Acta Cryst. (1990). A46, 352-359

Intensity-Based Domain Refinement of Oriented but Unpositioned Molecular Replacement Models

BY T. O. YEATES AND J. M. RINI*

Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

(*Received* 29 *August* 1989; *accepted* 21 *November* 1989)

Abstract

A program is described that performs least-squares group refinement of oriented molecular replacement models whose positions in the unit cell are unknown. The program (*INTREF*) is designed to produce improved models for use in a translation function by optimizing the orientations and relative translations of the model domains. The molecular contents of the asymmetric unit are refined as a small number of rigid bodies whose origins relative to each other may be unknown. More than one molecule in the asymmetric unit can be accommodated. The refinement seeks to minimize the residual error between the observed and calculated intensities that have been modified to produce the equivalent of a radial weighting in Patterson space. Calculated intensities include contributions from all symmetry-related molecules, enabling meaningful refinement in high-symmetry space groups. Derivatives of the intensities with respect to the rigid-body parameters are evaluated numerically using fast Fourier transforms and the shifts are obtained by non-linear least-squares analysis. Results with test cases show that the program is capable of adjusting the orientations and relative translations of protein domains to give models that more closely resemble the known structures. Consequently, the resulting models produce more accurate and more interpretable results in translation functions. The importance of including all crystallographically related molecules and of downweighting the contribution of the longer-radius region of the Patterson function is demonstrated.

Introduction

Experience with the molecular replacement method has shown that, while it is often possible to determine the orientation of a model, it is frequently more difficult to solve for its translational position (Lattman, 1985). The success of a translation search is critically dependent on both the structural similarity between the model and the unknown and the accuracy with which it is oriented (Rini, Hardman, Einspahr, Suddath & Carver, 1989; Cygler & Anderson, 1988*a*; Lattman, 1985). The existing methods for refining the orientation of molecular replacement models have been based on fine grid searches over the rotational parameters (*BRUTE*; Fujinaga & Read, 1987; Lattman & Love, 1972). If the model is sufficiently similar to the unknown, the accurately oriented model will yield the desired translational position. However, with multi-domain proteins, internal flexibility results in errors in the relative dispositions of domains and correspondingly poor models. Although it may be possible to determine the approximate orientation of such a model, the inherent errors can make it difficult, if not impossible, to determine the translational parameters. An extreme example arises with the Fab structures where the large range in elbow angles has made it generally necessary to treat the constant and variable domains as separate models (Cygler & Anderson, 1988*b*). However, by dividing the models into smaller pieces, difficulties arise because of poor signal-to-noise ratios. Clearly, it would be more desirable to improve the model by making rigid-body adjustments of the domains prior to performing the translation search.† Although *BRUTE* (Fujinaga & Read, 1987) can refine the orientation of one piece relative to a second stationary piece, both the orientation and translation of the second piece must be known. In this paper we describe an intensity-based procedure, *INTREF*, which is specifically designed to optimize oriented molecular replacement models whose positions in the cell are not yet known.

Methods

1. *Definition of the refinable parameters*

The contents of the crystallographic asymmetric unit can be divided into a small number of distinct pieces whose orientations and relative translations are to be optimized. The division occurs at two levels:

(*a*) *Groups.* These are pieces whose orientations are known, but whose relative positions are entirely unknown. The orientation of each group is refined with no contribution from cross vectors between

* To whom all correspondence should be addressed.

† Recently, a method has been developed which can refine multiple fragments by maximizing an intensity-based correlation coefficient (Brunger, 1989. Personal communication).

0108-7673/90/050352-08$03.00

groups. More than one molecule in the asymmetric unit can be accommodated and each would ordinarily be classified as a distinct group.

(*b*) *Parts.* Each group may be divided into a small number of 'parts', whose relative positions are known approximately. Typically, this occurs in multi-domain proteins when domain dispositions are expected to differ slightly from that in the molecular replacement model. Each domain could then be treated as a separate part; the orientation and relative translation of each part in each group would be refined.

The variables used for refining the orientations are small rotations about three orthogonal axes. A group with one part (*i.e.* a group which is not further subdivided) has three refinable parameters corresponding to its orientation. A group with two parts has nine refinable parameters; three rotational parameters for each of the two parts and three parameters for translating the second part relative to the first. The current version of the program can refine up to 21 parameters.

