We thank Professor Philip Coppens for very helpful discussions and a critical reading of this paper.

APPENDIX A

We want to prove that

$$\hat{L}[f * g] = [\hat{L}f] * g = f * [\hat{L}g].$$

Let

$$h(x, y) = f * g = \int_0^x du \int_0^y dv f(u, v) g(x-u, y-v)$$

$$\hat{L}(h) = \tilde{\sigma}^2 \int_0^x du \int_0^y dv\, h(u, v).$$

If h is expanded into its components f and g:

$$\hat{L}(h) = \tilde{\sigma}^2 \int_0^x du \int_0^u du' \int_0^y dv \int_0^v dv' f(u', v') \times g(u-u', v-v'),$$

we change variables (u, u') into $(u', \xi = u-u')$, (v, v') into $(v', \eta = v-v')$ and get

$$\hat{L}(h) = \tilde{\sigma}^2 \int_0^x du' \int_0^y dv' f(u', v') \times \int_0^{x-u'} d\xi \int_0^{y-v'} d\eta\, g(\xi, \eta)$$

$$= \int_0^x du' \int_0^y dv' f(u', v') \hat{L}g(x-u', y-v')$$

$$= f * [\hat{L}g]$$

$$= [\hat{L}f] * g.$$

References

AUTHIER, A. (1970). *Advances in Structure Research by Diffraction Methods*, pp. 1-52. Oxford: Pergamon Press.

BATTERMAN, B. W. & COLE, H. (1964). *Rev. Mod. Phys.* **36**, 681-717.

BECKER, P. (1977). *Acta Cryst.* A**33**, 667-671.

BECKER, P. J. & COPPENS, P. (1974). *Acta Cryst.* A**30**, 129-153.

BECKER, P. J. & COPPENS, P. (1975). *Acta Cryst.* A**31**, 417-425.

BECKER, P. & DUNSTETTER, F. (1984). *Acta Cryst.* A**40**, 241-251.

DARWIN, C. G. (1914). *Philos. Mag.* **27**, 315-333.

DARWIN, C. G. (1922). *Philos. Mag.* **43**, 800-824.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)

KATO, N. (1968). *J. Appl. Phys.* **39**, 2225-2237.

KATO, N. (1973). *Z. Naturforsch. Teil A*, **28**, 604-609.

KATO, N. (1976). *Acta Cryst.* A**32**, 453-466.

KATO, N. (1980*a*). *Acta Cryst.* A**36**, 763-769.

KATO, N. (1980*b*). *Acta Cryst.* A**36**, 770-778.

KATO, N. (1980*c*). In *Electron and Magnetisation Densities in Molecules and Crystals*, edited by P. BECKER, pp. 237-250. New York: Plenum.

KULDA, J. (1984). *Acta Cryst.* A**40**, 120-126.

OLEKHNOVICH, N. M., KARPEI, A. L., OLEKHNOVICH, A. I. & PUZENKOVA, L. D. (1983). *Acta Cryst.* A**39**, 116-122.

TAKAGI, S. (1969). *J. Phys. Soc. Jpn*, **26**, 1239-1253.

ZACHARIASEN, W. H. (1945). *Theory of X-ray Diffraction in Crystals.* New York: John Wiley.

ZACHARIASEN, W. H. (1967). *Acta Cryst.* **23**, 558-564.

Acta Cryst. (1988). A**44**, 270-282

The Use of Molecular-Replacement Phases for the Refinement of the Human Rhinovirus 14 Structure

BY EDWARD ARNOLD* AND MICHAEL G. ROSSMANN

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

(*Received* 8 *July* 1987; *accepted* 25 *November* 1987)

Abstract

The structure of human rhinovirus 14 has been refined, by the method of restrained least squares, to an R factor of 0·16 for various random samples between 6 and 3 Å resolution with $F > 3\sigma(F)$. As a first step the non-crystallographic symmetry parameters were optimized using the initial atomic model in a rigid-body refinement procedure. Phase determination by the molecular-replacement phase extension and refinement procedure was continued to 2·94 Å resolution, employing the improved non-crystallographic symmetry operators. The resultant structure-factor phases and weights, together with the measured amplitudes, constituted the X-ray observations used in the restrained refinement. The Hendrickson-Konnert program system [Konnert & Hendrickson (1980). *Acta Cryst.* A**36**, 344-350] was modified to incorporate non-crystallographic symmetry constraints and structure-factor phases as observations.

* Present address: Center for Advanced Biotechnology and Medicine (CABM) and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0759, USA.

0108-7673/88/030270-13$03.00

The non-bonded contacts between subunits related by non-crystallographic symmetry were also restrained.

Introduction

At moderately low resolution, the typically underdetermined set of X-ray diffraction observations can be supplemented with chemical knowledge for restrained refinement of macromolecules (Konnert & Hendrickson, 1980). In a structure with extensive non-crystallographic symmetry, such as an icosahedral virus, it is also reasonable to constrain the non-crystallographic symmetry to reduce the number of variable parameters (Silva & Rossmann, 1985; Rossmann, 1976).

In the molecular-replacement phase extension and refinement procedure, the non-crystallographic redundancy is exploited to obtain phases for the structure-factor amplitudes. The ultimate accuracy of these molecular-replacement phases depends on the extent of redundancy, percentage of solvent content, fraction of data collected, accuracy of the observed structure-factor amplitudes, and on the accuracy of the description of the non-crystallographic symmetry: orientation, position and molecular boundary (Arnold & Rossmann, 1986). The quality of the phase determination of human rhinovirus 14 (HRV14) was excellent as evidenced both by the ease with which the electron density map could be interpreted (Rossmann *et al.*, 1985) and by the remarkable agreement between the observed structure amplitudes with those obtained by back-transformation of the averaged map (Table 1 is derived from the optimizations described in this paper). The latter was close to the limit of the errors of the amplitudes [(Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987) and see below].

Many details implicit in the electron density are discarded in traditional crystallographic refinement of the atomic model with respect to structure-factor amplitudes. For instance, the atomic model frequently omits information about the nature of disordered regions, solvent structure and thermal broadening. Since the electron density represents these features more closely than the model, it is reasonable to use the phases from the molecular-replacement procedure as observations in the crystallographic refinement. So long as the correct model parameters are provided, the features of the eventual model should closely resemble the features of the electron density.

The structure determination of HRV14 has previously been detailed (Rossmann *et al.*, 1985; Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). HRV14 is a picornavirus with an external diameter of approximately 300 Å (Rueckert, 1986). The protein coat consists of 60 copies of each of four distinct polypeptide chains arranged in an icosahedrally symmetric fashion about a core of a single strand of RNA. Knowledge of the three-dimensional structure illuminated features of rhinoviral immunogenicity, receptor binding, assembly pathway, and evolutionary relations to other small icosahedral viruses of known structure (Rossmann *et al.*, 1986; Arnold, Luo, Vriend, Rossmann, Palmenberg, Parks, Nicklin & Wimmer, 1987). More recent experiments have successfully elucidated the atomic details of the binding of antiviral agents to a hydrophobic pocket in the VP1 β-barrel and yield suggestions about how they prevent the pH-mediated uncoating of rhinoviruses and other picornaviruses (Smith *et al.*, 1986). The initial HRV14 atomic model was used successfully to initiate phasing by the molecular-replacement procedure for Mengo virus, a member of the cardiovirus class of picornaviruses (Luo *et al.*, 1987).

