Acta Cryst. (1988). A**44**, 38-45

Application of the Molecular Replacement Method to Multidomain Proteins. 1. Determination of the Orientation of an Immunoglobulin Fab Fragment

BY M. CYGLER AND W. F. ANDERSON

MRC Group on Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

(*Received* 15 *April* 1987; *accepted* 11 *August* 1987)

Abstract

Multidomain proteins provide special problems in the application of the molecular replacement method of structure determination. The structure of the Fab fragment from the autoimmune poly(dT)-specific antibody HED10 has been determined using molecular replacement. An analysis of the effects of varying the model and the parameters used in the rotation function indicates that dividing the molecule into individual relatively rigid domains simplifies interpretation of the results, and that the optimal parameters depend on the molecule under study.

Introduction

The continually growing number of proteins whose three-dimensional structures have been determined increases the possibility that a new protein being investigated will have some features in common with one or more of the known structures. In addition, there is growing interest among biochemists and protein crystallographers in determining the changes in protein folding and/or packing caused by specific modifications of the amino-acid sequence. For these cases the application of the standard multiple isomorphous replacement technique (Blundell & Johnson, 1976) to determine phases, while it will give the final answer, may not be the fastest or most straightforward way to achieve this goal. A more direct approach is to utilize information from a closely related protein of known structure by application of the molecular replacement (MR) technique (Rossmann, 1972). If successful, this approach can decrease dramatically the time required to determine the protein structure and can make heavy-atom derivatives unnecessary. While the application of the MR method is computationally intensive, this is no longer an obstacle.

The task of positioning a model molecule in the unit cell involves six degrees of freedom: three to determine the orientation and three to determine the translation of the molecule. From a theoretical analysis of the properties of the Patterson function (Hoppe, 1957; Rossmann & Blow, 1962) it became obvious that such a task can be reduced to two consecutive three-dimensional problems. The first step is the determination of the correct orientation of the model and the second is the determination of the position of the correctly oriented model within the unit cell. The orientation of the molecule is determined by the comparison of the Patterson function of the unknown crystal with that of the model molecule in all possible orientations. The most widely used score function for analyzing the similarity of Patterson functions is the integral of their products over a volume around the origin of the unit cell. Fast algorithms that can be used to calculate this rotation function (RF) in both direct (Huber, 1965; Steigemann, 1974) and reciprocal (Crowther, 1972) space have been developed and have been successfully applied.

Despite many years of experimenting with the RF in a number of laboratories there is no clear understanding of the effect of various factors involved in the calculations on the final success or failure of the method. The current approach is to repeat the calculations many times, varying parameters that are considered important by the investigator.

There are numerous examples of successful applications of the MR method [*e.g.* lysozyme (Bott & Sarma, 1976); insulin (Dodson, Harding, Hodgkin & Rossmann, 1966); hemoglobin (Derewenda, Dodson, Dodson & Brzozowski, 1981); serine proteases (Fujinaga, Read, Sielecki, Ardelt, Laskowski & James, 1982; McPhalen, Svendsen, Jonassen & James, 1985); phospholipase A2 (Dijkstra, van Nes, Kalk, Brandenburg, Hol & Drenth, 1982); immunoglobulin pFc′ fragment (Phizackerley, Wishner, Bryant, Amzel, Lopez de Castro & Poljak, 1979); phycocyanin (Schirmer, Huber, Schneider, Bode, Miller & Hackert, 1986)] and some methodological and practical aspects have been the subject of a special symposium (Daresbury Study Weekend, 1985). The procedure is relatively straightforward in the case of a rigid molecule and success depends primarily on how well the model approximates the unknown protein.

Multidomain or multisubunit proteins that undergo a conformational change upon ligand binding [*e.g.* hemoglobin, hexokinase, arabinose binding protein, citrate synthase, *etc.* (Huber & Bennett, 1983)] or that are relatively flexible [*e.g.* immunoglobulins (Amzel

0108-7673/88/010038-08$03.00

& Poljak, 1979)] pose special difficulties in the application of the MR method. Even if the structure of each domain (or subunit) is unaltered, the changes in their relative orientation greatly influence the Patterson function.

