The molecular-replacement solution of an intermediate-sized helical polypeptide, antiamoebin I

C. F. Snook† and B. A. Wallace*

Department of Crystallography, Birkbeck College, University of London, London WC1E 7HX, England

† Present address: Department of Molecular Physiology and Biological Physics, University of Virginia Health Sciences Center, Charlottesville, VA 22906–0011, USA.

Correspondence e-mail: ubcg91c@ccs.bbk.ac.uk

The successful use of molecular-replacement methods for the solution of the intermediate-sized helical polypeptide anti-amoebin I required the careful consideration of a number of parameters and exhibited some unusual characteristics when compared with molecular-replacement solutions of globular proteins. High-resolution data were required owing to several features, including the comma-like shape of the molecule (which results in a pseudo-symmetric structure at low resolution), the relative uniformity of the structure in the direction along the helix axis and the small differences between the two independent molecules in the $P1$ asymmetric unit. Other parameters which were important for the solution of this relatively low solvent content closely packed cell included the radius of integration, the use of normalized structure factors and especially the choice of starting model.

1. Introduction

There have been a number of reports in the recent literature of the application of the molecular-replacement (MR) technique to increasingly difficult problems (see, for example, Turkenburg & Dodson, 1996). These papers fall into three categories: (i) combination of MR with another technique such as multiple isomorphous replacement (Brady *et al.*, 1991; Baker *et al.*, 1995; Sicheri & Yang, 1996) or with the incorporation of known structural information (Tickle & Driessen, 1996; Vitali *et al.*, 1996), (ii) the use of increasingly smaller fragments as models (Bernstein & Hol, 1997 and references therein) and (iii) the use of MR to solve large protein complexes at low resolution (Urzhumtsev & Podjarny, 1995). Recently, the information on structural repeats in B-DNA helices has been incorporated into the use of *AMoRe* for solution of a decameric nucleic acid structure (Baikalov & Dickerson, 1998). One class of structure which has proven to be particularly difficult to solve using MR has been protein structures consisting of a single helical motif, *i.e.* the antifreeze protein from the winter flounder (Sicheri & Yang, 1996). All of the above MR studies were performed on relatively large polypeptides or proteins. For small peptides which produce very high resolution data, direct-method procedures such as *SHELXS* (Sheldrick, 1986) or *SnB* (Miller *et al.*, 1994) have been most effective for structure solution. However, for intermediate-sized peptides ($\sim$10–30 residues), which may crystallize with more than one monomer in the unit cell and produce relatively low resolution data (for small molecules), direct methods are less likely to succeed or else involve protracted solutions. The use of molecular replacement in these cases can prove most fruitful when a reasonable starting model is available. Recently, a number of intermediate-sized

peptide structures have been solved using molecular replacement, including endothelin using a partial NMR model (Janes *et al.*, 1994), a gramicidin/potassium thiocyanate complex using a gramicidin/caesium chloride complex (Doyle & Wallace, 1997), a 24-mer peptitergent using a 24-mer polyalanine α-helix (Schafmeister *et al.*, 1993) and, most recently, two different structures of antiamoebin I (AAM) from *Emercelliopsis poonensis*, both using Leu1–zervamicin as the starting model (Karle *et al.*, 1998; Snook *et al.*, 1998). This paper will concentrate on the reciprocal-space techniques used for the AAM structure determined by Snook *et al.* (1998) and will examine some of the MR parameters important for its solution.

First, a general description of the structure, so that the nature of the problem can be seen: AAM is a member of the peptaibol family of polypeptides, which are characterized by the presence of a number of α-amino isobutyric acid residues (which tend to stabilize helices) and end in a C-terminal alcohol group rather than a carboxylic acid (Sansom, 1993; Whitmore *et al.*, 1997). In crystals grown from methanol, the unit cell contains two independent molecules of AAM, which are related by an approximate twofold rotation along the c axis. The molecules are very similar to each other, differing by only a root-mean-square deviation (r.m.s.d.) of 0.415 Å. Each molecule is essentially entirely helical in structure, with residues 1–9 forming an α-helix, residues 10–12 forming a 3_{10}-helix and residues 12–16 forming two overlapping type I β-turns (which are like a 3_{10}-helix with breaks in the hydrogen-bonding patterns at the two imino acids at positions 13 and 15). Owing to the presence of a third imino acid at position 10, there is a bend in the middle of the molecule, which results in it having an overall comma-like shape and appearing at low resolution to be symmetric about the centre of the molecule.