2. *An intensity-based error function*

When two groups with an unknown relative translation between them contribute to a reflection, the true intensity cannot be calculated accurately, since the relative phase between the two complex structure factors is unknown. Relying instead on a statistical analysis, the best estimate of the calculated intensity is given by

$$\langle|\mathbf{F}_1+\mathbf{F}_2|^2\rangle = |\mathbf{F}_1|^2+|\mathbf{F}_2|^2+2|\mathbf{F}_1||\mathbf{F}_2|\int_{\varphi=0}^{2\pi} P(\varphi)\cos(\varphi)\,\mathrm{d}\varphi, \quad (1)$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are the complex structure factors calculated from the two groups in space group $P1$ and $P(\varphi)$ is the probability distribution of the unknown relative phase φ between $\mathbf{F}_1$ and $\mathbf{F}_2$. Without regard to packing constraints, the distribution of φ is uniform, the integral equals zero, and the best estimate of the total intensity is simply the sum of the component intensities.* By induction, this generalizes to N groups as

$$\left\langle\left|\sum_{i=1}^{N}\mathbf{F}_i\right|^2\right\rangle=\sum_{i=1}^{N} I_i \quad (2)$$

where $I_i=|\mathbf{F}_i|^2$. Using this statistical approximation to the calculated intensities we can write the following simple residual error function:

$$E=\sum_{\mathbf{h}}\left[I_{\mathrm{obs}}(\mathbf{h})-s\sum_i\sum_j I_i(S_j\mathbf{h})\right]^2 \quad (3)$$

where $I_{\mathrm{obs}}(\mathbf{h})$ is the observed intensity of reflection $\mathbf{h}$, S_j is the jth symmetry operator of the crystallographic point group, and $I_i(S_j\mathbf{h})$ is the intensity calculated from group i in space group $P1$ for reflection $S_j\mathbf{h}$. The scale factor s, typically evaluated separately for each resolution shell, is chosen to minimize E. The reflection vector $\mathbf{h}$ ranges over the unique region of reciprocal space as dictated by the crystal symmetry. Since crystallographically related molecules are simply a special case of distinct groups, summing over the symmetry-related molecules serves to include their contributions to the calculated intensities.

Since the Patterson vectors between groups tend to be longer than the intra-group vectors, this approximation to the correct intensities is expected to produce a Patterson function which is more accurate near the origin. Therefore, the residual error function can be improved by modifying the observed and calculated intensities to downweight the unknown inter-group vectors in Patterson space. The weighting function in Patterson space was chosen to be the radially symmetric Gaussian function,

$$g(x)=\exp(-x^2/2\rho^2), \quad (4)$$

where ρ corresponds to the standard deviation of a one-dimensional Gaussian function. Its Fourier transform, which is the corresponding convolution function in reciprocal space,† is given by

$$G(\mathbf{k})=\exp[-2\pi^2\rho^2 d^{*2}(\mathbf{k})], \quad (5)$$

where $d^*(\mathbf{k})$ is the reciprocal-space length of the difference reflection vector $\mathbf{k}$. Since the function G falls off quickly in reciprocal space, the modified intensities may be approximated by considering only a small number of neighboring reflections in what would otherwise be a weighted sum over all reflections. *INTREF* only sums over the difference vectors, $\mathbf{k}$, whose weights $G(\mathbf{k})$ exceed 5% of the maximum $[G(0)=1{\cdot}0]$. For a typical value of 20 Å for ρ and a unit cell with 100 Å edges, the three indices of $\mathbf{k}$ would range from -1 to 1 and the number of difference vectors required for summation would be 27.

The residual error function becomes

$$E=\sum_{\mathbf{h}}\left\{\sum_{\mathbf{k}} G(\mathbf{k})I_{\mathrm{obs}}(\mathbf{h}+\mathbf{k}) - s\sum_{\mathbf{k}} G(\mathbf{k})\sum_i\sum_j I_i[S_j(\mathbf{h}+\mathbf{k})]\right\}^2. \quad (6)$$

The residual error is minimized by the non-linear least-squares method, calculating the partial derivatives of the second term of (6) with respect to the refinable rigid-body parameters. Derivatives are computed numerically by taking the difference between intensities calculated from the model before and after

* In Patterson space, this treatment corresponds to omitting cross vectors between groups.

† The function G plays a role similar to that of the interference function described by Rossmann & Blow (1962).

a small step in a given parameter has been made. A similar approach has been used to calculate derivatives in the reciprocal-space rigid-body refinement method of Huber & Schneider (1985). Structure factors are calculated with a fast Fourier transform (FFT) (Ten Eyck, 1973) in the space group $P1$ and the corresponding intensities are modified as described above to effect the radial weighting and to incorporate the crystal symmetry. Since the full symmetry is considered in the derivative calculation, reliable refinement can be achieved even as the crystal symmetry increases. The number of Fourier transforms required to obtain the necessary derivatives is equal to the number of refinable parameters plus the number of separate groups. The desired shifts in the parameters are obtained by solving the resulting normal equations. The parts of the model are moved accordingly and the refinement cycle is repeated. The scale factors are re-evaluated in each cycle so that, at convergence, both the rigid-body parameters and the scale factors have been optimized.