Immunoglobulin molecules are composed of two polypeptide chains. Both types of polypeptide chain consists of structural units, approximately 110 residues long (12·5 kDa), that are connected by flexible segments. All units share a common fold, a two-layer β-sheet sandwich with the same pattern of connectivity between β strands (Amzel & Poljak, 1979). An Fab fragment is composed of a light (L) chain, which consists of two units, and the first two units of the heavy (H) chain (separated proteolytically from the total of four in an IgG). The N-terminal units of both chains associate into the variable (V) domain containing the antibody recognition site (antigen binding site) while the other two units form the constant (C) domain. The units are designated by the symbols V_L and C_L for the light chain and V_H and C_H1 for the first two units of heavy chain. The structures of four Fab fragments have been reported in the literature and are all well refined [McPC603 (Segal, Padlan, Cohen, Rudikoff, Potter & Davies, 1974); NEW (Saul, Amzel & Poljak, 1978); KOL (Marquart, Deisenhofer, Huber & Palm, 1980); J539 (Suh, Bhat, Naira, Cohen, Rao, Rudikoff & Davies, 1986)].

The structure of the anti-poly(dT) immunoglobulin Fab fragment HED10 was solved by the MR method (Cygler, Boodhoo, Lee & Anderson, 1987). Since we encountered some difficulties in the process and because it provides an example of a flexible multi-domain protein, a detailed *post factum* analysis of the results has been undertaken.

HED10 monoclonal antibody is produced by a hydridoma cell line derived from autoimmune NZB/NZW mice (Lee, Lewis, Morgan, Mosmann & Singh, 1981) and is specific for single-stranded DNA with a preference for poly(dT) or poly(dB) sequences (Lee, Dorbroski & Mosmann 1982). It is an IgG2a with a κ-type light chain. The Fab fragment from this IgG crystallizes in the monoclinic space group $P2_1$ with cell dimensions $a = 64·3$, $b = 90·4$, $c = 42·6$ Å, $\beta = 97·1°$. The crystals diffract to approximately 2·2 Å resolution.

Results

Selection of a model structure

The analysis of structural data for Fab fragments (Amzel & Poljak, 1979) indicates the large degree of flexibility of the molecule at the V/C domain junction. Both domains possess approximate twofold symmetry. The values of the rotation angles that superimpose part of the heavy chain (V_H or C_H1) onto the corresponding part of the light chain (V_L or C_L), together with the translational screw components along the rotation axis, are given in Table 1. The data for the HED10 Fab fragment are based on the model refined to $R = 24·0\%$ at 2·4 Å resolution (Cygler, Boodhoo, Lee & Anderson, 1987 and unpublished data). Since these parameters are derived from the positions of all α-C atoms they should not change much upon further refinement.

Table 1. *Comparison of Fab fragments*

	V domain Rotation of V_H to V_L		C domain Rotation of C_H to C_L		Elbow angle*
Fab	ω(°)	t (Å)†	ω(°)	t (Å)	(°)
KOL	167·7	0·02	170·5	3·29	165·6
NEW	164·1	0·35	171·6	3·61	130·6
McPC603	173·0	0·20	170·7	2·66	132·0
J539	169·7	0·03	174·2	2·03	143·1
HED10	175·9	0·30	169·9	1·82	160·3

* Angle between pseudo-twofold axes of V and C domains.

† Translation component along the rotation axis. The axis points in the same direction relative to the domains in all compared molecules and the sign of the translational component is the same in all compared cases.

The rotation angle for the C domain is very similar in all four Fab fragments while the translational component changes by more than 2 Å, ranging from 1·4 to 3·7 Å. The V domain, on the other hand, has a very small translation component but the rotation angle differs by as much as 8°. The angle between the approximate twofold axes of V and C domains (elbow angle) varies from 130 to 166°, demonstrating the level of molecular flexibility.

Any one of the known Fab structures could be taken as the starting model. The amino-acid sequence homology with HED10 is expected to be somewhat higher for Fab McPC603 [mouse IgA(κ)] than for NEW or KOL [human IgG(λ)] (Kabat, Wu, Bilofsky, Reid-Miler & Perry, 1983). This makes Fab McPC603 the first choice. The strategy was to look for a peak in the RF which was well above the mean value of the map and significantly higher than the next highest one. When the results were not convincing other available Fab models, and later the V or C domains, were tried. The results of the calculations and a detailed analysis are presented for the Fab McPC603 molecule taken as a model.

Analysis of the RF results

A series of calculations have been performed in order to determine the influence of parameters such as: (*a*) the resolution range of the data used; (*b*) the outer integration radius for the Patterson function; (*c*) the number of strongest reflections used; and (*d*) which atoms of the model were included in the calculations.

