Table 5. *Comparison of results with 5 K data for the one-Gaussian and Howard's (1982) three-Gaussian peak-shape functions*

The errors in the lattice constants do not include the absolute error in the neutron wavelength.

	One-Gaussian	Three-Gaussian
R factor	0·057	0·053
Scan zero	0·082 (4)	0·005 (15)
Unit-cell parameters (Å)		
a	7·1689 (3)	7·1725 (7)
b	9·5631 (5)	9·5689 (12)
c	4·6766 (3)	4·6791 (6)

to the statistical limit for these data. The difference between the refined scan origins for the two peak-shape functions is very considerable, being about 6 standard deviations, and the unit-cell volume changes systematically for the 5 and 57 K data by about 0·22% (see Table 5). This shows that the new function should be taken as a standard in powder refinements, especially as there is no noticeable slowing down of the program. Users of *EDINP* are invited to request an update description from the first author to incorporate the new function.

The error estimates for those parameters affected by the asymmetry function (for the 5 and 57 K data) increase when the better function is used. This appears to be quite a usual result in least-squares refinements when an improved function is used.

Concluding remarks

The structure of solid CF_3Cl between 5 and 81 K is found to be orthorhombic $Cmc2_1$. A phase transition below 5 K is most unlikely, leaving only 81 K to the melting point at 92 K as a possible range in which a plastic phase might exist. The mean thermal linear expansivities along *x*, *y* and *z* in the ranges 5–57 and 57–81 K are 0·0011, 0·0006, 0·0006 and 0·0021, 0·0007, 0·0011 K^{-1} respectively, showing no anomaly in the upper range.

Structure solution has been achieved solely with the use of neutron powder diffraction data. The unit cell was solved by using diffraction peak positions estimated without correction for peak shift due to finite-size counter apertures, whereas the refinement has shown this shift to be significant. Future unit-cell analyses should certainly take this shift into account, and it is suggested that Howard's (1982) three-Gaussian function should be used throughout the profile refinement.

References

HEWAT, A. W. (1978). *The POWDER System for the Collation and Examination of Neutron Powder Diffraction Data.* Internal Report, ILL, Grenoble.
HOWARD, C. J. (1982). *J. Appl. Cryst.* **15**, 615–620.
IMMIRZI, A. & PERINI, B. (1977). *Acta Cryst.* A**33**, 216–218.
PAWLEY, G. S. (1980). *J. Appl. Cryst.* **13**, 630–633.
PAWLEY, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
VISSER, J. W. (1969). *J. Appl. Cryst.* **2**, 89–95.
WOLFF, P. M. DE (1968). *J. Appl. Cryst.* **1**, 108–113.

Acta Cryst. (1985). B**41**, 139–147

Structure of a Serine Protease from Rat Mast Cells Determined from Twinned Crystals by Isomorphous and Molecular Replacement

BY R. A. REYNOLDS,* S. J. REMINGTON, L. H. WEAVER, R. G. FISHER,† W. F. ANDERSON,‡ H. L. AMMON§ AND B. W. MATTHEWS

Department of Physics and Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, *USA*

(*Received* 7 *September* 1983; *accepted* 17 *April* 1984)

Abstract

The structure of rat mast cell protease II, a serine protease from the mucosal mast cells of rat small intestine, has been determined by a combination of single isomorphous replacement and molecular replacement. The crystals, space group $P3_1$, have two molecules per asymmetric unit, related by a local twofold axis of symmetry approximately parallel to **a**, resulting in pseudo space-group symmetry $P3_121$. The crystals are invariably twinned in such a way that the *hkl* and $kh\bar{l}$ reflections superimpose. Methods used to 'detwin' the data are described. An initial electron density map phased on a single heavy-atom derivative could not be interpreted, but provided the basis for the eventual determination of the structure. An automatic procedure was used to determine the approximate molecular boundaries. A search within

* Present address: Physics Department, Colby College, Waterville, Maine 04901, USA.
† Present address: Smith Kline and French Labs, Mail Code F32, 1500 Spring Garden Street, Philadelphia, PA 19101, USA.
‡ Present address: Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.
§ Present address: Department of Chemistry, University of Maryland, College Park, MD 20742, USA.

0108-7681/85/020139-09$01.50

this region for electron density corresponding to an α-chymotrypsin-like molecule clearly revealed the location of the molecule. Then a series of cycles of refinement yielded the three-dimensional structure. The R value is 24·0% for data to a resolution of 1·9 Å. As expected from its amino acid sequence homology, the structure of the mast cell protease resembles that of α-chymotrypsin, although there are surface loops where the two molecules differ substantially. Initial attempts to locate the molecule by rotation-function techniques, using α-chymotrypsin as a probe, were unsuccessful. It is shown that this failure was due to a combination of two factors, namely the inadequacy of the search model together with the relatively high symmetry of the crystals.

Introduction

Rat mast cell protease II (RMCP II) is a serine protease originating from the mucosal mast cells of rat small intestine (Woodbury, Gruzenski & Lagunoff, 1978). Originally thought to hydrolyze preferentially the apo form of certain pyridoxal-phosphate-requiring proteins (Katanuma *et al.*, 1975), substrate specificity studies have shown that this is most likely not the case and its function *in vivo* is, at present, uncertain (Yoshida, Everitt, Neurath, Woodbury & Powers, 1980). The protein consists of a single polypeptide chain of 224 amino acid residues with molecular mass 24 665 daltons. There are three disulfide linkages and the amino acid sequence is known (Woodbury, Katanuma, Kobayashi, Titani & Neurath, 1978). RMCP II has approximately 33% sequence identity with α-chymotrypsin, trypsin, and elastase (Woodbury, Katanuma *et al.*, 1978) and is therefore expected to have structural similarities with these enzymes.