Vector refinement

Normally in least-squares refinement the quantity

$$\sum_h w_h(|F_{h,\mathrm{obs}}|-|F_{h,\mathrm{calc}}|)^2$$

is minimized. This is equivalent to assuming that the phase computed from the atomic model is equal to the correct phase and can therefore be called 'scalar refinement'. In the case where previous phase information is available, we have chosen to minimize

$$\sum_h w_h[(A_{h,\mathrm{obs}}-A_{h,\mathrm{calc}})^2+(B_{h,\mathrm{obs}}-B_{h,\mathrm{calc}})^2],$$

in which the real and imaginary components of the structure factors have been separated. Therefore, without changing the number of normal equations, the number of X-ray observational equations has been doubled. This is equivalent to minimizing the magnitude of the vector difference between $F_{h,\mathrm{obs}}$ and $F_{h,\mathrm{calc}}$ and can therefore be called 'vector' refinement (Table 2). In the HRV14 case the 'observed' phases have been obtained from the measured amplitudes by an imposition of non-crystallographic symmetry and density modification constraints. However, they are treated as observations in this refinement procedure since they are not allowed to vary. Since the phases are not altered, this procedure optimizes the fit of the atomic model to the electron density map and, hence, corresponds to real-space model fitting, similar to a graphical procedure described by Jones & Liljas (1984). Whereas the scalar refinement deems a solution acceptable as long as calculated and observed amplitudes match, vector refinement requires a match of the observed and calculated amplitudes as well as phases (Fig. 1). This formalism differs from that of Rees & Lewis (1983), who explicitly separate the amplitude and phase information. In both cases, amplitudes and phases are refined simultaneously, but the ambiguity in the direction of phase shift when

Table 1. *Final molecular-replacement results at 2·94 Å resolution*

(i) Resolution ranges

Mean range (Å)	*R* factor (%)	Correlation coefficient	Number of reflections	Relative local scale factor	Phase change (°) from previous results
14·99	11·44	0·94	7 202	0·090	8·6
9·92	7·46	0·97	10 933	0·088	8·4
7·94	6·85	0·98	13 367	0·087	8·5
6·81	7·13	0·98	15 162	0·088	8·7
6·05	7·47	0·98	16 928	0·088	9·4
5·51	7·60	0·98	18 991	0·089	9·5
5·08	7·22	0·98	20 552	0·090	10·0
4·75	7·12	0·98	22 138	0·090	10·1
4·47	7·64	0·97	23 418	0·091	10·8
4·23	8·61	0·97	24 166	0·092	11·4
4·03	10·15	0·96	25 100	0·094	11·6
3·86	11·55	0·94	25 667	0·096	12·4
3·57	14·73	0·91	26 603	0·100	13·3
3·44	16·98	0·88	26 579	0·104	14·2
3·33	20·06	0·82	26 549	0·109	14·6
3·23	22·19	0·79	26 750	0·113	15·8
3·14	25·58	0·71	26 024	0·120	19·6
3·05	30·49	0·59	21 736	0·136	29·2
2·98	43·46	0·29	19 008	0·264	—
Overall:					
25·0–2·94 Å	13·48	0·93	423 123		

(ii) *F* ranges

	R factor (%)	Number of reflections	Phase change (°) from previous results
0–0·4 $\hat{F}$	34·73	9 437	32·5
0·4–0·8 $\hat{F}$	24·92	168 187	25·0
0·8–1·2 $\hat{F}$	13·94	132 168	12·7
1·2–1·6 $\hat{F}$	7·43	65 545	6·8
1·6– $\hat{F}$	5·23	47 886	4·4

In the final cycle, there was a total of 423 123 observed amplitudes included in the computations out of a total number of 613 723 Fourier coefficients. The latter included calculated amplitudes where no observed measurements were available.

The *R* factor was defined as

$$R = \frac{\sum |(|F_{obs}| - |F_{calc}|)|}{\sum |F_{obs}|} \times 100$$

where F_{calc} is the result of back transformation of the averaged map.

The correlation coefficient was defined as

$$C = \frac{\sum (\langle F_{obs}\rangle - |F_{obs}|)(\langle F_{calc}\rangle - |F_{calc}|)}{\left[\sum (\langle F_{obs}\rangle - |F_{obs}|)^2 \sum (\langle F_{calc}\rangle - |F_{calc}|)^2\right]^{1/2}}$$

where $\langle F_{obs}\rangle$ and $\langle F_{calc}\rangle$ are mean amplitudes in each resolution range.

See text to obtain a full definition of phase change with respect to previous results.

the difference is close to 180° is resolved in the present formulation. Scalar refinement is preferable when only poor phase information exists. Pure vector refinement is appropriate when exceedingly good phase information is available as it was for HRV14. 'Good' can be defined as occurring when the observed phase information is much better than that derived from the available model. In the present case it can be shown that the final molecular-replacement *R* factor of around 0·09 (Table 1) corresponds to a mean probable phase error of around 3·3°, whereas the initial HRV14 model had a mean phase difference from the observed phases of 18° (see section on convergence) in the 6 to 4 Å resolution range. A combination (Rossmann & Blow, 1961; Hendrickson & Lattman, 1970) of the observed and calculated model phases should be used only when the model phases are equal to or better than the available observed phase information, as would normally be the case for isomorphous-replacement phases (Rees, Lewis & Lipscomb, 1983; Watenpaugh, Sieker, Herriott & Jensen, 1973). In the vector refinement of HRV14, we selected the weights w_h on individual reflection observations derived from the molecular-replacement phase determination (see section on optimal phases). These weights reflect the reliability of the phase observation, and depend on a correlation of the observed amplitudes and those calculated in the molecular-replacement back-transformation process. Other choices of weighting schemes could account

Table 2. *Algebra of scalar and vector refinement*

	Scalar	Vector
Quantities to minimize	$\sum_h w_h(\lvert F_{h,\text{obs}}\rvert - \lvert F_{h,\text{calc}}\rvert)^2$	$\sum_h w_h[(A_{h,\text{obs}} - A_{h,\text{calc}})^2 + (B_{h,\text{obs}} - B_{h,\text{calc}})^2]$
Elements of normal matrix: N_{jk} (for parameters p_j, p_k)	$\sum_h w_h \frac{\partial\lvert F_{h,\text{calc}}\rvert}{\partial p_j}\frac{\partial\lvert F_{h,\text{calc}}\rvert}{\partial p_k}$	$\sum_h w_h\left[\frac{\partial A_{h,\text{calc}}}{\partial p_j}\frac{\partial A_{h,\text{calc}}}{\partial p_k} + \frac{\partial B_{h,\text{calc}}}{\partial p_i}\frac{\partial B_{h,\text{calc}}}{\partial p_j}\right]$
Gradient vector elements: E_k	$\sum_h w_h \frac{\partial\lvert F_{h,\text{calc}}\rvert}{\partial p_k}\Delta F_h$	$\sum_h w_h\left[\frac{\partial A_{h,\text{calc}}}{\partial p_k}\Delta A_h + \frac{\partial B_{h,\text{calc}}}{\partial p_k}\Delta B_h\right]$

Definitions

$$F_h = A_h + iB_h \qquad \alpha_{\text{calc}} = \tan^{-1}(B_{h,\text{calc}}/A_{h,\text{calc}}) \qquad \Delta F_h = \lvert F_{h,\text{obs}}\rvert - \lvert F_{h,\text{calc}}\rvert$$

$$A_{h,\text{obs}} = \lvert F_{h,\text{obs}}\rvert\cos\alpha_{\text{obs}} \qquad B_{h,\text{obs}} = \lvert F_{h,\text{obs}}\rvert\sin\alpha_{\text{obs}}$$

$$A_{h,\text{calc}} = \lvert F_{h,\text{calc}}\rvert\cos\alpha_{\text{calc}} \qquad B_{h,\text{calc}} = \lvert F_{h,\text{calc}}\rvert\sin\alpha_{\text{calc}}$$

$$\Delta A_h = A_{h,\text{obs}} - A_{h,\text{calc}} \qquad \Delta B_h = B_{h,\text{obs}} - B_{h,\text{calc}}$$

for individual estimates of phase reliability if that information were available.

Refinement of particle orientation and position

The initial model, derived from the molecular-replacement-phased electron density map, was used for a rigid-body least-squares refinement of the non-crystallographic symmetry operators. In the cubic HRV14 case, there are only two global non-crystallographic symmetry parameters which needed to be refined: the rotation of the virus particle about the crystallographic threefold axis (κ), and the position of the particle center on the threefold axis (x, x, x). In a more general case there would be six parameters that would need to be determined, as in the Mengo virus structure (Luo *et al.*, 1987).

A procedure which has been used extensively for the refinement of heavy-atom parameters (Rossmann, 1976) (see Appendix) was adapted to the refinement of the particle non-crystallographic symmetry operators. This has the advantage of factoring out and thereby separating the translational elements from the rotational elements. If there are m crystallographic symmetry elements and n non-crystallographic symmetry elements, then there will be mn operators of the kind $[R_{mn}]$ that generate the whole unit cell (square brackets signify a matrix). It can be shown that

$$[R_{mn}] = [T_m][\alpha][\rho_n][\Delta][\beta],$$

where $[T_m]$ is the mth crystallographic rotation operator, with a translation component d_m, $[\rho_n]$ is the nth known non-crystallographic operator with respect to orthogonal molecular axes, $[\alpha]$ and $[\beta]$ are the de-orthogonalizing and orthogonalizing matrices, and $[\Delta]$ is a matrix describing the tilt of the particle with respect to the unit-cell axes.

