beam path just downstream of the collimator. At the beginning of each fill, the incoming intensity was reduced according to the formula: 30% total by left and right (horizontal), and 20% by top and bottom (vertical) beam truncation. This scheme of beam-shape definition, when used with a 0·3 mm collimator, produced diffraction spots on a photograph at $F =$ 100 mm with dimensions roughly 300×300 μm. The

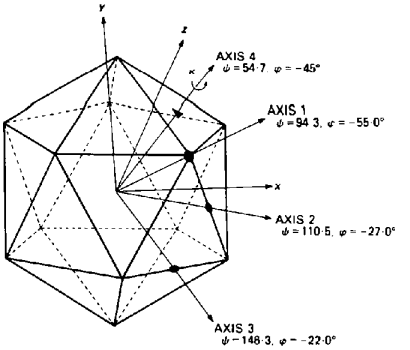


Fig. 3. Definition of κ angle describing reference virion orientation in crystal unit cell, as well as the specific symmetry axes used for icosahedral particle building.

* Additional tables, figures and explanatory notes as detailed in the text have been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 43483 (21 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 2. *Icosahedral symmetry axes used for particle generation and their orientation in the crystal cell*

The specific symmetry axes listed below are shown in Fig. 3.

Symmetry element		$\kappa = 0°$					$\kappa = -6°$				
		Polar coordinates*		Direction cosines			Polar coordinates*		Direction cosines		
Order	Fold	ψ (°)	φ (°)	u	v	w	ψ (°)	φ (°)	u	v	w
1	5	90·0	−58·28253	0·525731	0	0·850651	94·26385	−55·02155	0·571682	−0·074350	0·817100
2	2	108.0	−31·71747	0·809017	−0·309017	0·5	110·50375	−27·00656	0·834512	−0·350269	0·425325
3	2	144·0	−31·71747	0·5	−0·809017	0·309017	146·31115	−22·01381	0·514243	−0·832062	0·207912
4	3	54·73561	−45·0	0·577350	0·577350	0·577350	54·73561	−45·0	0·577350	0·577350	0·577350

* The polar coordinates ψ, φ are defined as in Rossmann & Blow (1962).

lead beam stop was sufficiently thin to permit a weak exposure of the direct beam on an oscillation photograph, which is essential for subsequent processing. The hutch temperature varied between 289 and 293 K for HRV14 crystal experiments.

One data exposure (two films per film pack) was recorded from each crystal and covered an oscillation range of 0·3° for high-resolution photographs. A representative oscillation photograph from the native data set collected at CHESS in May 1984 is shown in Fig. 4. The highest resolution reflections are 2·5 Å, at the corners, and the resolution cutoff at the edge along the spindle axis is 3·0 Å. Most crystals measured at least 0·3 mm from apex to apex through the center although many were larger. Initial orientation of the crystals was performed optically using cues from the octahedral morphology (Fig. 1). Two orthogonal stills of 10–15 s duration were recorded on Polaroid X-ray film (type 57, cassette had fluorescent screen to effect film exposure from X-rays). The first setting photograph was taken near the [011]-zone direction ($\varphi = 90°$) and the second close to the [100] zone ($\varphi = 0°$). The actual setting was usually within one degree of the intended range. Crystal exposure times for the oscillation data photographs were 13 min in May 1984. By November 1984, improvements in the X-ray optics at CHESS permitted the same or higher quality photograph to be obtained in 4 min.* The data exposure time was decided by the requirement that a subsequent still photograph on Polaroid should show similar resolution to a comparable still recorded before the oscillation photograph. To spread the radiation damage uniformly throughout the oscillation range, the spindle rotation rate was adjusted to allow at least 20 integral oscillations to be completed during the exposure.

Although some radiation damage occurred during the data exposures, the crystal generally looked intact immediately following the exposures. When the crystals were observed in the light microscope afterwards, a progression of macroscopic changes were seen. Within 10–15 min, the path that the beam had taken through the crystal became translucent and white, and turned brown several minutes later. A channel would then form as brown material oozed out into the capillary. Finally, the remainder of the crystal often collapsed and fragments of the crystal were strewn as far as 5 mm apart. If a crystal was exposed for even a short time and then not used immediately for a data photograph, the diffracting power of that crystal rapidly diminished. Apparently, both time and radiation dose contribute to damage from X-radiation.

Film processing, scaling and post-refinement

All data sets collected at synchrotron X-ray sources from HRV14 cubic crystals were processed using the Cyber 205 version of the oscillation-film-processing package developed at Purdue University (Rossmann, 1979) (see supplemental item No. 2*). Optical densities were measured and digitized using an Optronics System C-4100 photoscanner using a 50 μm raster. Rough knowledge of the crystal orientation is required to initiate successful processing. At least some of the reflections which are predicted must intersect with the reflections on the film for the refinement procedure to begin. Therefore, if the starting crystal orientation is wrong by more than the oscillation range plus the effective mosaicity in any direction, the procedure cannot work. For a 0·3° oscillation and a 0·07° mosaic spread, the error must be less than 0·5° in any direction. The initial crystal-orientation parameters used for processing the 0·3° oscillation HRV14 films were measured directly from each data exposure. Owing to the narrow oscillation range, each photograph could be treated as a still photograph: the approximate spindle axis setting, φ_y and a starting $\delta\varphi_x$ correction were determined by measurement of the [100] direction on the photograph [see Rossmann (1979) for nomenclature]. An approximate $\delta\varphi_z$ was also determined by measurement of the inclination angle between the spindle axis and the horizontal lattice lines which are parallel to

* By December 1985 the X-ray optics had been improved further requiring reduction of exposure times of 0·3° oscillation photographs taken with 0·2 mm diameter crystals to only 2 min.

* See deposition-footnote.

the spindle-axis direction [01$\bar{1}$]. Fiducial marks were not recorded on the films, but care was taken to orient the film in a standard orientation within the film cassettes, such that the same edge of the film was always parallel to the spindle direction. The direct-beam image provided an origin for predicting the positions of reflections on the film. A start obtained in this way was successful for processing roughly 70 film packs. To handle the more recalcitrant remaining films, an option to the film processing program was constructed which iterates the initial crystal orientation around the best guess in input-specified steps of $\delta\varphi_x$, $\delta\phi_y$ and $\delta\varphi_z$. The remaining films were analyzed in this way, using increments of the order of 0·5° in a range ±1·0° for each mis-setting angle. The 125 or so starts that this procedure generates are then evaluated at low resolution (<12 Å for HRV14 cubic crystals giving several hundred starting reflections). First the initially predicted reflection positions are refined, then the crystal orientation is refined in the convolution setting refinement procedure. One final round of refinement at the newly predicted orientation completes the evaluation. If the variance of observed *versus* calculated positions of reflections was low (less than 0·3 raster steps), if most predicted whole reflections were non-zero measurements, and if the agreement between intensities of symmetry-related reflections (if any) was acceptable, then a given solution was deemed possible and extended to higher resolution.