The use of the molecular-replacement technique for this problem was prompted by a number of factors. The crystals are densely packed (29% solvent content; Oliva, 1988), suggesting that soaking in heavy-atom solutions would be problematical. The introduction of any additives to the crystallization solution significantly reduced the size of crystals which could be grown, rendering the use of co-crystallization with heavy atoms useless. Previous attempts to solve the structure using MR techniques with a model helical structure (Oliva, 1988) had been unsuccessful. Attempts to solve the structure (Snook & Wallace, unpublished results) using the direct-methods programs *SHELXS* (Sheldrick, 1986) and *SnB* (Miller *et al.*, 1994), carried out in parallel with the MR studies, failed to find a solution. This was probably because of the large number of non-H atoms in the unit cell and the relatively low resolution of the data available (even though the data were 93.7% complete over the resolution range 25.0–1.21 Å, these are marginal conditions for solution of a molecule of this size by direct methods). This paper presents the parameters which had to be considered in order to achieve a successful MR solution for the structure of the peptide antiamoebin I. Details of the refinement, the structure and its functional characterizations have recently been published (Snook *et al.*, 1998; Duclohier *et al.*, 1998) and will not be repeated here.

Table 1
Data-collection statistics.

Values for the outer resolution shell (1.27–1.21 Å) are shown in parentheses.

Space group	$P1$
a (Å)	26.53
b (Å)	28.82
c (Å)	9.06
α (°)	88.8
β (°)	96.7
γ (°)	123.9
Molecules per asymmetric unit	2
Resolution range (Å)	25.0–1.21
Total reflections	7009
Unique reflections	6715
Multiplicity	1
R_{merge} (F) (%)	1.2
Mean $F/\sigma F$	13.53
Reflections with $F > 3\sigma$ (%)	86.5 (75.4)
Completeness (%)	93.7 (87.4)

2. The data and parameters to be tested

The data used (Oliva, 1988; Snook *et al.*, 1998) were collected at room temperature on a diffractometer using Ni-filtered Cu $K\alpha$ radiation. Data-collection statistics are listed in Table 1. The solvent content was calculated to be 29%, corresponding to a Matthews coefficient (V_m; Matthews, 1968) of 1.71 Å^3 Da^{-1}.

It was necessary to consider in detail the parameters used in both the rotation- and translation-function solutions. The size and characteristics of the model used, the radius of integration, the data-resolution limits, the effects of using structure factors or normalized structure factors and the role of noise in the maps, amongst other parameters, were all investigated.

In this paper, we define the signal-to-noise ratio as the height of the peak S above the mean map height with the symbol S/σ; according to this definition, the noise level of the map will always be 1.0. The peak-height difference ($\Delta S/\sigma$) will be used to compare the difference between the signal-to-noise ratios of two peaks, as suggested by Tickle (1992) for the analysis of translation-function solutions, and will be defined as

$$\Delta S/\sigma = (S_1/\sigma) - (S_2/\sigma) = (S_1 - S_2)/\sigma.$$

The peak-height difference is calculated as the difference between the highest (S_1) and the next highest (S_2) peaks. The only exception is when a known solution occurs further down the list of peaks. In that case, $\Delta S/\sigma$ is calculated so that S_1 is the peak of interest and S_2 is the maximum peak, not the next peak, and will result in a negative $\Delta S/\sigma$ value.

3. Self-rotation function

The self-rotation function calculations were carried out using the *CCP*4 versions of both *AMoRe* (Navaza, 1994) and *POLARRFN* (Collaborative Computational Project, Number 4, 1994), with both structure factors and normalized structure factors over a wide range of high-resolution limits. The low-resolution limit was set to 10.0 Å and the high-resolution limit was varied from 3.5 to 1.2 Å.

Table 2
Self-rotation functions using structure factors and normalized structure factors in the programs *POLARRFN* and *AMoRe* over a range of high-resolution limits, showing the values obtained for both the S/σ and the $\Delta S/\sigma$ criteria.

	Structure factors				Normalized structure factors			
	POLARRFN		*AMoRe*		*POLARRFN*		*AMoRe*	
Res†	S/σ	$\Delta S/\sigma$	S/σ	$\Delta S/\sigma$	S/σ	$\Delta S/\sigma$	S/σ	$\Delta S/\sigma$
1.2	5.023	0.496	4.864	2.396	11.563	1.082	16.613	9.960
1.3	5.050	0.525	4.861	2.062	11.896	1.085	16.410	9.779
1.5	5.007	0.507	4.841	1.614	9.437	1.058	13.480	7.328
1.7	4.967	0.489	4.819	1.293	8.031	1.038	10.614	4.478
1.9	4.933	0.482	4.792	0.599	6.661	0.998	8.650	3.109
2.1	4.876	0.462	4.716	0.726	6.462	1.000	7.911	3.086
2.3	4.746	0.352	4.546	0.477	6.188	0.833	6.864	2.082
2.5	4.478	0.168	4.191	0.397	4.801	0.618	5.373	1.157
2.7	4.458	0.117	4.016	0.224	4.817	0.519	4.373	0.643
2.9	4.246	0.150	4.134	0.243	4.247	0.448	4.718	0.504
3.5	4.211‡	−0.275	4.052	0.328	4.185‡	−0.260	4.124	0.237

† Res is the high-resolution limit. ‡ This solution was not the top solution.