The calculations described here were conducted with Crowther's fast rotation function. The version of the program available to us used Bessel functions up to the 30th order (Dodson, 1979), restricting the

Table 2. *Rotation-function peaks (as multiples of the r.m.s. of the map) with Fab McPC603 as a model*

Resolution (Å)	Integration radii	Number of reflections Crystal	Model	V-domain peak*		C-domain peak*		First spurious peak
All atoms included in the model								
10-3·5	4-18	2422	2406	3·59	1	—	>15	3·54
10-3·5	4-21	2422	2406	4·09	1	3·47	4	3·71
10-4·0	4-24	1606	2019	4·73	1	4·23	2	4·13
10-4·5	4-27	971	1409	5·16	1	4·72	2	4·20
10-5·0	4-27	578	1049	4·40	3	5·87	1	4·60
10-5·0	4-30	578	1049	4·46	3	6·04	1	4·90
Backbone and β-C atoms only								
10-3·5	4-21	2422	1618	3·41	4	3·87	2	3·98
10-4·0	4-24	1606	1363	—	>10	4·21	1	4·05
10-4·5	4-27	971	1402	—	>15	4·02	1	3·91
α-C atoms only								
10-3·5	4-21	2422	2473	3·87	7	3·97	4	4·30
10-4·0	4-24	2422	2000	3·66	9	4·10	3	4·20
10-4·5	4-27	971	1116	3·73	10	3·90	4	4·07
Sum of Patterson functions of V and C domains								
10-4·0	4-24	1606	1023	7·31	1			4·23

* With sequential number of the peak in the list of rotation-function peaks sorted in descending order.

ratio of the outer integration radius to the maximum resolution to be less than or equal to 6. The program also limited the number of reflections to 2500. The model molecule was positioned in an orthogonal unit cell with $P1$ symmetry and with the length of the edges being chosen as the maximum dimension of the molecule in the particular direction, extended by the maximum integration radius used in the calculations. In this way no intermolecular vectors were present in the model Patterson function within the integration radius from the origin.

Complete Fab fragment as a model

When the intact Fab McPC603 was used as a model there were, in general, two strong peaks present (Table 2). In retrospect, they correspond to the correct orientation of V and C domains respectively. The best results were obtained with the data in the 10 to 4·5 Å resolution shell and an outer integration radius of 27 Å. From the data presented in Table 2 it is obvious that higher-resolution data contribute more to the peak corresponding to the variable domain, while the lower-resolution data contribute more to the peak for the constant domain (see also Tables 3 and 4). The value of the outer integration radius is very important. Too small a value (relative to the protein size) causes the solution to become indistinguishable from spurious peaks. Another important finding is that when data up to 3·5 Å resolution are included, the peak corresponding to the C domain decreases to the size of spurious peaks, with only one peak (for the V domain) that stands out from the rest. This could lead to the erroneous conclusion that HED10 and 603 have very similar overall shapes and approximately the same elbow angle.

The influence of the completeness of the model was also investigated. First, all side-chain atoms beyond the β carbon were removed from the model. This had a dramatic effect on the RF peaks (Table 2). Only the peak corresponding to the C domain was among the highest peaks of the map and the discrimination between the correct and spurious peaks was very poor. The V-domain peak is not among the ten strongest peaks. In the extreme case of including only α-C atoms in the model there are still local maxima corresponding to the correct rotation angles. However, the spurious peaks are higher (Table 2) and it would be impossible to interpret these results correctly without prior knowledge.

In summary, the results of the above searches could be somewhat confusing. There seem to be one or two main peaks. While the interpretation of the latter case might be that they correspond to the correct orientation of V and C domains, the assignment cannot be made unequivocally. The results obtained with Fab KOL and Fab NEW as starting models were even more difficult to interpret as there were more peaks of similar height.