After the determination of the crystal orientation parameters *via* the convolution setting refinement procedure, reflection intensities were measured using a variable profile-fitting procedure to a maximum

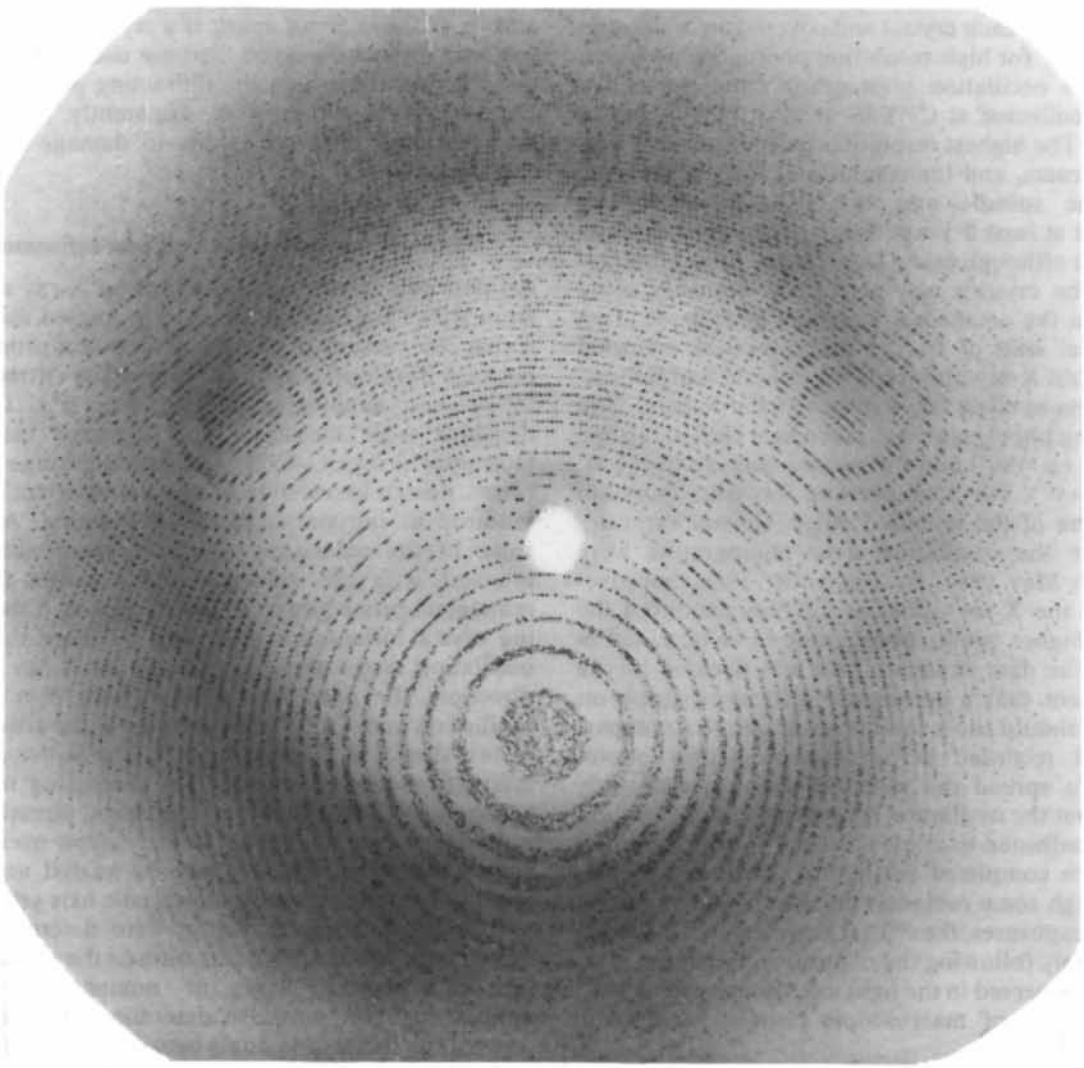


Fig. 4. 0·3° oscillation photograph of HRV14 crystal taken at CHESS. The highest-resolution reflections are 2·5 Å at the corners and resolution cutoff at the edge, along the spindle axis, is 3·0 Å.

resolution of 2·6 Å for the HRV14 native data set. Supplemental item No. 3* summarizes the characteristic parameters for a typical 'A' film (the stronger of a two-film pack).

Improvements were made in the Cyber 205 version of the integration algorithm (Rossmann, 1984). These included procedures for correcting overlapped reflections, measuring overloaded reflections, flagging suspicious measurements of reflections according to a rejection hierarchy and treating the polarization of synchrotron radiation. The detection of overlapped reflections was based on a look-up table derived from the integration mask ('*MBOX*'), which differentiates among peak–peak, peak–background and background–background overlap. Crude estimates, independent of overlap, are determined in the first profile-fitting cycle. In subsequent cycles optical densities relating to a given reflection are corrected using the profile and approximate intensities of nearby reflections. About four cycles were used to achieve reasonable convergence of the intensity estimates. Estimates for 'overloaded' reflections, which contain some optical density values which are beyond the linear range of the film, were made by using the non-overloaded portions of the peak area. The scale used for fitting the profiles of these reflections was determined from the non-overloaded optical densities in the integration area. Weak and otherwise rejected reflections (supplemental item No. 3) were retained but flagged to indicate the nature and severity of the problem. In this way, for instance, weak observations may be distinguished from entirely unmeasured reflections. The estimates of weak and strong reflections are valuable in the molecular replacement procedure in assessing the reliability of calculated amplitudes for reflections without a reliable observed amplitude measurement.

The polarization factor used in evaluating the synchrotron films was

$$P=\tfrac{1}{2}(1+\cos^2 2\theta)-\tfrac{1}{2}J'(\cos 2p)(\sin^2 2\theta),$$

with the same nomenclature as in Kahn, Fourme, Gadft, Janin & André (1982). Here J' is the degree of polarization of the synchrotron radiation and was assumed to be 0·903. When $J'=0$ there is no polarization, while when $J'=1$ there is complete plane polarization. It is improbable that $J'<0{\cdot}8$, although the value varies according to the synchrotron source, its optical system and the position of the crystal in the beam. Nevertheless, error due to a wrong choice of J' is unlikely to cause an error of greater than 1·0% in P at 3·5 Å resolution, or 1·6% at 2·8 Å resolution.

The wavelength of the synchrotron radiation was determined by post-refinement of the apparent cell dimensions relative to the reference value established from Cu $K\alpha$ radiation (Table 1) (Vriend *et al.*, 1986). After the wavelength was determined in this way from a subset of films, the films were processed using the updated estimate of the wavelength.

Initial measurements of intensities to 5 Å resolution on a given film were used to determine the hand of the reciprocal axial system by scaling against a small set of standard films. This established whether the film should be indexed as $I(hkl)$ or $I(hlk)$. Processing of that film was then extended to 2·6 Å resolution. The effective isotropic-mosaic-spread value, m, used in integration was 0·075° for the HRV14 native data set. The weaker, or 'B', film of the two-film pack was then processed to 4 Å resolution using the same crystal orientation matrix.

Improvements were made in the Cyber 205 version of the film–film scaling procedure (Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979) to take into account the differential horizontal and vertical crossfire of synchrotron radiation (Vriend *et al.*, 1986) and to incorporate the flagged reflections from the processing procedure. First, the A (strong) and B (weak) films in a film pack were scaled together using the middle intensity range. Then linear scale factors were determined using whole reflections with $F^2>4\sigma(F^2)$. Cycles of scale-factor determination alternated with post-refinement (Rossmann *et al.*, 1979) of individual crystal cell dimensions, crystal orientation parameters and the effective horizontal and vertical mosaic spreads [see supplemental item No. 4* and Vriend *et al.* (1986)]. Finally, the scaled reflections with $F^2>1\sigma(F^2)$ and with a partiality greater than 0·5 were averaged. Included here also were the reflections whose intensities had been flagged as questionable in the film processing procedure. These were weighted according to the degree of uncertainty associated with the given type of problem. The weights were adjusted by inspection of the R factors appropriate to each type of problem (see supplemental item No. 5*). Overlapped reflections and reflections overloaded only in the peak area were worthy of inclusion in the data set, while other kinds of errors gave poorer results. Observations that had substantial disagreement (*cf.* Rossmann *et al.*, 1979) from the overall mean were rejected. 83 film packs were eventually included in the final data set, giving an R factor, defined as

$$\left(\sum_h\sum_i|(F_h^2-F_{hi}^2)|/\sum_h\sum_i F_h^2\right)\times 100,$$

of 11·0% for 505 896 unique reflections from a set of around $2{\cdot}7\times10^6$ observations.

Results for seven typical films are shown in supplemental item No. 6*. The average number of rejected

* See deposition footnote.

* See deposition footnote.