All calculations within these limits produced peaks corresponding to a 180° rotation, but only when the high-resolution data limit was set to include data greater than 3.0 Å was a peak detected which was greater than the noise (Table 2). This peak is indicative of the presence of a non-crystallographic twofold rotation, thereby agreeing with the density calculations, which suggested the presence of two molecules in the asymmetric unit of the *P*1 space group. For the calculations using structure factors, the values of S/σ were very similar at all resolutions. However, at low resolution a number of additional peaks were visible on the $\kappa = 180°$ section of the polar coordinate maps, which meant that the $\Delta S/\sigma$ value was considerably lower in calculations which only used data at a resolution lower than 1.7 Å (Table 2). This is because there are many possible pseudo-solutions at low resolution arising from the internal symmetries present in the molecule: firstly, there is the comma-like shape of the molecule which, at low resolution, would appear to have an internal twofold rotation, thereby producing additional peaks; secondly, the helix symmetry would appear to produce regular peaks corresponding to the helix repeat. Only when high-resolution data is included will these false solutions disappear relative to the 'true' solution. These observations demonstrate why it was useful to use both S/σ and $\Delta S/\sigma$ as criteria.

The advantage of using normalized structure factors can be clearly seen in Table 2, with both programs showing an improvement in the magnitude of both the signal-to-noise ratios (S/σ) and the peak-height differences ($\Delta S/\sigma$) of the solutions. *AMoRe* produced the largest peaks when normalized structure factors with increasingly higher resolution data above 2.5 Å resolution were used. The results obtained with *AMoRe* using normalized structure factors were considerably more favourable than those obtained with *POLARRFN* (especially for the $\Delta S/\sigma$ criterion), although both methods produced the same angular values for the rotation solution. All further calculations were undertaken using normalized structure factors.

4. Cross-rotation function

4.1. Models

At the time of undertaking this study, there were only two structures of long peptaibols known, alamethicin Rf-30 (ALM; Fox & Richards, 1981) and Leu1–zervamicin (ZL; Karle *et al.*, 1991). Several models derived from each of the three independent monomers in the ALM crystal structure were tested (Wallace & Snook, unpublished results), but they were all less successful than models derived from the ZL structure. This is probably because (as we now know) the structure of AAM more closely resembles that of ZL (the r.m.s.d.s for the backbone atoms of ZL and the AAM chains *A* and *B* are 0.696 and 0.648 Å, respectively, but they range between 1.202 and 1.351 Å for AAM and the different *A*, *B* and *C* chains of ALM), as calculated by *LSQMAN* (Kleywegt & Jones, 1994). This is the cumulative result of the bend which is present in the middle segment of the molecule for both ZL and AAM and is caused by the presence of the imino acids at positions 10, 13 and 15 in these structures. Of the equivalent positions in alamethicin (residues 14, 17 and 19), only residue 14 is an imino acid; hence, the bend is much less pronounced.

Thus, the ZL crystal structure (Karle *et al.*, 1991) formed the basis for the models (Table 3) used in the molecular replacement reported here. Model I included all the backbone atoms of ZL and all the side chains which were equivalent, with all the non-identical side chains reduced to the shortest equivalent side chain, *i.e.* alanine. Model I contained the most atoms and thus would produce the maximum scattering power. However, if the structures of ZL and AAM differed significantly, this model would potentially be most adversely affected. Model II, which included residues 6–16 of ZL, again with non-conserved side chains reduced to the shortest equivalent side chain, was constructed for two reasons: firstly, to allow for flexibility at the glycine at position 6, which might cause the N-terminus to be displaced relative to the C-terminus starting at this point, and secondly, to allow for the possible existence of a left-handed helix at the N-terminus, which had been originally proposed for AAM based on NMR data (Das *et al.*, 1986) and which would produce a completely different structure for the N-terminus to that found in ZL. Model IV was constructed as a conservative minimal model, incorporating only atoms found in the highly homologous seven C-terminal residues. Model III was envisaged as a model with an intermediate number of residues between those found in models II and IV. All four models, although only differing in a small number of atoms, were tested, as previous experience with an intermediate-sized peptide (Janes *et al.*, 1994) had shown that a successful molecular-replacement solution of a molecule in this size range was highly dependent on the size and close similarity of structure of the starting model.

4.2. Initial search parameters

The cross-rotation searches were carried out using normalized structure factors in *AMoRe* (Navaza, 1994; Collaborative Computational Project Number 4, 1994) and using *RFCOR* (Collaborative Computational Project Number

Table 3
Models used for molecular replacement: ZL is the Leu1–zervamicin sequence, AAM is the antiamoebin I sequence.

Models I, II, III and IV refer to the different model sequences used in the MR tests. The non-standard amino-acid abbreviations are U, α-methyl alanine; J, α-ethylalanine; O, γ-hydroxyproline.