Fab individual domains as models

Experience from other laboratories with the MR method shows that a model comprising half of the asymmetric unit contents may be enough to obtain the correct solution (Derewenda, Dodson, Dodson & Brzozowski, 1981; Wang, Bode & Huber, 1985). Since the results of the RF, with the entire Fab fragment as a model, were not easily interpretable the calculations were repeated with individual domains. Table 3 presents the results of the RF calculations using the C domain as a model. The results are

Table 3. *Rotation-function peaks (as multiples of r.m.s. of the map) with C domain of Fab McPC603 as a model*

Resolution (Å)	Integration radii	Number of reflections* Crystal	Model	Correct peak†		First spurious peak
All atoms included						
10–3·5	4–21	2422	2297	5·79	1	4·06
10–4·0	4–21	1606	1863	5·66	1	4·81
10–4·0	4–24	1606	1863	6·08	1	4·45
15–4·0	4–24	1685	2180	5·59	1	4·43
10–4·5	4–27	971	1320	5·55	1	5·02
10–4·0	4–24	2015	2222	6·02	1	4·54
10–4·0	4–24	1606	1863	6·08	1	4·45
10–4·0	4–24	1272	994	5·98	1	4·77
10–4·0	4–24	518	517	6·26	1	4·93
10–4·0	4–24	361	276	6·11	1	4·67
4–3·5	4–21	816	434	—	>15	4·04
5–4·0	4–24	1028	1001	3·87	6	4·29
7–5·0	4–24	423	559	4·13	1	3·65
10–7·0	4–24	155	303	4·17	2	4·23
20–10	4–24	94	422	—	>15	3·09
		B crystal	*B* model			
10–4·0	4–24	−10	0	5·98	1	4·82
10–4·0	4–24	−30	−10	5·98	1	4·77
10–4·0	4–24	−50	−30	5·28	1	4·80
α-C only						
10–3·5	4–21	2422	2410	5·21	1	4·37
10–4·0	4–24	1606	1737	5·27	1	4·57
10–4·5	4–27	971	1051	4·85	1	4·01
10–5·0	4–24	578	577	3·41	3	3·46
10–5·0	4–30	578	577	4·00	3	4·11

* The number of reflections used was changed by altering the minimum F used.
† With sequential number of the peak in the list of rotation-function peaks sorted in descending order.

extremely clear. There is only one strong peak, about $1{\cdot}5\sigma$ above the highest spurious peak. The number of included reflections seems to make very little difference in the case of a good model and a very few of the strongest reflections are enough to get the correct answer. Analysis of the results as a function of resolution shows that there is little contribution to the peak from the higher-resolution shell, 4–3·5 Å, and that the data between 10 and 4 Å contribute almost equally. In only one shell is the correct peak the highest one. On the whole, it is, however, the only one that is consistent among the various shells. Finally, the dependence on the additional temperature factor (B), which has the effect of sharpening the data, shows that modest values are the best. A very high negative B value stresses the high-resolution reflections, where there is little useful information in the rotation-function sense, and lowers the peak height.

To examine how incomplete a good model can be we have tested the extreme situation where only the α-C atoms of the C domain were used in the calculations. The correct orientation shows up unequivocally (Table 3) and the peak in the RF is nearly 1σ larger than the first spurious peak. The information, however, comes mainly from the intermediate resolution and the peak diminishes to the size of the spurious ones for 10–5 Å resolution. It is worth noting that α-C atoms form about 12% of all the atoms of the C domain, or only 6% of the atoms in the asymmetric unit.

Similar calculations have been done for the V domain of Fab McPC603 and the results are presented in Table 4. The residues of the complementarity-determining regions (CDR) were not included in the model. In this case the best resolution is 15–4 Å and the best choice for the integration radius is 24 Å. The highest peak was $0{\cdot}5\sigma$ to 1σ above the next one and the results were easy to interpret. Increasing the integration radius results in lowering the height of the peak, but this is partly due to limiting resolution to 4·5 Å.

Contrary to the results for the C domain, including more reflections increased the discrimination ratio of the correct solution. When, however, too many weak reflections were included the discrimination began to decrease (data not shown).

The resolution dependence shows that most of the contribution to the peak came from the 5–4 Å shell and neither lower- nor higher-resolution data contributed much.

The influence of the CDR regions of the model on the peak height was also checked. The results (Table 4) showed that it was better to include them, though the effect was not very pronounced. As was the case for the C domain, the V-domain α-C atoms are sufficient to obtain the correct solution.