Results

1. *Test case with an ideal model*

The performance of the program was evaluated in this test case using the refined model of the photosynthetic reaction center from *Rb. sphaeroides* (Yeates, Komiya, Rees, Allen & Feher, 1987). Ideal 'observed' structure factors were calculated from the refined model in the true cell ($P2_12_12_1$) with uniform atomic temperature factors of 20 Å^2. A starting model was generated by rotating the L subunit by 1·7° about the crystallographic a axis and translating the M subunit by 1 Å along the b axis. The H subunit was left in its refined position. Five cycles of *INTREF* refinement were performed treating the model as one group with three parts, thereby refining the rotation and inter-domain translations of each piece. Data between 8 and 4 Å resolution were used and the Patterson radius ρ was chosen to be 25 Å. The residual error was reduced to 66% of its starting value and the r.m.s. difference between atomic positions (with the centers of mass of the models superimposed) was reduced from 0·711 to 0·07 Å. The importance of including the radial weighting was evaluated by repeating the refinement in the absence of the weighting scheme. The rate of convergence was similar, but the final error in atomic positions was 0·22 Å. In order to assess the importance of including the crystallographically related molecules, the observed data were expanded to $P1$ and the refinement was repeated in the absence of the symmetry terms. In this case, 12 refinement cycles were required to achieve the same degree of convergence and the final r.m.s. error was 0·75 Å. For comparison, the model was also refined using the observed diffraction data including the symmetry terms and radial weighting as described above. The final atomic r.m.s. error was 0·23 Å. These tests show that *INTREF* is capable of refining accurately the orientations and relative translations of the three domains. As expected, including symmetry-related molecules and downweighting the longer vectors in Patterson space improves the accuracy of the final result.

The convergence properties of the method are expected to be improved by exponential damping of higher-resolution terms (*i.e.* with a temperature factor) and by using relatively large perturbations in evaluating derivatives. Both techniques achieve a smoothing of the error function. Evaluating a derivative numerically with a large step is equivalent to taking the average value of the (analytical) derivative over that range. While we have not established the optimal combination of these parameters or its dependence on the resolution range of the data, steps of about a quarter of the expected error seem to perform well. In addition, an elevated atomic temperature factor of 100 Å^2 has proven to be useful in the initial refinement cycles.

To test the radius of convergence with respect to rotational errors, a series of starting models was generated by rotations about the x-y diagonal. For the case of a 16° error, five cycles of *INTREF* refinement (angular step size of 4°, 8·0–4·0 Å resolution, B = 100 Å^2) reduced the angular error to 5·7°. Subsequent refinement with a smaller step size of 0·2° and a temperature factor of 30 Å^2 refined the model to within 0·10° of the correct orientation; similar attempts to refine a model that was misoriented by 20° were unsuccessful. For comparison, an attempt was made to refine the same starting model (16° error) using a smaller rotational step size of 0·2°. Following five cycles of refinement, the angular error was only reduced to 13·6°. After 15 additional cycles, the angular error was reduced to 1·0°. Therefore, it appears that the speed of convergence is improved by using large perturbations to evaluate derivatives, at least for the first several cycles of refinement; smaller step sizes are required in the final stages in order to obtain accurate derivatives.

2. *Test case with an imperfect two-domain model* ($P2_12_12_1$)

A pea lectin–trisaccharide complex (Rini, Hardman, Einspahr, Suddath & Carver, 1989) provided a test case using observed structure-factor amplitudes and an imperfect two-domain model. The structure was solved using the native pea lectin coordinates as model (Einspahr, Parks, Suguna, Subramanian & Suddath, 1986), but parallel attempts to solve the structure with concanavalin A (Con A) (Hardman, Agarwal & Freiser, 1982) had also been made. The two molecules are structurally very similar and have

40% sequence homology. The twofold rotation axis relating the monomers in Con A is crystallographic while the twofold relationship between monomers in pea lectin is non-crystallographic and approximate. Although the rotation-function solutions obtained with Con A, using either the intact dimer or the individual monomers, agreed to within a few degrees of that found for the corresponding pea lectin models, these rotational inaccuracies led to uninterpretable results in the T_1 translation function (Crowther & Blow, 1967). The Con A model was used here to determine whether *INTREF* refinement could produce an improved model that would result in a successful solution of the translational parameters.

As a control, the Con A dimer, oriented by the fast rotation function (Crowther, 1972) using data from 6·0 to 3·0 Å resolution, was run in the translation search from *X-PLOR* (Brunger, 1988). The search is based on the correlation between observed and calculated diffraction intensities and was run using data in the 8·0–4·0 Å resolution range. The cell was sampled from 0 to $\frac{1}{2}$ along each of the unit-cell directions on a 0·5 Å grid. The highest feature in the translation map (Fig. 1*a*) lies on a strong streak in the *z*-axis direction and shows little discrimination over several other peaks in the map. Although a streak is expected parallel to a twofold rotation or screw axis, in this case it is particularly prominent. The extensive β sheet which continues across the dimer interface is oriented such that the β strands are parallel to the *z* axis. Since the plane of the sheet is inclined at an angle of approximately 45° to the *x* axis, the 3 Å repeat in the *x* direction presumably corresponds to misalignment of the model by one β strand. The solution corresponds to the fractional shift vector (0·297, 0·363, 0·383) which places the center of gravity of the model 3·0 Å from the true position.