Model	1				5					10					15				Number of atoms in the model
ZL	X	L	I	Q	J	I	T	U	L	U	O	Q	U	O	U	P	F	OH	—
AAM	X	F	U	U	U	J	G	L	U	U	O	Q	J	O	U	P	F	OH	119
I	X	A	A	A	A	A	G	A	A	U	O	Q	A	O	U	P	F	OH	102
II							G	A	A	U	O	Q	A	O	U	P	F	OH	74
III									A	U	O	Q	A	O	U	P	F	OH	65
IV											O	Q	A	O	U	P	F	OH	54

4, 1994) to analyze the peaks of the cross-rotation search with the self-rotation map. Searches were carried out using 5° search grids over 180° for the $P1$ space group. The radius of integration was initially set, based upon contemporary recommendations (Tickle & Driessen, 1996), at 6.9 Å, which is 77% of the smallest unit-cell dimension. The data used initially were in the range 10.0–2.5 Å, the limits being chosen based upon a brief survey of the literature. These initial parameters failed to elicit the expected (based on the self-rotation function) pair of peaks related by 180° for any of the models tested.

4.3. Resolution limits of the data

The high-resolution limit was raised gradually to 1.2 Å to optimize the peak height. The highest $\Delta S/\sigma$ values were obtained for models II and III, with slightly poorer results for model I and the worst results for model IV. The solutions for model II over a range of high-resolution limits with an integration radius of 3.5 Å (see §4.4) are illustrated in Fig. 1. It should be noted that the expected pairs of solutions were only visible above the background when the resolution limit included data beyond 1.7 Å, with the maximal $\Delta S/\sigma$ obtained using data to 1.2 Å resolution. Therefore, the full data set to 1.2 Å resolution was then used to explore the effects of the radius of integration.

4.4. Radius of integration

The shape of the molecule needs to be borne in mind when choosing the radius of integration. Owing to the small unit-cell size and the compact unit cell resulting from the low solvent content, the choice of the radius of integration had to take into consideration the possibility of vectors arising from molecules in neighbouring unit cells. By analogy with other peptaibol structures, we suspected that the molecule would be almost entirely helical. Based on the report of MR studies on the helical flounder antifreeze protein (Sicheri & Yang, 1996), it was decided to vary the integration radius by 0.5 Å increments from 2.0 Å to the limit of the smallest unit-cell dimension (9.0 Å) in an attempt to obtain a larger $\Delta S/\sigma$. The results are shown in Fig. 2. Variation of the radius produced two peaks, a broad one at around 70–80% of the smallest unit-cell dimension (*i.e.* at ~6.5 Å) and a sharper one at ~3.5 Å. The broadness of the peak at ~6–7 Å is probably a consequence of the inclusion of origin peaks from neighbouring unit cells. The locations of these peaks may be purely coincidental, but 3.5 Å corresponds to the approximate radius of an α-helix and the broad 7 Å peak to the diameter. Thus, these two peaks could be indicative of the presence of a helical structure.

4.5. Choice of model

Parameter optimizations were carried out on all four models in order to determine which would be the best one to use in the translation function. The $\Delta S/\sigma$ ratio and the $\Delta S/\sigma$ ratio per residue for the top two solutions (molecules A and B in the asymmetric unit) in the model were used to compare each of the models, although this analysis may further benefit from the more rigorous treatment recently proposed by Tsuchiya & Takenaka (1998).

It was found that for the 6.5 Å radius of integration (Table 4), model III gave the best $\Delta S/\sigma$ per residue values (0.312 and 0.311 for solutions A and B, respectively), but for the 3.5 Å radius of integration, model II was the best (0.156 and 0.154 for solutions A and B, respectively). The rotation angles for both sets of solutions were very similar, so it appeared that the solutions were the same for both models. Model II produced the most consistent $\Delta S/\sigma$ values for all radii of integration and the highest values for the best-defined (3.5 Å) radius-of-integration peak. Thus, for the final rotation solution, model II was used with normalized structure factors and data-resolution limits between 10.0 and 1.2 Å and a radius of integration of 3.5 Å. *AMoRe* produced two solutions at
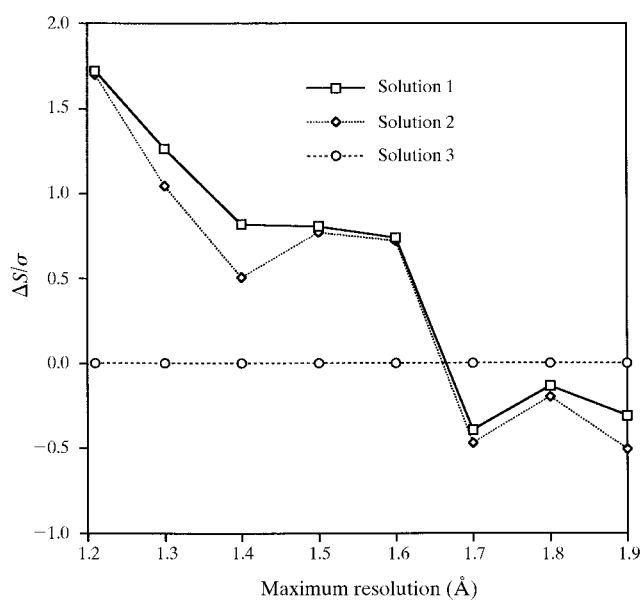