Table 4. *Rotation-function peaks (as multiples of r.m.s. of the map) with V domain of Fab McPC603 as a model*

Resolution (Å)	Integration radii	Number of reflections* Crystal	Model	Correct peak†		First spurious peak
All atoms included, except for CDR regions						
10-3·5	4-21	2422	2120	4·21	1	4·08
10-4·0	4-21	1606	1695	5·11	1	4·76
10-4·0	4-24	1606	1695	5·48	1	4·70
15-4·0	4-24	1685	1950	6·44	1	5·28
10-4·0	4-24	1607	1695	5·48	1	4·70
10-4·0	4-24	1085	1060	5·08	1	4·65
10-4·0	4-24	537	559	5·00	1	4·80
10-4·0	4-24	266	285	4·76	1	4·46
4-3·5	4-21	816	425	1·6	>15	3·92
5-4·0	4-24	1028	1020	5·16	1	4·70
7-5·0	4-24	423	443	2·25	>15	4·35
10-7·0	4-24	155	232	1·7	>15	3·18
20-10	4-24	94	355	1·8	>15	3·13
α-C only						
10-3·5	4-21	2422	2135	—	>15	4·04
10-4·0	4-24	1606	1905	3·79	5	4·41
10-4·5	4-27	971	1121	4·25	1	4·15
15-4·5	4-27	1050	1351	5·31	1	4·72
10-5·0	4-30	578	598	4·16	11	5·50
All atoms with CDR regions included						
10-3·5	4-21	2422	2077	3·95	1	3·86
10-4·0	4-24	1606	2105	5·27	1	4·49
10-4·5	4-27	971	1629	4·99	1	4·40

* The number of reflections used was changed by altering the minimum F used.
† With sequential number of the peak in the list of rotation-function peaks sorted in descending order.

Table 5. *Results of rotation searches with individual units of Fab McPC603 (peak heights given as multiples of r.m.s. of the map)*

Resolution (Å)	Integration radii	Number of reflections* Crystal	Model	Correct peak†		First spurious peak
V_L domain						
15-4·0	4-24	1685	2127	3·52	6	4·55
15-4·0	6-24	1685	2127	3·26	4	4·07
V_H domain						
10-4·0	4-21	1606	1553	3·59	2	3·69
15-4·0	4-24	1685	1757	4·21	1	3·74
15-4·0	6-24	1685	1757	4·72	1	3·81
C_L domain						
15-3·0	4-18	2324	2357	4·53	1	3·90
15-4·0	4-18	1685	2263	4·39	1	4·18
15-4·0	4-15	1685	2263	3·86	3	3·99
15-4·0	4-24	1685	2263	3·42	7	3·87
C_H1 domain						
15-4·0	4-21	1685	1799	3·01	5	3·41
15-4·0	6-21	1685	1799	3·24	4	3·48
15-4·0	4-24	1685	1799	3·09	7	3·61

* The number of reflections used was changed by altering the minimum F used.
† With sequential number of the peak in the list of rotation-function peaks sorted in descending order.

Individual units

The correct determination of the orientation of a model containing only the α-C atoms of one domain leads to the question of how small a part of the molecule can be used successfully as a model. Schirmer *et al.* (1986) were able to analyze properly the RF based on a model containing only one-third of the asymmetric unit. The correct analysis was, however, possible only because some additional information about the orientation of the molecule was available. We have performed calculations with the individual units (one-fourth of the asymmetric unit) as models and have found that the expected peaks appeared in the RF map (Table 5). In all cases the best results were obtained with data from a broad resolution range, 15-4 Å (15-3 Å for C_L). Only the V_H and C_L units gave, under appropriate conditions,

Table 6. *Results of rotation searches with the V domain of Fab KOL (peak heights given as multiples of r.m.s. of the map)*

Resolution (Å)	Integration radii	Number of reflections Crystal	Model	Correct peak*		First spurious peak
Unmodified						
15-4·0	4-24	1102	1010	4·09	5	4·20
10-4·0	4-24	1606	1888	—	>15	4·07
Modified†						
15-4·0	4-24	1102	927	4·94	1	4·49
10-4·0	4-24	1606	1916	3·26	9	4·14

* With sequential number of the peak in the list of rotation-function peaks sorted in descending order.
† See text for description of modification of the model.

highest peaks in the map which exceeded the spurious peaks by $0{\cdot}7\sigma$.

Fab KOL used as a model

When the V and C domains of Fab KOL were used as models in the RF calculations the results were somewhat surprising. The correct orientation could easily be interpreted for the C domain, although the RF peak was less pronounced than in the case of Fab McPC603 (data not shown). The V domain, on the other hand, when used as a model did not reveal the correct orientation. Under the best set of conditions the appropriate peak was the fifth highest in the map (Table 6).