We describe the structure determination of RMCP II, which involved several unusual aspects. The observed data had to be analyzed and corrected for twinning, a phenomenon which occurred to a greater or lesser degree in all crystals of the protein. The actual solution of the structure involved a combination of the methods of single isomorphous replacement and molecular replacement. We describe the method used to locate the molecular boundary and to determine the position and orientation of the molecules in the unit cell.

Crystallization

The protein was crystallized from 23% saturated ammonium sulfate, in a range of pH from 7·0 to 8·0, as described by Anderson, Matthews & Woodbury (1978). The space group is $P3_1$ with cell dimensions $a = b = 78{\cdot}2$ Å, $c = 96{\cdot}8$ Å, but the distribution of intensities is pseudo $P3_121$ (*i.e.* the true Laue symmetry is $\bar{3}$ but the observed symmetry is pseudo $\bar{3}m$). There are two molecules per asymmetric unit and it was therefore inferred that there must be an axis of local twofold symmetry close to the x axis, *i.e.* close to the crystallographic twofold axis in the higher-symmetry space group. The crystals were merohedrally twinned about the same axis. The approximate coincidence of an axis of twinning with a local twofold axis is not without precedence, occurring, for example, in α-chymotrypsin (Blow, Rossmann & Jeffrey, 1964). By symmetry, RMCP II has additional local twofold axes close to the y axis and to the bisector of x and y. Crystallization in the presence of various solvents using different buffers and pH values did not eliminate the twinning. Also, individual crystals could not be mechanically separated into untwinned domains.

Data collection

A preliminary set of three-dimensional data was measured by precession photography (Matthews, Klopfenstein & Colman, 1972). Rotation-function calculations (Rossmann & Blow, 1962; Crowther & Blow, 1967), based on this data set, confirmed the presence of a local twofold symmetry axis approximately parallel to **a**, but did not reveal the orientation of an α-chymotrypsin-like molecule in the unit cell. It was therefore necessary to resort to isomorphous replacement.

A preliminary search for heavy-atom derivatives was made using projection data collected by precession photography. The most promising candidate was obtained with sodium mersalyl. The derivative was prepared by soaking a crystal for 2 d in 5 mM sodium mersalyl followed by a 4 h period in 0·5 mM sodium mersalyl prior to mounting.

Three-dimensional data were collected for native crystals and for the sodium mersalyl derivative to 1·9 Å using oscillation photography and camera geometry described by Schmid *et al.* (1981). An oscillation angle of 1·5° per film pack was used for the native crystals, while 2° per film pack was used for the derivative. The crystals were rotated about the x axis through a total range of 90°. The X-ray source was a graphite-monochromatized Elliot GX-21 rotating-anode generator operated at 39 kV, 130 mA, and the exposure per film pack was 3 h for the native data and 4·5 h for the derivative. Typical crystal size was $1{\cdot}0 \times 0{\cdot}6 \times 0{\cdot}5$ mm, substantially larger than the cross section of the X-ray beam ($0{\cdot}35 \times 0{\cdot}45$ mm). Films were evaluated using the methods of Rossmann (1979) and Schmid *et al.* (1981). Because of the twinning of the crystals there was often a substantial variation in the optical density profile from one reflection to the next (Fig. 1). This variation made profile fitting impossible and intensities were therefore measured by summing of density values within a

suitable size box, with appropriate background subtraction.*

Correction for twinning

The data were 'detwinned' using programs of Fisher & Sweet (1980) based on the principles of Britton (1972). The twinning causes a superposition of the reflections hkl and $k, -h-k, l$ which is in turn related to $kh\bar{l}$ by the $P3_1$ symmetry equivalence and Friedel's law. If the (desired) intensities from an untwinned crystal are $I(hkl)$ and $I(kh\bar{l})$, and x is the volume fraction of a crystal that is twinned, then the resultant (observed) intensities are:

$$J(hkl) = (1-x)I(hkl) + xI(kh\bar{l})$$
$$J(kh\bar{l}) = (1-x)I(kh\bar{l}) + xI(khl), \quad (1)$$

from which

$$I(hkl) = \frac{(1-x)J(hkl) - xJ(hk\bar{l})}{(1-2x)}. \quad (2)$$

The data-collection strategy was chosen to ensure that twin-related reflections would occur in pairs on the same film (Fig. 1). It was therefore possible to estimate a twinning fraction for each film pack separately. Plots were made of the number of reflections calculated to have negative intensities as a function of the twinning fraction (Fig. 2). In the absence of experimental error, the desired twinning fraction is the largest value that predicts no negative intensities. This would result in a plot in which the twinning fraction is given by the intercept of the line and the twin fraction axis (*cf.* Fig. 2*a*). Errors in intensity measurements result in some calculated negative intensity measurements, even with the correct twinning fractions. The effects of errors become more serious as the twinning fraction increases (Fisher & Sweet, 1980). Examples of well determined and poorly determined twin fractions are shown in Figs. 2(*a*) and 2(*b*). In practice we chose as a first estimate for the twinning fraction the value at which the curve changed from monotonically decreasing to essentially horizontal (Fig. 2*b*). In general, the procedure works well for the low twin fractions (less than about 15%) and becomes less reliable for intermediate values. The equations for recovering the 'untwinned' reflection intensities become indeterminate as the twin fraction approaches 50% [equation (2)]. For this reason, the many crystals that had estimates of twinning of 33% or greater had to be discarded.