Thus, for instance, if ξ is a variable related to the tilt matrix given by $[\Delta]$,

$$\frac{\partial}{\partial\xi}[R_{mn}] = [T_m][\alpha][\rho_n]\left(\frac{\partial}{\partial\xi}[\Delta]\right)[\beta]. \qquad (1)$$

The operations $[R_{mn}]$ relate a point with fractional coordinates $\mathbf{x}$ to the point $\mathbf{x}_{mn}$ according to

$$\mathbf{x}_{mn} = [R_{mn}]\mathbf{x} + [T_m]\mathbf{S} + \mathbf{d}_m,$$

where $\mathbf{S}$ defines the origin of the reference particle and $\mathbf{d}_m$ corresponds to the translational component of the crystallographic rotation operator $[T_m]$.

In the special case of HRV14, ξ would be the rotation κ (Table 3) about that icosahedral threefold axis which is superimposed on the crystallographic body diagonal threefold axis and $\mathbf{S}$ corresponds to the particle position (x, x, x). The progress of the refinement of the global non-crystallographic symmetry operators is charted in Table 3. A subset of 992 reflections between 3·7 and 4·0 Å resolution was used for the initial computations, and within three iterations (cycles 1–3) the overall orientation had

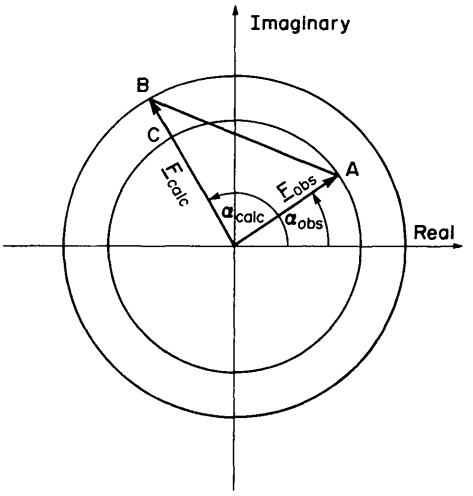


Fig. 1. Scalar refinement minimizes the sum of distances $(CB)^2$ over all reflections. Vector refinement minimizes the sum of the distances $(AB)^2$ over all reflections. Scalar refinement can give poor results when α_{calc} is far from α_{obs}, assuming that α_{obs} is a reliable estimate of the 'true' phase.

Table 3. *Refinement of HRV14 global non-crystallographic symmetry operators by rigid-body least-squares refinement based on the initial atomic model*

Cycle number	κ(°)	x	Resolution range (Å)	Number of reflections	R factor (%) Initial	Final
1-3	−6·0	0·00060	4·0-3·7	992	31·44	29·95
4	−6·10	0·00042	4·0-3·7	992	29·95	31·64
5-6	−6·097	0·000402	4·0-3·3	5349	31·64	31·53
Converged value	−6·086	0·000389	4·0-3·3	5349	31·53	

$R = \frac{\sum|(|F_{obs}| - k|F_{calc}|)|}{\sum|F_{obs}|} \times 100$, where F_{calc} are obtained by structure-factor calculations derived from the initial atomic model.

changed by −0·1° and the particle position had made a net movement of 0·15 Å. For the final refinement (cycles 4-6) of κ and x, a subset of roughly 5000 reflections between 3·3 and 4·0 Å resolution was randomly selected, and in three cycles yielded convergence at $\kappa = -6{\cdot}086°$ and $x = 0{\cdot}000389$. A different subset with 2000 reflections was used to check the convergence and led to negligible changes in the parameters. The procedure requires few reflections since there are only two parameters.

The optimal phases from molecular replacement

After the particle orientation and position had been refined, the molecular-replacement phase refinement by symmetry averaging and reconstruction was continued (Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). Starting with phases obtained from the previous molecular-replacement refinement and extension to 3·0 Å resolution ($\kappa = -6{\cdot}0°$, $x = 0{\cdot}0006$), we carried out four cycles of molecular-replacement phase refinement at 3·0 Å resolution followed by four cycles at 2·94 Å resolution. In contrast to the previous molecular-replacement refinements, calculated amplitudes for those reflections which had not been observed were included in the Fourier syntheses (Rayment, 1983; Arnold & Rossmann, 1986) and the weights used for these terms were the average correlation coefficients for all observations in that resolution range. For the observed amplitudes, weights w_h corresponding to the geometric mean of Sim (1959, 1960) and exponential weighting (Rayment, 1983) were employed, such that

$$w_h = (w_{Sim} w_{exp})^{1/2},$$

where $w_{exp} = \exp(-|\Delta F|/F_{obs})$. Various tests indicated that the mixed weighting scheme gave better results than either independent scheme alone. The Sim scheme gave suitably low weights to poorly correlated observations while tending to give too high a weight for most Fourier coefficients. The exponential weighting scheme discriminates well among the better observations, but where the error between F_{obs} and F_{calc} (ΔF) is large this scheme can give unacceptable results. For instance, if $\Delta F = F_{obs}$ then the weight is 0·37, which is unreasonably large for an essentially undetermined phase. These features of both schemes were detected from an R-factor distribution as a function of the weights. A combination of the two schemes as given above alleviates these problems and yields the best overall molecular-replacement statistics. Developed for the work with HRV14, this scheme has also been successfully applied to the molecular-replacement phase determination of black beetle virus (Hosur, Schmidt, Tucker, Johnson, Gallagher, Selling & Rueckert, 1987), Mengo virus (Luo *et al.*, 1987), polyoma virus (Delatore, Rossmann, Baker, Rayment, Murakami & Caspar, 1987), cowpea mosaic virus (Stauffacher *et al.*, 1988), alfalfa mosaic virus (Fita & Rossmann, 1987), and lactate dehydrogenase from *B. stearothermophilus* (Piontek, Chakrabarti, Rossmann & Zuber, 1988).

The results for the final cycle of molecular replacement at 2·94 Å were compared with the previous results, prior to the refinement of particle orientation and position (Table 1). The earlier molecular-replacement refinement had used a pure exponential weighting scheme and had not used calculated amplitudes where observed amplitudes were missing. The R factor dropped by about 7% in most resolution shells (Arnold, Vriend, Luo, Griffith, Kamer, Erickson, Johnson & Rossmann, 1987). The low R factor, in spite of the constraints imposed by the high order of non-crystallographic symmetry (Rees, 1983), shows that the non-crystallographic symmetry is precise, although some perturbations can be expected at the contacts between particles.

The changes in phases compared with the previous results show a steadily increasing trend from 8° at low resolution to nearly 30° at 3 Å resolution. A final skew-averaged Fourier map was computed for model building which incorporated all observed data between 25 and 3 Å resolution. Upon qualitative comparison of this new map with the earlier 3·08 Å resolution map (which had been the basis of the initial atomic model), there was surprisingly little difference in spite of the dramatically improved agreement statistics. However, it was possible to detect a small improvement in details, such as increased flattening of aromatic rings.

Restraint of non-bonded contacts between subunits related by non-crystallographic symmetry

A considerable proportion of the short (<4·0 Å) non-bonded contacts within a macromolecular structure with closed point symmetry such as a tetrameric enzyme or virus are located at the interfaces between subunits. Depending on the arrangement of subunits relative to crystal symmetry elements, some of these contacts may involve subunits related by non-crystallographic symmetry. To maximize use of prior chemical knowledge in a refinement of a macromolecule with such contacts, it is desirable to restrain the minimum approach distance in a manner analogous to restraint of other van der Waals contacts.

A provision was made in the program *PROTIN*, which develops the stereochemical dictionary, to read the viral symmetry operators and generate the symmetry-related atoms within a single virion. These contacts can then be used to supplement the geometric restraints employed in the refinement program. A similar procedure has recently been reported by Sheriff (1987).

Flexible combination of X-ray and geometric restraints

The restrained refinement program *PROLSQ* (Konnert & Hendrickson, 1980) was modified to allow the separate accumulation of the contributions to the normal equations representing the geometric restraints and selected X-ray data sets. These could then be readily combined with different weights and solved by means of the rapid conjugate-gradients algorithm. Typically, contributions from about 30 000 reflections were used in deriving the normal equations. The relative weight of the geometric portion was adjusted to give atomic shifts that retained physically reasonable geometry (Hendrickson, 1981) except in certain intermediate cycles (see below).

