MEGAW, H. D. (1973). *Crystal Structures: A Working Approach*, p. 243. Philadelphia: W. B. Saunders.
NYMAN, H. & ANDERSSON, S. (1979). *Acta Cryst.* A**35**, 580–583.
O'KEEFFE, M. & HYDE, B. G. (1981). *Structure and Bonding in Crystals*, edited by M. O'KEEFFE and A. NAVROTSKY, ch. 10. New York: Wiley.
O'KEEFFE, M. & HYDE, B. G. (1985). *Struct. Bonding (Berlin)*, **61**, 77–144.
RIBBE, P. H. & GIBBS, G. V. (1971). *Am. Mineral.* **56**, 1155–1163.
ROBINSON, K., GIBBS, G. V. & RIBBE, P. H. (1973). *Am. Mineral.* **58**, 43–49.
SCHUBERT, K. (1964). *Kristallstrukturen Zweikomponentiger Phasen*, pp. 305–307. Berlin: Springer Verlag.
SUDARSANAN, K. & YOUNG, R. A. (1969). *Acta Cryst.* B**25**, 1534–1543.
SUDARSANAN, K. & YOUNG, R. A. (1978). *Acta Cryst.* B**43**, 1401–1407.
VEGAS, A. (1985). *Acta Cryst.* C**41**, 1689–1690.
WENK, H. R. & RAYMOND, K. N. (1973). *Z. Kristallogr.* **137**, 86–105.
WHITE, T. J. & HYDE, B. G. (1982*a*). *Phys. Chem. Miner.* **8**, 55–63.
WHITE, T. J. & HYDE, B. G. (1982*b*). *Phys. Chem. Miner.* **8**, 167–174.
WHITE, T. J. & HYDE, B. G. (1983). *Acta Cryst.* B**33**, 10–17.
WONDRATSCHEK, H., MERKER, L. & SCHUBERT, K. (1964*a*) *Z. Kristallogr.* **120**, 393–395.
WONDRATSCHEK, H., MERKER, L. & SCHUBERT, K. (1964*b*) *Z. Kristallogr.* **120**, 478.
ZEMANN, J., EFFENBERGER, H. & PERTLIK, F. (1982). *Österr. Akad. Wiss. Anzeig.* pp. 61–62.

Acta Cryst. (1991). B**47**, 23–29

Crystallization and Preliminary Structure Analysis of an Insect Virus with $T = 4$ Quasi-Symmetry: *Nudaurelia capensis* ω Virus

BY JEAN CAVARELLI,* WU BOMU,† LARS LILJAS,‡ SANGSOO KIM,§ WLADEK MINOR, SANJEEV MUNSHI, STEVE MUCHMORE, TIM SCHMIDT AND JOHN E. JOHNSON¶

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

AND DONALD A. HENDRY

Department of Microbiology, Rhodes University, Grahamstown, 6140 *South Africa*

(*Received* 14 *March* 1990; *accepted* 15 *August* 1990)

Abstract

We report the crystallization of *Nudaurelia capensis* ω virus, a pathogen of the pine emperor moth. The icosahedral particle has $T = 4$ quasi-equivalent symmetry and is approximately 410 Å in diameter. Triclinic crystals ($a = 414{\cdot}0$, $b = 410{\cdot}7$, $c = 420{\cdot}1$ Å, $\alpha = 59{\cdot}1$, $\beta = 58{\cdot}9$, $\gamma = 64{\cdot}0°$) diffracting X-rays to 2·7 Å resolution have been analyzed using a partial data set collected at high resolution (2·7 Å) using conventional oscillation photography and a more complete data set collected at low resolution (50–9 Å) using a Siemens area detector. The pseudo-rhombohedral symmetry of the crystals created significant problems in processing the unaligned oscillation photographs. Successful processing of the films depended on the use of an auto-indexing procedure followed by systematic scaling tests of four, nearly equivalent, reduced cells. The unique cell from the area-detector data collection was identified by scaling the four possible choices with the film data. The particle orientation was determined using the rotation function.

Introduction

Nudaurelia capensis ω virus (NωV) is a spherical RNA virus that was first isolated from larvae of the pine emperor moth (Hendry, Hodgson, Clark & Newman, 1985). The virus has been isolated only from larvae infected in the wild and it has not been successfully propagated in insect cell lines. NωV particles appear spherical in the electron microscope with an estimated diameter of 410 Å. The capsid consists of multiple copies of a protein subunit of molecular weight 65 kD. The particles contain two RNA molecules with molecular weights of about 2·3 and 0·8 mD. The physical and biological properties of NωV classify it as a member of the family Tetraviridae whose prototype is *Nudaurelia capensis* β virus (NβV). Although NωV and NβV are both

* Present address: Laboratoire de Cristallographie Biologique, IBMC, CNRS, 15 rue Rene Descartes, 67084 Strasbourg, CEDEX, France.
† Present address: Institute of Biophysics, Academia Sinica, Beijing 100080, People's Republic of China.
‡ Permanent address: Department of Molecular Biology, University of Uppsala Biomedical Center, S-751 24 Uppsala, Sweden.
§ Present address: Lucky Central Research Institute, Science Town, PO Box 10, Dae Jeon, Chung-Nam, Korea.
¶ Author to whom correspondence should be addressed.

