Acta Cryst. (1995). D**51**, 282–289

Use of Iron Anomalous Scattering with Multiple Models and Data Sets to Identify and Refine a Weak Molecular Replacement Solution: Structure Analysis of Cytochrome *c'* from Two Bacterial Species

BY EDWARD N. BAKER,* BRYAN F. ANDERSON AND AARON J. DOBBS

Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand

AND ELEANOR J. DODSON

Chemistry Department, University of York, Heslington, York YO1 5DD, England

(*Received* 6 *September* 1994; *accepted* 8 *November* 1994)

Abstract

The structure of cytochrome *c'* from two bacterial species, *Alcaligenes sp* and *Alcaligenes denitrificans*, have been determined from X-ray diffraction data to 3.0 Å resolution using the anomalous scattering of the single Fe atom in each to identify and refine a weak molecular-replacement solution. Molecular-replacement studies, with the program *AMORE*, used two isomorphous data sets (from the two species), two independent search models (the cytochromes *c'* from *Rhodospirillum molischianum* and *Rhodospirillum rubrum*), both with and without side chains, and two different resolution ranges (10.0–4.0 and 15.0–3.5 Å) to generate a large number of potential solutions. No single solution stood out and none appeared consistently. The Fe-atom position in each structure was then determined from its anomalous-scattering contribution and all molecular-replacement solutions were discarded which did not (i) place the Fe atom correctly and (ii) orient the molecule such that a crystallographic twofold axis generated a dimer like those of the two search models. Finally, electron-density maps phased solely by the Fe-atom anomalous scattering were calculated. As these were combined and subjected to solvent flattening and histogram matching (with the program *SQUASH*), correlation with the remaining molecular-replacement solutions identified one as correct and enabled it to be improved and subjected to preliminary refinement. The correctness of the solution is confirmed by parallel isomorphous-replacement studies.

Introduction

The method of molecular replacement, which combines a rotation search (Rossmann & Blow, 1962) and a translation search (Crowther & Blow, 1967), has frequently provided a powerful method for determination of the structures of homologous proteins. It is not always successful, however, and is likely to have difficulties when available search models are poor, the structure is symmetrical, the space group is of high symmetry or diffraction data are incomplete. Anomalous scattering is often used to supplement isomorphous-replacement data for phase determination by SIRAS (single isomorphous replacement with anomalous scattering) or MIRAS (multiple isomorphous replacement with anomalous scattering) methods (Blundell & Johnson, 1976). It is also the basis of phase determination by multiwavelength anomalous dispersion (MAD) methods (Hendrickson, 1991). The former method utilizes the anomalous signal from added heavy atoms while the latter uses atoms intrinsic to the protein structure (*e.g.* Fe or Se atoms).

Cytochrome *c'* is a small electron-transfer protein, with 126–128 amino-acid residues and a single haem, which is found in a variety of denitrifying bacteria (Bartsch, 1978). It is normally dimeric in solution and is noted for the unusual spin state of the Fe atom (Maltempo, Moss & Cusanovich, 1974). Three-dimensional structures are available for the cytochromes *c'* from two bacterial species, *Rhodospirillum molischianum* (Weber, Howard, Xuong & Salemme, 1981; Finzel, Weber, Hardman & Salemme, 1985) and *Rhodospirillum rubrum* (Yasui, Harada, Kai, Kasai, Kusunoki & Matsuura, 1992); these have been refined at 1.67 and 2.8 Å, respectively. The two structures are both four-helix bundles (Fig. 1), but differ considerably in the nature of their monomer–monomer interactions (Yasui Harada, Kai, Kasai, Kusunoki & Matsuura, 1992). Moreover, members of the cytochrome *c'* family generally show only a low level of sequence identity, typically 25–30% (Ambler, Bartsch, Daniel, Kamen, McLellan, Meyer & Van Beeuman, 1981).

Here we report the successful structure solution for two other cytochromes *c'*, from *Alcaligenes sp* and *Alcaligenes denitrificans*. These two proteins crystallize isomorphously with one another (Norris, Anderson, Baker & Rumball, 1979), but molecular replacement

* To whom correspondence should be addressed (air mail).

with either of the two available search models proved extremely difficult. Eventually, the anomalous-scattering signal from the single Fe atom in each structure, together with density-modification methods, the availability of two isomorphous data sets and knowledge of the likely dimeric structure, provided a surprisingly effective means of identifying and refining the correct solution.

Crystallization and data collection

Cytochrome c' was isolated from cultures of *Alcaligenes sp* (*Asp*, NCIB 11015, formerly known as *Pseudomonas denitrificans*) and *Alcaligenes denitrificans* (*Ade*, NCTC 8582) and crystallized as previously described (Norris, Anderson, Baker & Rumball, 1979). The crystals from the two species are closely isomorphous, with cell dimensions $a = 54.4$, $b = 54.4$, $c = 181.1$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ (*Asp*), and $a = 54.6$, $b = 54.6$, $c = 181.4$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ (*Ade*), space group either $P6_122$ or $P6_522$, with one monomer in the asymmetric unit.