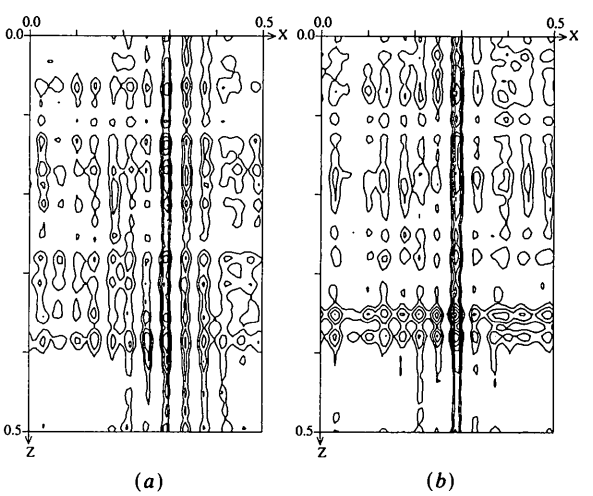

Fig. 1. A *y* section of the translation search from *X-PLOR* (Brunger, 1988) for the pea lectin–trisaccharide complex using (*a*) Con A, oriented by the fast rotation function, as the search model. The highest peak in the translation-function map is at (0·297, 0·363, 0·383), 3·0 Å from the true position. The peak is 5·6σ above the mean and is 0·3σ over the next highest peak. (*b*) Translation function using the same model as in (*a*) after performing *INTREF* refinement. The highest peak in this map is at (0·289, 0·349, 0·350), 1·3 Å from the true position. The peak is 7·8σ above the mean and is 0·8σ over the next highest peak. In both (*a*) and (*b*) the maps are contoured in 1σ intervals starting at 1σ above the mean of the maps.

The same oriented but unpositioned model of Con A was refined by *INTREF* as a single group. Each monomer was defined as a separate part, thereby optimizing the orientations as well as the relative separations of the two domains. Fifteen cycles of *INTREF* refinement were performed using data in the 8·0 to 4·0 Å resolution range with a Patterson radius, ρ, of 20 Å. Shift increments of 0·5° and 0·25 Å were used in calculating derivatives. The resulting reorientations about the *x*, *y* and *z* axes were 3·1, −6·7, 2·4° and −2·5, −0·9, −1·7° for the two monomers, while the magnitude of the relative translational shift was 1·9 Å. The resulting *INTREF*-refined Con A dimer was then run in the *X-PLOR* translation search exactly as described for the unrefined model. The highest peak in the translation map (Fig. 1*b*) is 0·8σ higher than the next peak and corresponds to the fractional shift vector (0·289, 0·349, 0·350). This solution places the center of mass of the model only 1·3 Å from the expected position, in contrast to the 3·0 Å discrepancy obtained with the original model.

These results illustrate two important consequences of *INTREF* refinement. Firstly, the translational accuracy obtained with the *INTREF*-refined model places it within the radius of convergence typically obtained with phased rigid-body refinement procedures (Scheringer, 1963; Sussman, Holbrook, Church & Kim, 1977; Huber & Schneider, 1985; Yeates & Rees, 1988). Therefore, the *INTREF*-refined molecular replacement model is expected to refine smoothly as the structure determination proceeds. Derewenda (1989) has given an example of the difficulties encountered during atomic refinement due to errors in the rotation and translation parameters of the molecular replacement model. Secondly, the signal-to-noise ratio in the translation function is improved with the refined model and the correct peak can be identified with much greater certainty. As illustrated in this example, the highest peak in the translation map was the correct peak only after *INTREF* refinement of the model.

3. *An imperfect four-domain model*

A Fab′-peptide complex (Fab B13I2) solved by molecular replacement (Stanfield, Fieser, Lerner & Wilson, 1990) was used as a test case for a multi-domain protein in a high-symmetry space

group ($P6_322$). Since considerable difficulty was encountered in solving this structure, it serves as a challenging test of the *INTREF* refinement procedure. The models used for the variable and constant domains came from coordinates for Fab KOL (Marquart, Deisenhofer, Huber & Palm, 1980; PDB: Bernstein *et al.*, 1977) and Fab 17/9 (Rini & Wilson, unpublished) respectively. The variable domain from the KOL model had been superimposed onto that of Fab 17/9 to facilitate interpretation of the translation results. Each of the domains was then rotated to the orientation determined by the fast rotation function (Crowther, 1972) computed using 10 to 5 Å data for the variable domain and 10 to 4 Å data for the constant domain.