0108-7681/91/010023-09$03.00

isolated from larvae of the pine emperor moth they are serologically distinct and appear different in negative stained electron micrographs. NβV has been extensively characterized and a low-resolution structure has been determined using electron micrographs of negatively stained particles (Finch, Crowther, Hendry & Struthers, 1974) and frozen hydrated particles (Olson, Baker, Bomu, Johnson & Hendry, 1987). NβV is an icosahedral virus containing one subunit type of molecular weight 61 kD and the particles display $T = 4$ quasi-symmetry with an average diameter of 395 Å. The biophysical similarities between NβV and NωV make it virtually certain that NωV is also a $T = 4$ virus although electron micrographs of the particles do not show the surface features characteristic of this quasi-symmetry that are obvious in NβV particles.

Large quantities of NωV can be isolated from dead larvae following outbreaks of the virus infection in the wild. Typically 40 mg of virus can be isolated from 100 g of larvae. The readily producible large quantities of NωV, the novel $T = 4$ quasi-symmetry of the capsid and the large subunit size make NωV an interesting subject for crystallization and study by X-ray diffraction methods.

We report in this paper the preparation of high-quality crystals of NωV and the collection of a partial data set using photographic film at high resolution and a Siemens area detector at low resolution. The data were analyzed using the rotation function and the particle packing was established. The unusual triclinic crystal form required novel methods for properly processing and indexing each film.

Materials and methods

Virus production

The virus was purified from larvae of the pine emperor moth obtained in the George region of South Africa. The purification of virus has been described (Hendry, Hodgson, Clark & Newman, 1985); yields of 0·4 mg of virus per gram of larvae tissue are typical. The virus is purified on a 40% caesium chloride gradient using a 50 m*M* tris-HCl buffer at pH 7·5 and then pelleted and resuspended in 0·07 *M* sodium acetate buffer at pH 5·0 just prior to crystallization.

Crystallization

NωV was crystallized using the sitting-drop method of vapor diffusion (McPherson, 1982). The reservoir solution was prepared using 0·075 *M* morpholinopropanesulfonic acid (MOPS) buffer at pH 7·0 with polyethylene glycol (PEG) 8000 at 2%, $CaCl_2$ at 0·25 *M* and NaN_3 at 10^{-3} *M*. The virus solution was at 8–10 mg ml^{-1} in 0·07*M* sodium acetate buffer at pH 5·0. The crystallization drops consisted of 10 μl of the virus solution mixed with 40 μl of the reservoir solution. This mixture was allowed to reach vapor equilibrium with the reservoir solution (20 ml). Tabular shaped crystals appeared in 2–4 weeks.

Data collection

Data from 30–2·7 Å resolution were collected using photographic film on beam line X12-C at the National Synchrotron Light Source (Brookhaven National Laboratory). Monochromatic radiation ($\lambda = 1{\cdot}220$ Å) was used to record 0·3° oscillation photographs with a crystal-to-film distance of 120 mm and exposure times of 7–13 min. One to three photographs were obtained with each crystal. The films were processed in two stages. First the auto-indexing procedure developed by Kim (1989) was used to establish the orientation of the crystal. The orientation matrix from this program was then used in the film-processing program developed by Rossmann (1979) to measure the diffraction intensities. The data were scaled and post-refined (Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979).

Data between 150 and 8·5 Å resolution were collected on a Siemens area detector (Blum, Metcalf, Harrison & Wiley, 1987). A GX20 rotating-anode X-ray generator equipped with a 200 μm focal cup and operating at 35 kV 40mA was used to produce Cu $K\alpha$ radiation. Franks focusing mirrors and 0·25 mm collimator were used to define the X-ray beam. The detector was placed 280 mm from the crystal (a helium path was used to reduce scatter from air) and all data were collected at a detector setting of $2\theta = 0°$. The data acquisition and camera movements were controlled by a software package developed at Harvard (Durbin, Burns, Moulai, Metcalf, Freymann, Blum, Anderson, Harrison & Wiley, 1986) and slightly modified locally (Minor & Bolin, 1990). Data were collected at room temperature using an oscillation angle of 0·05° per frame and 6 min exposure time. A partial data set was measured from a tetragonal lysozyme crystal prior to making measurements with the virus crystals. The lysozyme data were used to refine all the camera parameters associated with the data-collection configuration and this simplified the determination of orientation and unit-cell parameters for the virus crystals. After collecting 10 to 20 frames of data the crystal quality was analyzed by the program *ELOP* (Minor & Bolin, 1990). Up to 10 frames were merged by *ELOP* to form an electronic pseudo-oscillation photograph which was displayed on a MicroVAX II GPX workstation. The three-dimensional profile of any diffraction peak could be analyzed interactively. This allowed crystals with large mosaic

spread or imperfections (multiple crystals) to be readily identified. 160 to 500 frames were collected from each crystal. The data were later processed using the *XENGEN* software (Howard, Gilliland, Finzel, Poulos, Ohlendorf & Salemme, 1987).

Rotation function

The self rotation function was computed using the Rossmann–Blow algorithm (Rossmann & Blow, 1962).

Results

Crystals of NωV diffract X-rays to 2·7 Å resolution (Fig. 1). The crystal morphology and initial film processing suggested that the NωV crystals were rhombohedral; however, processing results, assuming trigonal symmetry, were only marginally acceptable and the scaling of symmetry-equivalent data was poor. An alternative method for processing the films was attempted using the auto-indexing program developed by Kim (1989). This program determines three independent reciprocal-lattice vectors to describe the photographic diffraction pattern. A reduced cell with a conventional set of lattice constants could then be derived as described below. Using this procedure axial lengths were not equal to each other nor were the interaxial angles, suggesting that the cell was not trigonal. Using a triclinic cell with lattice constants derived from Kim's program, the diffraction pattern was predicted with much greater precision than with the trigonal cell and the film-processing statistics were acceptable. Based on the fit of the reciprocal vectors to the lattices in different crystals a reduced primitive unit cell containing one virus particle was identified (Table 1). Attempts to define a *C*-centered monoclinic cell were unsuccessful and there was no evidence of 2/*m* symmetry in the diffraction pattern.