Since the backbone conformations of the V_L and V_H units in McPC603 and KOL are very similar (r.m.s. for α-C atoms excluding CDR regions are 1·60 and 0·73 respectively), there are two possible explanations of the observed differences in the RF results. One is that the positions of side-chain atoms may differ significantly. Another is that the association of V_H and V_L is significantly different in these two molecules (there is a 5·5° difference in the rotation along the pseudo-twofold axis; see Table 1).

To determine which of these two possibilities (or both) takes place we have first calculated the RF using only the V_H or the V_L domain of Fab KOL as a model. The V_H domain gave the expected peak (best conditions are: resolution 15-4 Å, integration radius 24 Å, approximately 800 strong reflections only), while the V_L unit did not. These results are similar to the ones obtained for Fab McPC603.

To investigate the second possibility a modified Fab KOL V-domain model was created by transforming V_H and V_L units separately onto the corresponding units of Fab McPC603. In this modified model the V_L/V_H association was exactly the same as in Fab McPC603. Under favorable conditions (Table 6) the expected peak appears as the highest one in the RF map. However, it is not as pronounced as the McPC603 peak. Therefore, both the positions of side-chain atoms and the variation of the association of V_L and V_H affect the RF results.

Comparison of the Patterson function of the crystal with the sum of the Patterson functions of the V and C domains

The Patterson function of two rigid domains contains two categories of maxima: those corresponding to vectors between atoms of the same domain and those corresponding to vectors between atoms from different domains. If the domain orientations are known, but their relative translation is not, the best approximation to the Patterson function of the intact molecule is the sum of the Patterson functions corresponding to each of the two domains. The superposition of the sum of the Patterson functions with the unknown Patterson function should give better results than the Patterson function of each domain separately. If the expected hinge motion between the two domains is not very large, one domain can be rotated relative to the other in small steps and the appropriate RF can be calculated, using for F

$$F_c = (F_{c1}^2 + F_{c2}^2)^{1/2}$$

where F_{c1} corresponds to the first domain and does not have to be recalculated, and F_{c2} corresponds to the second domain and has to be recalculated after each small rotation. The results of such calculations are shown in Table 2. For the correct relative orientation of the V and C domains of Fab McPC603 the correct peak in the RF was much higher than for each separate domain and was almost 3σ above the spurious peaks. This method was also applied to Fab KOL, but without much success. We attribute this to the fact that the V domain of Fab KOL does not give the expected peak in the RF (see above).

Discussion

The protein that we have used in the tests described in this paper had two features that were of importance for the application of the MR method. First, it was a two-domain molecule displaying large flexibility in the domain association. The relative orientation of the two domains could differ by as much as 30-40°. Second, it was an all-β-structure protein, which may be a more difficult case for the MR method. The Fab

fragments that were used as models in the process were expected to show different degrees of amino-acid homology to HED10. Although only part of the amino-acid sequence of the latter has been determined, the compilation of sequence data on immunoglobulins (Kabat *et al.*, 1983) suggested that there was at least 70% homology between the C domains of McPC603 and HED10 and the homology to NEW or KOL was expected to be lower. The homology of the V domain of HED10 to McPC603, NEW and KOL is expected to be relatively low.

Our experience with the RF clearly indicates that the utilization of a rigid domain as a starting model rather than a whole molecule makes the interpretation of the results easier. When the model closely resembles the molecule in the crystal, just the α-C atoms of one domain (Tables 3 and 4) contain enough structural information to determine the correct solution. This observation was also reported by Schierbeek, Renetseder, Dijkstra & Hol (1985) in their test calculations where the actinidin molecule was used as a model for papain. The α-C atoms form approximately 13% of the non-hydrogen atoms of actinidin. In that case, leaving a quarter of the α-C atoms out of the calculations was enough to lose the correct solution among the spurious peaks. Results with individual domains show that even as little as one quarter of the molecule (V_H or C_L unit) can provide the highest peak in the map.