Table 3. *Final native data set – percentage of data at different resolutions*

Resolution	Unique total	Percent $F^2>1\sigma$	Percent $F^2>10\sigma$
∞–30	359	63	72
30–15	3 443	86	85
15–10	9 107	84	85
10–7·5	17 535	83	78
7·5–5	70 026	81	70
5–3·5	184 857	78	60
3·5–3·0	133 096	63	28
3–2·75	60 638	36	8
2·75–2·6	26 835	20	4
Total	505 896	57·7%	(out of possible 876 011 to 2·60 Å resolution)

reflections is about 3% for full and roughly 10% for partial observations. A useful factor for determining the quality of a given film was found to be $g_i = (\sum_j R_{ij} n_{ij})/\sum_j n_{ij}$ (see supplemental item No. 6). On the average each reflection was measured about twice (see supplemental item No. 7*). The percentage of data collected in different resolution shells is shown in Table 3. The data are at least two-thirds complete to 3·0 Å resolution, but then drop off rapidly to 2·6 Å resolution. Indeed, beyond 3·0 Å there are few data since most of the reflections beyond 3 Å were off the edge of the film.

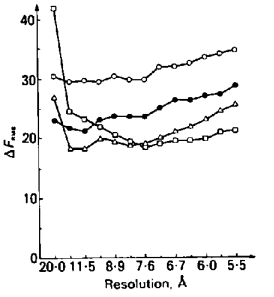
Analysis of isomorphous heavy-atom derivatives

A number of heavy-atom-derivative data sets of HRV14 cubic crystals were collected and analyzed. Degradation due to radiation damage precluded the possibility of using precession photography for surveying derivative conditions. However, the measured intensities from a single oscillation photograph scaled to the native data set provided a signature for a strongly substituted compound.

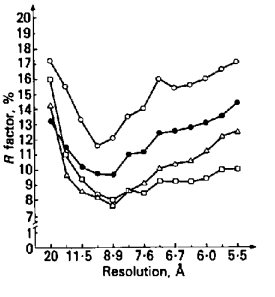
A rapid survey of 27 possible heavy-atom derivatives was made at CHESS in May 1984. Small crystals were soaked in wells containing 0·5, 5 and 50 m*M* concentration of the candidate heavy-metal reagent. After 24 h of soaking, several reagents showed potential, namely those for which crystals still diffracted after soaking at 5 m*M* but not at the higher reagent concentration. A partial low-resolution (5 Å) data set was therefore collected from crystals soaked in 5 m*M* $KAu(CN)_2$.

For the 1 m*M* $KAu(CN)_2$ derivative data set, collected at CHESS in November 1984, two crystals were soaked per hanging drop containing 15 μl of a 1 m*M* $KAu(CN)_2$ solution (in 0·5% PEG 8000, 10 m*M* tris, pH 7·3). The soaking crystals were kept at room temperature (approximately 298 K) for 12 h and then placed in a refrigerator at 277 K for another 12 h. The soaking was terminated by mounting and draining each crystal. The procedure used for collecting data from crystals soaked in 5 m*M* $KAu(CN)_2$ was analogous except that the entire soak was 12 h at room temperature.

Scaling of native and derivative data was performed with mean local scaling in resolution ranges. Each of the derivative data sets exhibited increasing differences beyond a given resolution suggesting lack of isomorphism (Fig. 5). For example, the 1 m*M* $KAu(CN)_2$ data followed approximately a scattering-factor curve only to around 6 Å resolution. The lack of isomorphism for the 5 m*M* $KAu(CN)_2$ was also indicated by a decrease in cell dimension of nearly 2 Å as determined by post-refinement. The size of the differences should also be essentially independent of the F magnitude at a given resolution. However, for each HRV14 data set, within a given resolution range, the small number of reflections with the largest F magnitudes were found to have anomalously large differences. Hence, reflections beyond a given resolution and larger than $k\hat{F}$(native) were omitted in computing difference Patterson functions. For example, the $KAu(CN)_2$ difference Patterson function was based on resolution limits of 15–6 Å and $k = 2{\cdot}6$. In addition, differences larger than



(*a*)



(*b*)

Fig. 5. (*a*) Magnitude of differences between heavy-atom derivative and native data sets. Note the increase of differences at higher resolution showing lack of isomorphism. (*b*) R factor relating heavy-atom derivative differences to native amplitudes as a function of resolution.

$$R = \sum ||F_N| - k|F_H|| / \sum |F_N| \times 100.$$

○ $KAu(CN)_2$ 5 m*M*/12 h; ● $KAu(CN)_2$ 1 m*M*/June 1984; △ $K_2UO_2F_5$ 10 eq/8 days; □ $KAu(CN)_2$ 1 m*M*/Nov 1984.

* See deposition footnote.

$0{\cdot}8\hat{F}$(native) were also rejected. In general, about 2% of the difference terms were omitted in this manner. Most of these rejections were reflections with large F, not large differences.

Within the group of accepted differences, we had some indications, from reciprocal-space (Rossmann, Arnold & Vriend, 1986) and Patterson search methods, that the larger differences ($\Delta F > 2\hat{\Delta F}$) contained a stronger signal. The non-crystallographic symmetry heavy-atom vector search procedure (Argos & Rossmann, 1974, 1976) was adapted for the Cyber 205. Difference Patterson functions were computed over one-eighth of a full cell ($0 \le u \le \frac{1}{2}, 0 \le v \le \frac{1}{2}$, $0 \le w \le \frac{1}{2}$) which contains three asymmetric units. The Patterson functions were searched for possible heavy-atom sites in a variety of ways. Only the virus orientation need be known for an intraparticle (self-vector) search. The particle position must be assumed, or can be explored, when considering vectors between particles (see supplemental item No. 8*). Two criteria were used for evaluating the search results. The C_1 criterion (Argos & Rossmann, 1974, 1976),

$$C_1 = \sum_i^N P_i,$$

where P_i are the individual Patterson densities, sums the total of all the vector density values and should have a relatively large value for a correct search position. A unit vector density criterion, C_2, was devised where

$$C_2 = (1/M) \sum_{i=1}^{M} P_i/n_i.$$

For each of M look-up positions the density P_i is divided by the number of vectors n_i which fall on that grid point. Vectors within 25 Å of the Patterson function origin were rejected for both criteria. The latter criterion was found to be more reliable at special search positions (such as an icosahedral threefold axis of symmetry) where a large number of vectors overlap in the corresponding Patterson. The unit vector density criterion was also useful in comparing results from intra- and interparticle searches as it is reasonable to expect that the unit vector density value should not degrade significantly for a more complex, but still correct, search model.

A grid spacing of 2·65 Å was used in computing 15–8 Å resolution Patterson functions (168 intervals along a cell edge) and a spacing of 2·23 Å per grid step was used for the 15–6 Å resolution Patterson functions. A complete three-dimensional intra- (self) particle search was computed for the 5 mM $KAu(CN)_2$ derivative with 15–8 Å data. Positions were generated for an icosahedral asymmetric unit, with ψ and φ increments of 1°, for sections of constant radius R from 70 to 170 Å, spaced every 2 Å. Promising sites were located on sections at 108, 110 and 122 Å radius. These sections were then utilized for a search incorporating both the intra- and interparticle vectors for a given particle position (x, x, x) in which the parameter x was varied over the entire possible range ($0 \le n \le 0{\cdot}250$) in increments of 0·002. The highest value for both C_1 and C_2 corresponded to a site A positioned at $\psi = 90°$, $\varphi = 21°$, $R = 110$ Å (close to the icosahedral threefold axis) with x close to 0·000. A finer search in which the value of x was varied from 0·0000 to 0·0020 in increments of 0·0002 led to a 'best' particle position value of $x = 0{\cdot}0004$. The result is consistent with the observation of the approximate face-centering condition for the very-low-resolution data (spacing >30 Å). The particle orientation parameter κ was now checked by repeating inter- (cross) particle searches with $-7 \le \kappa \le -5°$ in steps of 0·1°. The highest peak for the A site was obtained for $\kappa = -6{\cdot}0°$, thus confirming the value obtained from the locked rotation function.

A full three-dimensional search for single sites, using both intra- (self) and inter- (cross) particle vectors, was computed next. In addition to the A site, a number of plausible second sites showed search values at roughly one-half of the height of the A site. A novel type of heavy-atom search was then attempted by holding the A site fixed and searching for a second site utilizing the entire unique vector set for the pair of sites. This search indicated that the site, B, at $R = 122$ Å, $\psi = 69°$ and $\varphi = 7°$ (Table 4) was also probably correct. The non-crystallographic-symmetry feedback search procedure (Rossmann *et al.*, 1986) was adapted for this type of search and gave consistent results (Table 4).