For *Asp* cytochrome c', data were collected to a maximum resolution of 2.8 Å, using a Xuong–Hamlin area detector (Xuong, Nielsen, Hamlin & Anderson, 1985) with graphite-monochromated Cu $K\alpha$ radiation from a Rigaku RU200 rotating-anode generator. The two detectors were placed at distances of 117.04 cm ($\theta = 24.0°$) and 110.80 cm ($\theta = 15.0°$) giving essentially complete coverage of Friedel pairs. Each frame was 0.1° and counts were accumulated for 45 s per frame. For *Ade* cytochrome c', data were collected to a maximum resolution of 1.86 Å, using a Rigaku R-AXIS II imaging-plate system, with graphite-monochromated Cu $K\alpha$ radiation from a Rigaku RU200 rotating-anode generator operated at 50 kV 100 mA. A total of 30 frames, of 2.0° each, were collected, giving a total rotation range of 60°. The exposure time for each frame was 50 min and the crystal-to-detector distance was 119.8 cm. For each data collection only a single crystal of cytochrome c' was required.

Statistics of the data collection and data processing are given in Table 1, and the completeness and quality of each data set are indicated in Fig. 2. The *Asp* data were processed using the supplied detector software (Howard, Nielsen & Xuong, 1985), while for the *Ade* data the supplied R-AXIS software (Higashi, 1990; Rossmann, 1979) was used. Both data sets are characterized by a high level of completeness (to 3.0 Å for *Asp* and 2.25 Å for *Ade*), low R_{merge} values, and a reasonably high level of redundancy in the measured data (redundancy factors 3.8 and 6.7, respectively). R values for the agreement of Friedel pairs are comparable with the R_{merge} values. The data sets for the two species scale together well, with an R value of ~0.15 over the whole resolution range, consistent with the high degree of isomorphism in the unit-cell dimensions.

Table 1. *Data-collection statistics*

	Alcaligenes sp	*Alcaligenes denitrificans*
Resolution (Å)	2.82	1.86
Number of observations	15232	60806
Unique reflections	4038	9113
Completeness (%)	94 (3.0 Å)	93 (2.25 Å)
R_{merge}*	0.016	0.032
R_F†	0.024	0.028

* $R_{merge} = \sum |I - \bar{I}| / \sum \bar{I}$, where the summation is over all redundant measurements (but not Friedel pairs).

† $R_F = \sum |I^+ - I^-| / \sum (I^+ + I^-)$ where I^+ and I^- are Friedel pairs, and the summation is performed after merging redudant measurements.

Molecular replacement

Search models

The two structures available for use as search models were the 1.67 Å model for *Rhodospirillum molischianum* (*Rmo*) cytochrome c' (Finzel, Weber, Hardman & Salemme, 1985; coordinates 1CCY from the Brookhaven Protein Data Bank) and the 2.8 Å model for *Rhodospirillum rubrum* (*Rru*) cytochrome c' (Yasui, Harada, Kai, Kasai, Kusunoki & Matsuura, 1992; coordinates kindly supplied by Dr M. Yasui). Both form dimers, and in both cases there is one dimer in the crystal asymmetric unit. Sequence identity between the two species is low, however, (~25%) and superposition of their monomer units shows a correspondingly imprecise agreement; r.m.s. agreement of 2.5 Å when 90% of C_α atoms are superimposed (omitting sites of insertions/deletions). The agreement is worst in the loops joining the four helices, but even if just the C_α atoms of the four helices are compared the r.m.s. agreement is still only 1.35 Å (for 71 C_α atoms), since the angular dispositions of the helices differ by small amounts (1.0–8.3°) in the two structures (Yasui, Harada, Kai, Kasai, Kusunoki & Matsuura, 1992).

The amino-acid sequence of *Ade* cytochrome c' is not known, but the high isomorphism between *Ade* and *Asp* crystals suggests that the two structures are very similar. They differ substantially from both the known structures, however, with the *Asp* sequence (Ambler, 1973) having only 29% identity with *Rmo* and 27% with *Rru*. There was no reason to choose either of these known structures in preference, therefore, and each was used as the basis for the construction of separate search models. Search models of several types were used (Table 2). Side chains were either all truncated to C_β atoms or were retained only where they were identical, or similar in geometry. In the latter case, 'similarity' equated Phe with Tyr (minus O_η), Ser with Thr (minus $C_{\gamma 2}$), Thr with Val, Asn with Asp, Gln with Glu, Val with Ile (minus $C_{\delta 1}$), Lys with Arg (as far as C_δ) and Leu with Phe or Tyr (as far as $C_{\delta 1}$ and $C_{\delta 2}$); all other non-identical side chains were truncated to Ala. Models were constructed both with and without the inter-helix connecting loops, but the haem group was retained in all models.