The two most important parameters in the RF for a particular model are the resolution range of the data and the outer integration radius. The positions and heights of the RF peaks are much more dependent on these parameters in the case of a poor (Table 2) or incomplete (Table 6) model than in the case of a good model (Tables 3 and 4). From literature reports on the application of the MR method the resolution range proven to be successful in different cases varied greatly. In some cases a narrow resolution range was used [*e.g.* 4·5–3·7 Å for lactate dehydrogenase (Musick & Rossmann, 1979); 3·0–2·3 Å for γ-IV crystallin (Driessen & White, 1985), while others have used a wider range (Dijkstra *et al.*, 1982; Derewenda, Dodson, Dodson & Brzozowski, 1981). To a large degree the best resolution range depends on the quality of the starting model. In our case neither high- (above 4 Å) nor low- (below 10–15 Å) resolution data were extremely useful (Tables 3 and 4) and the contribution from narrow shells depended on the starting model. For a good model (the C domain) the whole 10–4 Å range was useful. For a somewhat poorer model (the V domain) the 5–4 Å range contained the most useful information. The results for pancreatic phospholipase A2 (Dijkstra *et al.*, 1982) demonstrated the importance of data in the 6–4 Å resolution range. Interestingly, the best results for individual units as well as for the V domain of KOL (modified and unmodified) were obtained with the 15–4 Å resolution range. Omitting the 15–10 Å range had a disastrous effect on the height of the correct peaks.

The best integration radius depends on the size and shape of the molecule and the packing in the crystal. When many close contacts exist between neighboring molecules, the RF is very sensitive to the choice of the integration radius and a relatively small value has to be chosen [*e.g.* 13 Å for despentapeptide insulin (Bi Ru-chang, Cutfield, Dodson, Dodson, Reynolds & Tolley, 1983)]. Normally a value representing 50–70% of the average dimension of the model molecule seems to be a satisfactory choice. Too small a value may cause the correct peak to disappear among spurious peaks (Table 2; Schierbeek, Renetseder, Dijkstra & Hol, 1985). The differences in the optimal value of the integration radius observed here (Table 5) for the V_H and C_L units may be rationalized on the basis of tighter association of the units in the C than the V domain.

It should be emphasized that the inherent limitation of the version of the RF program used in this study restricts the ratio of the integration radius to the high-resolution limit to be less than or equal to 6·0. Consequently we were unable to explore fully the influence of larger integration radii on the RF peaks. For the 4·0 Å resolution data the integration radius was thus limited to 24 Å or less. Calculations with larger integration radii (27 and 30 Å) had to be performed with altered high-resolution limits on the data (4·5 and 5·0 Å respectively). Since the 4–5 Å resolution shell was very important for the RF results, the observed changes are the sum of two effects that are difficult to separate. Because both resolution and integration radius are crucial for the RF results, an appropriate compromise of the resolution range and the maximum integration radius has to be chosen. This will depend on the particular problem at hand. The combination of 4 Å resolution with a 24 Å integration radius is a good starting point for proteins with molecular weights between 20 and 100 kDa.

The number of reflections included in the calculations was not as important as the above-mentioned parameters. Two observations should, however, be mentioned here. For a good model (C domain) the results using as few as 300 reflections are as good as they are for 2000 reflections. For a poorer model (V domain), including more reflections is beneficial unless too many weak reflections are added and the noise level of the map is increased.

A general conclusion from our results and those of others is that although the correct peak may not be the highest feature in the RF calculated with different parameters, it appears most consistently on all the maps.

We would like to conclude with a note of caution in the interpretation of the rotation-function results and stress the importance of repeating the calculations with various sets of parameters. It is evident

that, in the case of flexible molecules, much clearer results are obtained by using a smaller but correct fragment, rather than a larger one that may differ significantly in a global sense from the unknown protein.

This work was supported by the Medical Research Council of Canada through the Group on Protein Structure and Function. The authors thank Dr D. R. Davies for providing them with coordinates of the J539 Fab fragment.

References

AMZEL, L. M. & POLJAK, R. J. (1979). *Ann. Rev. Biochem.* **48**, 961-997.

BI RU-CHANG, CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., REYNOLDS, C. D. & TOLLEY, S. P. (1983). *Acta Cryst.* B**39**, 90-98.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

BOTT, R. & SARMA, R. (1976). *J. Mol. Biol.* **106**, 1037-1046.

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

CYGLER, M., BOODHOO, A., LEE, J. S. & ANDERSON, W. F. (1987). *J. Biol. Chem.* **262**, 643-648.

DEREWENDA, Z. S., DODSON, E. J., DODSON, G. G. & BRZOZOWSKI, A. M. (1981). *Acta Cryst.* A**37**, 407-413.

DIJKSTRA, B. W., VAN NES, G. J. H., KALK, K. H., BRANDENBURG, N. P., HOL, W. G. J. & DRENTH, J. (1982). *Acta Cryst.* B**38**, 793-799.