Estimates of the twinning fraction determined from successive exposures of the same crystal were compared for internal consistency. Where these estimates agreed, the initial twinning fraction was redetermined using all the data collected from the crystal. For one native and one derivative crystal, the different crystal alignment from one exposure to the next was associated with an obvious change in the estimated twin fraction. This change occurred because the X-ray beam did not bathe the whole crystal. Indeed, we tried to select a part of the crystal for which the twin fraction was low. In these two cases the data were separated into consistent subsets.

One needs to distinguish a 5% twin from a 95% twin. This is equivalent to interchanging the left and right halves of a given film (see Fig. 1). There is also, in this space group, the need to distinguish the absolute sense of the crystallographic z axis. These problems of consistent indexing were resolved during the merging of the data.

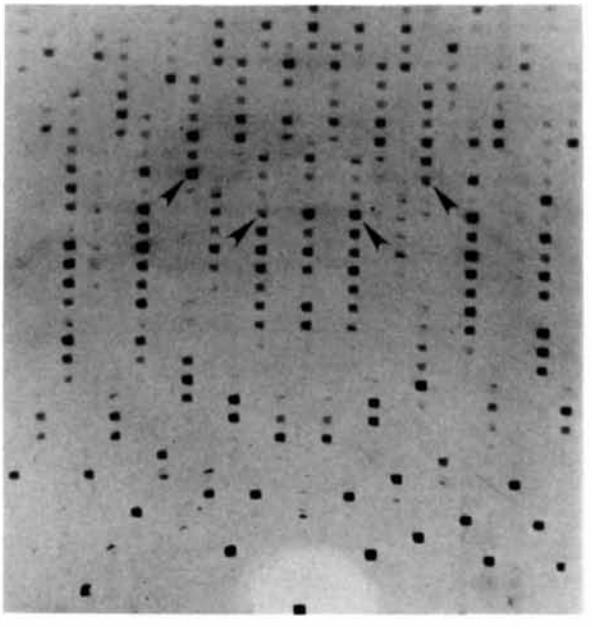

Fig. 1. Section of an oscillation photograph (x axis horizontal). Twin-related reflections occur in symmetrical positions on the left and right halves of the film. Two pairs of twin-related reflections in which the consequences of the twinning can easily be seen are indicated by arrowheads. In this case the amount of twinning is low, about 5%. The film cassette is cylindrical with radius 80 mm. The reflections measure about 0·7 × 0·7 mm, and the spacing between the non-centered vertical rows is 3·1 mm. The film was momentarily exposed to the direct beam (bottom center). It is the rectangular profile of the direct beam, and not the shape of the crystal, that defines the reflection profiles. Reflections at the top of the figure have Bragg spacings of 3·0 Å.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 3RP2, R3RP2SF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Cambridge, Melbourne or Osaka. The data have also been deposited with the British Library Lending Division as Supplementary Publication No. SUP 37012 (2 microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 1. *Merging statistics for crystals with and without correction for twinning*

$R_{merge} = 100 \sum |I - \bar{I}| / \sum I$ where the intensities are averaged over different films. Crystals numbered (4*a*), (4*b*), *etc.* are exposures of different regions of the same crystal.

Crystal number	Number of film packs	Initial twin fraction	Final twin fraction	R_{merge} without detwinning	R_{merge} after detwinning
Native data					
(1)	14	5%	5%	11·2	7·7
(2)	25	3	3	16·5	9·0
(3)	14	5	4·5	22·9	11·7
(4*a*)	8	16	17	16·1	8·4
(4*b*)	3	23	21	15·6	8·1
Derivative data					
(1*a*)	10	5	5	—	10·4
(1*b*)	6	9	9	—	9·4
(1*c*)	4	15	14	—	10·3
(1*d*)	4	22	21	—	11·7
(2)	25	24	24	—	11·1

Table 2. *Overall reflection statistics*

Crystal	Total number of films	Total number of reflections measured	Number of independent reflections	R_{merge}	Completeness of data to 1·9 Å
Native	64	79 878	31 427	9·2	60·2%
Derivative	49	58 502	24 230	10·5	46·4

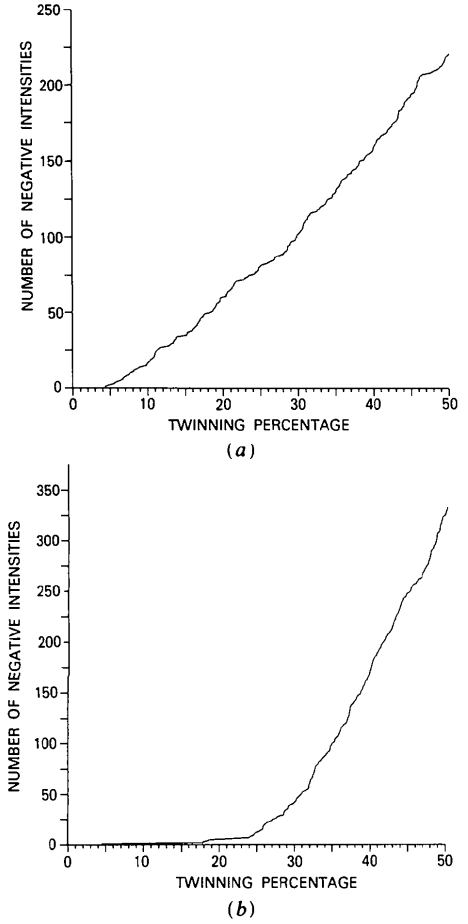

Fig. 2. Plots used to determine the approximate fraction of twinning, after Britton (1972) (see text). (*a*) Crystal with estimated twinning of about 4%. (*b*) Crystal with estimated twinning of about 24%.