Table 2. *Search models for molecular replacement.*

Model name*	Residues	Atoms†	Comments
Mol1	3–125	648	All side chains truncated at C_β
Mol2	3–28; 40–63; 80–125	632	Loops 1–2 and 2–3 deleted Similar or identical side chains retained (51 side chains)
Rub1	1–122	647	All side chains truncated at C_β
Rub2	1–25; 38–60; 72–122	629	Loops 1–2 and 2–3 deleted Similar or identical side chains retained (52 side chains)

* Models Mol1 and Mol2 derived from *Rhodospirillum molischianum* cytochrome c', Rub1 and Rub2 from *Rhodospirillum rubram* cytochrome c'.

† Including the haem group.

Molecular-replacement methods

Initial molecular-replacement attempts using the program *ALMN*, from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994) gave no convincing solution for the rotation function; many peaks of similar height were obtained, with the highest peaks varying according to the resolution range or search model employed. Attempts to use *X-PLOR* (Brünger, 1990) with Patterson-correlation refinement were also unsuccessful.

A systematic attempt to find the correct molecular-replacement solution was then made with the program *AMORE* (Navaza, 1994). In this program the rotation search follows a similar philosophy to those of programs such as *ALMN* and *MERLOT*, performed in reciprocal space, but using a modified fast-rotation-function algorithm, in which a different approach to the calculation of spherical harmonics is adopted, using radial quadrature rather than recursive iteration. It also has the advantage of being able to rapidly test a large number of potential solutions, because the initial structure factors are calculated on a very fine grid, such that subsequent structure-factor sets can be generated by interpolation. [This approach is also used in *MERLOT*, following earlier ideas of Lattman (Lattman & Love, 1970).]

In the rotation search four models were used (Table 2), each with two different resolution ranges (10.0–4.0 Å and 15.0–3.5 Å). Individual atomic B factors from the original structural models were retained. The same calculations were performed for both *Asp* and *Ade* data sets, giving a total of 16 separate rotation-function experiments. All of the top rotation solutions (about 50 in each case) were tested in the translation function, which was performed in each of the two possible enantiomeric space groups, $P6_122$ and $P6_522$. The expectation was that the correct solution would appear in most, if not all, of the experiments, and that it would be identified by its consistency, even if it was not necessarily the top solution in each case. Again, however, no clear solution was obtained. Many solutions appeared more than once, but none was found as a strong solution for each model, each data set, and both resolution ranges. Correlation coefficients were in the range 0.15 to 0.25, and crystallographic R factors were all higher than 0.56.

Two known features of the cytochrome c' structure were then used to screen for the most probable solution. These were the position of the single Fe atom and the likely location of the dimer twofold axis.

The Fe-atom position was obtained from a Patterson map calculated using the observed anomalous differences as coefficients; a similar anomalous difference Patterson was calculated for both the *Asp* and *Ade* data sets, and the similarity of the two maps (Fig. 3) gave strong support to the correctness of the solution. For both species the Fe-atom position was found at (0.148, 0.704, 0.211) when space group $P6_122$ was assumed, or (0.556, 0.704, 0.211) in space group $P6_522$. This gave an Fe–Fe distance, across the nearest crystallographic twofold axis, of 24.6 Å, consistent with distances of 24–25 Å found in the two known cytochrome c' dimers (Weber, Howard, Xuong & Salemme, 1981; Yasui, Harada, Kai, Kasai, Kusunoki & Matsuura, 1992).

The second criterion, the position of the monomer–monomer twofold axis, was weaker since although both known cytochrome c' structures are dimeric and use the same helix (helix 1) in dimerization, the mode of association differs in detail, and their monomer–monomer twofold axes differ in angular disposition by ~12° when the respective monomers are superimposed.

The solutions that placed the Fe atom within 6 Å of the correct position are listed in Table 3. Two immediate conclusions were that most of these best solutions were in space group $P6_522$ rather than $P6_122$, and that the 'truncated/conserved' model gave a larger number of acceptable solutions, with lower R factors and higher correlation coefficients. The number of acceptable solutions was further reduced when the second criterion was applied, *i.e.* that the expected monomer–monomer twofold axis should lie close to a crystallographic twofold axis. To test this, each model had included two dummy atoms, defining the twofold axis found in either the *Rmo* or *Rru* structure, and solutions were rejected if the line joining these atoms was not within 15° of a crystallographic twofold axis. Solution 7, with a well placed Fe atom and a high correlation coefficient, was eliminated by this test, as were others, many of which placed the monomer–monomer twofold axis close to the crystallographic sixfold. Two 'best' solutions remained, *i.e.* numbers 3 and 11, both of which placed the Fe atom ~2.4 Å from its correct position.