A limited (20 000 to 30 000 reflections) subset of data yielded similar convergence as a larger or different data subset (Silva & Rossmann, 1985). The non-crystallographic redundancy in the refinement of southern bean mosaic virus (Silva & Rossmann, 1985) was tenfold, while here the redundancy is 20-fold. Silva & Rossmann (1985) used subsets of between 14 and 50% of the possible data. Here we use subsets of between 4 and 25%. Non-overlapping subsets gave closely similar parameter shifts and agreement statistics for HRV14 (the R factors varied by less than 0·005). Hence, different non-overlapping subsets were used in consecutive cycles to save computing time and to extend the dependence of the refinement to the majority of the observed data. Since the structure is constrained by the non-crystallographic symmetry, the ratio of total observations to refinable parameters is increased here by the non-crystallographic redundancy. In addition, the use of phase observations again doubles this ratio. The increase in total information content due to the phase information is difficult to assess because the phases were derived from a combination of the X-ray amplitudes and the non-crystallographic symmetry.

The path of vector refinement

Table 4 summarizes the course of refinement. The resolution of the data used in the refinement was intermittently varied from 4 Å (to permit larger changes in parameters) to 3 Å (to improve accuracy). As the refinement progressed more emphasis was given to the higher-significance data in an attempt to increase the precision of the refinement. The initial refinement used 12 to 3·9 Å (cycles 1–3) and then 6 to 4 Å (cycles 4–5) resolution X-ray data while permitting the positional parameters of the 6264 atoms in the HRV14 protomer to vary. Cycles 1 to 5, in which no selection criteria for significance were used, dropped the R factor from 26 to 18% for the 6–4 Å resolution data. In the fourth cycle the X-ray data were given high weight relative to the geometrical restraints, permitting significant movement of the atoms and temporary distortion of the model. In particular, it was seen that planar groups cannot change orientation unless the planarity restraints were relaxed, for otherwise the atoms are required to remain in the same plane rather than in any plane. The restraints permit only sliding or lateral movement of the atoms in a plane as an ensemble instead of swiveling of the entire group about the free torsion angle. In the fifth cycle the geometric restraints were applied with sufficient weight to bring the mean deviations close to the values suggested for initial refinement (Hendrickson, 1981). Although the fifth cycle resulted in raising the R factor approximately 1%, the resulting model was more chemically reasonable. The protein model was then re-examined in its entirety and a substantial number of adjustments were made, largely to side chains such as valine, threonine and leucine.

After the initial refinement of the protein positional coordinates (following cycle 9 - Table 4), 213 solvent waters and two ions were included in the refinement model. The occupancies of the waters were allowed to vary (cycles 10–16). Attempts to refine the positional coordinates of the waters after cycle 9 led to unreasonably large shifts and therefore positional refinement of waters was abandoned in all subsequent refinement. A large number of water molecules were identified in the 3 Å resolution native electron density map (map *A* in Table 4). Although 213 of these were built into the map and included in the model, these positions were abandoned (after cycle 26) when a skew-averaged difference Fourier map (map *B* in Table 4) was computed with

Table 4. *Progress of HRV14 refinement*

Model description	Variables*	Resolution selection (Å)	Amplitude selection	Number of reflections per cycle	Density (%)†	Cycle number	*R* factor‡ Normal	*R* factor‡ Vector	R.m.s. ΔX (Å) after cycle since last rebuilding of model	Type of map§	Comments
6264 atoms in protomer	*xyz*	12-3·9	$>1\sigma(F)$	24 000	10	1-2	0·264	0·427	0·213		
	xyz	12-3·9	$>1\sigma(F)$	24 000	10	3	0·216	0·337	0·263		
	xyz	6-4	$>1\sigma(F)$	65 000	38	4	0·197	0·301	0·339		Geometric restraints relaxed to allow gross movement especially of planes.
	xyz	6-4	$>1\sigma(F)$	26 000	15	5	0·171 after 0·181	0·253 0·272	0·344	*A*	Geometric restraints tightened to yield quality model. Entire model reexamined; about 100 side chains rebuilt.
Rebuilt protomer, 6268 atoms	*xyz*	6-4	$>1\sigma(F)$	23 000	14	6	0·190	0·287	0·110		Intersubunit contacts introduced as additional restraints.
	xyz	6-4	$>1\sigma(F)$	27 500	16	7	0·181	0·270	0·130		
	xyz	6-3	$>1\sigma(F)$	65 000	13	8	0·250	0·385	0·199		
	xyz	6-3	$>1\sigma(F)$	30 000	6	9	0·220	0·345	0·226	*A*	Intermediate coordinate set deposited with Brookhaven Protein Data Bank; about 20 side chains rebuilt; initial waters built into molecular-replacement map.
6268 protein atoms, 213 waters, 2 ions	*Q*'s for waters	6-4	$>10\sigma(F)$	43 000	25	10-16	0·177 after 0·163	0·265 0·250		*A*	Solvent atoms introduced as interpreted from the molecular-replacement map.
6268 protein atoms, 193 waters, 2 ions	*xyz*, *B* for protein atoms, *B* for solvent atoms	6-4	$>10\sigma(F)$	28 000	17	17-19	0·167 after 0·128	0·256 0·189	0·194		20 waters with $Q<0·2$ were discarded.
	xyz	6-3	$>10\sigma(F)$	30 000	6	20-26	0·144 after 0·129	0·217 0·189	0·257	*B*	Difference Fourier map computed based on 170 000 terms with 6-3 Å resolution with F_{calc} computed by omitting solvent model. 275 new solvent sites found by centroid calculation; about 10 side chains rebuilt.
6268 protein atoms, 275 waters, 2 ions	*xyz*, *B* for protein atoms, *Q* for waters	6-3	$>10\sigma(F)$	22 000	4·4	27-31	0·123 after 0·126	0·182 0·186	0·083		New solvent model utilized; geometric constraints tightened.
Same except 2 waters discarded	*B* for all atoms	6-3	$>9\sigma(F)$	23 500	4·7	32-36	0·130 after 0·128	0·191 0·188		*C*	Final vector difference Fourier map computed; final coordinates deposited with Brookhaven Data Bank

* *xyz*, *B* refer to atomic positional and temperature parameters. *Q* refers to occupancy of water molecules.

† Density refers to the percentage of data included in each cycle. Different random subsets were selected by choosing every *n*th reflection in a list of reflections ordered by Miller indices.

‡ The *R* factor is given at start of cycle unless otherwise indicated, where

$$R_{normal}=\frac{\sum|(|F_{obs}|-k|F_{calc}|)|}{\sum|F_{obs}|},\quad R_{vector}=\frac{\sum|(F_{obs}-kF_{calc})|}{\sum|F_{obs}|}.$$

k is the overall scale factor applied to the F_{calc} values in order to scale them to the F_{obs} values.

The mean difference in degrees between observed and calculated phases is given by $(R_{vector}\times 100)/(\pi\sqrt{2})$.

§ Types of maps which were examined:

A $w_{MR}F_{obs}\exp(i\alpha_{MR})$ coefficients. 12-3 Å resolutions; 400 000 terms.

B $w_{MR}(|F_{obs}|-k|F_{calc}|)\exp(i\alpha_{MR})$ where F_{calc} is based on the atomic model omitting the solvent atoms. 6-3 Å resolutions; 170 000 terms.

C $w_{MR}[|F_{obs}|\exp(i\alpha_{MR})-k|F_{calc}|\exp(i\alpha_{calc})]$ where F_{calc} and α_{calc} are based on all the atoms in the model and solvent. 6-3 Å resolutions; 150 000 terms.