As a control, the translation function described by Harada, Lifchitz, Berthou & Jolles (1981), as implemented by D. J. Filman, was run for each domain independently. Since the packing terms were not included in the analysis, these results correspond strictly to the numerator of the function. Data in the resolution range 8–4 Å were included and a value of 25 Å was used for ρ. The cell was sampled on a 0·5 Å grid from 0 to 1, 0 to 1, 0 to $\frac{1}{2}$, in the unit-cell directions *a*, *b* and *c* respectively. Fig. 2(*a*) shows a section of the translation-function map obtained using the variable domain model. The peak is $6{\cdot}9\sigma$ over the mean and corresponds to the correct solution as determined from the solved structure. It is the highest peak in the map and is $0{\cdot}9\sigma$ greater than the next. In contrast, the search performed with the constant domain model does not yield interpretable results. The section of the map expected to contain the correct peak is shown in Fig. 3(*a*) for comparison. These results are similar to those found by Stanfield & Wilson (unpublished) using the *BRUTE* (Fujinaga & Read, 1987) translation function. They were only able to position the constant domain by performing a computationally intensive six-dimensional grid search over a limited set of rotation and translation parameters using *BRUTE*.

The same starting model as described above was then refined by *INTREF*. Because of the uncertainty in the relative translational position of the variable and constant domains, each was initially defined as a separate group made up of a single part. *INTREF* refinement was run for five cycles using shift increments of 0·5° and 0·25 Å for calculating derivatives followed by five cycles with shift increments of 0·2° and 0·1 Å. Each cycle of refinement required 8 min CPU on a Convex C2 computer. Both the light and heavy chains of each domain (V_L, V_H, C_L, C_H) were then refined in both orientation and relative position by running five additional cycles using shift increments of 0·2° and 0·1 Å. The cumulative angular changes were 6·9, 3·4, 4·1, and 3·0° for V_L, V_H, C_L, and C_H respectively. The relative translational position of the light and heavy chains changed by 1·6 and 0·6 Å for the variable and constant domains respectively.

The resulting *INTREF*-refined variable and constant domain models were then rerun in the translation function under conditions identical to those described for the unrefined models. Fig. 2(*b*) shows the same region of the translation search map as shown for the unrefined variable domain. As before, the peak is the highest feature in the map but has now become $10{\cdot}4\sigma$ over the mean and $3{\cdot}0\sigma$ higher

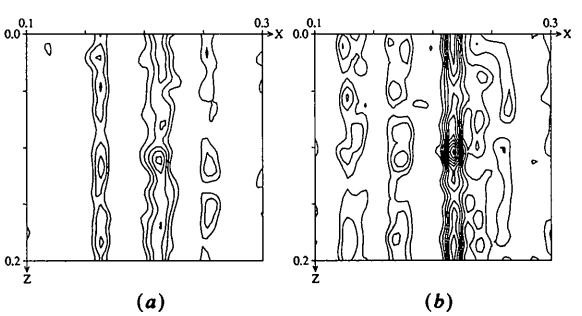

Fig. 2. Partial *y* section of the full symmetry translation function of Harada, Lifchitz, Berthou & Jolles (1981) for the Fab B13I2–peptide complex using (*a*) the KOL variable domain, oriented by the fast rotation function, as the search model. The highest peak in the translation-function map corresponds to the true position. The peak is $6{\cdot}9\sigma$ above the mean and is $0{\cdot}8\sigma$ over the next highest peak. (*b*) Translation function using a model obtained from the one in (*a*) by refining the orientations and relative dispositions of the light and heavy chain domains in *INTREF*. The correct peak is now $10{\cdot}4\sigma$ above the mean and is $3{\cdot}0\sigma$ above the next highest peak. In both (*a*) and (*b*) the maps are contoured in 1σ intervals starting at 1σ above the mean of the maps.

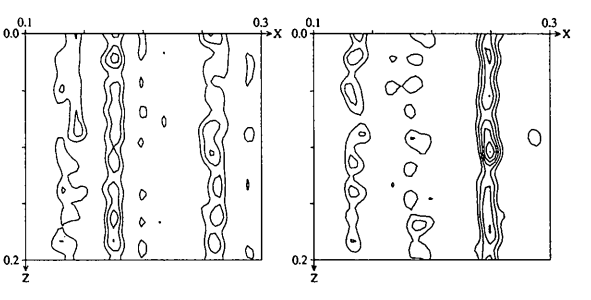

Fig. 3. Partial *y* section of the full-symmetry translation function of Harada, Lifchitz, Berthou & Jolles (1981) for the Fab B13I2–peptide complex using (*a*) the Fab 17/9 constant domain oriented by the fast rotation function as the search model. There is no interpretable peak on the section. (*b*) Translation function using a model obtained from the one in (*a*) by refining the orientations and relative dispositions of the light and heavy chains in *INTREF*. The highest peak in the translation map is now the correct peak. The peak is $6{\cdot}1\sigma$ above the mean and is $0{\cdot}5\sigma$ above the next highest peak. In both (*a*) and (*b*) the maps are contoured in 1σ intervals starting at 1σ above the mean of the maps.