Phasing from anomalous scattering

Next, the Fe-atom anomalous scattering was used to generate electron-density maps which could then be used to test the likely correctness of the molecular-replacement solutions. Although the anomalous signal is weak for a single Fe atom ($f'' = 3.4$ electrons for Cu $K\alpha$ radiation) in a molecule of ~1000 atoms, and should give an average intensity difference of ~2% (Crick &

Table 3. *Potential molecular-replacement solutions after screening on* Fe-*atom positions*

Solution*	Rotation angles†			Translations‡			cc§	R¶	ΔFe**
	α	β	γ	trx	try	trz			
(*a*) Space-group $P6_122$; polyalanine model									
1. ADEJ41	27.9	90.0	135.0	0.464	0.312	0.396	16.1	59.0	5.14
(*b*) Space group $P6_122$; truncated/conserved model									
2. ASPC351	12.7	77.0	52.3	0.586	0.122	0.095	37.4	55.0	5.09
(*c*) Space group $P6_522$; polyalanine model									
3. ADEC355	29.0	55.1	245.0	0.489	0.580	0.227	20.5	62.0	2.41
4. ADEC355	32.0	50.4	249.0	0.489	0.573	0.228	18.9	62.0	2.76
5. ASPJ355	40.1	70.5	238.3	0.388	0.996	0.112	18.1	59.0	5.32
6. ASPJ45	42.6	43.2	247.6	0.484	0.683	0.236	20.1	59.0	5.85
(*d*) Space group $P6_522$; truncated/conserved model									
7. ADEC355	28.6	78.8	245.8	0.635	0.453	0.469	38.8	56.0	0.76
8. ADEJ45	37.2	85.8	165.8	0.492	0.840	0.140	30.3	55.0	1.01
9. ASPJ45	38.2	85.8	167.2	0.498	0.841	0.140	29.7	55.0	1.01
10. ASPJ355	2.2	89.9	169.2	0.784	0.329	0.302	30.7	55.0	2.36
11. ASPJ355	41.3	56.1	243.4	0.454	0.571	0.214	40.5	52.0	2.43
12. ADEJ45	54.1	71.2	9.0	0.770	0.571	0.423	27.4	56.0	3.45
13. ADEC45	39.4	39.7	256.1	0.835	0.454	0.064	29.5	55.0	3.82
14. ASPC45	40.3	39.5	254.2	0.833	0.450	0.064	30.1	55.0	3.84
15. ASPJ355	25.5	90.4	348.0	0.787	0.424	0.042	33.1	55.0	4.16
16. ASPJ355	36.6	92.0	244.4	0.612	0.842	0.130	30.7	55.0	5.04
17. ASPJ45	34.4	89.8	247.7	0.620	0.837	0.130	25.0	56.0	5.45
18. ASPJ355	10.0	77.4	52.8	0.258	0.385	0.420	29.5	56.0	5.51

* The code accompanying each solution indicates the MR experiment it derives from: ADE = *Ade* data, ASP = *Asp* data; J = *Rru* model, C = *Rmo* model; 35 = 15–3.5 Å data, 4 = 4–10 Å data; 1 = $P6_122$, 5 = $P6_522$.

† Eulerian angles (°).

‡ Fractional coordinates.

§ Correlation coefficient (%).

¶ Crystallographic R factor, $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ for the given resolution range (%).

** Difference (Å) between Fe-atom position and that derived from anomalous-difference Patterson.

Magdoff, 1956), two independent but isomorphous data sets (for the two *Alcaligenes* species) were available.

Best phases (Blow & Crick, 1959) to a resolution of 3.0 Å were calculated for each species using only the Fe-atom anomalous scattering; this gave a mean figure of merit of 0.04 and 0.05 for the *Asp* and *Ade* data sets, respectively. Attempts at refinement, with *MLPHARE*, were unsuccessful, probably because of the weakness of the signal, and instead each data set was placed on an approximately absolute scale through a Wilson plot (Wilson, 1949) and an Fe-atom occupancy of 1.0 was assumed. Electron-density maps calculated using these phases were, not surprisingly, of poor quality but the two maps (for *Asp* and *Ade*) did show a reasonable agreement (correlation coefficient 0.63). In an attempt to reduce noise, a combined electron-density map was also calculated, simply by adding the two individual maps, point by point, with equal weight, using the program *OVERLAPMAP*. This effectively doubled the multiplicity of the observed data, thus increasing the phasing accuracy and reducing noise.

Phases were improved, using the solvent-flattening and histogram-matching procedures in the program *SQUASH* (Cowtan & Main, 1993). Further improved phases were also obtained from the combined (*Asp* + *Ade*) map by carrying out a Fourier transform on this map and then treating the resulting phases with *SQUASH*. The electron-density map calculated using these *SQUASH*ed, combined phases, with amplitudes from the *Asp* data set, to a resolution of 3.0 Å, was used in the initial stages of structure rebuilding; two views of density from the map are shown in Fig. 4.