The procedure for establishing the reduced cell required the following steps. The nomenclature for the different matrices is from Rossmann (1979).

(1) The *A* matrix relating crystallographic indices to the coordinates of reciprocal-lattice points in an orthogonal coordinate system based on the camera was determined by the auto-indexing program, $\mathbf{X} = [A]\mathbf{h}$.

(2) The product $G = [A]^T[A]$ is a matrix which contains information only about the unit-cell parameters and is devoid of crystal-orientation information. The cell parameters can be extracted from this matrix and a reduced cell can be identified by finding the three shortest non-coplanar translations in the lattice (Table 1).

(3) A new *A* matrix was then determined from the original *A* matrix and the matrix determined in the cell-reduction procedure described in the previous step.

(4) The film was then processed assuming the reduced lattice and the orientation of this lattice relative to the camera coordinate system. Invariably the final orientation of the lattice determination during the refinement portion of the processing program was nearly identical to the orientation determined by the auto-indexing program.

(5) The processed data were then added to the data set and the validity of the cell assignment was determined by the scaling results.

In the course of the data processing it was apparent that the unique cell was not always identified because of statistical variation in the lattice constants. There are four nearly identical cells related to the cell noted above by various vector permutations and additions (Table 1). For some films it was necessary to systematically apply each permutation until the correct cell was identified by good scaling with previously measured films. As the data base of successfully scaled films increased this procedure was routinely used because significant numbers of reflections measured on each new film were already present in the data set. Seventy-one films were incorporated into the data set. The percentage of

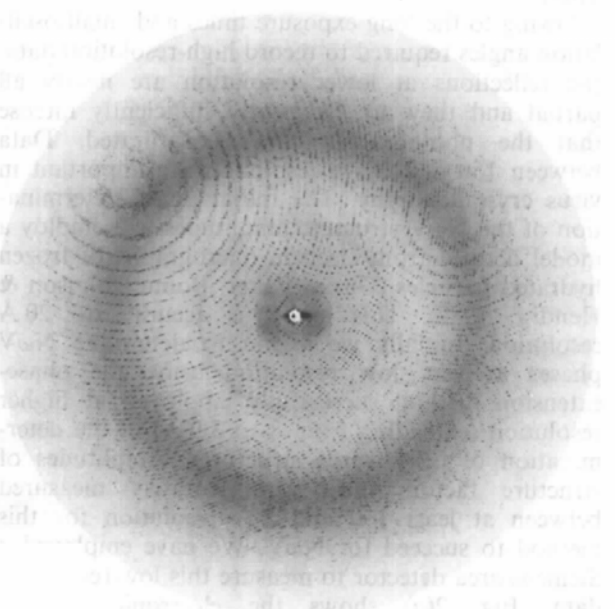


Fig. 1. A 0·3° oscillation photograph from a crystal of NωV. The diffraction pattern was obtained on the X12-C beam line at the National Synchrotron Light Source at Brookhaven National Laboratory. The wavelength was 1·220 Å and the exposure time was 12·5 min. Data were measured to 2·7 Å resolution. A total of 8752 whole and 20 451 partial reflections are predicted to occur on this photograph. Measurements above background were made for 5601 whole and 8644 partial reflections. A total of 71 comparable photographs were processed.

Table 1. *Reduced unit-cell choices in the NωV triclinic crystal*

The relationship is given between four unit cells with nearly identical lattice constants, but different indices. The first cell was chosen for all crystallographic calculations. The cell dimensions were averages from the film.

Reciprocal axes			Real axes			Real cell dimensions
a^*	b^*	c^*	a	b	c	414·0 Å 410·7 Å 420·1 Å 59·1° 58·9° 64·0°
$-a^*$	$-b^*$	$a^*+b^*+c^*$	$-a+c$	$-b+c$	c	410·1 Å 410·1 Å 420·1 Å 59·8° 59·3° 64·0°
$-b^*$	$-a^*$	$-c^*$	$-b$	$-a$	$-c$	410·7 Å 414·0 Å 420·1 Å 58·9° 59·1 64·0°
b^*	a^*	$-a^*-b^*-c^*$	$b-c$	$a-c$	$-c$	410·1 Å 410·0 Å 420·1 Å 59·3° 59·8 64·0°

Table 2. *Number of independent reflections in each resolution range after scaling* 71 *oscillation films*

The scaling R^* factor including whole reflections and partial reflections with a partiality greater than or equal to 0·5 was 14·03% (761 052 observations, 683 047 unique reflections). The scaling R value for only whole reflections was 13·08% (299 192 observations, 282 467 unique reflections).

Resolution	No. of unique reflections	% of theoretically observable data
30–15 Å	4851	18
15–10 Å	13008	17
10–8 Å	17419	17
8–6 Å	45555	16
6–5 Å	56498	16
5–4 Å	123727	15
4–3 Å	290563	13
3·2·65 Å	131426	7
Total	683047	12·7

$^*R = 100\sum_h\sum_i |F_h^2 - F_{hi}^2| / \sum_i\sum_h F_{hi}^2$ where F_h^2 is the average of equivalent reflections and F_{hi}^2 is the intensity of a particular reflection.

complete data in each resolution range is shown in Table 2.