Figure 1
The effects of changing the high-resolution limits on the two best cross-rotation solutions for model II produced by *AMoRe*. The radius of integration was 3.5 Å and the low-resolution limit was 10.0 Å.

Table 4
The values of $\Delta S/\sigma$ and $\Delta S/\sigma$ per residue for each model at the two radii of integration (3.5 and 6.5 Å) which gave rise to the peaks in Fig. 2.
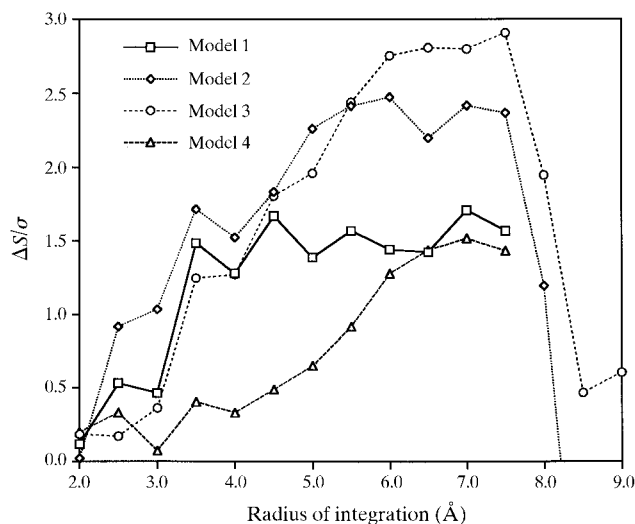
A and B designate solutions for the two different monomers in the asymmetric unit.

Model	Radius of integration 3.5 Å		Radius of integration 6.5 Å	
	$\Delta S/\sigma$	$\Delta S/\sigma$ per residue	$\Delta S/\sigma$	$\Delta S/\sigma$ per residue
IA	1.486	0.093	1.426	0.089
IB	1.337	0.084	1.165	0.073
IIA	1.716	0.156	2.200	0.200
IIB	1.696	0.154	2.029	0.184
IIIA	1.248	0.139	2.808	0.312
IIIB	1.132	0.126	2.795	0.311
IVA	0.403	0.058	1.439	0.206
IVB	0.347	0.050	1.150	0.164

$\alpha_1 = 273.7$, $\beta_1 = 95.5$, $\gamma_1 = 198.8°$ and $\alpha_2 = 94.7$, $\beta_2 = 89.0$ and $\gamma_2 = 85.8°$, both of which had signal-to-noise ratios of 1.7 (in this case, solution 3 was used as the next highest peak for the calculation of $\Delta S/\sigma$). That the two peak heights for the A and B solutions are so similar is because of the structural similarity of the two monomers in the asymmetric unit.

4.6. Effects of map noise

The rotation functions were calculated for each value of the radius of integration using a series of B_{add} values, a means of effectively increasing the B factors input in the model. Values from 0 to 20 Å^2 over the data range 10.0–1.2 Å were tested. Although the rotation functions are calculated using normalized structure factors, *AMoRe* (Navaza, 1994; Collaborative Computational Project, Number 4, 1994) treats them as structure factors. The use of larger temperature factors would tend to smear out the sharp peaks produced by the normalized structure factors, simulating noise in the map. The effects of increasing the B_{add} value were a marked decrease in magni-
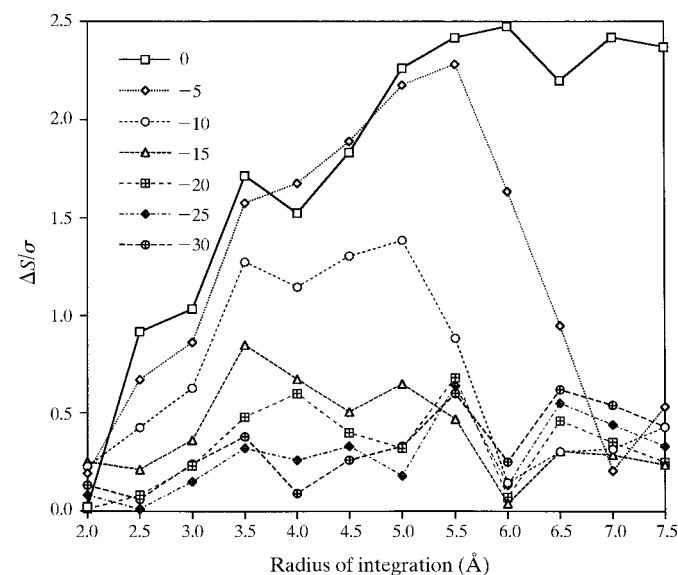
tude and a shift in position of the broad peak originally observed at a radius of integration of around 6–7 Å, as illustrated in Fig. 3. In contrast, the peak at 3.5 Å remained, decreasing in magnitude to a lesser extent. This observation can be explained if one assumes that the broadness of the peak around 6–7 Å results from signals from neighbouring unit cells.