Structure determination

The screened molecular-replacement calculations had identified two top possibilities as molecular-replacement solutions, solutions 3 and 11 in Table 3. These were now tested against the electron-density maps phased by the Fe-atom anomalous scattering (above).

The rotational and translational parameters of the two MR solutions were first applied to both the *R. molischianum* and *R. rubrum* models, Mol2 and Rub2 (Table 2). These were then subjected to five cycles of rigid-body refinement with the program *X-PLOR* (Brünger, 1990), with five rigid groups comprising the four helices and the haem group. Final R factors, for data to 3.0 Å resolution, were 0.57 to 0.58 for solution 3 and 0.55 to 0.57 for solution 11 (separate refinements were carried out using both models and both *Asp* and *Ade* data).

A quantitative measure of the agreement between each of the two MR solutions and the electron-density maps phased by the anomalous scattering was obtained by comparing the latter with maps calculated from the MR models (F_c maps), and expressing the agreement as the correlation coefficient (program *OVERLAPMAP*). The results of these comparisons are summarized in Table 4.

Table 4. *Electron-density map correlations*

Code for maps: F_c = calculated map from model; Asp.z and Ade.z = anomalous-phased maps from *Asp* and *Ade* data, respectively; Asp.zsq and Ade.zsq = anomalous-phased maps from *Asp* and *Ade* data from phase improvement with *SQUASH*; Comb.z = combined anomalous-phased map; Comb.zsq = combined anomalous-phased map after *SQUASH*. nd = Not determined.

Model	Correlation coefficients					
	F_c/Asp.z	F_c/Ade.z	F_c/Asp.zsq	F_c/Ade.zsq	F_c/Comb.z	F_c/Comb.zsq
Solution 3	0.004	−0.010	−0.012	−0.004	−0.002	−0.002
Solution 11	0.189	0.172	0.199	0.184	0.200	0.242
(before *X-PLOR*)	0.138	0.102	0.136	0.112	0.164	
Solution 11, first rebuild, $R = 0.404$.	0.220	0.231	nd	nd	0.243	0.303
Solution 11, second rebuild, $R = 0.364$.	0.251	0.277	nd	nd	0.282	0.354

Solution 3 was clearly shown to be incorrect, as its F_c map had zero correlation with the anomalously phased maps. Solution 11, on the other hand, showed a positive correlation of 0.17 to 0.18; moreover this figure was increased for the *SQUASH*ed and combined maps, and gave a correlation coefficient of 0.24 with the final map calculated using *SQUASH*ed combined anomalous phases. This was taken as evidence that solution 11 was, indeed, correct. The usefulness of the preceding rigid-body refinement was shown by the fact that correlation coefficients obtained from the unrefined model were significantly lower, 0.11 to 0.16.

At this point some rebuilding into the final *SQUASH*ed combined anomalously phased map was carried out, on the assumption that the solution 11 model was correct. The map proved to be of surprisingly good quality, considering the weakness of the original phase information, and a quick inspection showed that the MR solution fitted well. Most of the haem group was in density (Fig. 4), including the propionate groups, and the two thioether linkages were among the strongest features of the map. No density was present at the Fe-atom position, presumably because of the use of the Fe-atom anomalous scattering in phasing, this being 90° out of phase with the real scattering contributions. Adjacent to the haem, density could be seen for the proximal His ligand and for Leu16 on the distal side (the equivalent of Met18 in *R. molischianum* cytochrome c'; Weber, Howard, Xuong & Salemme, 1981). The density for the four helices was basically good and many side chains could be placed. A tentative tracing of the three inter-helix loops was made, but little confidence could be placed in this.

Two rebuilds of the model were made, after each of which it was subjected to a round of simulated-annealing refinement using *X-PLOR* (Brünger, 1990). This led to an R factor of 0.36 for data in the range 10.0 to 3.0 Å and a correlation coefficient between the maps of 0.35. Extension of the resolution to 2.0 Å and refinement by restrained least squares (*PROLSQ*) further reduced the R factor to 0.32.

The structural model derived by these procedures has since been shown to be basically correct, and the *Asp* structure has been refined to an R factor of 0.185 using a 1.8 Å data set collected with synchrotron radiation (Dobbs, Anderson & Baker, 1995, manuscript in preparation). The loops were not correctly placed, however, until further phase information, from a single isomorphous heavy-atom derivative became available; the phase information from the derivative also confirmed the validity of the phasing described above.