than the next highest peak. This increased signal-to-noise ratio allows one to identify the correct peak with much greater confidence. Fig. 3(*b*) shows a section of the map computed with the *INTREF*-refined constant domain model in the vicinity of the correct peak. The model now results in a peak at the expected position which is the highest feature in the map. It is $6{\cdot}1\sigma$ over the mean and $0{\cdot}4\sigma$ higher than the next highest peak. Although the relative translation and consequently the vectors describing the interactions between the variable and constant domains were ignored, the corrections made in the orientations and relative dispositions of the chains in each domain were meaningful. This is evidenced by the fact that when taken individually the refined domains resulted in interpretable and more significant results in the translation function. These results are sufficient to position the contents of the asymmetric unit and therefore to solve the structure.

Since the KOL variable domain model had originally been superimposed on the variable domain of Fab 17/9, the error in relative translation between domains is primarily a result of the difference in elbow angle between the two structures. It was for this reason that the inter-domain translation was not originally refined. To determine whether *INTREF* could correct the relative translational errors between the two domains, and thereby assemble a more complete model, the *INTREF*-refined model from above was refined as a single group. Each of the variable and constant domains was initially treated as a single part. Seven cycles of *INTREF* refinement were then performed starting with a translational shift increment for derivative calculations of 2·0 Å which was then reduced to 0·4 Å. The orientations and relative translations of the light and heavy chains for each domain were then refined to convergence by running several cycles with shift increments of 0·4° and 0·2 Å. The resulting incremental changes in orientation were 1·1, 1·0, 3·9, and 2·1° for V_L, V_H, C_L and C_H respectively. The translational shifts for V_H, C_L, and C_H, relative to V_L (fixed by convention), are 0·32, 3·97 and 3·89 Å. The large concerted shift in relative translation shown by the two constant domain chains illustrates the magnitude of the corrections which can be achieved by this method. The composite reassembled model was then run in the translation search as described for the individual domains. As shown in Fig 4, the correct peak is now $14{\cdot}7\sigma$ over the mean and $4{\cdot}9\sigma$ greater than the next highest peak. An analysis of the cumulative shifts are presented in Table 1. Shown is a pairwise comparison between the *INTREF*-refined model, the solved structure and the initial model. The results indicate that the corrections made by the refinement are highly correlated with those required to superimpose the four individual chains of the initial Fab model onto those of the solved structure. In all cases, the errors associated with the individual chains have been reduced. The remaining angular and translational errors correspond to similar deviations in mean atomic positions for a molecule of this size.

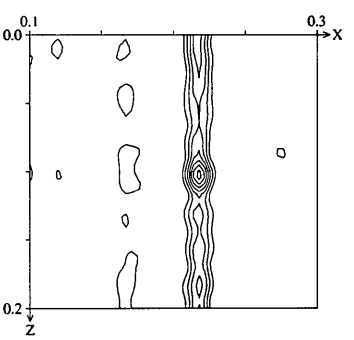

Fig. 4. Partial *y* section of the full-symmetry translation function of Harada, Lifchitz, Berthou & Jolles (1981) for the Fab B13I2–peptide complex with the complete *INTREF*-refined model. The highest peak in the map corresponds to the true solution. The peak is $14{\cdot}7\sigma$ above the mean and is $4{\cdot}9\sigma$ above the next highest peak. The map is contoured in 2σ intervals starting at 2σ above the mean of the map.

Table 1. *Pairwise comparisons of the INTREF-refined model, the solved structure and the initial model. The orientational change and the shift in center of mass required to superimpose each of the V_L, V_H, C_L and C_H chains is given for each pair of models*

Models	Domain	Angular difference (°)	Translational difference (Å)
Unrefined model *vs* solved structure	V_L	5·0	0·00*
	V_H	6·9	1·12
	C_L	5.2	4·88
	C_H	5·3	5·01
Unrefined model *vs* *INTREF*-refined model	V_L	7·5	0·00*
	V_H	5·6	1·26
	C_L	4·6	4·45
	C_H	5·1	4·67
INTREF-refined model *vs* solved structure	V_L	3·1	0·00*
	V_H	4·3	0·20
	C_L	3·2	0·79
	C_H	1·1	0·47

* Center of mass fixed by convention.