Search results for the 1 mM $KAu(CN)_2$ derivative with 15–6 Å data provided an independent confirmation of the interpretation of the 5 mM $KAu(CN)_2$ results. An intra- (self) particle search of the earlier more limited 1 mM $KAu(CN)_2$ data set (Table 1) had not indicated strongly the presence of even the A site; however, when the inter- (cross) vectors were included, with $x = 0{\cdot}0004$, there was a significant peak for the A site. Further searches using the more accurate and more complete 1 mM $KAu(CN)_2$ data set, with 15–6 Å resolution reflections, led to a 'best' particle position of $x = 0{\cdot}0006$, giving a net displacement of 0·46 Å from the crystal cell origin. This value was used for all subsequent calculations in both the heavy-atom analysis and the molecular replacement real-space averaging procedures.

The phasing start for non-crystallographic symmetry averaging was provided by the 5 mM and the more complete 1 mM $KAu(CN)_2$ derivative data sets. The A site was used for phasing both the 1 mM and 5 mM $KAu(CN)_2$ data, and the additional B site was used for the 5 mM data. The non-crystallographic symmetry heavy-atom least-squares phasing and

* See deposition footnote.

Table 4. *Subset of double-site heavy-atom vector search results (top) and HRV14 peak explorations with the non-crystallographic symmetry feedback search procedure (search for B site) (bottom)*

All unique vectors involving the pair of test sites were used; $x = 0{\cdot}0004$.

ψ (°) φ (°) R(Å)	(Site *A*) 89·5 21 110	(Site *B*) 69 7 122	(Incorrect test site) 86 18 126
(Site *A*) 89·5 21 110	9616/2·21 1551	19 354/0·89 8240	13 716/0·57 7253
(Site *B*) 69 7 122		3295/0·81 2336	7296/0·33 9081
(Incorrect test site) 86 18 126			3816/0·80 2216

The results of the vector search are given as $\frac{C_1/C_2}{N}$ for each pair of sites, where C_1 and C_2 are the criteria described in the text and N is the number of vectors falling to distinct grid points. Elements along the diagonal correspond to single-site searches.

(i) 15–7 Å gold 5 m*M* single-site search, $x = 0{\cdot}0000$, $R = 122$ Å:

ψ (°)	6·5	φ (°) 7·0	7·5
68·5	159	171	153
69·0	157	160	136
69·5	139	129	98

(ii) Same as (i) but assuming *A* site at $\psi = 89°$, $\varphi = 21°$, $R = 110$ Å:

ψ (°)	6·5	φ (°) 7·0	7·5
68·5	678	810	829
69·0	642	728	703
69·5	529	558	490

Note enhancement of signal owing to cross interactions with the *A* site. For definition of search criterion see Rossmann *et al.* (1986).

refinement program (Rossmann, 1976) was adapted for the Cyber 205 and used for the phasing calculations, with heavy-atom positions, relative occupancies and particle positions obtained from the vector search results. The actual occupancies relative to the assumed native F scale were adjusted by setting $(\hat{f}/|\widehat{F_H - F}|) = 1$ where f is the structure-factor contribution of the heavy atoms alone. Although the ideal value of this ratio is $\sqrt{2}$, its value declines with increasing error in the differences; the target value of 1 was chosen to minimize the lack of closure error. Phases and figures of merit were computed using the Hendrickson–Lattman algorithm (Hendrickson & Lattman, 1970). A 'best' Fourier synthesis was then computed with reflections phased between 15 and 6 Å resolution, which was the source for non-crystallographic symmetry averaging. The first averaged map showed good contrast between the regions expected to contain ordered protein density (radius approximately 100–150 Å) as opposed to the solvent regions and the presumably disordered RNA-containing core of the particle. One section of this map is shown in Fig. 6.

Phases were refined by two cycles of molecular replacement real-space averaging (see below) using 25–6 Å data. A skew-averaged electron density map was computed for viewing the icosahedral asymmetric unit which even showed some hints of β barrels. The molecular replacement phases and weights were then used for computing isomorphous difference Fourier syntheses for each derivative data set. A skew-averaged icosahedral asymmetric unit was examined for each of the difference Fouriers, and the 5 m*M* $KAu(CN)_2$ data set revealed the presence of a '*C*' site (Fig. 7).

The relative occupancies of the sites in the 5 m*M* $KAu(CN)_2$ data set were adjusted according to the peak heights in the difference Fourier synthesis used

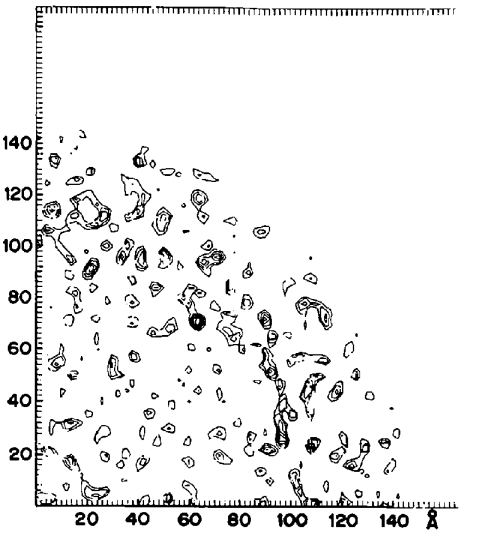


Fig. 6. A section through the averaged 6 Å resolution electron density map based on partial double isomorphous replacement phasing. Note the higher density contrast between 150 and 100 Å radius. The density outside a radius of 170 Å was set to zero in the averaged map.

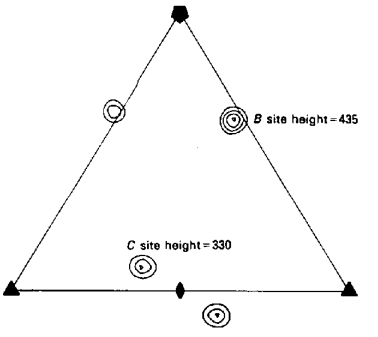


Fig. 7. A skew-averaged difference electron density map section for the 5 m*M* $KAu(CN)_2$ compound, showing the *C* site which had not been part of the initial phasing model.

for locating the *C* site. The updated heavy-atom model was used to provide the starting phases for six cycles of molecular replacement phase refinement using data between 25 and 5 Å resolution. The heavy-atom parameters (Rossmann *et al.*, 1985) used for phasing are shown in supplemental item No. 9 and the phasing statistics in supplemental item No. 10.* Roughly half of the reflections to 5 Å resolution were phased with isomorphous replacement data, of which the majority were obtained from a single derivative. The lack of isomorphism was reflected by the low average figure of merit of 0·35.

Real-space molecular replacement averaging

1. *Background*

The presence of multiple copies of an object in an asymmetric unit of a crystal places powerful constraints on phasing. The molecular replacement technique exploits this information. Poor images produced by isomorphous replacement (or any other) phasing method can be averaged. After setting the solvent regions to a constant value (usually zero), the cell, reconstructed from the averaged images, can be Fourier back-transformed to give an improved set of phases. These can then be used for another averaging cycle in conjunction with the observed amplitudes. Estimates of some phases with marginally improved resolution can also be obtained by the back-transformation. In this way, the resolution of phase information can be gradually extended (Fig. 8).

Decisions involved in the molecular replacement procedure are required at steps 0, 1 and 2 (Fig. 8). Step 0 requires the selection of amplitudes, phases and weights to be used for computing the initial electron density model by Fourier transformation. For instance, this might consist of a set of observed native amplitudes, isomorphous replacement phases and figures of merit. In step 1, the density model is then modified by averaging the non-crystallographically related copies of the repeating motif and reconstructing the cell with the averaged copy. In this step, the boundaries of the molecular envelope must be defined and should not overlap. (In the special case in which there is no non-crystallographic symmetry, this envelope provides the phasing power by the solvent flattening procedure.) The density for solvent regions may also be adjusted to various levels to contrast with the molecular density: in general, there may be distinct 'solvent' regions such as, for example, the RNA-containing core of picornaviruses and the solvent continuum surrounding the virions in the crystal. The modified density model can be transformed into a new set of calculated structure-factor amplitudes and phases. Step 2 involves the construction and specification of a new phasing model. An augmented set of structure-factor amplitudes may be constructed from the newly calculated amplitudes, either by producing modified observed amplitudes or by filling in data which were not collected or were poorly measured. Corresponding phases may be constructed by combination of the newly calculated phases with the pre-existing phase information. New weights should reflect the apparent reliability of the phasing information and are usually obtained from the correlation of the calculated and observed structure factors. During a phase extension (or 'intension') procedure, new reflections at the edge of the highest (or lowest) resolution are gradually added to the process in progressive cycles.