Discussion

Four aspects of the structure analysis deserve special comment. The failure of the molecular-replacement calculations to produce an unequivocal solution was unexpected and points to some unfavourable aspects of the problem. The use of additional information from a variety of sources was critical to the final identification of the correct solution, and the value of even very weak phase information (in this case the Fe-atom anomalous scattering) was highlighted. The accuracy of the measured data was also of critical importance to making use of the anomalous-scattering signal.

The failure of the molecular-replacement method to give a clear solution can be attributed to a number of factors. The two available search models had a relatively low sequence identity with the *Alcaligenes sp* cytochrome c', ~29% if pairwise comparisons are made, but only 10% if all three species are compared. The four-helix bundle structure also has a considerable amount of internal symmetry, at low resolution and without side chains, and this may be the reason why multiple solutions, with rather similar rotation angles, were often obtained. The correct rotation solution (41, 56, 243) was present among the top solutions in most of the rotation searches, but only in one case did it lead to the correct translation. Moreover in most rotation searches there were a number of other peaks of similar magnitude, with angles within 10–15° of the one that subsequently proved to be correct; some examples can be seen in Table 3. It is noteworthy that the truncated/conserved model, in which structurally homologous side chains were included as far as possible, gave significantly better results than the polyalanine model. The inclusion of some side chains, in addition to the haem group, may have been essential to

break some of the internal symmetry in the molecule. Attempts were made to use a homology model, constructed by modelling the *Alcaligenes sp* side chains into the *R. molischianum* structure, but this was no more successful as a search model.

The extent to which deficiencies in the search model handicapped the MR searches can be examined in retrospect. Comparison of the final refined *Alcaligenes sp* structure (Dobbs, Anderson & Baker, 1995, manuscript in preparation) with the two other cytochrome c' structures shows that when the four helices are superimposed on those of the *Rmo* and *Rru* structures (71 C_α atoms) the r.m.s. deviations are 1.56 and 1.44 Å, respectively. Inclusion of the haem reduces these figures slightly to 1.27 and 1.19 Å, but the agreement is still not good, and it appears that helices 2 and 3, in particular, are somewhat differently oriented in the *Asp* structure. Inclusion of all main-chain atoms of the helices, plus the other atoms included in the truncated/conserved search model used for molecular replacement (conserved side chains and some loop atoms) gives r.m.s. deviations of 1.77 and 1.82 Å for the *Rmo* and *Rru* structures, respectively.

The conclusions from these comparisons are that the *Asp* structure differs significantly from both models and that neither model is appreciably better than the other. This is consistent with the failure of either model to give better results in the molecular–replacement calculations. It is also clear why the truncated/conserved model gave better results than the polyalanine model since the 1–2 and 2–3 inter-helix loops have different conformations in the final structure from those in either search model.

Using the final *Asp* structure as search model (with all side-chain atoms included) gives a clear rotation solution (peak of 4.1σ, with the next highest peak 2.4σ). The correct rotation solution is also given by a truncated/conserved model (3.5σ; next highest 2.7σ) and a polyalanine model (3.2σ; next highest 2.8σ), and all three models give the correct translation, with peak heights 7.9σ, 6.8σ and 5.3σ, respectively (next highest 3.6σ, 2.9σ and 2.3σ). Although the symmetry of the four-helix bundle may have added an extra complication (especially in the rotation function), our conclusion is that the main source of difficulty in the molecular-replacement analysis was the substantial deviation of both search models from the target structure.

In circumstances where no clear molecular-replacement solution emerges, as in the present case, it is important to use as much additional information as possible, to reduce the number of possible solutions. This may include the use of a variety of search models, if

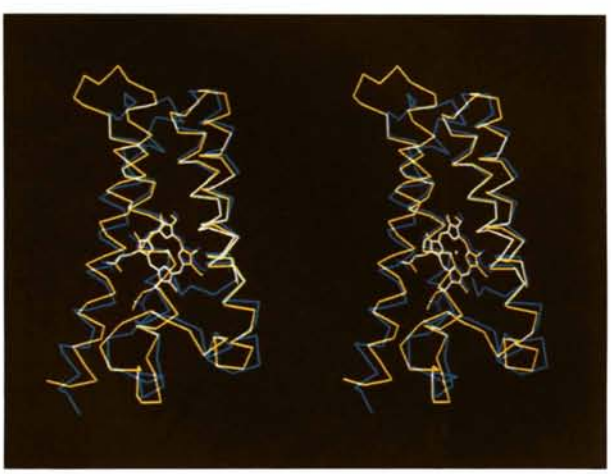

Fig. 1. Superposition of the four-helix bundle monomer units of *Rmo* (blue) and *Rru* (yellow) cytochromes c'. Superposition based on the haem groups and four helices (main-chain atoms only).