$w(|F_{obs}| - k|F_{calc}|) \exp(i\alpha_{MR})$ coefficients (Fig. 2), where the F_{calc} are dependent on the atomic model excluding the earlier solvent atoms. This map showed 275 solvent atoms in more clearly resolved positions with 192 of these within 0·5 Å of earlier water sites. An automatic search procedure was used to locate and position peaks (by a centroid computation) in the difference map, but only those peaks which were in chemically reasonable positions with respect to the protein model or other water sites were used in the subsequent refinement (cycles 27–36). In addition to the water molecules, there was electron density on the five- and threefold axes which was interpreted as possible Ca^{2+} and Mg^{2+} ions, respectively, corresponding to the identification of similar ions in other RNA viruses (*cf.* Abdel-Meguid, Yamane, Fukuyama & Rossmann, 1981). The refined occupancies and stereochemical environments are consistent with these assignments. While occupancies and thermal parameters will be exceedingly difficult to differentiate at 3 Å resolution, nevertheless occupancies were refined, in contrast to normal practice, in light of their physical significance. The occupancies and isotropic thermal parameters of the solvent atoms were varied in alternate sets of five cycles (cycles 27–36). Two water molecules whose occupancies decreased to less than 0·2 were discarded after cycle 31. A final vector difference Fourier electron density map (map *C* in Table 4) showed some additional small peaks in chemically reasonable positions for apparently weakly substituted water molecules, but these are not included in the final atomic coordinates deposited with the Brookhaven Protein Data Bank.*

Independent temperature factors were introduced and refined simultaneously with positional parameters for protein atoms in cycles 17 to 31. Finally, the temperature factors for protein and solvent atoms were refined, without change in positional parameters, in cycles 32–36, this being the only available mode for refining the water temperature factors without their positional parameters. The mean temperature factor for the protein atoms was found to be 14 $Å^2$ with values varying from 4 to 45 $Å^2$. The mean temperature factor for the water molecules was 19 $Å^2$, ranging from 3 to 38 $Å^2$. Their occupancies ranged from 1·0 to 0·24. The antigenic sites NImIA, IB, II and III on the viral surface showed unusually large thermal motion relative to other parts of the viral surface. A full analysis of the structure will be published elsewhere.

Table 5. *Comparison of initial and final model quality*

Target σ value in parentheses.

	Initial	Final
Number of reflections	23 500	23 500
Selection	$F > 9{\cdot}2\sigma$	$F > 9{\cdot}2\sigma$
Resolution (Å)	6–3	6–3
Modulus *R* factor	0·262	0·128
Vector *R* factor	0·400	0·188
Model (per icosahedral asymmetric unit)		
Number of waters and ions	0	275
Number of protein atoms	6264	6268
Geometrical restraints		
Distances		
Number	17 482	17 494
$>4\sigma$	349	116
σ_{12} (0·030) (Å)	0·032	0·036
σ_{13} (0·030) (Å)	0·070	0·052
σ_{14} (0·050) (Å)	0·103	0·063
Planes		
Number	1070	1074
$>2\sigma$	288	50
σ (0·020) (Å)	0·021	0·011
Chiral volumes		
Number	1000	1000
$>2\sigma$	44	97
σ (0·150) ($Å^3$)	0·155	0·186
Nonbonded contacts		
Intra-protomer van der Waals contacts:		
Number	4584	5001
σ_1 single torsion (0·500) (Å)	0·254	0235
σ_2 multiple torsion (0·500) (Å)	0·364	0·254
σ_3 possible H-bond (0·500) (Å)	0·349	0·215
Inter-protomer van der Waals contacts:		
Number	—	238
σ_2 multiple torsion *i.e.* non-H-bonding (0·500) (Å)	—	0·244
σ_3 possible H-bond (0·500) (Å)	—	0·168
Torsion angles		
Number	1979	1983
Planar (0, 180°)		
σ (3) (°)	4·0	4·8
Staggered (±60°, 180°)		
σ (30) (°)	28·5	25·1
Orthonormal (±90°)		
σ (30) (°)	33·0	28·0
Isotropic thermal parameters		
Number	—	15 206
Δ_{12} main chain bond (1·00) ($Å^2$)	—	0·81
Δ_{13} distance (1·50) ($Å^2$)	—	1·31
Δ_{12} side-chain bond (1·00) ($Å^2$)	—	0·92
Δ_{13} distance (1·5) ($Å^2$)	—	1·52

The adherence of the initial and final model to accepted geometry is shown in Table 5. The final geometric restraint discrepancies are close to their target deviations. However, the chiral volume and planar torsion angle deviations from ideality have become worse. This indicates that a larger weight should have been placed on these restraints during the final cycles of refinement. The *R* factor as a function of resolution is plotted in Fig. 3. The final scalar *R* factor is 0·16 for a 26 000 reflection subset with $F > 3\sigma(F)$ between 6 and 3 Å resolution, a result found to be representative of other subsets. Similar

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 4RHV, R4RHVSF), and are available in machine readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP37024 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

selection for data between 10 and 3 Å gave R factors of 0·197, 0·192 and 0·145 when $|F|$ was greater than 2σ, 3σ and 10σ, respectively.

Features of the final difference Fourier map

An additional peak with the approximate density corresponding to a water molecule was located adjacent to the amino terminus of VP3 (Fig. 4) in the final vector difference map (map *C* in Table 4). The native (map *A* in Table 4) and difference Fourier maps show this density to be in close contact (too close for a solvent molecule) with the terminal amino group of VP3, suggesting that it probably corresponds to a covalent modification. The shape and size of this unassigned density indicates that it probably is a single non-hydrogen atom, possibly an oxygen atom (for a nitrosyl or hydroxylamine modification), a carbon atom (for a methyl or methyleneimine modification) or a nitrogen atom (for a hydrazinyl modification). The stereochemical environment of the modification is consistent with an *N*-methyl group, which has been previously identified as a post-translational modification in a number of other proteins, primarily in ribosomal proteins in which the α-amino groups of methionine (Chen, Brosius, Wittmann-Liebold & Schafer, 1977; Brauer & Wittmann-Liebold, 1977) and alanine (Chen *et al.*, 1977) have been found to be monomethylated. Multiple methylations of amino termini have been identified in a ribosomal protein (Lederer, Alix & Hayes, 1977), and in a protozoan cytochrome (Pettigrew & Smith, 1977).

Other residual features of the native and difference Fourier electron density maps were not as straightforward to interpret. An additional two and three residues could plausibly be built into weak density corresponding to the amino-terminal extensions of VP1 (from 15 to 16) and VP2 (from 5 to 7), respectively. Because of the abrupt fall off of the electron density signal at the edge of the well ordered portion these extra residues were not included in the refinement. Similarly, the less-ordered residues between 25 and 29 of VP4 were also omitted from the refinement.

An unusual arrangement of density was located just below the protein shell around the fivefold axis of icosahedral symmetry, near the amino terminus of VP3 and surrounded by the extra residues of the VP4 amino terminus (25 to 29). This density, substantially weaker than the ordered protein, consists of a ball of density at the radius of residues 25 to 29 in VP4, below which are two plates of density perpendicular to and centered about the fivefold axis of icosahedral symmetry. Although the identity of this portion of the virion is not clear, it seems possible that this density could be associated with an N-terminal modification of VP4 in HRV14 which, by analogy with poliovirus 1 Mahoney (Chow, Newman, Filman, Hogle, Rowlands & Brown, 1987), is probably myristylated at the amino terminus. Indeed, an analogous feature is present in the Mengo virus electron density map at roughly the same place. In the Mengo virus map, it appears almost as if the 'myristic acid' chains form a bundle reminiscent of, but inverted with respect to, the β-cylinder formed by the amino termini of VP3 (Luo *et al.*, 1987).

Some of the features of the weaker unassigned density located near the internal surface of the HRV14 protein coat may be due to the single strand of RNA inside the virion. Since the 7500 nucleotide RNA strand cannot have icosahedral symmetry, it is unlikely that the RNA will show strong continuous features in an inherently icosahedrally averaged experimental system, such as the HRV14 cubic crystals. However, it is conceivable that there are repeat-

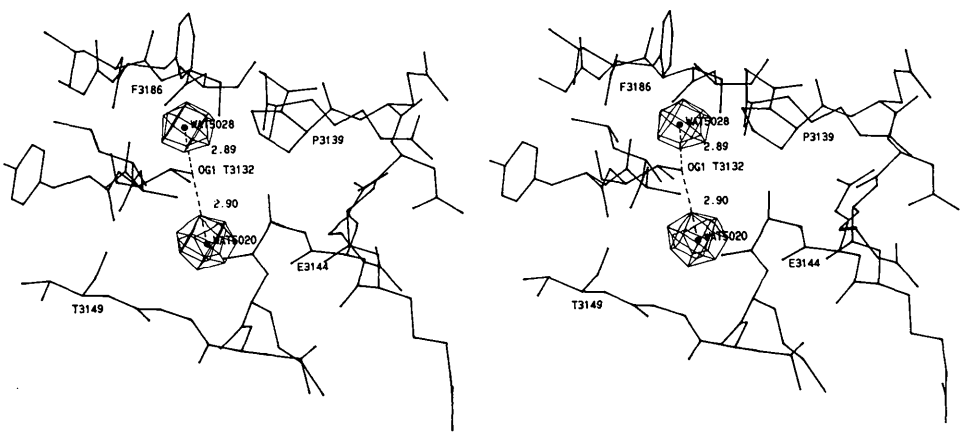


Fig. 2. Solvent omit difference map using molecular-replacement phases. Observe the well resolved water molecules in chemically reasonable positions. The following plausible hydrogen bonds have been omitted for clarity:

WAT5020	VP3, Al45 (O)	3·08 Å	WAT5028	VP3, A130 (O)	3·15 Å
WAT5028	VP3, S188 (Oγ)	2·86 Å	WAT5028	VP3, S188 (N)	3·15 Å.

ing motifs of RNA secondary structure which can be induced to have an approximate icosahedral order by interaction with the inner surface of the icosahedrally symmetric protein shell. There is unassigned density parallel to and at a distance of 3·4 Å from tryptophan 38 in VP2. This density could correspond to an aromatic ring that is stacked with the tryptophan ring in some of the icosahedral asymmetric units. Trials showed that a purine base can be well accommodated by this density. None of the missing amino-terminal residues in VP1, VP2 and VP4 are aromatic, supporting the possible assignment of this density as a purine base of the RNA. Other unassigned density on the inner surface appeared as relatively strong peaks with no contacts of less than 4 Å with polar groups in the ordered protein shell, and thus could be due to phosphates of the RNA. Yet other peaks were associated with basic residues on the internal surface. Alternatively, any of these peaks might be partially ordered residues at the amino termini of VP1, VP2 and VP4.

Convergence and precision

Rees & Lewis (1983) found, using various test calculations, that convergence in their formulation of vector refinement was faster than scalar refinement provided the phase error in the 'observations' was less than 20°. They also found that the rate of convergence is retarded if the error in the 'observed' phases was greater than 35°. It is easy to show that the mean phase difference between the 'observed' and calculated phases is given by $(R_{\text{vector}} \times 180)/(\pi\sqrt{2})$. Thus, the mean phase difference in degrees at the beginning of the HRV14 refinement was 17° and at the end was 9°, well within the limits of rapid convergence suggested by Rees & Lewis (1983). It is unlikely that the formulation of vector refinement used here will substantially alter the conclusions of Rees & Lewis. However, in light of the unusually accurate 'observed' phases derived from non-crystallographic symmetry constraints, the present case provided an opportunity for an application to a real structure.

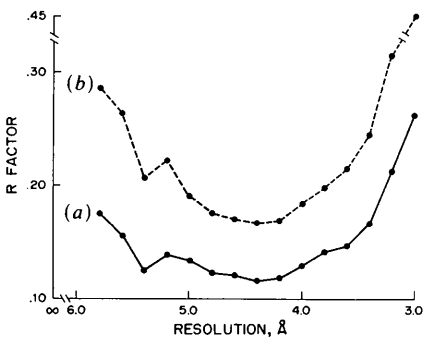


Fig. 3. Distribution of R factor with resolution for (a) the scalar R factor and (b) the vector R factor, where

$$R_{\text{scalar}} = \frac{\sum|(|F_{h,\text{obs}}| - k|F_{h,\text{calc}}|)|}{\sum|F_{h,\text{obs}}|}$$

and

$$R_{\text{vector}} = \frac{\sum|(F_{h,\text{obs}} - kF_{h,\text{calc}})|}{\sum|F_{h,\text{obs}}|}$$

and the $F_{h,\text{calc}}$ values are derived from the atomic model. A randomly selected set of 26 000 reflections was used for this calculation subject to rejection criterion $w_{MR} < 0{\cdot}8$ and resolution limits between 6 and 3 Å (<12% of the observations were rejected in the resolution range). Selection by means of $w_{MR} < 0{\cdot}8$ was found to be roughly equivalent to a rejection based on $|F| < 3\sigma(F)$. For the reflections in this subset, $R_{\text{scalar}} = 0{\cdot}157$ and $R_{\text{vector}} = 0{\cdot}235$.

Table 6. *Comparison of the rate of convergence for vector versus scalar refinement applied to the initial atomic model using data between 6 and 4 Å resolution*

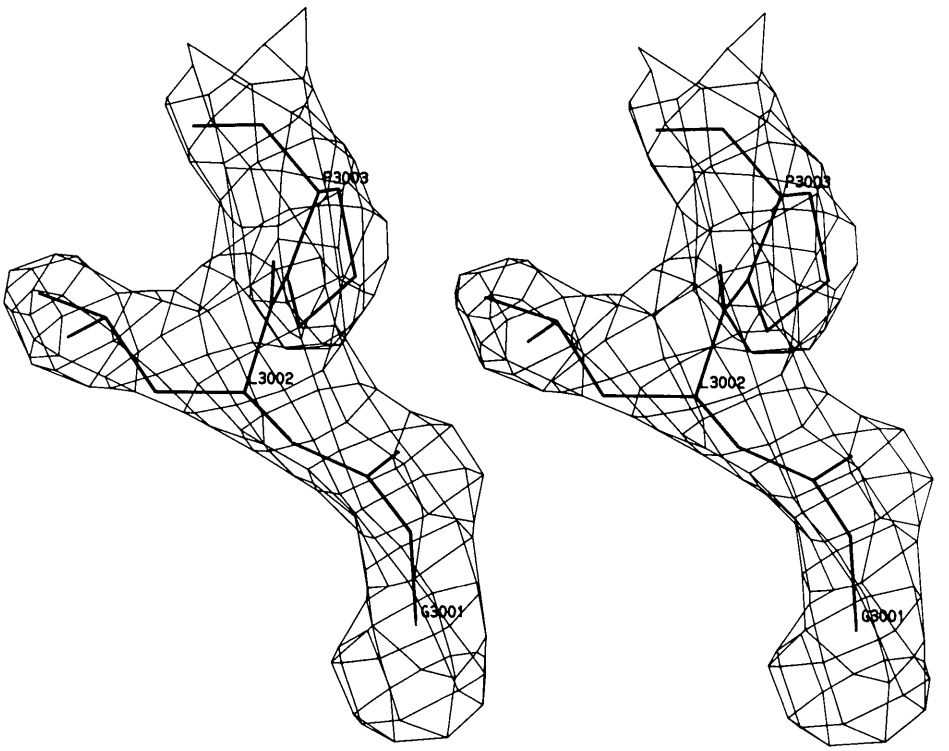
Cycle number	0 (initial model)	1	2	3*	3*†
Number of reflections used in cycle $[F > 3\sigma(F)]$		35 000	35 000	21 000	—
Vector method					
R_{scalar}	0·26	0·207	0·187	0·180	0·273
R_{vector}	0·42	0·310	0·284	0·268	0·413
Scalar method					
R_{scalar}	0·26	0·210	0·199	0·186	0·280
R_{vector}	0·42	0·337	0·308	0·292	0·421

* After cycle 3, r.m.s. shifts from initial-model coordinates were 0·39 and 0·37 Å for the vector and scalar methods, respectively. Stereochemical deviations from ideality were comparable.

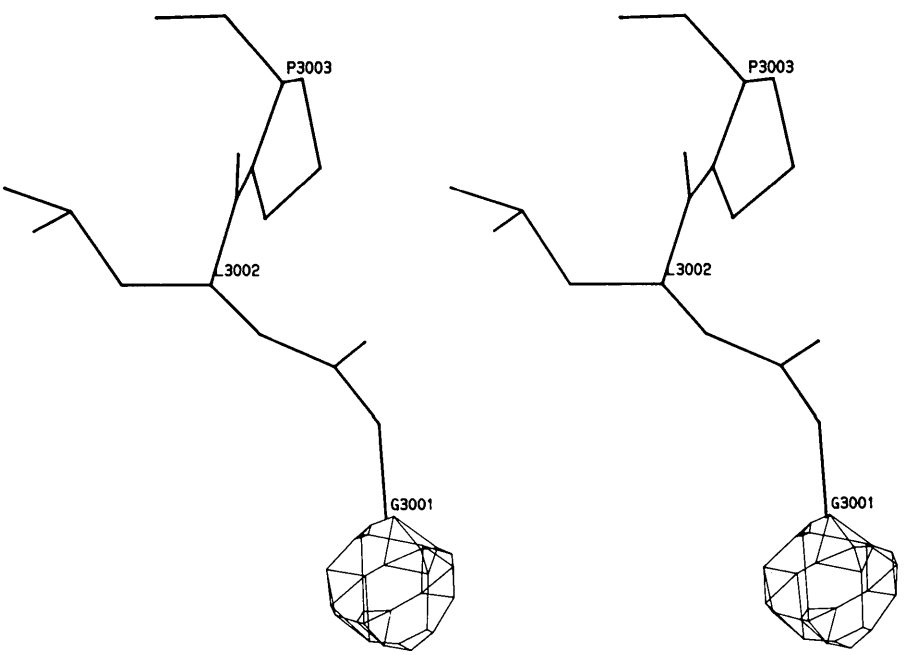
† R factors for data between 4·0 and 3·8 Å resolution.