These results illustrate the improvements associated with *INTREF* refinement. As shown, the translation-function results are dramatically improved with only small adjustments in the dispositions of the light and heavy chains of each domain. The resulting constant domain model, run independently, yields the translational parameters necessary to solve the structure. Similarly, the correct peak using the *INTREF*-refined variable domain alone is significantly enhanced. In both cases these improvements resulted from refinement which did not include cross vectors between the constant and variable domains. Including these vectors enabled the individually refined domains to be reassembled and their relative

translational disposition to be corrected by almost 4·0 Å. The remarkable increase in signal-to-noise ratio in the resulting translation map clearly illustrates the gains which can be made by assembling the component parts of a multi-domain protein. Furthermore, subsequent phased refinement of the structure is expected to be facilitated by this reduction of errors between the model and the unknown. This example, in space group $P6_322$, shows that the procedure performs well even in high-symmetry space groups.

Concluding remarks

The program *INTREF* was designed to produce improved molecular replacement models to be used in translation-function analysis. We have addressed the problem of multi-domain proteins where internal flexibility among structural homologues necessitates refinement of the dispositions of the individual domains. When refined simultaneously in *INTREF*, the individual domains have been shown to produce more interpretable translation functions. In addition, the relative translation between domains can be refined, allowing a more complete model to be assembled from its component domains. As expected, a complete refined model gives a translation function with a much higher signal-to-noise ratio than could be obtained with a partial model. The results suggest that the use of *INTREF* may lead to structure solutions in cases where only poor models are available.

INTREF refinement incorporates several distinguishing features which are summarized as follows:

(*a*) Scattering from all the molecules in the cell is included, to the extent that this is possible in the absence of absolute translational information. This is accomplished by adding the *P*1 intensities calculated from each of the groups in the asymmetric unit as well as their symmetry-related mates.

(*b*) The observed and calculated intensities have been modified to effect the equivalent of a radial weighting in Patterson space. This serves to partially offset errors in the calculated intensities resulting from the absence of Patterson vectors between groups, between independent molecules and between symmetry-related molecules.

(*c*) A least-squares procedure is used to refine all the rigid-body parameters simultaneously. This approach is expected to be superior to optimizing only a few parameters at a time, as would be required in a grid search. Another advantage of the least-squares implementation is that reasonable approximations are obtained for the second partial derivatives with respect to the rigid-body parameters. Consequently, convergence is relatively rapid. The computing time required depends primarily on the number of groups and parts in the asymmetric unit, and is relatively independent of the crystal symmetry.

(*d*) The derivatives of the error function are evaluated numerically by making small perturbations of user-defined magnitude in the refinable parameters. If very small perturbations are chosen, the derivatives obtained approach those that would be calculated analytically. When the starting errors in the rigid-body parameters are large, the speed of convergence is improved by evaluating the derivatives with larger perturbations and by damping higher-resolution terms.

INTREF is written in standard Fortran 77 and is available upon request. As currently dimensioned, the program requires 4 Mbytes of memory.

Discussions with Dr D. J. Filman and encouragement from Drs I. A. Wilson and J. M. Hogle are appreciated. We are indebted to Dr J. H. Arevalo, and D. Fremont for assistance in testing the program and to Dr R. Stanfield for supplying the diffraction data and rotation function results used in the B13I2 test case. This work was supported in part by NIH grant AI-20566 to J. M. Hogle and NIH grant AI-23498 to I. A. Wilson. TOY is supported by NIH training grant T32NS07078 and JMR is supported by a fellowship from the Medical Research Council of Canada. This is publication 6018-MB from the Research Institute of Scripps Clinic.

References

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

BRUNGER, A. T. (1988). *X-PLOR* Manual, Version 1.5. Yale Univ., New Haven, USA.

CROWTHER, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173-178. New York: Gordon & Breach.

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

CYGLER, M. & ANDERSON, W. F. (1988*a*). *Acta Cryst.* A**44**, 300-308.

CYGLER, M. & ANDERSON, W. F. (1988*b*). *Acta Cryst.* A**44**, 38-45.

DEREWENDA, Z. S. (1989). *Acta Cryst.* A**45**, 227-234.

EINSPAHR, H., PARKS, E. H., SUGUNA, K., SUBRAMANIAN, E. & SUDDATH, F. L. (1986). *J. Biol. Chem.* **261**, 16518-16527.

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.

HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* A**37**, 398-406.

HARDMAN, K. D., AGARWAL, R. C. & FREISER, M. J. (1982). *J. Mol. Biol.* **157**, 69-86.

HUBER, R. & SCHNEIDER, M. (1985). *J. Appl. Cryst.* **18**, 165-169.

LATTMAN, E. E. (1985). *Methods Enzymol.* **115**, 157-171.

LATTMAN, E. E. & LOVE, W. E. (1972). *Acta Cryst.* B**26**, 1854-1857.