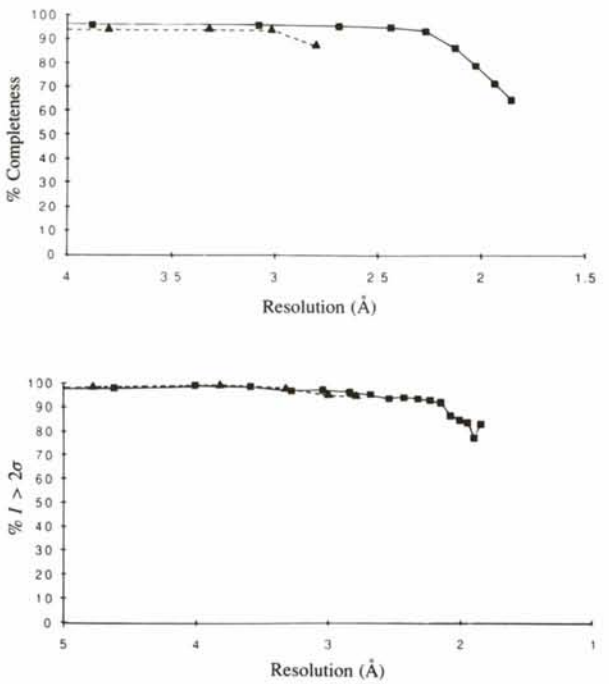

Fig. 2. Plots showing (*a*) the percentage completeness of the data set and (*b*) the percentage of reflections with $I > 2\sigma_I$ for *Alcaligenes sp* cytochrome c' (triangles, broken line) and *Alcaligenes denitrificans* cytochrome c' (squares, full line).

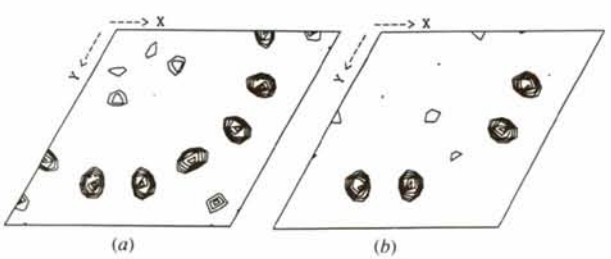

Fig. 3. Equivalent Harker sections ($z = \frac{1}{3}$) of the anomalous difference Patterson maps for (*a*) *Alcaligenes sp* cytochrome c' and (*b*) *Alcaligenes denitrificans* cycochrome c'.

more than one is available, and any chemical or other information. In the cytochrome c' analysis, the knowledge of the Fe-atom position and the ability to place the monomer–monomer twofold axis were essential in picking out potential solutions that could be tested. They were still not sufficient to define an unequivocal solution, however. In particular, an experiment in which the Fe atom of one of the search models was placed on the true Fe-atom position (derived from the anomalous difference Patterson) and the model was rotated round this position, still did not give a convincing solution in either of the two possible space groups. In retrospect it may perhaps have been possible to place the search model so that the Fe atom was correctly positioned and orient the monomer so that the crystallographic twofold created a reasonable dimer, but the problem of verifying the solution and refining it satisfactorily may still have remained. Given the significant deviations of the actual structure from either of the search models, especially in the inter-helix loops, it may not have been possible to refine from such an approximate solution.

The key factor in identifying the correct solution in present case (once the number of solutions that were feasible had been reduced to manageable proportions) was the phase information that could be derived solely from the Fe-atom anomalous scattering. This in turn depended on the availability of very accurate diffraction data. Despite the weakness of the anomalous signal, it gave sufficient phase information to allow the detection of a molecular boundary, and density modification was then able to extend and refine the phase set. The anomalously phased map shown in Fig. 4 is of remarkably good quality, given the weakness of the original phase information, and provided a powerful