A number of tests were performed in an attempt to compare the power of vector *versus* scalar refinement of HRV14. In any such comparison, there is an inherent difficulty in matching the weights of the X-ray terms relative to the stereochemical restraints. For a given cycle of refinement in these tests, the X-ray contributions to the normal equations for the vector refinement were multiplied by the ratio of the scalar R factor to the vector R factor, lest the shifts in the vector case be much larger. This scheme gave plausible results in that the r.m.s. shifts in coordinate positions were comparable (differed by less than 10%) for the two methods. A complete test of the power of the vector refinement would require repeating the entire refinement process without use of the phase information.

A test of the relative rate of convergence was made using the initial atomic model and three cycles of refinement employing 21% of the theoretically possible X-ray data between 6 and 4 Å resolution. Assaying in terms of the crystallographic R factor, a similar rate of convergence was observed (see Table 6) for both the scalar and vector refinement methods, although the vector method was slightly faster. The relative weights given to the different stereochemical restraints were the same for both methods and the



(*a*)



(*b*)

Fig. 4. The amino terminus of VP3: (*a*) native molecular-replacement electron density map; (*b*) solvent omit difference map. The contour level is arbitrary.

resultant overall stereochemical deviations from ideality were very similar. The results after three cycles of refinement were also comparable, again with the vector method showing a slight advantage. The R factor for data beyond the limits used for refinement is also consistently lower for the vector refined model. A further assay would have been to measure the deviation of the resultant structure from the true structure, but the fully refined model would be biased towards the vector observations. The scalar method also indirectly resulted in lowering the vector R factor, indicating convergence to a similar minimum. Although these tests did not conclusively demonstrate a faster rate of convergence for vector refinement, the procedure minimizes the actual error (AB on Fig. 1). Since the number of 'observations' is essentially doubled, vector refinement gives the potential of reducing the number of reflections needed to attain a given level of precision (Badger *et al.*, 1988).

We thank Jim Griffith, Greg Kamer, Ming Luo and Gerrit Vriend for their dedicated participation in the structure determination of HRV14 (Rossmann *et al.*, 1985), a necessary prerequisite for this work, and also for their helpful comments. We also thank Jeff Bolin for many useful discussions, Abelardo Silva and Ignacio Fita for implementing *PROLSQ* on the Cyber 205, and some astute referees for probing comments and observations. We are grateful to Sharon Fateley and Kathy Shuster for help in preparation of the manuscript. This work was supported by grants from NSF and NIH to MGR. EA was supported by a Postdoctoral Fellowship from NIH.

APPENDIX

Some errors occurred in the original paper (Rossmann, 1976) on the refinement of non-crystallographic symmetry parameters. Below is reproduced the structure-factor algebra with appropriate corrections. The structure-factor expression of reflection **h** for the J atoms in the compound H can then be written as

$$\mathbf{f}_H = \sum_{j=1}^{J} Z_j \exp(-B_j\sigma^2)\left[\sum_{m=1}^{M}\sum_{n=1}^{N} \exp(2\pi i\mathbf{h}\,.\,\mathbf{x}_{mn})\right]$$

where $\sigma^2 = [(\sin\theta)/\lambda]^2$ (using the usual notation for Bragg angle, wavelength and temperature factor). Therefore

$$\mathbf{f}_H = \sum_{j=1}^{J} Z_j \exp(-B_j\sigma^2) \times\left(\sum_{m=1}^{M}\sum_{n=1}^{N} \exp\{2\pi i\mathbf{h}\,.\,([R_{mn}]\mathbf{x}_j + [T_m]\mathbf{S} + \mathbf{d}_m)\}\right)$$

$$= \sum_{j=1}^{J} Z_j \exp(-B_j\sigma^2) \times\left(\sum_{m=1}^{M} \exp\{2\pi i\mathbf{h}\,.\,([T_m]\mathbf{S}+\mathbf{d}_m)\} \times \sum_{n=1}^{N} \exp(2\pi i\mathbf{h}\,.\,[R_{mn}]\mathbf{x}_j)\right).$$

Now set

$$\theta_m = 2\pi\mathbf{h}\,.\,([T_m]\mathbf{S}+\mathbf{d}_m)$$

and

$$\varphi_{mn} = 2\pi\mathbf{h}\,.\,[R_{mn}]\mathbf{x}_j.$$

These definitions can be rewritten as

$$\theta_m = 2\pi[\mathbf{h}'_m\,.\,\mathbf{S} + \Phi_m], \qquad \varphi_{mn} = 2\pi\mathbf{h}'_{mn}\,.\,\mathbf{x}_j \tag{2}$$

where

$$\mathbf{h}'_m = [T_m]^T\mathbf{h}, \qquad \mathbf{h}'_{mn} = [R_{mn}]^T\mathbf{h}, \qquad \Phi_m = \mathbf{h}\,.\,\mathbf{d}_m.$$

Hence

$$a_H = \sum_j Z_j \exp(-B_j\sigma^2)\left(\sum_m \cos\theta_m \sum_n \cos\varphi_{mn} - \sum_m \sin\theta_m \sum_n \sin\varphi_{mn}\right)$$

$$b_H = \sum_j Z_j \exp(-B_j\sigma^2)\left(\sum_m \sin\theta_m \sum_n \cos\varphi_{mn} + \sum_m \cos\theta_m \sum_n \sin\varphi_{mn}\right).$$

Hence

$$\partial a_H/\partial S_1 = \sum_j Z_j \exp(-B_j\sigma^2) \times\left[\sum_m (-2\pi h'_m)\sin\theta_m \sum_n \cos\varphi_{mn} - \sum_m (2\pi h'_m)\cos\theta_m \sum_n \sin\varphi_{mn}\right],$$

and similarly for S_2 and S_3, which are the y and z components of **S**;

$$\partial a_H/\partial x_j = \sum_j Z_j \exp(-B_j\sigma^2) \times\left[\sum_m \cos\theta_m \sum_n (-2\pi h'_{mn})\sin\varphi_{mn} - \sum_m \sin\theta_m \sum_n (2\pi h'_{mn})\cos\varphi_{mn}\right],$$

and similarly for y_j and z_j, which are the y and z components of $\mathbf{x}_j$;

$$\partial a_H/\partial\xi_1 = \sum_j Z_j \exp(-B_j\sigma^2) \times \left[-\sum_m \cos\theta_m \sum_n (\partial\varphi_{mn}/\partial\xi_1) \sin\varphi_{mn} - \sum_m \sin\theta_m \sum_n (\partial\varphi_{mn}/\partial\xi_1) \cos\varphi_{mn} \right],$$

and similarly for ξ_2 and ξ_3, which are the Eulerian or polar adjustment angles. Similar expressions can be written with respect to b_H. The derivatives with respect to Z_j and B_j are straightforward.

Finally, from (2),

$$\partial\varphi_{mn}/\partial\xi = 2\pi\mathbf{h} \cdot \frac{\partial}{\partial\xi}[R_{mn}] \cdot \mathbf{x}_j,$$

and from (1) (in the body of the paper),

$$\frac{\partial}{\partial\xi}[R_{mn}] = [T_m][\alpha][\rho_n]\left(\frac{\partial}{\partial\xi}[\Delta]\right)[\beta].$$

References

ABDEL-MEGUID, S. S., YAMANE, T., FUKUYAMA, K. & ROSSMANN, M. G. (1981). *Virology*, **114**, 81-85.

ARNOLD, E., LUO, M., VRIEND, G., ROSSMANN, M. G., PALMENBERG, A. C., PARKS, G. D., NICKLIN, M. J. H. & WIMMER, E. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 21-25.