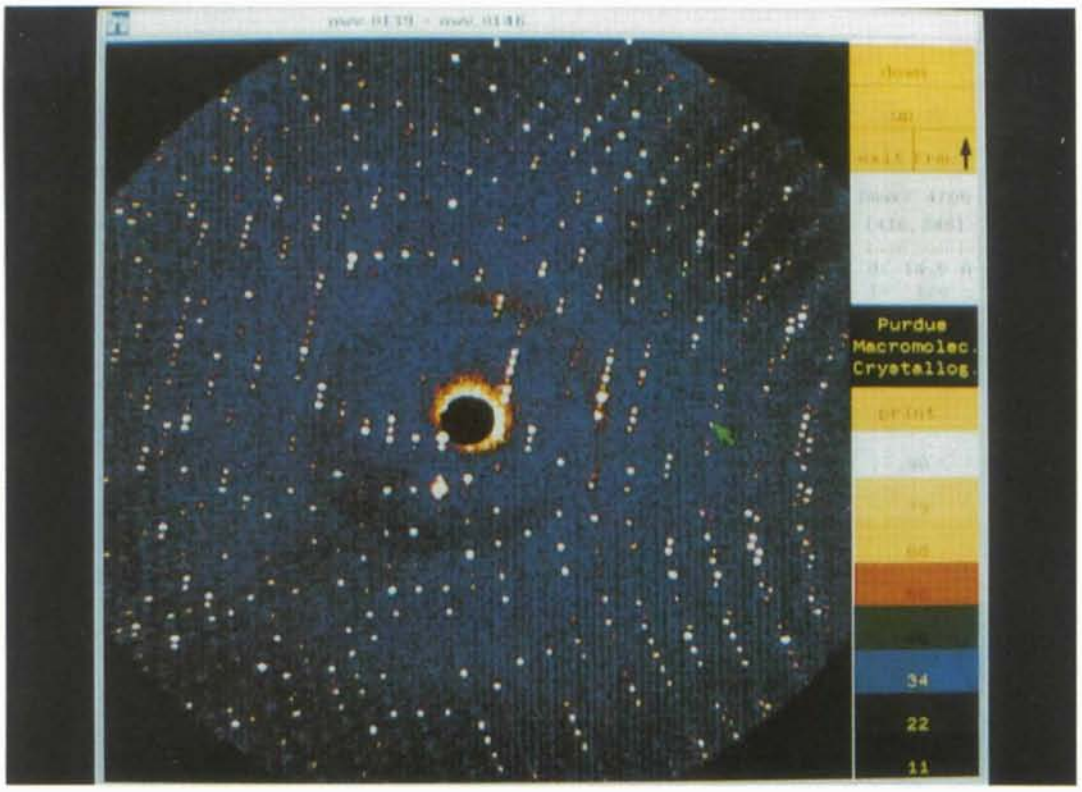
Owing to the long exposure times and small oscillation angles required to record high-resolution data, the reflections at lower resolution are nearly all partial and they are frequently sufficiently intense that the photographic film is saturated. Data between 150 and 8·5 Å resolution are important in virus crystallography. The initial phase determination of the NωV structure amplitudes will employ a model derived from electron micrographs of frozen hydrated particles (Olson, Baker, Bomu, Johnson & Hendry, 1987). This model is accurate to 20 Å resolution. Initially we intend to determine NωV phases at very low resolution using the phase-extension method successfully employed at higher resolution (extending from 8 to 3·0 Å) in the determination of other virus structures. Amplitudes of structure factors must be accurately measured between at least 100 and 8 Å resolution for this method to succeed for NωV. We have employed a Siemens area detector to measure this low-resolution data. Fig. 2(*a*) shows the electronic pseudo-oscillation photograph of NωV that can be obtained using double mirror focusing and a crystal-to-detector distance of 280 mm. The signal-to-background ratio, although very good, is artificially worsened by the fact that the background on the photograph is a sum of the backgrounds in the ten individual frames whereas only a few frames contributed to the intensity of the diffracted spots. Fig. 2(*b*) shows the spatial separation of the spots around the 9 Å resolution region. The green window shows an analysis that can be performed on individual diffraction maxima. The quality of the auto-indexing procedure and of the lattice-refinement process was analyzed using the programs developed by Minor (Minor & Bolin, 1990). Fig. 3 presents the observed and predicted diffraction spots displayed by the program *QUAL*.

The problems in determining the correct cell choice encountered with the photographic data were also found in the cell assignment for the area-detector data. The cell for each crystal was checked against the film data set and the correct cell was identified by the indexing that gave the best scaling. After identifying the unique cell all the area detector data were scaled together and the results are shown in Table 3.