5. Translation function

Owing to the improvement seen with normalized structure factors in the rotation functions, it was decided to use normalized structure factors for the translation-function calculations. This narrowed the choice of program used to *TFFC* (Collaborative Computational Project, Number 4, 1994), which is based on the T2 function (Tickle, 1992).

As with the rotation functions, the radius of integration and the high resolution limit parameters were varied to obtain the maximum $\Delta S/\sigma$ value. Changing the radius of integration from 2.5 to 9.0 Å in 0.5 Å increments produced no variation in the resulting translation peak. Because of this invariance, the radius of integration was set to that of the rotation solution, namely 3.5 Å, for testing the resolution limits. Increasing the high-resolution data limit did produce a significant increase in the translation peak height and large changes in the position of the peak. By far the best result was obtained with a high-resolution limit of 1.2 A, which produced a $\Delta S/\sigma$ of 2.41 (Table 5).

The starting R factor for this model was 48.0% (25.0–8.0 Å, 140 atoms), which reduced to 42.6% after rigid-body refinement. The final R factor for the refined structure (using all data to 1.4 Å and a full-matrix least-squares refinement (*SHELXL*93; Sheldrick, 1993) was 15.4% (Snook *et al.*, 1998).

Figure 2
The values of $\Delta S/\sigma$ as a function of radius of integration for all four models, for the common peak solution around $\alpha = 273$, $\beta = 95$ and $\gamma = 98°$. All calculations were carried out using normalized structure factors with data in the resolution range 10.0–1.2 Å using the program *AMoRe*.

Figure 3
The effects of different values of map noise (B_{add}) in *AMoRe* on the top cross-rotation solution for model II. Notice the rapid decline of the peaks which occurs with a radius of integration greater than 50% of the minimum unit-cell dimension (4.5 Å).

Table 5
The effect of using different data-resolution limits in the T2 translation function of *TFFC* with a radius of integration of 35 Å, showing the variability of the peak position and height and the necessity for using high-resolution data.

Res is the high-resolution limit (in Å); x, y, z are fractional coordinates of the highest peak.

Res	x	y	z	$\Delta S/\sigma$
1.2	0.872	0.007	0.801	2.41
1.3	0.455	0.229	0.591	0.43
1.4	0.909	0.139	0.523	0.04
1.5	0.871	0.236	0.773	0.16
1.6	0.474	0.192	0.442	0.49
1.7	0.424	0.222	0.705	0.56
1.8	0.924	0.181	0.091	0.17
1.9	0.561	0.347	0.727	0.05

6. Comparison of the initial models with final structure

The success of this molecular-replacement study relied on the availability of a reasonable starting model derived from the ZL structure. The equivalent residues in model II differed from those in the final structure by r.m.s.d.s of 0.489 and 0.392 Å for the backbone atoms of chains *A* and *B*, respectively (Fig. 4*a*). For all atoms, the r.m.s.d.s are 2.020 and 2.050 Å, respectively. The differences between the AAM structures and the whole ZL structure were 0.696 and 0.648 Å for the backbone atoms of chains *A* and *B*, respectively, and 1.032 and 1.280 Å for all atoms. These calculations indicate that AAM deviates most from ZL in its N-terminal segment and this can be seen in Fig. 4(*b*).