2. *Fourier computations*

Fourier syntheses were computed using Ten Eyck's *P*1 forward and back fast Fourier transformation (FFT) routines (Ten Eyck, 1973) which were adapted for the Cyber 205 (see supplemental item No. 11*). The finely sampled Fourier electron density map which was used as a source for averaging was computed at one-fifth to one-sixth of the resolution of the data. It was stored in packed planes and contained up to $768 \times 768 \times 192 = 1{\cdot}132 \times 10^8$ points, for the most finely sampled cell, at 3 Å resolution, with h(max)

* See deposition footnote.

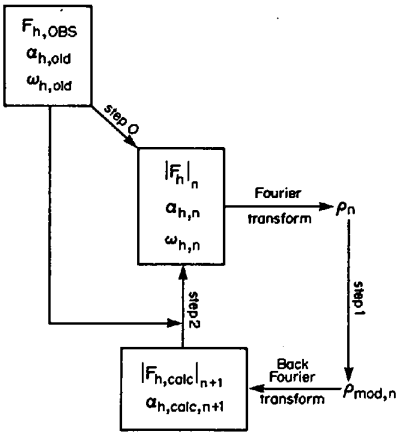


Fig. 8. Scheme for phase improvement and extension using real-space molecular averaging and density modification. Decisions that affect the course of the procedure are required at steps 0, 1 and 2. Step 0 specifies the initial phasing model. In step 1, it is necessary to define the nature and extent (molecular envelope) of the non-crystallographic symmetry as well as the solvent density level. In step 2, the next phasing model is specified in terms of the amplitudes, phases and weights to initiate the next cycle.

* See deposition footnote.

of 148. The largest cell used for $P1$ back-transformation, at 3·0 Å resolution, had 360 grid intervals per axis. The only results retained for a given molecular replacement cycle were the F_{calc}'s computed from the back-transformation of the averaged map which had been reduced to a standard crystallographic asymmetric unit. Contouring of the electron density maps was done using an adaptation of the NCAR (McArthur, 1983) graphics package on the Cyber 205. Skew-averaged views of an icosahedral asymmetric unit, bounded by planes that intersect at two adjacent threefold axes and a neighboring common fivefold axis, were plotted in sections perpendicular to an icosahedral twofold axis of symmetry.

3. *Conditions used for step* 1: *density modification*

The algorithm used for real-space averaging was the double-sort procedure (Bricogne, 1976; Johnson, 1978) which has been rewritten for the Cyber 205.* An averaged map computed from an electron density map based solely on the heavy-atom phases was scrutinized using different possible envelopes. Use of the largest non-overlapping spherical envelope with 157 Å radius resulted in apparent truncation of density near the icosahedral fivefold axes of symmetry. Planes which were tangent to each of 12 neighboring interparticle contacts at a distance of 157 Å from the particle center were used to limit a larger sphere with 170 Å radius. The plane normal directions were described by direction cosines obtained from the vectors between adjacent particle centers. Using this combination of spheres and planes to define the envelope led to no serious density truncation. Although the planes used for envelope definition did not adhere to icosahedral symmetry, they represented a convenient choice for minimizing density truncation at points of contact between virions, while maintaining a more liberal envelope elsewhere. This scheme of envelope definition was used in all following molecular replacement phase refinement and extension except that the spherical radii were contracted from 80–170 Å to 104–163 Å after extension had proceeded beyond 4·5 Å (no ordered density was visible inside 104 Å or outside 163 Å).

Large negative density 'ghosts' were present at the heavy-atom sites during the initial cycles of molecular replacement at 5 Å resolution. These were removed by setting to zero all densities lower than the minimum protein density in the modified and averaged electron density maps. The 'ghosts' had been completely exorcized within several cycles.

* A faster alternative procedure has now been implemented (Luo *et al.*, 1987). It depends on a coarser grid sampling, compensated by nonlinear interpolation of electron density, permitting the storage of larger portions of the electron density map in fast memory, thereby avoiding time-consuming sorting (Hogle *et al.*, 1985; Nordman, 1980).

4. *Conditions used for step* 2: *structure amplitude, phase and weight selection*

The phases used for computing the next electron density map were those obtained from the calculated structure factors from the previous cycle. Combinations of phase probabilities between current calculated phases and earlier estimates (*e.g.* isomorphous replacement phases) were not used. The weights placed on the structure amplitudes were

$$\omega = \exp\left(-\left||F_{obs}|-|F_{calc}|\right|/|F_{obs}|\right)$$

(Rayment, 1983). Several tests using Sim weights (Sim, 1959, 1960) gave similar rates of convergence. F_{calc}'s were not used to substitute missing reflections because the data were reasonably complete. The F_{calc}'s were found to correlate well with the structure amplitudes of isomorphous derivatives, especially for those terms that were large but missing from the native data set. However, later tests showed that use of F_{calc}'s for the missing data gave dramatic improvement in the phasing statistics (the R factor dropped by 4%).

The calculated structure factors from the back Fourier transformation were sifted into a standard cubic asymmetric unit and merged with the observed structure factors. The observed and calculated amplitudes were divided into twenty resolution ranges from 1/Res(inner) to 1/Res(outer) in equal increments of $(1/\text{Res})^2$. Mean local scale factors were computed in each range to put the observed and calculated structure factors on the same relative scale. Statistics such as the correlation coefficient

$$C = \sum_h (F_{h,obs} - \langle F_{h,obs}\rangle)(F_{h,calc} - \langle F_{h,calc}\rangle) \times \left[\sum_h (F_{h,obs} - \langle F_{h,obs}\rangle)^2 \times \sum_h (F_{h,calc} - \langle F_{h,calc}\rangle)^2\right]^{-1/2}$$

and the R factor

$$R = \left(\sum_h |F_{h,obs} - F_{h,calc}|\right) \Big/ \sum_h F_{h,obs}$$

were tabulated for each resolution range, and also as a function of $|F_{obs}|$. The detailed results for the cycle of molecular replacement at 3·42 Å from which the 3·5 Å map was calculated are shown in supplemental item No. 12* and in Table 5 for the final cycle at 3·0 Å resolution from which the final 3·08 Å resolution map was calculated.

The phases used for the new reflections in a phase-extension step were originally the single isomorphous replacement (SIR) phases from the 1 mM $KAu(CN)_2$ derivative to 4·3 Å resolution. This provided starting

* See deposition footnote.

Table 5. *Final molecular replacement results at* 3·08 Å *resolution*

(i) Resolution ranges

Mean range	*R* factor	Correlation coefficient	Number of reflections	Relative local scale factor
15·18	17·8	0·89	6831	0·030
10·10	14·9	0·92	10 354	0·029
8·09	13·9	0·93	12 548	0·029
6·94	13·8	0·93	14 330	0·029
6·17	13·8	0·94	16 002	0·029
5·62	14·3	0·93	17 775	0·029
5·19	14·3	0·93	19 290	0·029
4·84	14·6	0·93	20 786	0·029
4·56	15·0	0·92	22 102	0·030
4·32	15·8	0·91	22 962	0·030
4·11	16·8	0·90	23 892	0·031
3·94	17·8	0·88	24 237	0·032
3·78	18·7	0·87	25 033	0·033
3·64	20·0	0·85	25 382	0·034
3·51	21·5	0·82	25 457	0·036
3·40	23·4	0·78	25 582	0·038
3·30	25·5	0·73	25 527	0·040
3·20	27·3	0·69	24 430	0·042
3·12	30·3	0·59	14 274	0·048
3·04	40·0	0·35	11 543	0·086

(ii) *F* ranges

F ranges	*R* ranges	Number of reflections*
0–0·4 $\hat{F}$	43·8	9942
0·4–0·8 $\hat{F}$	30·0	150 558
0·8–1·2 $\hat{F}$	19·3	122 507
1·2–1·6 $\hat{F}$	12·9	61 763
1·6–∞ $\hat{F}$	9·9	43 574

* Owing to selection of reflections for scaling, not all terms used in the calculation of electron density are used in the statistics. The final molecular replacement computations included 407 711 terms.

phases for about half of the reflections. Later it was shown that the SIR phases beyond 5 Å resolution were essentially random. In fact, phase extension from 5·0 to 4·3 Å was subsequently repeated without resort to the SIR phases, giving a solution with mean phase change of less then 5° from the final phases in any resolution range. Beyond 4·3 Å resolution, the extension phases were obtained from back-transformation of the averaged and reconstructed cell.