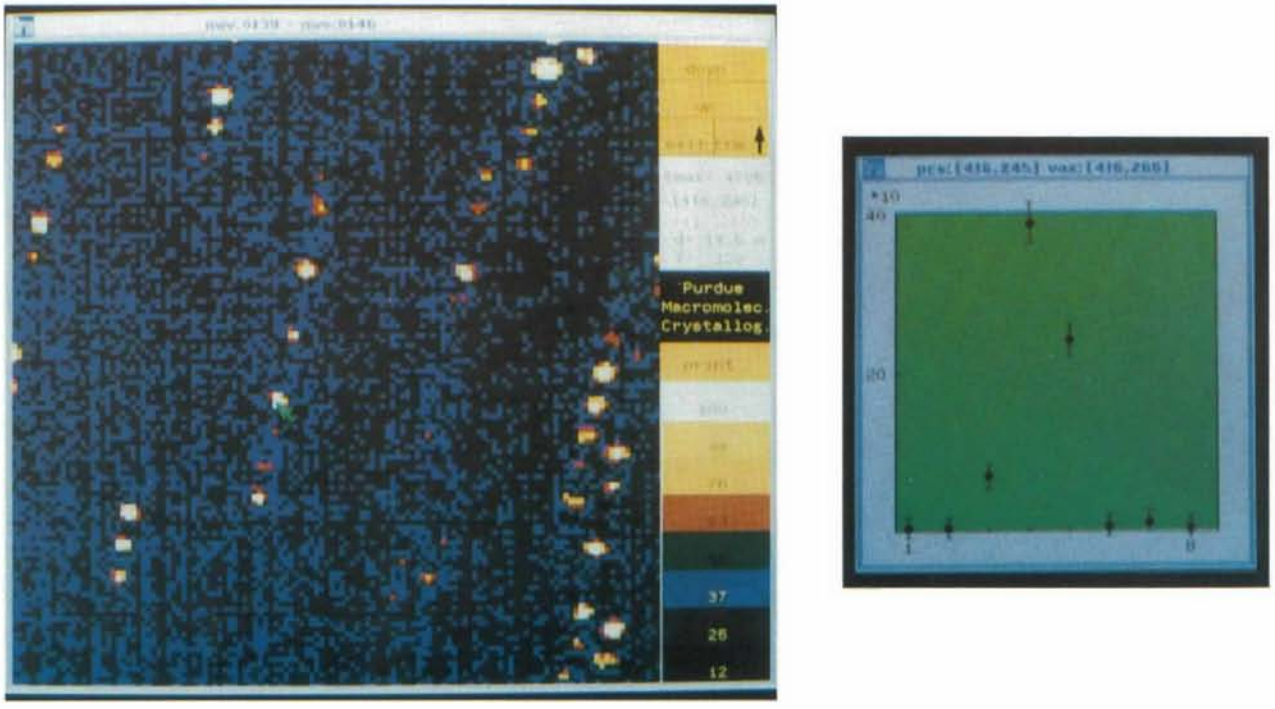
Rotation functions were computed using only the film data (Fig. 4) and only the area-detector data, and the results were similar. In spite of the partial data set used in the calculation, exceptionally clear rotation functions were produced. The presence of one icosahedral particle in the asymmetric unit of the crystal cell requires 31 unique non-crystallographic rotational symmetry axes (15 twofold axes, 10 threefold axes, 6 fivefold axes). Peaks for all 31 positions were identified in rotation functions computed with $\kappa = 180$, 120, 72 and 144°. The symmetry elements of an icosahedron were fitted to this constellation by the least-squares method (Fig. 5). Using the orientation from the rotation function a packing diagram was constructed. Fig. 6 shows the relation between the icosahedral symmetry elements and the unit-cell axes.

Discussion

The results obtained in this study demonstrate that semi-automated procedures can be used to process data from crystals with a unit cell displaying an exceptional degree of similarity in axial parameters and pseudo-symmetry. This study has also demonstrated the feasibility of using an area detector for the collection of low-resolution virus data. This data is useful for overlapping resolution ranges analyzed by electron microscopy and X-ray diffraction. What is more, data in the 100 to 8 Å range can be used to



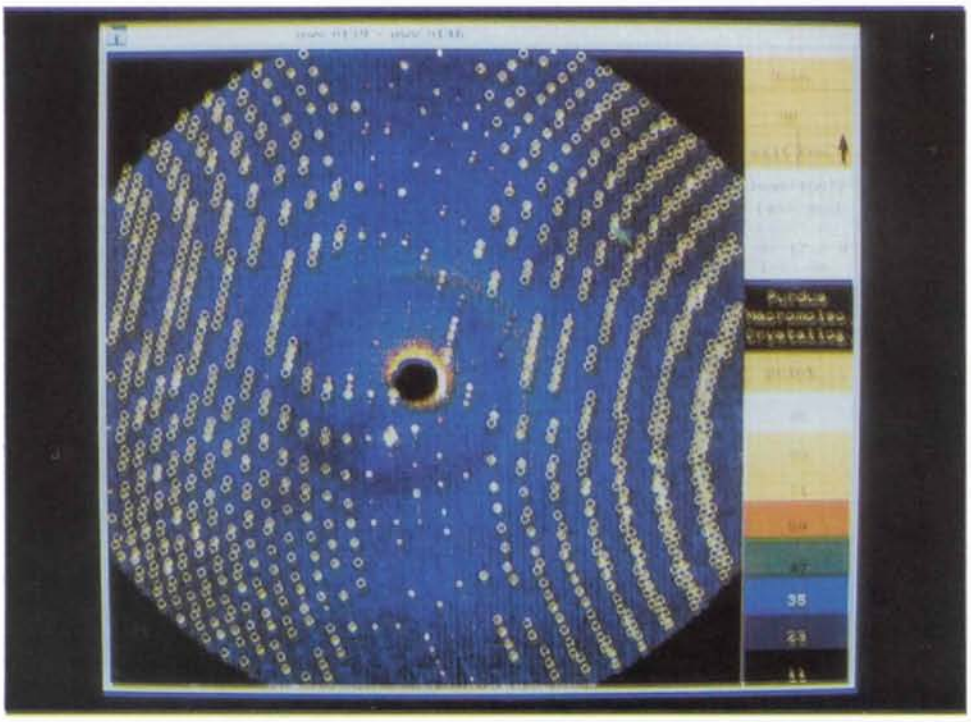
(*a*)