MARQUART, M., DEISENHOFER, J., HUBER, R. & PALM, W. (1980). *J. Mol. Biol.* **141**, 369-391.

RINI, J. M., HARDMAN, K. D., EINSPAHR, H., SUDDATH, F. L. & CARVER, J. P. (1989). *Trans. Am. Crystallogr. Assoc.* Proc. of the Symposium on Molecular Recognition and Carbohydrate-Protein Interactions, Univ. of Washington, Seattle, Washington, USA, 28 July 1989. In the press.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.

SCHERINGER, C. (1963). *Acta Cryst.* **16**, 546-550.

STANFIELD, R. L., FIESER, T. M., LERNER, R. A. & WILSON, I. A. (1990). Submitted.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. (1977). *Acta Cryst.* A33, 800-804.
TEN EYCK, L. F. (1973). *Acta Cryst.* A29, 183-191.
YEATES, T. O., KOMIYA, H., REES, D. C., ALLEN, J. P. & FEHER, G. (1987). *Proc. Natl Acad. Sci. USA*, 84, 6438-6442.
YEATES, T. O. & REES, D. C. (1988). *J. Appl. Cryst.* 21. 925-928.

Acta Cryst. (1990). A46, 359-365

Real-Atom Grid Approximation (RAGA) – a New Technique of Crystal Structure Analysis using only Amplitudes without Determination of Phases of Reflections*

BY G. N. RAMACHANDRAN†

Mathematical Philosophy Group, Indian Institute of Science, Bangalore 560012, *India*

(*Received* 15 *August* 1989; *accepted* 27 *November* 1989)

Abstract

This article deals with the development of a new technique, RAGA (real-atom grid approximation), for crystal structure analysis in the initial and intermediate stages. It is particularly suited to equal-atom structures of non-centrosymmetric crystals, and especially those of the lowest symmetry $P1$, which are the most difficult to solve by conventional methods. The electron-density distribution is approximated by a set of atoms, all having the same form factor, but variable 'masses', m_i, over a grid forming a sublattice of the unit cell. In the associated computer program *RAGA*, the subroutine *GRLS*, for grid least-squares refinement, reduces the R value between the actual F structure and the approximated G structure, thus leading to a continuous sequence of structures with smaller and smaller R values. The quantities used are all in real space, although the refinement makes use of the Fourier transforms of the two structures F and G. It starts with a low resolution of the order of one third of the largest unit-cell dimension with a large temperature factor in order to wipe out intensities of reflections beyond this order of resolution, and proceeds in stages to higher resolutions, reducing the value of B in the process, and this leads to electron-density information at a resolution of twice this order. A two-dimensional example of an equal-atom structure with symmetry $P1$ is given, all atoms of which could be developed starting from a completely flat background as input. RAGA can also be used for the intermediate stages of further refinement in which atoms at unknown atomic sites can be developed using information about the known atomic sites. RAGA thus has the potential to be developed as a valuable additional tool in the armoury of direct methods.

Genesis and principles

In X-ray crystallography, the structure factors $F(hkl) = |F| \exp(i\alpha)$ of the reflections hkl are the Fourier coefficients of the electron-density function $\rho(xyz)$ over the unit cell. Since only the intensities $I(hkl) = |F(hkl)|^2$ are available, the standard methods of proceeding from $I(hkl)$ to $\rho(xyz)$ are based on the determination of the phases $\alpha(hkl)$ by some suitable method, either of an experimental or of a theoretical nature. [For a brief account, see Chapters 3 and 4 of Dunitz (1979) and, for an extensive survey see Schenk, Wilson & Parthasarathy (1987).]. The theory behind the experimental techniques, such as the use of the presence of heavy atoms, of isomorphous crystals and of anomalous-dispersion data, is dealt with from a unified point of view in Ramachandran & Srinivasan (1970). The theoretical techniques behind *ab initio* phase determination for non-centrosymmetric crystals are mostly based on the well known tangent formula, which has been expressed as an algorithm by Main, Lessinger, Woolfson, Germain & Declercq (1977). Fourier techniques and direct methods have been combined by Beurskens *et al.* (1983) in the form of the computer program *DIRDIF*.

It is usually only in the final stages of refinement that a least-squares procedure for improving the fit with intensity data is applied for obtaining more accurate coordinates of the atoms in the structure. Among the theoretical methods, the direct methods are the ones most widely used for the initial stages of phase determination for non-centrosymmetric crystals, particularly for equal-atom structures. They are also the most difficult to solve, particularly in the case of crystal structures belonging to the space group of lowest symmetry, $P1$.

* This paper is Matphil Report No. 76 of the Mathematical Philosophy Group.

† Present address: 5, Gita, 10A Main Road, Malleswaram West, Bangalore 560055, India.

0108-7673/90/050359-07$03.00