The initial twinning fractions were refined by optimizing the agreement between the data sets from different crystals. The detwinned data from the crystal with the lowest estimated twinning fraction were used as a reference data set. The remaining data were detwinned and merged, crystal by crystal, with the reference set. Inspection of the merging statistics as a function of small adjustments in the twinning fraction for each crystal allowed refinement of the twin fractions. The resultant changes in the twin fractions were small (Table 1). A comparison of the data-merging statistics determined with, and without, twinning correction (Table 1) shows that the application of the detwinning procedure is clearly worthwhile. The overall statistics for all data used in the structure determination are shown in Table 2.

Electron density map

A difference Patterson map calculated with the three-dimensional data to 3·5 Å resolution clearly indicated two major heavy-atom sites. Heavy-atom refinement, coupled with the usual difference Fourier techniques, led to the location of four more heavy-atom sites. The heavy-atom sites could be grouped in pairs, with the members of each pair having similar occupancy, consistent with the presumed non-crystallographic symmetry axis (Tables 3 and 4).

Using the single isomorphous replacement (SIR) data we calculated an electron density map at 3·5 Å resolution. The map confirmed the local non-crystallographic symmetry, but could not be interpreted in detail. Using the transformation that related one set of heavy-atom sites to the other, the SIR electron density map was averaged about the local symmetry

Table 3. *Heavy-atom parameters*

Heavy-atom site	x	y	z	Occupancy	Temperature factor (Å^2)
(1)	0·7839	0·7230	0·0205	87·8	20·0
(2)	0·7095	0·7968	−0·0206	78·3	31·2
(3)	0·6514	0·5118	0·0256	27·1	38·4
(4)	0·4967	0·6466	−0·0395	34·7	46·0
(5)	0·1857	0·5154	0·0867	14·4	43·1
(6)	0·5344	0·1979	−0·1015	17·2	45·8

Table 4. *Heavy-atom refinement statistics [the quantities have their usual meanings (e.g. see Matthews, 1977)]*

Resolution (Å)	6·0–3·5
Number of centrosymmetric data	0
Number of non-centrosymmetric data	7477
Mean figure of merit	0·31
Mean isomorphous difference ($F_{PH} - F_P$)	74·8
r.m.s. lack of isomorphism (E)	64·5
'Phasing power' (r.m.s. F_H)/E	1·17

element. This map appeared to be an improvement over the unaveraged map, but was still of poor quality, with little indication of the course of the polypeptide backbone.

Location of the molecule in the SIR map

At this point, we attempted to locate the expected serine-protease-like structure within the electron density map. Because of the poor quality of the map it was necessary to devise an automatic search procedure. Our approach was in two parts, first to locate the molecular boundary, and then to search for an α-chymotrypsin-like molecule within this region.

The location of the molecular boundary was based on the expectation that the electron density within the molecular boundary has greater excursions, both positive and negative, if the $F(000)$ term is omitted, than the density corresponding to solvent, which is relatively featureless. A program was written to calculate a smoothed 'modulus' electron density in which each density value was the sum of the absolute values of all density points in a box of $n \times n \times n$ density values centered on the point in question. Fig. 3 shows a section of the resulting map, with $n = 17$, which indicates the molecular boundary quite clearly. A very similar result would be obtained by calculating the root-mean-square density within the parallelepiped. Recently B. C. Wang (personal communication) has independently developed a similar method of determining the molecular boundary. Wang evaluates a weighted average of the electron density above a threshold value. The main difference between Wang's method and the one we have used is that we make use of both negative and positive density values. We believe that this may be important in cases where the solvent density is very high. It is possible, for example for crystals grown from concentrated phosphate solutions, that the electron density of the solvent may exceed the average electron density within the protein boundary. In such cases the function that Wang uses would tend to give higher values in the solvent region than within the protein, thus interchanging the apparent locations of protein and solvent. The essential difference between the protein and solvent is not that the protein density is higher than that of the solvent, but that the solvent density should be uniform, whereas the density within the protein boundary fluctuates from high to low values. The procedure that we adopted is based on this distinction.

The next step was to determine the orientation and position of the molecule within the molecular boundary. As a search motif we used that part of the known α-chymotrypsin backbone (Matthews, Sigler, Henderson & Blow, 1967) that was believed to be structurally conserved in RMCP II (Greer, 1981). This corresponds to 58% of the α-chymotrypsin structure. The function calculated, for all possible rotations and translations, was the sum of the electron