(*a*)   (*b*)
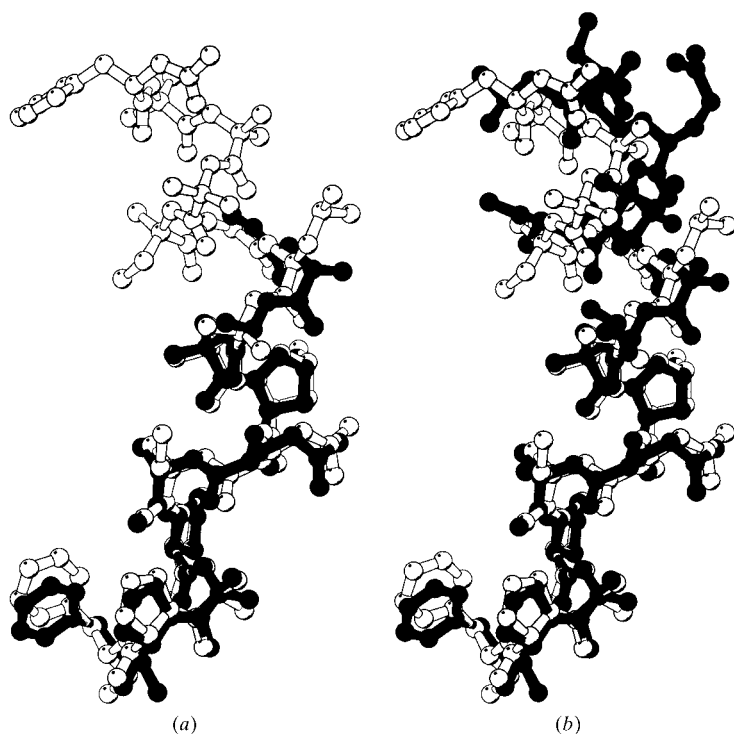
Figure 4
Overlay of the final AAM structure (in white) (*a*) with model II, the model that was used for the MR solution, and (*b*) with ZL, both drawn using the graphics program *BOBSCRIPT* (Esnouf, 1997).

The poorer signals with models I and IV probably arose from two different factors: model I differs from model II by inclusion of five N-terminal residues which, from the comparison of the r.m.s.d.s of the backbone atoms above, clearly deviate significantly from the final structure, hence the poorer fit. Model IV differs from the other models in that it is the shortest, containing only the residues in the relatively straight C-terminal portion of the molecule after the bend at position 10. There are two possible reasons why this model was not as good as the others: (i) there were not enough distinct features in the model to orientate it within the density, similar to the problem encountered with the antifreeze protein (Sicheri & Yang, 1996), and (ii) it did not contain enough atoms to produce a reasonable signal. However, it should be noted that protein structures have been solved with a smaller proportion of atoms in the model than this (Bernstein & Hol, 1997). Models II and III produced roughly comparable solutions, as the additional two amino acids in model II are of similar (but not identical) structure to the AAM structure in this region (Fig. 4*a*).

High-resolution data were apparently needed for several reasons: (i) the *A* and *B* monomers were very similar in structure, (ii) each monomer itself was pseudosymmetric at low resolution, so that a 180° rotation about its centre produces a similarly shaped molecule (Fig. 5), and (iii) the internal helical symmetry in the molecule meant that there were a number of pseudorotation and pseudotranslation solutions owing to the helical repeat (especially if a short search structure which was essentially a straight helix, such as model IV, was used). Longer models which included the bend in the middle tended to break this internal symmetry.

7. Conclusions

The solution of the antiamoebin I structure illustrates the utility of using MR for solving structures of this size which may normally be considered to be difficult. In this case, normalized structure factors needed to be used to obtain the best signal-to-noise ratios.

The results of this molecular replacement suggest that, although using as little as 4% of the final model has been reported to result in a successful MR solution (Chantalat *et al.*, 1996), the proportion of the final model which is required for effective solution varies with the structure. This variation would appear to depend upon two factors. The first is the size of the polypeptide being solved. It may be that the larger the protein, the smaller portion of the final model can be used. The other factor is the topology of the molecule. If the portion of the structure used is small or pseudosymmetric enough that it can be orientated into many parts of the target density, then the rotation function will give a large number of false peaks. This was demonstrated when the starting model used was shortened to below the helix bend.
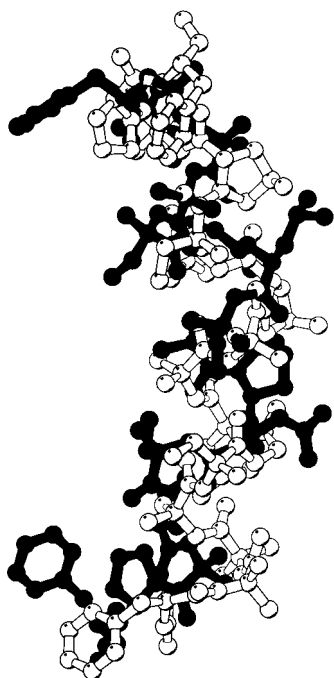
Figure 5
Overlay of AAM chain *A* with itself rotated by 180° about an axis centered on the Hyp10 N atom parallel to the *a* axis, showing the similarity in shape which made the rotation solution difficult when only using low-resolution data. This figure was produced using the graphics program *BOBSCRIPT* (Esnouf, 1997).

An increase in the signal for the rotation functions was obtained by first using more of the high-resolution data. A second improvement was achieved with the reduction in the integration volume, so that the lengths of the vectors used were shorter than would be normally used for globular proteins. This had the effect of reducing the background noise, thereby increasing the $\Delta S/\sigma$ for the solution peaks.