The optimal rate for phase extension was decided by consideration of the '*G*' function (Rossmann & Blow, 1962; Arnold & Rossmann, 1986). The *G* function represents a diffraction integral over the volume of the molecular envelope with an argument of *HR*. Here *H* is the reciprocal distance between the reflection whose phase is to be determined and its nearest already phased contribution, while *R* is the outer radius of the approximate molecular envelope. Since the largest values of *G* occur for $HR < 0{\cdot}725$, the limiting number *n* of reciprocal-lattice points used in any resolution increment will be given by $(n/a)R < 0{\cdot}725$. But $R = 155$ Å and $a = 445$ Å for cubic HRV14 crystals, hence $n < 2{\cdot}1$. Values actually used for extension varied between $2{\cdot}3 < n < 2{\cdot}6$. Following any cycle of phase extension, the phasing statistics for reflections in the current highest-resolution shell are relatively poor since all of the reflections beyond the resolution edge are omitted from the computations (Arnold & Rossmann, 1986).

5. *Phase refinement and extension from* 5 *to* 3 Å *resolution*

The double isomorphous replacement phases were used to initiate six cycles of molecular replacement phase refinement at 5 Å resolution. A plot of the correlation coefficients as a function of resolution for the successive cycles is shown in supplemental item No. 13.* Beyond 7 Å resolution, the initial correlation coefficients were of the order of 0·2, emphasizing the low quality of the isomorphous replacement phasing, attributable both to the partial data sets used in heavy-atom phasing (much of which was SIR – supplemental item No. 10) and also to the lack of isomorphism which had been detected in the scaling process.

The 5 Å resolution skew-averaged map was plotted at 1·5 Å intervals (Fig. 9). Note the tremendous contrast between the solvent-containing area in the center of the triangular face and the protein regions. This area has not been solvent flattened in any cycle and therefore its appearance is totally unbiased. Similarly, the region which contains RNA, below the inner surface of the protein shell, was seen to be mostly level. A number of polypeptide chains are clearly resolved in Fig. 9. These physically reasonable features were convincing evidence of the correctness of the solution.

Phase extension from 5–3 Å resolution was initiated with the molecular replacement refined phases from 25–5 Å resolution. In the first two cycles all data from 25–3 Å were included. However, only the phases

* See deposition footnote.

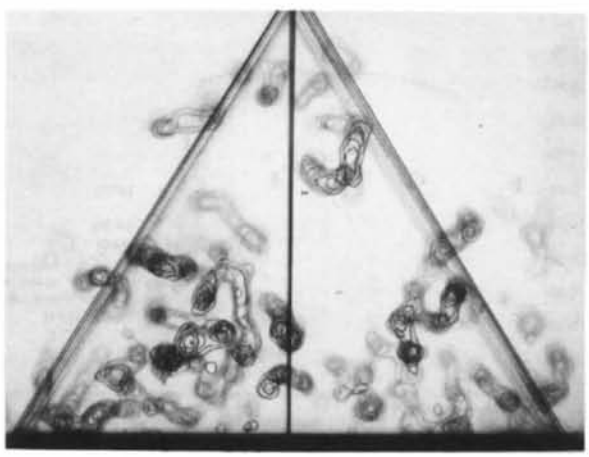


Fig. 9. External portion of 5 Å resolution skewed averaged map (sections 135–150 Å).

of reflections close in resolution to the refined nucleus of phases were seen to change significantly from cycle to cycle. The computations were therefore limited to 4·5 Å resolution data since the process was much faster with a smaller grid. After 12 cycles, using data from 25–4·5 Å resolution, the phasing statistics for the new data from 5–4·5 Å resolution were as favorable as those for the previously refined data. Another skew-averaged map was computed from these phases and the map showed plausible increase in detail relative to the 5 Å resolution map. Data were then included in one step to 4 Å resolution and 12 more cycles were computed. At this time there began to appear a discontinuity in the correlation plot *versus* resolution: the front had advanced to 4·4 Å resolution, but a trough between the front and the next peak at 4·0 Å resolution became relatively deeper in successive cycles. The situation became even worse on a further extension to 3·5 Å resolution (Fig. 10). It seemed possible that the peaks at higher resolution in the correlation coefficient plots represented a solution to the phasing which was consistent with the non-crystallographic symmetry constraints but were anticorrelated with the lower-resolution phasing solution. The resolution was, therefore, reduced back to the edge of the front and was then extended in increments of only 2·5 reciprocal-lattice units at a time (see above) to avoid the build up of alternative solutions. Two to three refinement cycles were performed between successive extension steps. As can be seen from the correlation plots for selected cycles of phase extension from 4·3 to 3·0 Å resolution (Fig. 11), progress was far more rapid than it had been when larger-resolution increments had been used (Fig. 10). After the final 3·0 Å resolution cycle of molecular replacement (Fig. 11), the correlation coefficients to 3·5 Å resolution maintain a value of 0·82, and then fall off to slightly under 0·6 at 3·08 Å resolution. By comparison with the correlation coefficients obtained for SBMV and TBSV molecular replacement phase determinations, in which no values of the correlation coefficient were above 0·8 beyond 7 Å resolution (Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1981), the HRV14 correlation coefficients are much higher.

The results of the molecular-replacement phase-extension procedure were also assayed for physical reasonableness by examining skew-averaged maps computed for the entire icosahedral unit between a radius of 100 to 164 Å at the 4·5, 4·0, 3·5 and 3·0 Å stages. The 3·5 Å averaged map was computed using phases obtained when extension had proceeded to 3·42 Å resolution. The combination of the high redundancy ($N = 20$), the excellent quality of the synchrotron X-ray data, the well determined non-crystallographic symmetry parameters and the ability to calculate many cycles of molecular replacement phase refinement and extension conspired to make the 3·5 Å map very simple to interpret (Fig. 12), even for more than 800 amino acids in four separate polypeptide chains within one icosahedral asymmetric unit.

After completion of the structure determination, a successful attempt was made to initiate phase extension from 12 to 9 Å and 'intension' from 15 to 26 Å resolution. Comparison with the final set showed a mean phase change of less than 5°. This suggests that complete *ab initio* phase determination, starting only with an acentric envelope filled with uniform density (Rayment, Baker, Caspar & Murakami, 1982; Johnson, Akimoto, Suck, Rayment & Rossmann, 1976), would be viable. Alternatively, in light of the considerable structural similarity among small RNA viruses (Rossmann *et al.*, 1985; Hogle *et al.*, 1985), a better phasing start might be any one of the known virus structures. In general, the most difficult problem that remains is an adequate solution of the particle translation problem.