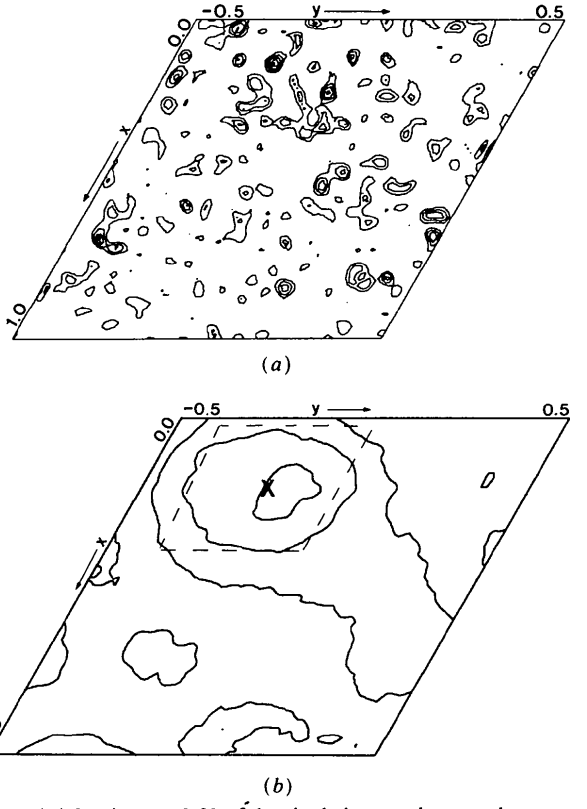

Fig. 3. (*a*) Section $z = 0{\cdot}23$ of the single isomorphous replacement electron density map, resolution 3·5 Å. (*b*) Section $z = 0{\cdot}22$ of the smoothed modulus density function (see text) intended to show the molecular boundary. Contours are drawn at increments of σ, beginning at 2σ, where σ is the r.m.s. value throughout the unit cell. The broken line indicates the region within which the initial search was made for correspondence with α-chymotrypsin. The cross indicates the location subsequently determined for the center of the molecule (see Fig. 4).

density at the α-carbon positions. The most efficient mode of carrying out this six-dimensional search is first to rotate the test motif to a given orientation and then translate the motif through the volume of interest. Because we had not determined the absolute configuration of the electron density distribution it was necessary to carry out the search in each of the enantiomorphic maps. Coordinates for α-chymotrypsin (Birktoft & Blow, 1972) were obtained from the Protein Data Bank (Bernstein *et al.*, 1977). Fig. 4 depicts the section of the six-parameter search that contained the major peak. The peak height is $11{\cdot}2\sigma$ and was close to the expected molecular center. The actual calculation was not as lengthy as we had anticipated, and, as a check, we subsequently carried out a coarse search through the full asymmetric unit (coordinate increments of 2·0 Å and angular increments of 15°). In space group $P3_1$, this coarse search required 68·0 h c.p.u. time on a VAX 11/780 computer, and revealed two major peaks, one the expected peak and the other corresponding to the twofold-related monomer. The two major peaks were of height 10σ and all other peaks were 5σ or less. In the enantiomorphic space group, $P3_2$, no peak was above 5σ. Thus, although the SIR map was of poor quality, it was sufficient to reveal the correct enantiomorph and the orientation and position of the probe molecule in a convincing manner.

Refinement

Having determined an initial orientation, the model was expanded to include 72% of the α-chymotrypsin structure. The additional residues correspond to loop regions of RMCP II that have similar length and/or residue character as α-chymotrypsin and were therefore expected to have a similar conformation in both α-chymotrypsin and RMCP II (Greer, 1981). The

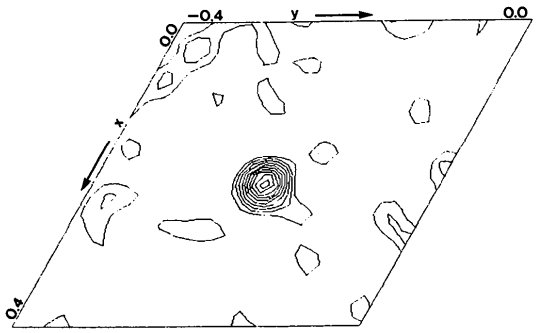

Fig. 4. Portion of the six-dimensional search function (section $z = 0{\cdot}22$) matching 58% of the backbone of α-chymotrypsin with the averaged SIR electron density map of RMCP II. Contour levels are drawn at increments of $1{\cdot}1\sigma$ where σ is the r.m.s. value throughout the search volume.

Table 5. *Summary of structure determination and refinement*

Type of refinement	Number of cycles	Resolution limits (Å)	Residual
Initial configuration	—	6·0-3·5	50·9%
CORELS	4	5·5-4·5	47·2
EREF	16	5·5-3·0	36·4
Model building	1	5·0-3·0	40·4
EREF	14	5·0-2·5	32·5
Model building	1	5·0-2·5	36·7
EREF	12	5·0-2·0	28·5
Model building	1	5·0-2·0	31·5
EREF	7	5·0-1·9	24·0

RMCP II dimer was generated by use of the local symmetry. Structure factors calculated for this model had an R factor of 50·9% for data from 6·0 to 3·5 Å resolution. Rigid-body refinement of the two monomers with *CORELS* (Sussman, Holbrook, Church & Kim, 1977) reduced the R factor to 47·2% (data from 5·5 to 4·5 Å) and resulted in a net movement of the model of about 1 Å. The model was then subjected to the Jack & Levitt (1978) refinement procedure (*EREF*) resulting, after 16 cycles, in a residual of 35·5% to 3·5 Å resolution.