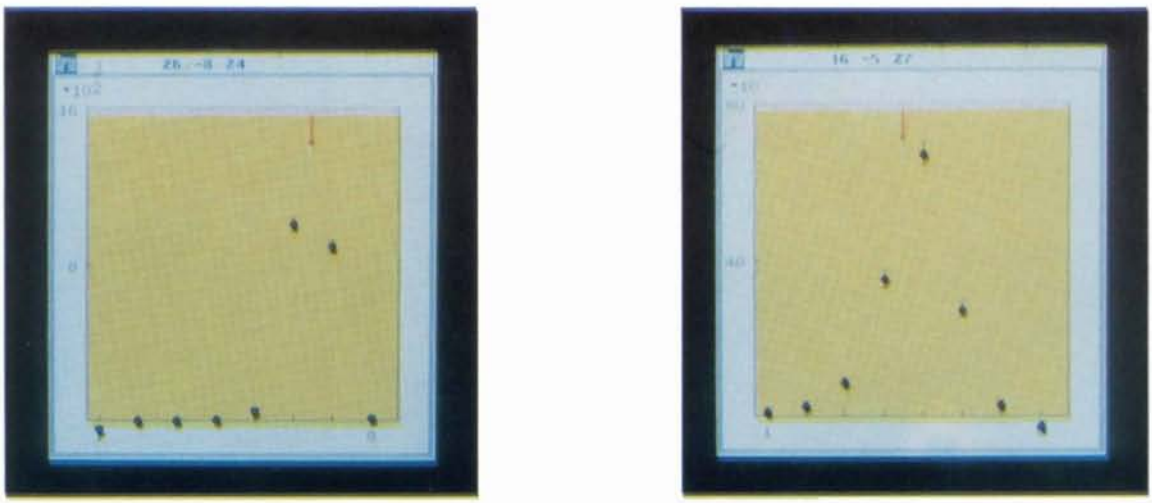
(*b*)

Fig. 2. (*a*) An electronic pseudo-oscillation photograph of an NωV crystal. The picture contains ten individual frames of 0·05° each, corresponding to an oscillation angle of 0·5°. Note that the background on the 'photograph' is a sum of the backgrounds in the ten individual frames whereas only a few frames contributed to the intensity of each diffraction maxima. (*b*) Separation of the diffraction spots on a pseudo-oscillation photograph (expanded diffraction pattern). The green window, activated by the green pointer in the left window, shows the profile of a particular diffraction spot in the third dimension (frame number or equivalent ω angle). The integrated intensity (together with full error analysis) of the selected spot is calculated in three dimensions.

[*To face p.* 26



(*a*)



(*b*)

Fig. 3. Predicted and observed spots for an oscillation photograph displayed by the program *QUAL*. The upper window (*a*) gives information about the difference between the observed and calculated positions in two dimensions. The two bottom windows (*b*), activated by the green pointer on the upper window, show the experimental profile of the selected spot as well as the calculated peak position (red arrow). The region near the vertical rotation axis is inaccessible for automatic integration because the reflection profiles are too broad.

Table 3. *Number of independent reflections after scaling of area-detector data from* 14 *crystals*

The R value for all reflections is 17·5% for 231 907 observations.

Resolution	No. of unique reflections	% of theoretically observable data
50–40 Å	213	26
40–30 Å	912	40
30–20 Å	4456	47
20–15 Å	9713	53
15–12 Å	16511	55
12–10 Å	24629	55
10–9 Å	19900	50
9–8 Å	24817	39
Total	100358	45

solve the basic crystallographic problems associated with molecular or isomorphous replacement for virus studies. This will permit all the preliminary crystallographic work to be completed in the laboratory. Synchrotron radiation and photographic film are required for collecting high-resolution data. The successes reported for other virus structures using molecular averaging to extend the resolution of phases indicates that a structure at 3 Å should be solvable using only the high-resolution native data. The use of the area detector in the laboratory should allow synchrotron time to be used more efficiently.

We thank Bonnie McKinney for technical assistance, Sharon Fateley for help in preparation of the manuscript, Jeff Bolin for helpful discussions regarding the area detector, Bob Sweet, the staff at NSLS and Paul Sehnke for help with the data collection, and Tim Baker and Norm Olson for their helpful discussions. We thank Andrew Fisher for preparing Fig. 6. This work was supported by National Institutes of Health Grant GM34220 to JEJ, National Science Foundation Grant 8509549A3-DIR to JEJ for computing and graphics facilities, and a grant from the Lucille P. Markey Foundation. SM was supported by National Institutes of Health Biophysics Training Grant GM0826. LL was supported by the Swedish National Science Research Council.