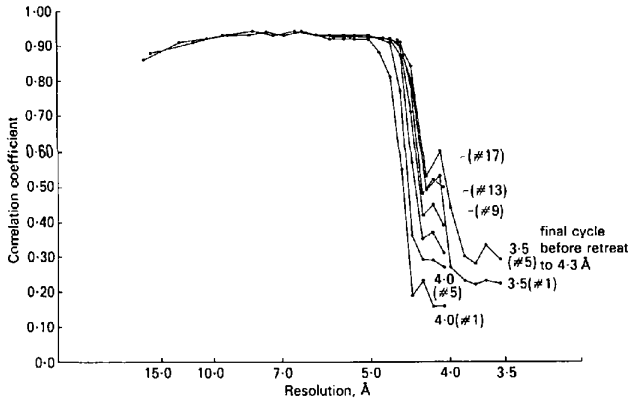


Fig. 10. 'Bumps' in the correlation coefficient on a too rapid phase extension to 3·5 Å resolution. Cycle number at given resolution shown in parentheses [*e.g.* (#1) for cycle 1].

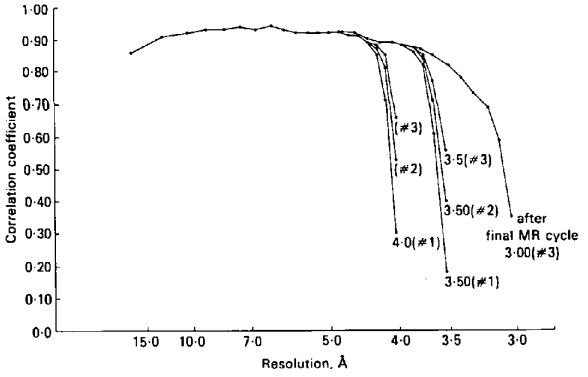


Fig. 11. Correlation coefficients plotted against resolution for selected phase-extension steps from 4·3 to 3·0 Å resolution.

Map interpretation and model building

All four distinct polypeptide chains in the icosahedral asymmetric unit were completely traced in the 3·5 Å resolution mini-map in only two days. Aside from the disordered N termini of VP1, VP2 and VP4, there was not one single break in the main-chain density. The carbonyl oxygens of most residues in the 3·08 Å resolution map are identifiable as are many water molecules. The identities of the chains were distinguished by matching short stretches of amino acid sequence with the electron density features. Protein VP2 was identified first: the large protrusion on the surface [now known as the VP2 'puff' (Rossmann *et al.*, 1985)] was assumed to be an antigenic site. The approximate residue number was counted backward from the C terminus, and this region was around 100 from the end. Inspection of the details of the immunogenic sites (see Table 3 of Rossmann *et al.*, 1985) showed that only 'NIm-II' (neutralizing immunogen site II) was at this distance from the C-terminal residue in VP2.

A complete atomic model was built for the ordered portions of viral proteins VP1 (17–289), VP2 (5–262), VP3 (1–236) and VP4 (25–68) using *FRODO* (Jones, 1978, version adapted for an Evans & Sutherland PS300 at Rice University) with respect to the 3·08 Å resolution map. The excellent clarity of this map can be seen in Fig. 13. The map used for model building was skew-averaged and was computed using 1 Å grid spacing. In total, 6320 atoms were built into 811 amino acid residues in the four distinct polypeptide chains.

On completion of the model building, the orientation and position of the virion were refined by a least-squares procedure based on the atomic positions (Rossmann, 1976) and yielded $\kappa = -6{\cdot}086°$ and $x = 0{\cdot}00039$. A further eight cycles of molecular replacement were then performed using these new particle parameters and incorporating calculated amplitudes for those reflections that had not been observed, while extending the resolution to 2·94 Å. The overall molecular replacement R factor dropped by about 7% in most resolution ranges; the correlation coefficients were greater than 0·95 to 4·0 Å resolution and greater than 0·90 to 3·5 Å resolution, while the overall correlation coefficient for all data to 2·94 Å resolution was 0·94.

Table 6. *Randomly selected sample of reflections comprising 1/15th of observed amplitudes between 10·0 and 3·0 Å resolution*

Resolution range (Å)	R	Number of reflections
10·0–8·44	37·4	537
8·44–7·43	38·5	660
7·43–6·72	39·7	682
6·72–6·18	36·5	775
6·18–5·75	35·7	819
5·75–5·40	33·1	842
5·40–5·11	29·3	948
5·11–4·86	26·6	938
4·86–4·64	25·3	1073
4·64–4·45	25·7	1121
4·45–4·28	26·7	1121
4·28–4·13	26·3	1083
4·13–4·00	27·5	1237
4·00–3·87	27·4	1161
3·87–3·76	28·7	1187
3·76–3·66	29·8	1280
3·66–3·56	30·1	1314
3·56–3·47	31·2	1307
3·47–3·39	31·7	1264
3·39–3·32	33·0	1200
3·32–3·25	35·2	1268
3·25–3·18	34·3	1268
3·18–3·12	36·2	1221
3·12–3·06	38·4	1217
3·06–3·00	37·9	878
Overall	31·6	26401

The R factor based on the initial atomic positions with the updated particle position parameters but without waters is shown as a function of resolution in Table 6. The overall R factor from 10 to 3 Å resolution was 31% using a global temperature factor of 15 Å^2. This initial R factor was 26% in the 6 to 4 Å resolution range and dropped to 18% on refinement using the Hendrickson–Konnert procedure (Konnert & Hendrickson, 1980) modified to include the molecular replacement phases as observations. Two cycles of refinement to date (March 1986) have involved roughly 70 000 phased reflections while maintaining geometrical restraints suggested for initial refinement (Hendrickson, 1981).

Concluding remarks

The structure determination of HRV14 has shown more clearly than any previous result that, given sufficient accuracy in the structure-amplitude measurements, it is possible to solve the structures of

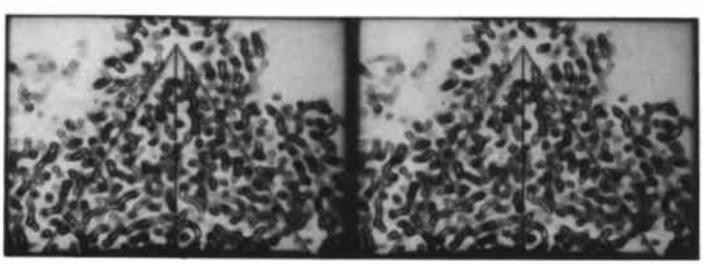


Fig. 12. Stereo view of a portion of the 3·5 Å resolution map showing sections through the β sheets in VP1, VP2 and VP3.

highly complex biological assemblies by use of non-crystallographic symmetry (Rossmann & Blow, 1962, 1963) and solvent flattening (Bhat & Blow, 1982; Schevitz, Podjarny, Zwick, Hughes & Sigler, 1981; Wang, 1985). While real-space averaging has been used extensively for phase improvement within a crystal (Buehner, Ford, Moras, Olsen & Rossmann, 1974; Biesecker, Harris, Thierry, Walker & Wonacott, 1977; Bloomer, Champness, Bricogne, Staden & Klug, 1978; Harrison *et al.*, 1978; Liljas *et al.*, 1982) or between crystals (Fletterick & Steitz, 1976; Varghese, Laver & Colman, 1983), it has been used only rarely for *ab initio* phase determination either at low resolution (Johnson *et al.*, 1976; Rayment *et al.*, 1982; Rossmann & Henderson, 1982) or for limited phase extension at higher resolution (Nordman, 1980; Gaykema *et al.*, 1984). The lack of use of non-crystallographic constraints for phase extension can be attributed to the requirement of extensive computing resources (*e.g.* the availability of the Cyber 205 supercomputer in the present work) and the need for high-accuracy massive data-gathering techniques (*e.g.* a synchrotron source). In light of the success with HRV14, the structure determination of poliovirus (Hogle *et al.*, 1985) and Mengo virus (Luo *et al.*, 1987), we have examined elsewhere (Arnold & Rossmann, 1986) the relative importance of the various factors contributing to phase determination using molecular replacement.

The results presented here bode well for many structure determinations of exceptional biological interest by techniques similar to those described here.