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

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

BADGER, J., MINOR, I., KREMER, M. J., OLIVEIRA, M. A., SMITH, T. J., GRIFFITH, J. P., GUERIN, D. M. A., KRISHNASWAMY, S., LUO, M., ROSSMANN, M. G., MCKINLAY, M. A., DIANA, G. D., DUTKO, F. J., FANCHER, M., RUECKERT, R. R. & HEINZ, G. A. (1988). *Proc. Natl Acad. Sci. USA*. In the press.

BRAUER, D. & WITTMANN-LIEBOLD, B. (1977). *FEBS Lett.* **79**, 269-275.

CHEN, R., BROSIUS, J., WITTMANN-LIEBOLD, B. & SCHAFER, W. (1977). *J. Mol. Biol.* **111**, 173-181.

CHOW, M., NEWMAN, J. F. E., FILMAN, D., HOGLE, J. M., ROWLANDS, D. J. & BROWN, F. (1987). *Nature (London)*, **327**, 482-486.

DELATORE, D., ROSSMANN, M. G., BAKER, T. S., RAYMENT, I., MURAKAMI, W. T. & CASPAR, D. L. D. (1987). *J. Mol. Biol.* Submitted.

FITA, I. & ROSSMANN, M. G. (1987). Unpublished results.

HENDRICKSON, W. A. (1981). *Refinement of Protein Structures*, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER, pp. 1-8. Warrington: SERC Daresbury Laboratory.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136-143.

HOSUR, M. V., SCHMIDT, T., TUCKER, R. C., JOHNSON, J. E., GALLAGHER, T. M., SELLING, B. H. & RUECKERT, R. R. (1987). *Proteins: Structure, Function, Genet.* **2**, 167-176.

JONES, T. A. & LILJAS, L. (1984). *Acta Cryst.* A**40**, 50-57.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344-350.

LEDERER, P., ALIX, J. H. & HAYES, D. (1977). *Biochem. Biophys. Res. Commun.* **77**, 470-480.

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.

PETTIGREW, G. W. & SMITH, G. M. (1977). *Nature (London)*, **265**, 661-662.

PIONTEK, K., CHAKRABARTI, P., ROSSMANN, M. G. & ZUBER, H. (1988). In preparation.

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

REES, D. C. (1983). *Acta Cryst.* A**39**, 916-920.

REES, D. C. & LEWIS, M. (1983). *Acta Cryst.* A**39**, 94-97.

REES, D. C., LEWIS, M. & LIPSCOMB, W. N. (1983). *J. Mol. Biol.* **168**, 367-387.

ROSSMANN, M. G. (1976). *Acta Cryst.* A**32**, 774-777.

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., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H. J., JOHNSON, J. E., KAMER, G., LUO, M., VRIEND, G., MOSSER, A. G., PALMENBERG, A. C., RUECKERT, R. R. & SHERRY, B. (1986). *Molecular Evolution of Life*, edited by H. BALTSCHEFFSKY, H. JÖRNVALL & R. RIGLER, pp. 313-323. Cambridge Univ. Press.

ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641-647.

RUECKERT, R. R. (1986). *Fundamental Virology*, edited by B. FIELDS & D. KNIPE, pp. 357-390. New York: Raven Press.

SHERIFF, S. (1987). *J. Appl. Cryst.* **20**, 55-57.

SILVA, A. M. & ROSSMANN, M. G. (1985). *Acta Cryst.* B**41**, 147-157.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.

SMITH, T. J., KREMER, M. J., LUO, M., VRIEND, G., ARNOLD, E., KAMER, G., ROSSMANN, M. G., MCKINLAY, M. A., DIANA, G. D. & OTTO, M. J. (1986). *Science*, **233**, 1286-1293.

STAUFFACHER, C. V., USHA, R., SCHMIDT, T., HARRINGTON, M., ARNOLD, E., KAMER, G., VRIEND, G. & JOHNSON, J. E. (1988). In preparation.

WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943-956.

Acta Cryst. (1988). A**44**, 283-285

The RED Extinction Model. I. An Upgraded Formalism

BY JIŘÍ KULDA

Nuclear Physics Institute, Czechoslovak Academy of Sciences, 250 68 Řež near Prague, Czechoslovakia

(*Received* 9 *March* 1987; *accepted* 26 *November* 1987)

Abstract

Some modifications are introduced into the extinction correction formulae based on the RED (random elastic deformation) model developed earlier by the author [Kulda (1987). *Acta Cryst.* A**43**, 167-173]. The scattering cross section, which includes a correction for primary extinction, has been extended to allow for a more general angular variation of the effective deformation gradient. The originally proposed $\cos\theta$ dependence is included as one of the limiting cases corresponding to a pure lattice-plane misorientation. A modified expression has been derived for the angular reflection curve profile which describes properly the broadening due to the finite dimensions of the diffracting region. This version of the RED model containing three free parameters has been employed in experimental tests reported in paper II [Kulda (1988). *Acta Cryst.* A**44**, 286-290].

1. Introduction

In a recent paper (Kulda, 1987), hereafter referred to as paper *A*, we have proposed the random elastic deformation (RED) model for extinction treatment in real crystals. Unlike the traditional approaches, reviewed, for example, by Becker (1977), which use the concept of mosaic structure, RED is based on a stochastic sequence of elastically deformed domains. It is just this feature which in the end makes it possible to cover primary extinction effects more adequately by a direct use of the dynamical diffraction theory (Kulda, 1984). Secondary extinction is treated in a conventional way by solving the Hamilton-Zachariasen intensity coupling equations.

Paper *A* was confined to development of the RED concept and to comparison of its basic characteristics with other existing extinction models. The aim of the present paper is to provide working formulae that could be built into a least-squares refinement program and used in practical performance tests of the RED model (Kulda, 1988). As the most important step in this direction we shall derive a modified expression for the angular distribution function $w_{\rm RED}(\varepsilon)$, taking into account broadening caused by finite dimensions of the reflecting regions and allowing for a more general deformation type.

2. Generalized deformation gradient

In paper *A*, for the sake of simplicity, we set $|\partial\Delta\theta/\partial s_0| = \cos\theta/R$, thus considering the component of elastic bending that brings about pure lattice-plane misorientation as the only source of deformation. In the course of practical tests this assumption – though realistic for a particular crystal – proved too limiting for general use. As pointed out already in the discussion of paper *A* any real deformation consists of both misorientation components and components exhibiting changes of the lattice parameter. In the latter case with the diffraction vector parallel to the atomic displacement **U** we would have $|\partial\Delta\theta/\partial s_0| = \sin\theta/R$. Therefore it appears reasonable to introduce a more flexible model $|\partial\Delta\theta/\partial s_0| = C(c,\theta)/R$, which could with the help of a free parameter c allow for both situations. Furthermore, the mosaic width parameter α [*cf.* equations (13) and (14) of paper *A*, here referred to as (*A*13) and (*A*14)] is angle independent within the Becker & Coppens (1974, 1975) treatment which is known to be exact in the limit of pure secondary extinction (Kato, 1976). Thus also the possibility $C(c,\theta) = \text{constant}$ should be covered by our improved formula. The simplest choice satisfying all three requirements is

$$C(c,\theta) = [c + (1-2c)\cos^2\theta]^{1/2} \qquad (1)$$

where $c \in \langle 0, 1\rangle$ is an additional free parameter of the RED model.

The generalized angular distribution function $w(\varepsilon)$ is now written in analogy with (*A*11) as

$$w_{\rm RED}(\varepsilon) = \{R/[\bar{t}C(c,\theta)]\} \times \exp(-\pi\{R\varepsilon/[\bar{t}C(c,\theta)]\}^2). \qquad (2)$$

This expression, with $C(c,\theta)$ given by (1), can be interpreted as a joint distribution function of two superimposed random walk sequences of the same type (*A*7) differing just in the deformation gradients $|\partial\Delta\theta/\partial s_0|_1 = \cos\theta/R$ and $|\partial\Delta\theta/\partial s_0|_2 = \sin\theta/R$ and in the numbers of permitted states N_1, N_2. Putting $c = N_2/(N_1+N_2)$ we see that (2) is just a convolution of the distribution functions

$$w_i(\varepsilon) = (\bar{t}_i|\partial\Delta\theta/\partial s_0|_i)^{-1} \times \exp[-\pi\varepsilon^2/(\bar{t}_i|\partial\Delta\theta/\partial s_0|_i)^2].$$

0108-7673/88/030283-03$03.00