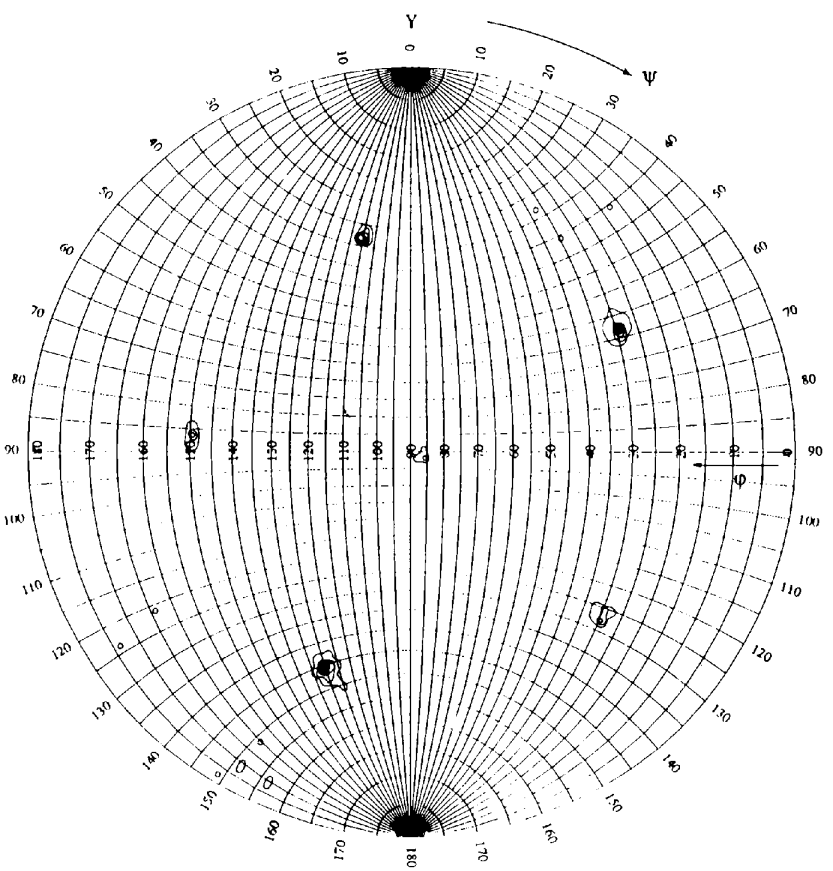


Fig. 4. A representative rotation function ($\kappa = 72°$) for NωV computed with the partial data set recorded on photographic film. The resolution of data used was 9–6 Å with a radius of integration of 120 Å. A total of 964 'large terms' (representing 2% of the measured data in the resolution range) were used to represent the second Patterson. The rotation function is scaled so that the largest peak is 10. The first contour is at 2 with additional contours at intervals of 2. Note that the Z axis projects out of the page and the contours are on the back of the stereographic sphere since φ was computed from 0 to 180° (see Fig. 5).

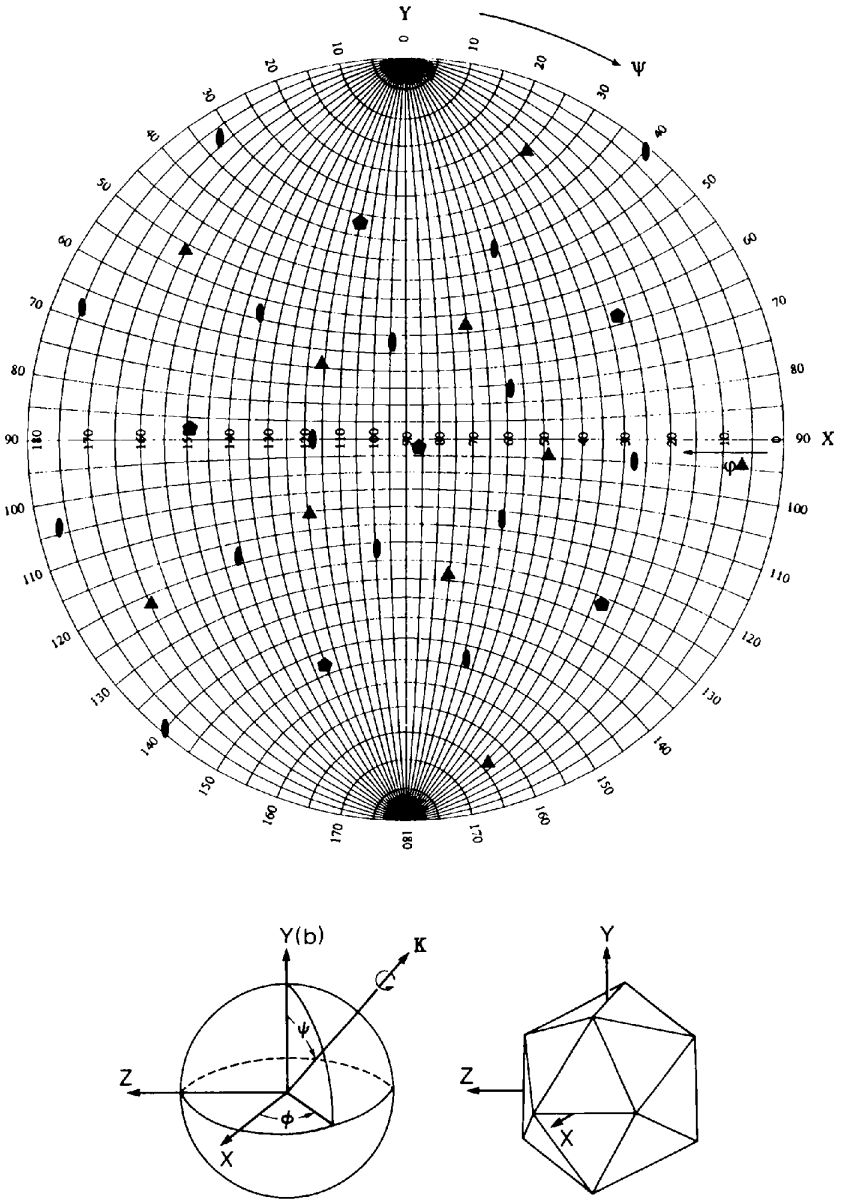


Fig. 5. The top of the figure shows a stereographic net with the constellation of symmetry elements for the NωV particle in the triclinic crystal cell. The orthogonal coordinate system defined by Rossmann & Blow (1962) was used, and the relation between the spherical polar coordinate system and the orthogonal coordinate system of the rotation function is shown in the lower left figure. The lower right figure shows an icosahedron in a standard orientation. The symmetry elements of the icosahedron in the standard orientation are oriented in the crystal cell by the following operations: $\mathbf{X}_{\text{crystal}} = [\alpha][\rho]\mathbf{X}_{\text{standard orthogonal}}$, where ρ is a matrix to rotate coordinates within the orthogonal system, α is a matrix to convert the orthogonal to the crystal cell (Rossmann & Blow, 1962),