We thank Roland R. Rueckert for suggesting HRV as a subject for structural analysis and his continuous encouragement and support. We thank him and Barbara Sherry for their help in educating us in the techniques of animal virology, and for their enthusiastic contributions in the interpretation of the structure. We also thank Libby Frankenberger for her dedicated work in propagating and purifying HRV14 as well as her assistance in data collection. We are most grateful to Sharon Wilder for her outstanding assistance in the organization of the Purdue structural laboratory over many years and in the preparation of this manuscript. We thank Kathy Shuster and Bill Boyle for preparation of the figures and Jun Tsao for constructing various physical models of HRV14. We are most grateful for all the dedicated help we have received from Keith Moffat, Wilfried Schildkamp, Robert Hunt, Don Bilderback, Boris Batterman, Aggie Sirrine and all the CHESS staff and operators, as well as Gale Rhodes and Diana Delatore for helping us on our visits to CHESS. We thank Hans Bartunik and Klaus Bartels for their assistance in the use of the DESY synchrotron in Hamburg and thank Tibor den Ouden for his participation in data collection. We are most grateful to the Purdue University Computer Center staff (in particular Saul Rosen, John Steele, Tom Putnam and Paul Townsend) for their generous help and encouragement of our use of the Cyber 205 supercomputer. We have benefited greatly by the helpful and stimulating discussions with our other colleagues in the Purdue structural groups including Jeffrey Bolin, Abelardo Silva, Ignacio Fita, Celerino Abad-Zapatero, R. Usha, M. V. Hosur, Cynthia Stauffacher, Patrick Argos and J. K. Mohana Rao. We thank Richard Colonno (Merck Sharp & Dohme Co.) for his interest in the work and providing the HRV14-RNA sequence prior to publication. We thank Timothy S. Baker for sharing the VAX 11/750 resource with us and Tim Schmidt for maintaining the X-ray equipment at Purdue University. The work was supported by a NIH grant to MGR and a NSF grant for supercomputer time to MGR. A postdoctoral Walter Winchell–Damon Runyon fellowship supported EA for some of the time. A Purdue University Showalter Foundation grant was awarded to MGR to equip the cell-culture laboratory and a recent grant from the Merck Sharp & Dohme Co. contributed to the salary of one technician involved in the virus propagation.

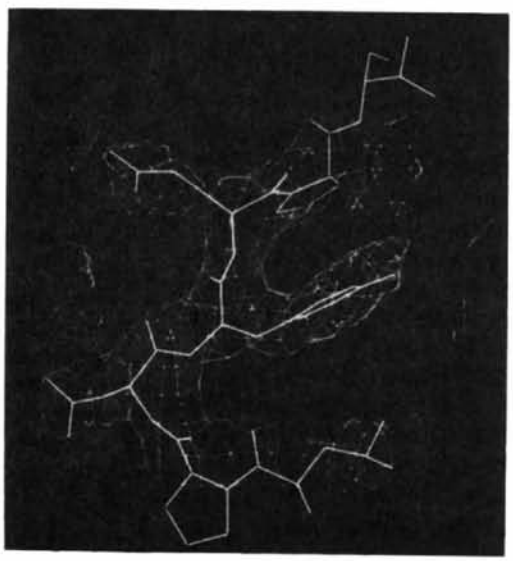


Fig. 13. Portion of VP3 showing fit of sequence *P-N-Y-E-P-T* (residues 25–30) into the 3·08 Å resolution map.

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.

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1981). *Acta Cryst.* B**37**, 2002–2018.

ARGOS, P. & ROSSMANN, M. G. (1974). *Acta Cryst.* A**30**, 672–677.

ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* B**32**, 2975–2979.

ARNOLD, E., ERICKSON, J. W., FOUT, G. S., FRANKENBERGER, E. A., HECHT, H. J., LUO, M., ROSSMANN, M. G. & RUECKERT, R. R. (1984). *J. Mol. Biol.* **177**, 417-430.
ARNOLD, E. & ROSSMANN, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489-5493.
BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst.* A**38**, 21-29.
BIESECKER, G., HARRIS, J. I., THIERRY, J. C., WALKER, J. E. & WONACOTT, A. J. (1977). *Nature (London)*, **266**, 328-333.
BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature (London)*, **276**, 362-368.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832-847.
BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W. & ROSSMANN, M. G. (1974). *J. Mol. Biol.* **82**, 563-585.
CALLAHAN, P. L., MIZUTANI, S. & COLONNO, R. J. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 732-736.
CROWTHER, R. A. (1969). *Acta Cryst.* B**25**, 2571-2580.
ERICKSON, J. W., FRANKENBERGER, E. A., ROSSMANN, M. G., FOUT, G. S., MEDAPPA, K. C. & RUECKERT, R. R. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 931-934.
FINCH, J. T. & KLUG, A. (1959). *Nature (London)*, **183**, 1709-1714.
FLETTERICK, R. J. & STEITZ, T. A. (1976). *Acta Cryst.* A**32**, 125-132.
GAYKEMA, W. P. J., HOL, W. G. J., VEREIJKEN, J. M., SOETER, N. M., BAK, H. J. & BEINTEMA, J. J. (1984). *Nature (London)*, **309**, 23-29.
HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368-373.
HENDRICKSON, W. A. (1981). *Refinement of Protein Structures*, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER, pp. 1-8. Daresbury: Science and Engineering Research Council.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136-143.
HOGLE, J. M., CHOW, M. & FILMAN, D. J. (1985). *Science*, **229**, 1358-1365.
JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 576-577.
JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394-400.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268-272.
KAHN, R., FOURME, R., GADFT, A., JANIN, J. & ANDRÉ, D. (1982). *J. Appl. Cryst.* **15**, 330-337.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-350.
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, G., MINOR, I., ARNOLD, E., ROSSMANN, M. G., BOEGE, U., SCRABA, D. G., DUKE, G. M. & PALMENBERG, A. C. (1987). *Science*, **235**, 182-191.
MCARTHUR, G. R. (1983). Tech. Note NCAR-TH/166-IA, May. National Center for Atmospheric Research, Boulder, Colorado.
MAIN, P. (1967). *Acta Cryst.* **23**, 50-54.
MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 67-72.
NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747-754.
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. (1976). *Acta Cryst.* A**32**, 774-777.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225-238.
ROSSMANN, M. G. (1984). *Biological Systems: Structure and Analysis*, edited by G. P. DIAKUN & C. D. GARNER, pp. 28-40. Daresbury: Science and Engineering Research Council.
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., ARNOLD, E. & VRIEND, G. (1986). *Acta Cryst.* A**42**, 325-334.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39-45.
ROSSMANN, M. G. & HENDERSON, R. (1982). *Acta Cryst.* A**38**, 13-20.
ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570-581.
SCHEVITZ, R. W., PODJARNY, A. D., ZWICK, M., HUGHES, J. J. & SIGLER, P. B. (1981). *Acta Cryst.* A**37**, 669-677.
SCHILDKAMP, W. & BILDERBACK, D. (1986). *Nucl. Instrum. Methods.* In the press.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.
STANWAY, G., HUGHES, P. J., MOUNTFORD, R. C., MINOR, P. D. & ALMOND, J. W. (1984). *Nucl. Acids Res.* **12**, 7859-7875.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.
VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). *Nature (London)*, **303**, 35-40.
VRIEND, G., ROSSMANN, M. G., ARNOLD, E., LUO, M., GRIFFITH, J. P. & MOFFAT, K. (1986). *J. Appl. Cryst.* **19**, 134-139.
WANG, B. C. (1985). *Meth. Enzymol.* **115**, 90-112.

Acta Cryst. (1987). A**43**, 361-369

Three-Beam Diffraction in a Finite Perfect Crystal

BY GUNNAR THORKILDSEN

Rogaland Distriktshøgskole, Ullandhaug, 4000 Stavanger, Norway

(*Received 5 March* 1986; *accepted 17 November* 1986)

Abstract

By using the Laplace transform on the Takagi-Taupin equations for three coupled waves in a perfect crystal it has been possible to obtain general boundary-value Green functions for the wave fields D_o, D_h and D_g. For a crystal shaped as a parallelepiped the integrated power P_h is calculated in the kinematical limit by suitable integrations over one divergence angle and over the entrance and exit surfaces. The result, which is expressed as a function of the deviation from the Bragg condition for the third wave, is continuous through the three-beam point, and gives the expected asymmetry associated with the invariant phase of the product of the three structure factors involved. The asymptotic behaviour is the same as that obtained

0108-7673/87/030361-09$01.50