A $(2F_o - F_c)$ electron density map was calculated by combining the SIR and calculated phases (Hendrickson & Lattman, 1978), then averaged about the local symmetry axis. This map and current model were transferred to an MMSX graphics system for model building (Molnar, Barry & Rosenberger, 1976). The amino acid sequence of the model was changed to that of RMCP II and several obvious discrepancies between map and model were corrected, using *FRODO* (Jones, 1978). Few new residues were added at this point, as the map quality was still poor in the loop regions where the conformation of RMCP II and α-chymotrypsin differ. A second series of refinement cycles, working to higher resolution, resulted in an R factor of 32·8% (5-2·5 Å). A second model-building cycle was initiated, again using an averaged $2F_o - F_c$ map, calculated with combined phases. This map was substantially better than previous maps and allowed all the remaining residues of the RMCP II molecule to be placed. After crystallographic refinement this model had a residual of 28·5% to 2 Å resolution, confirming the overall correctness of the structure determination. Table 4 summarizes the course of the refinement, including additional cycles of model building that have been completed to this point. The present coordinates have been deposited in the Brookhaven Data Bank (Bernstein *et al.*, 1977). The model building included in Table 5 is for electron density maps averaged about the local symmetry axis. To complete the refinement, and to define possible structural differences between the two molecules in the asymmetric unit, it will be necessary to check the model against the unaveraged electron density. This

is currently in progress and will be reported elsewhere.

Results

This paper deals with the methods used to determine the structure of RMCP II. A detailed description of the molecule and its relation to the other known serine proteases will be deferred pending the completion of the refinement.

As expected from the amino acid sequence homology, the overall folding of RMCP II is similar to that of the mammalian pancreatic serine proteases. Fig. 5 is a superposition of the backbone of RMCP II on that of α-chymotrypsin. At the center of the molecule, the two backbones correspond closely, but in some surface loops there are substantial differences. In particular, the loop in α-chymotrypsin that includes His 40 is three residues longer in RMCP II and has moved by about 10 Å. If the backbones of α-chymotrypsin and RMCP II are superimposed by the method of Rossmann & Argos (1976), 206 α-carbons are found to be stereochemically 'equivalent', with an r.m.s. positional discrepancy of 2·0 Å. If the criterion for 'equivalence' is made more rigorous, then 152 α-carbons are found to superimpose within 0·76 Å. The parts of the two molecules that agree best are those that tend to be structurally conserved in other serine proteases (James, Delbaere & Brayer, 1978; Greer, 1981). In general, the comparison of the structure of RMCP II with α-chymotrypsin supports the sequence alignment of the two molecules proposed by Woodbury, Katanuma *et al.* (1978) except that residues Thr 52–His 58 of RMCP II should be aligned with Val 65–Phe 71 of α-chymotrypsin [*cf.* Table IV of Woodbury, Katanuma *et al.* (1978), noting that RMCP II was referred to at that time as GSP]. A detailed comparison of the structure and sequence of RMCP II together with implications of the active-site geometry for substrate specificity will be given in a subsequent report.

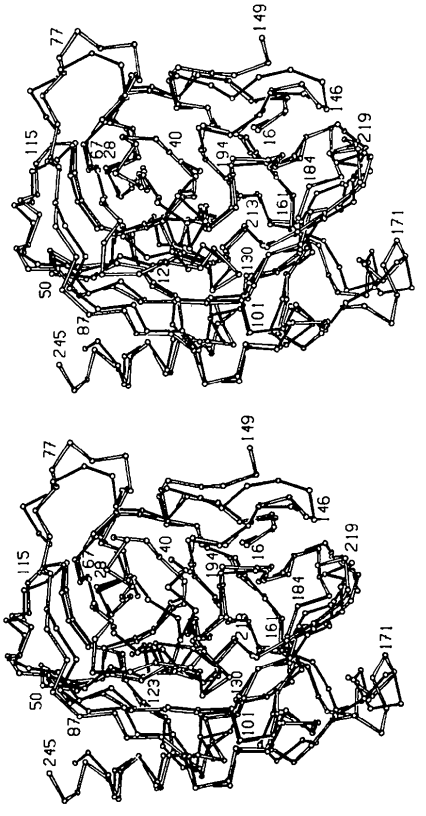

Fig. 5. Stereodrawing of the backbone of RMCP II (drawn solid) superimposed on that of α-chymotrypsin (open connections with numbering). The numbering scheme follows the standard convention for the serine proteases (Hartley, 1970). The active site region is formed by the depression in the lower front of the molecule.

From a crystallographic viewpoint, one feature of the structure determination that was disconcerting was the failure of the rotation function to locate successfully an α-chymotrypsin-like molecule in the asymmetric unit. Subsequent to solving the structure of RMCP II we repeated these calculations using the methods of Rossmann & Blow (1962), Crowther & Blow (1967) and Huber (1969). Each of these methods yielded approximately the same result.

The cause of the failure is therefore not due to the particular algorithm used, but might be due either to the inadequacy of α-chymotrypsin as a search model for RMCP II or to the presence of intramolecular vectors in the Patterson function, or to some combination of these. In order to distinguish between these possibilities we performed a series of test calculations.

First, instead of using α-chymotrypsin as the search molecule, we used one monomer of the refined RMCP II structure itself (*i.e.* one half of the asymmetric unit). The rotation function (Fig. 6*a*) yielded a large maximum at the expected position (0, 0, 0) and also at the position (0, 180, 0) generated by the local twofold axis. We then used a molecule of chymotrypsin in the same orientation as RMCP II (defined by the transformation superimposing 152 α-carbons to 0·76 Å) as a search motif, and repeated the rotation-function calculation (Fig. 6*b*). There are large peaks far from the 'expected' rotations of (0, 0, 0) and (0, 180, 0) but no significant features at the 'correct' locations. The calculation was repeated after removing those residues of chymotrypsin that are not structurally equivalent to residues of RMCP II, but this did not significantly affect the results. (Fig. 6*b* is essentially equivalent to our initial rotation-function calculation but transformed into a convenient angular system so that the relation between the calculation and the 'expected' result is obvious.) At face value, these tests suggest that the initial rotation-function search failed because α-chymotrypsin is not a sufficiently good approximation to RMCP II, but a subsequent test showed that this is not necessarily so. In this test we placed an RMCP II dimer in a $P1$ unit cell with dimensions 78×78×96 Å, *i.e.* the same dimensions as the native