$$[\rho] = \begin{bmatrix} 0{\cdot}46534824 & 0{\cdot}07509195 & -0{\cdot}88156861 \\ -0{\cdot}00050272 & 0{\cdot}99691778 & 0{\cdot}08332828 \\ 0{\cdot}88511270 & -0{\cdot}03777157 & 0{\cdot}46370584 \end{bmatrix}, \quad [\alpha] = \begin{bmatrix} 1{\cdot}1996059 & 0{\cdot}0000000 & 0{\cdot}0000000 \\ -0{\cdot}2789580 & 1{\cdot}0000000 & -0{\cdot}5935200 \\ -0{\cdot}4752850 & 0{\cdot}0000000 & 1{\cdot}1628700 \end{bmatrix},$$

$$[\alpha][\rho] = \begin{bmatrix} 0{\cdot}55823451 & 0{\cdot}09008075 & -1{\cdot}05753493 \\ -0{\cdot}65564740 & 0{\cdot}99838847 & 0{\cdot}05403021 \\ 0{\cdot}80809796 & -0{\cdot}07961351 & 0{\cdot}95822597 \end{bmatrix}.$$

The matrix ρ was determined by a least-squares fit of the standard icosahedral symmetry elements to the 31 independent symmetry elements determined using rotation functions calculated for κ = 180, 120, 72 and 144°. Note that as in Fig. 4, the Z axis projects out of the page and the symmetry elements are on the back (below the page) of the spherical stereographic net.

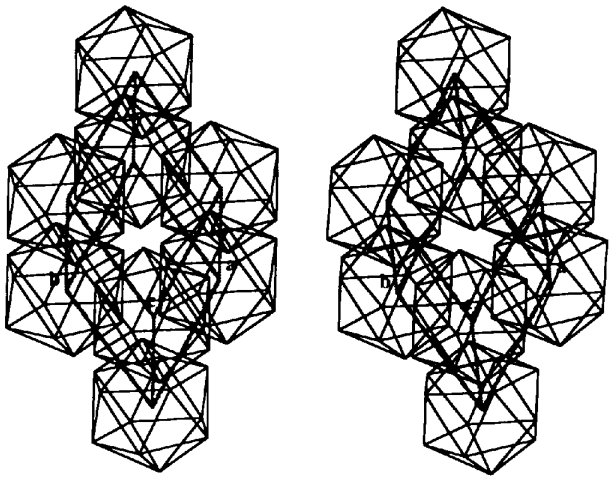


Fig. 6. A packing diagram showing the unit cell defined in line 1 of Table 1 and the particles oriented according to the rotation function in Fig. 5.

Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12-C in the National Synchrotron Light Source. This facility is supported by the US Department of Energy, Office of Health and Environmental Research.

References

BLUM, M., METCALF, P., HARRISON, S. C. & WILEY, D. C. (1987). *J. Appl. Cryst.* **20**, 235–242.
DURBIN, R. M., BURNS, R., MOULAI, J., METCALF, P., FREYMANN, D., BLUM, M., ANDERSON, J. E., HARRISON, S. C. & WILEY, D. C. (1986). *Science*, **232**, 1127–1132.
FINCH, J. T., CROWTHER, R. A., HENDRY, D. A. & STRUTHERS, J. K. (1974). *J. Gen. Virol.* **24**, 191–200.
HENDRY, D., HODGSON, V., CLARK, R. & NEWMAN, J. (1985). *J. Gen. Virol.* **66**, 627–632.
HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. R. (1987). *J. Appl. Cryst.* **20**, 383–387.
KIM, S. (1989). *J. Appl. Cryst.* **22**, 53–60.
MCPHERSON, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: John Wiley.
MINOR, W. & BOLIN, J. T. (1990). In preparation.
OLSON, N. H., BAKER, T. S., BOMU, W., JOHNSON, J. E. & HENDRY, D. A. (1987). *Proceedings of the 45th Annual Meeting of the Electron Microscopy Society of America*, edited by G. W. BAILEY, pp. 650–651. Baltimore, MD: Electron Microscopy Society of America.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
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.

Acta Cryst. (1991). B**47**, 29–40

Automated Conformational Analysis from Crystallographic Data. 1. A Symmetry-Modified Single-Linkage Clustering Algorithm for Three-Dimensional Pattern Recognition

BY FRANK H. ALLEN* AND MICHAEL J. DOYLE

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

AND ROBIN TAYLOR

ICI Agrochemicals, Jealott's Hill Research Station, Bracknell, Berkshire RG12 6EY, England

(*Received* 2 *November* 1989; *accepted* 12 *September* 1990)

Abstract

Single-linkage cluster analysis is used to identify discrete conformational subgroups for a chemical fragment from crystal structure data. Fragment conformations are defined by N_t torsion angles for N_f occurrences of the fragment in the Cambridge Structural Database. Conformational analysis is complicated by (*a*) the 2D topological symmetry of the fragment, giving rise to permutations of torsion-angle sequences, and (*b*) by the presence of 3D enantiomers in the original crystal structures. All steps in the single-linkage algorithm are modified to use fragment symmetry to obtain the optimum torsional overlap of all fragments. Thus, all symmetry equivalents of a given conformation are grouped into the same cluster and the final set of clusters represents an asymmetric unit of conformational space. Principal-component analysis is used to provide a visual mapping of the clustering process. The complete procedure is shown to be effective when applied

* Author for correspondence.

0108-7681/91/010029-14$03.00