DODSON, E. (1979). Fast rotation function program. Univ. of York, England.

DODSON, E., HARDING, M. M., HODGKIN, D. C. & ROSSMANN, M. G. (1966). *J. Mol. Biol.* **16**, 227-241.

DRIESSEN, H. P. C. & WHITE, H. (1985). *Molecular Replacement.* Proceedings of the Daresbury Study Weekend, 15-16 February 1985, edited by P. A. MACHIN, pp. 27-32. Warrington: SERC Daresbury Laboratory.

FUJINAGA, M., READ, R. J., SIELECKI, A., ARDELT, W., LASKOWSKI, M. JR & JAMES, M. N. G. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4868-4872.

HOPPE, W. (1957). *Acta Cryst.* **10**, 750-751.

HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.

HUBER, R. & BENNETT, W. S. JR (1983). *Biopolymers*, **22**, 261-279.

KABAT, E. A., WU, T. T., BILOFSKY, H., REID-MILER, M. & PERRY, H. (1983). *Sequences of Proteins of Immunological Interest.* US Department of Health and Human Services, Public Health Service, National Institute of Health.

LEE, J. S., DOMBROSKI, D. F. & MOSMANN, T. R. (1982). *Biochemistry* **21**, 4940-4945.

LEE, J. S., LEWIS, J. R., MORGAN, A. R., MOSMANN, T. R. & SINGH, B. (1981). *Nucleic Acids Res.* **9**, 1707-1721.

MCPHALEN, C. A., SVENDSEN, I., JONASSEN, I. & JAMES, M. N. G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7242-7246.

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

MUSICK, W. D. L. & ROSSMANN, M. G. (1979). *J. Biol. Chem.* **254**, 7611-7620.

PHIZACKERLEY, R. P., WISHNER, B. C., BRYANT, S. H., AMZEL, L. M., LOPEZ DE CASTRO, J. A. & POLJAK, R. J. (1979). *Mol. Immunol.* **16**, 841-850.

ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon and Breach.

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

SAUL, F., AMZEL, L. M. & POLJAK, R. J. (1978). *J. Mol. Biol.* **253**, 585-97.

SCHIERBEEK, A. J., RENETSEDER, R., DIJKSTRA, B. W. & HOL, W. (1985). *Molecular Replacement.* Proceedings of the Daresbury Study Weekend, 15-16 February 1985, edited by P. A. MACHIN, pp. 16-21. Warrington: SERC Daresbury Laboratory.

SCHIRMER, T., HUBER, R., SCHNEIDER, M., BODE, W., MILLER, M. & HACKERT, M. L. (1986). *J. Mol. Biol.* **188**, 651-676.

SEGAL, D. M., PADLAN, E. A., COHEN, G. H., RUDIKOFF S., POTTER, M. & DAVIES, D. R. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 4298-4302.

STEIGEMANN, W. (1974). PhD Thesis, Technische Univ. Munich, Federal Republic of Germany.

SUH, S. W., BHAT, T. N., NAIRA, M. A., COHEN, G. H., RAO, D. N., RUDIKOFF, S. & DAVIES, D. R. (1986). *Proteins Struct. Funct. Genet.* **1**, 74-80.

WANG, D., BODE, W. & HUBER, R. Y. (1985). *J. Mol. Biol.* **185**, 595-624.

Acta Cryst. (1988). A**44**, 45-51

On Integrating the Techniques of Direct Methods and Isomorphous Replacement. A New Probabilistic Formula for Triplet Invariants

BY C. GIACOVAZZO, G. CASCARANO AND ZHENG CHAO-DE*

Centro Interdipartimentale di Cristallografia, c/o Dipartimento Geomineralogico, Campus Universitario, Via Amendola, 70124 *Bari, Italy*

(*Received* 22 *April* 1987; *accepted* 11 *August* 1987)

Abstract

The mathematical technique recently used [Hauptman (1982). *Acta Cryst.* A**38**, 289-294] for integrating direct methods and isomorphous replacement techniques is reconsidered. The atomic positions are assumed to be the primitive random variables instead of the reciprocal vectors. A new probabilistic formula for estimating three-phase invariants given six magnitudes has been obtained which differs from the corresponding result of Hauptman. The first applications

* Permanent address: Institute of Physics, Academia Sinica, Beijing, China.

0108-7673/88/010045-07$03.00