RMCP II crystals, but with one dimer rather than three per unit cell, and calculated the rotation function using the full chymotrypsin model as the search motif (again, about one half of the asymmetric unit). Section $\theta_3 = 0$ is presented in Fig. 7. There is a large peak at the expected position plus one generated by the local twofold axis, showing that in this spacious, low-symmetry, packing arrangement, α-chymotrypsin can serve as a good probe for the RMCP II structure. In a final test we retained the full symmetry by placing a dimer of RMCP II in a $P3_1$ unit cell, but increased the cell dimensions to $110 \times 110 \times 130$ Å. This corresponds to a 2·7-fold increase in the unit-cell volume. In this case a rotation-function calculation using a full chymotrypsin molecule as probe did not have significant peaks at the 'expected' positions and had 'spurious' peaks elsewhere (data not shown). Thus, the 'success' of the rotation-function calculation shown in Fig. 7 is to be attributed to the reduction in symmetry rather than the increase in the volume of the unit cell. In other words it appears to be the superposition of the intramolecular vectors from a large number of the symmetry-related molecules rather than the intermolecular vectors between these molecules that prevents the successful application of the rotation function in this instance.

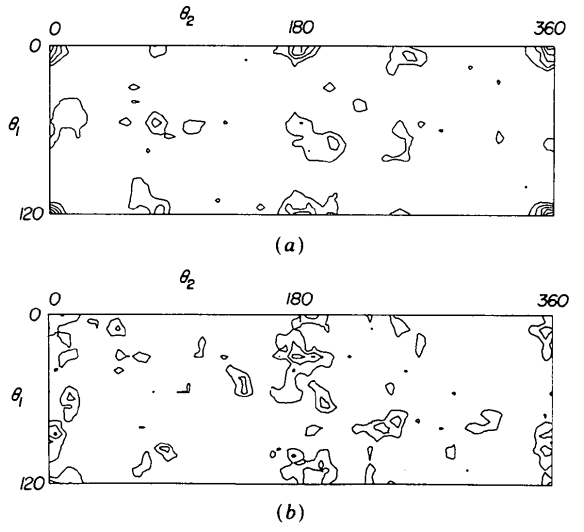

Fig. 6. Tests of rotation-function calculations made with the observed structure amplitudes for RMCP II. Data from 6 to 3·5 Å resolution were included and contours are drawn at increments of one standard deviation above the mean values. The section shown is $\theta_3 = 0$. (*a*) Search motif a single molecule of RMCP II (half of the asymmetric unit). 'Expected' peaks are at (0, 0, 0) due to the superposition of the RMCP II monomer on its own image, and at (0, 180, 0) due to superposition on the monomer related by non-crystallographic twofold symmetry. (*b*) The same calculation as in (*a*) but with search motif a molecule of α-chymotrypsin pre-aligned to correspond to RMCP II. The same peaks are 'expected' as for (*a*).

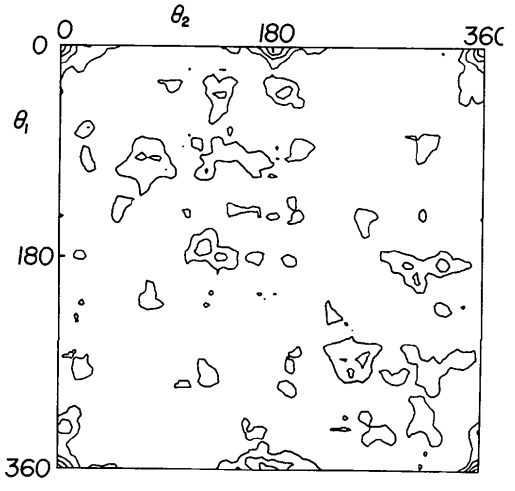

Fig. 7. Rotation-function calculation made with structure factors calculated for a hypothetical structure having the same unit cell as RMCP II but only one dimer (instead of three) per unit cell. The search molecule is α-chymotrypsin. Data from 6 to 3·5 Å included. Contours plotted at increments of one standard deviation above the average value. Section $\theta_3 = 0$. 'Expected' peaks are at (0, 0, 0) and (0, 180, 0).

The conclusions we draw from these experiments are as follows. (1) An accurate structural model seems to be required for a successful rotation-function search in a high-symmetry space group. (2) A less accurate structural model may be adequate in a lower-symmetry space group.

Apparently our failure in locating chymotrypsin in the RMCP II cell was due to the combination of the inadequacy of the search model together with the relatively high symmetry (pseudo $P3_121$) of the RMCP II crystals.

We thank Drs R. G. Woodbury and H. Neurath for supplying the protein and for continued helpful discussions.

The work was supported in part by grants from the National Institutes of Health (GM20066), the National Science Foundation (PCM-8014311) and the M. J. Murdock Charitable Trust.

References

ANDERSON, W. F., MATTHEWS, B. W. & WOODBURY, R. G. (1978). *Biochemistry*, **17**, 811-819.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SCHIMANOUCHI, T. & TASUNMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

BIRKTOFT, J. J. & BLOW, D. M. (1972). *J. Mol. Biol.* **68**, 187-240.

BLOW, D. M., ROSSMANN, M. G. & JEFFERY, B. A. (1964). *J. Mol. Biol.* **8**, 65-78.