The radius of integration did not have a significant effect on the translation function, whereas the high-resolution limit of the data did have a significant effect. The apparent lack of effect of the radius of integration may arise from the formulation of the T2 translation function (Harada *et al.*, 1981) used in the *TFFC* (Collaborative Computational Project, Number 4, 1994) program used. In this formulation, the search for more than one molecule requires the whole of the unit cell to be searched. The translation function relies on the normalized structure factors, so that increasing the high-resolution limit would have an effect upon the translation function.

It is interesting to note that another structure of AAM published contemporaneously with our structure (Karle *et al.*, 1998) was solved using a different MR strategy, starting model and refinement technique. Even though the crystals were themselves quite different, being formed in octanol rather than methanol, and were in a different space group, the resulting structures were remarkably similar, differing only by 0.25 Å r.m.s.d. for the polypeptide main-chain atoms, thus suggesting a robust solution.

In summary, this paper describes the use of molecular replacement to solve the structure of a molecule in the intermediate size range and shows that the success of the procedure is highly dependent on the nature of the molecular model and also on the parameters used.

References

Baikalov, I. & Dickerson, R. E. (1998). *Acta Cryst.* D**54**, 324–333.
Baker, E. N., Anderson, B. F., Dobbs, A. J. & Dodson, E. J. (1995). *Acta Cryst.* D**51**, 282–289.
Bernstein, B. E. & Hol, W. G. J. (1997). *Acta Cryst.* D**53**, 756–764.
Brady, R. L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E. J. & Dodson, G. G. (1991). *Acta Cryst.* B**47**, 527–535.
Chantalat, L., Wood, S. D., Rizkallah, P. & Reynolds, C. D. (1996). *Acta Cryst.* D**52**, 1146–1152.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Das, M. K., Ragothama, S. & Balaram, P. (1986). *Biochemistry*, **25**, 7110–7117.
Doyle, D. A. & Wallace, B. A. (1997). *J. Mol. Biol.* **266**, 963–977.
Duclohier, H., Snook, C. F. & Wallace, B. A. (1998). *Biochim. Biophys. Acta*, **1415**, 255–260.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Fox, R. O. & Richards, F. M. (1981). *Nature (London)*, **300**, 325–330.
Harada, Y., Lichitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* A**37**, 398–406.
Janes, R. W., Peapus, D. H. & Wallace, B. A. (1994). *Nature Struct. Biol.* **1**, 311–319.
Karle, I. L., Flippen-Anderson, J. L., Agarwalla, S. & Balaram, P. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5307–5311.
Karle, I. L., Perozo, M. A., Mishra, V. K. & Balaram, P. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5501–5504.
Kleywegt, G. J. & Jones, T. A. (1994). *Jnt CCP4/ESRF–EACBM Newslett. Protein Crystallogr.* **31**, 9–14.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miller, R. J., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Oliva, G. (1988). PhD Thesis, University of London.
Sansom, M. S. P. (1993). *Eur. Biophys. J.* **22**, 105–124.
Schafmeister, C. E., Miercke, L. J. W. & Stroud, R. M. (1993). *Science*, **262**, 734–738.
Sheldrick, G. M. (1986). *SHELXS*86, *Program for the Solution of Crystal Structures.* University of Göttingen, Germany.
Sheldrick, G. M. (1993). *SHELXL*93, *Program for the Refinement of Crystal Structures.* University of Göttingen, Germany.
Sicheri, F. & Yang, D. S. C. (1996). *Acta Cryst.* D**52**, 486–498.
Snook, C. F., Woolley, G. A., Oliva, G., Patthabi, V., Wood, S. P., Blundell, T. L. & Wallace, B. A. (1998). *Structure*, **6**, 783–792.
Tickle, I. J. (1992). *Proceedings of the CCP4 Study Weekend. Molecular Replacement*, edited by E. J. Dodson, S. Glover & W. Wolf, pp. 20–32. Warrington: Daresbury Laboratory.
Tickle, I. J. & Driessen, H. P. C. (1996). *Methods in Molecular Biology*, Vol. 56, *Crystallographic Methods and Protocols*, edited by C. Jones, B. Mulloy & M. Sanderson, pp. 173–203. Totowa, NJ: Humana Press.
Tsuchiya, D. & Takenaka, A. (1998). *Acta Cryst.* D**54**, 151–153.
Turkenburg, J. P. & Dodson, E. J. (1996). *Curr. Opin. Struct. Biol.* **6**, 604–610.
Urzhumtsev, A. & Podjarny, A. (1995). *Acta Cryst.* D**51**, 888–895.
Vitali, J., Martin, P. D., Malkowski, M. G., Olsen, C. M., Johnson, P. H. & Edwards, B. F. P. (1996). *Acta Cryst.* D**52**, 453–464.
Whitmore, L., Snook, C. F. & Wallace, B. A. (1997). *Peptaibol Database*, http://www.cryst.bbk.ac.uk/peptaibol/welcome.html.