BRITTON, D. (1972). *Acta Cryst.* A**28**, 296-297.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.

FISHER, R. G. & SWEET, R. M. (1980). *Acta Cryst.* A**36**, 755-760.

GREER, J. (1981). *J. Mol. Biol.* **153**, 1027-1042.

HARTLEY, B. S. (1970). *Philos. Trans. R. Soc. London Ser. B*, **257**, 77-87.

HENDRICKSON, W. A. & LATTMAN, E. E. (1978). *Acta Cryst.* B**26**, 136-143.

HUBER, R. (1969). *Crystallographic Computing Procedures*, edited by F. R. AHMED, pp. 96-102. Copenhagen: Munksgaard.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931-935.
JAMES, M. N. G., DELBAERE, L. T. J. & BRAYER, G. D. (1978). *Can. J. Biochem.* **56**, 396-402.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268-272.
KATANUMA, N., KOMINAMI, E., KOBAYASHI, K., BANNO, Y., SUZUKI, K., CHICHIBY, K., HAMAGUCHI, Y. & KATSUNUMA, T. (1975). *Eur. J. Biochem.* **52**, 37-50.
MATTHEWS, B. W. (1977). *The Proteins*, Vol. 3, 3rd ed., edited by H. NEURATH & R. L. HILL, pp. 403-590. New York: Academic Press.
MATTHEWS, B. W., KLOPFENSTEIN, C. K. & COLMAN, P. M. (1972). *J. Phys. E*, **5**, 353-359.
MATTHEWS, B. W., SIGLER, P. B., HENDERSON, R. & BLOW, D. M. (1967) *Nature (London)*, **214**, 652-656.
MOLNAR, C. E., BARRY, C. D. & ROSENBERGER, F. U. (1976). Tech. Mem. No. 229, Computer Systems Laboratory, Washington Univ., St Louis.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225-238.
ROSSMANN, M. G. & ARGOS, P. (1976). *J. Mol. Biol.* **105**, 76-96.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
SCHMID, M. F., WEAVER, L. H., HOLMES, M. A., GRÜTTER, M. G., OHLENDORF, D. H., REYNOLDS, R. A., REMINGTON, S. J. & MATTHEWS, B. W. (1981). *Acta Cryst.* A**37**, 701-710.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. (1977). *Acta Cryst.* A**33**, 800-804.
WOODBURY, R. G., GRUZENSKI, G. M. & LAGUNOFF, O. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2786-2789.
WOODBURY, R. G., KATANUMA, N., KOBAYASHI, K., TITANI, K. & NEURATH, H. (1978). *Biochemistry*, **17**, 811-819.
YOSHIDA, N., EVERITT, M. T., NEURATH, H., WOODBURY, R. G. & POWERS, J. (1980). *Biochemistry*, **19**, 5799-5804.

Acta Cryst. (1985). B**41**, 147-157

The Refinement of Southern Bean Mosaic Virus in Reciprocal Space

BY ABELARDO M. SILVA* AND MICHAEL G. ROSSMANN

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

(*Received* 26 *June* 1984; *accepted* 5 *October* 1984)

Abstract

The restrained least-squares refinement procedure of Konnert & Hendrickson [*Acta Cryst.* (1980), A**36**, 344-350] has been applied to 2·8 Å resolution, native, southern bean mosaic virus diffraction data. The initial model, based on a multiple-isomorphous-replacement map which had been improved by three cycles of real-space molecular-replacement averaging, gave an overall R factor of 40·7%. The R factor decreased to 25·5% after 65 refinement cycles interspersed with model checks on a computer graphics system using the *FRODO* program [Jones (1978). *J. Appl. Cryst.* **11**, 268-272]. The refinement was made possible by the vectorization of the program code, the use of non-crystallographic symmetry as a constraint and the selection of partial data sets. The latter was possible as a consequence of the non-crystallographic symmetry. A total of 34 water molecules per icosahedral asymmetric unit were identified. A region of relatively low, but continuous, electron density can be seen near a series of basic residues on the internal viral surface. This might be a small portion of RNA with partial icosahedral symmetry.

I. Introduction

(*A*) *The virus*

Southern bean mosaic virus (SBMV) is a small icosahedral plant virus of diameter varying from 328 Å at the fivefold axes to about 285 Å at the 'quasi'-threefold axes. The triangulation number of the capsid is $T = 3$, according to the nomenclature of Caspar & Klug (1962). There is only one type of coat protein (molecular weight 28 200), consisting of 260 amino acids. The amino acid sequence was determined by Hermodson, Abad-Zapatero, Abdel-Meguid, Pundak, Rossmann & Tremaine (1982). The encapsidated RNA has an approximate molecular mass of $1{\cdot}4 \times 10^6$ daltons which accounts for about 21% of the total molecular mass of the particle. Measurements of the Ca^{2+} and Mg^{2+} content vary from 80 to 280 and from 120 to 380 ions per virion, respectively (Abdel-Meguid, Yamane, Fukuyama & Rossmann, 1981; Hsu, Sehgal & Pickett, 1976; Hull, 1978).

The molecular structure of the cowpea strain of SBMV has been determined by X-ray diffraction techniques at a resolution of 2·8 Å (Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980, 1981; Rossmann, Abad-Zapatero, Hermodson & Erickson, 1983). The space group of the crystals is $R32$ with hexagonal

* Present address: Departamento de Física, Facultad de Ciencias Exactas, U.N. de La Plata, C. C. No. 67, 1900 La Plata, Argentina.

0108-7681/85/020147-11$01.50