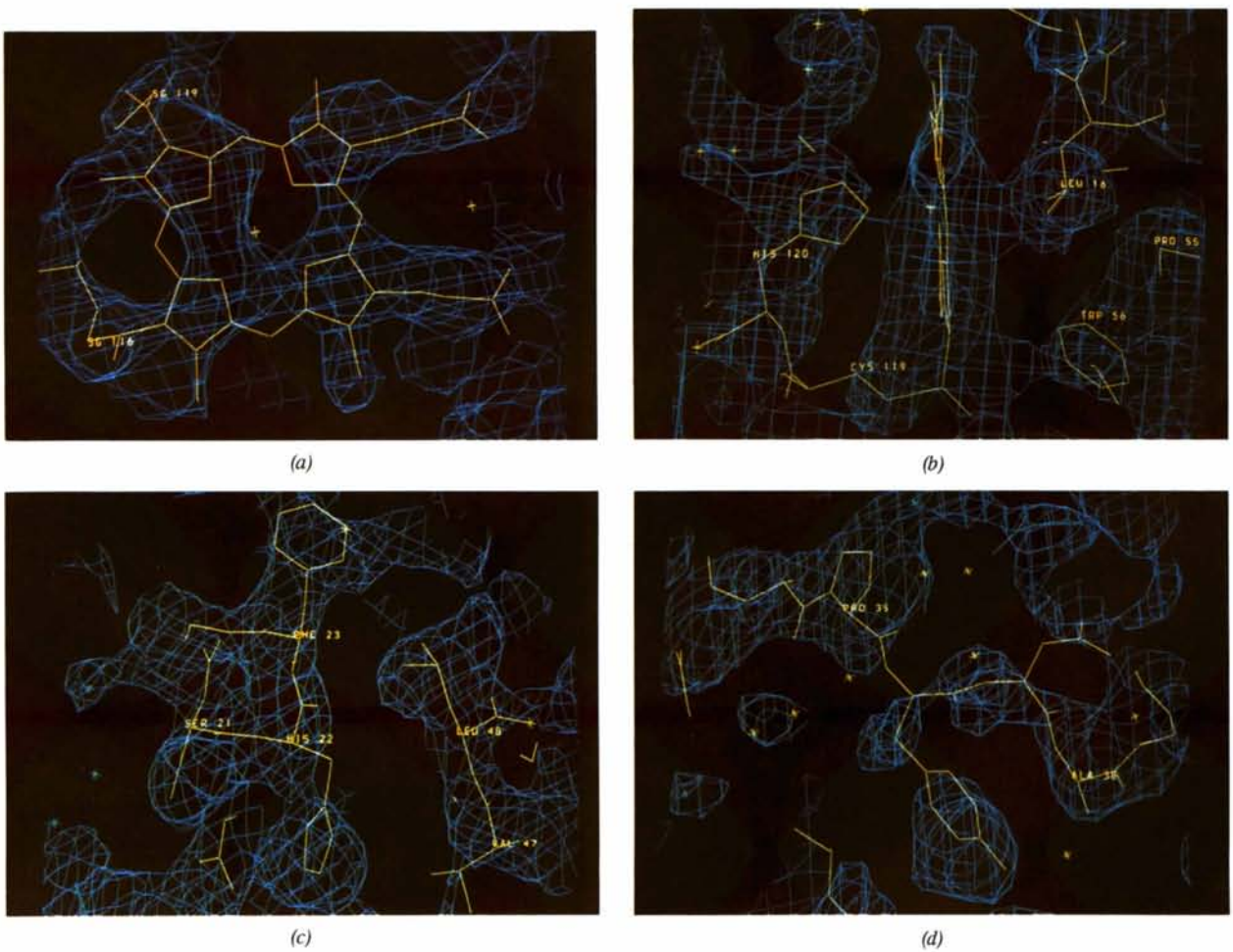

Fig. 4. Some representative views of the electron-density map phased from the Fe-atom anomalous scattering, after combination of the maps for the two species and density modification with *SQUASH*. Contoured at 1.0σ, and with the final refined model superimposed. (*a*) View perpendicular to the haem plane; note the absence of density at the Fe-atom position (see text) and the density for the thioether and propionate groups. (*b*) View parallel to the haem plane, with the Fe-atom ligand, His120 and Leu16, which approaches on the distal side. (*c*) A turn of helix 1 where it packs against helix 2. (*d*) Part of the helix 1–helix 2 connecting loop; this could not be traced with any certainty but in retrospect density was present for most of it.

external test of the MR solutions. It was also sufficiently good to allow significant rebuilding of the structure, both in small shifts of the helices and building parts of the loops, and in the inclusion of many side chains in their correct positions.

Crucial to making the most of the anomalous-scattering information was the noise reduction that resulted from the combination of the electron-density maps for the two different (but isomorphous) species of cytochrome c' and the use of the density-modification procedures in the program *SQUASH*. Averaging of poorly phased electron-density maps for the same protein in different unit cells has previously been used to help solve other protein structures (*e.g.* Varghese, Laver & Colman, 1983; Moore, Gulbis, Dodson, Demple & Moody, 1994), and density-modification methods are widely used. A special feature of our analysis was that phasing came from a single position and that in such a case a considerable proportion of reflections will have no phase information at all (because the Fe-atom anomalous contribution is too small to be significant). It is this class of reflection that may be rescued by density modification. Furthermore, an advantage when phasing from anomalous differences is that reflections of this type are randomly distributed through reciprocal space, not concentrated at higher resolution as is the case when phasing by isomorphous differences, as in the single isomorphous-replacement (SIR) method.

We are grateful to Drs Rick Faber and Hale Nicholson for help with data collection and molecular-replacement calculations. AJD gratefully acknowledges the award of a Graduate Assistantship by Massey University. ENB also receives research support as an International Research Scholar of the Howard Hughes Medical Institute.

References

AMBLER, R. P. (1973). *Biochem. J.* **135**, 751–758.
AMBLER, R. P., BARTSCH, R. G., DANIEL, M., KAMEN, M. D., MCLELLAN, L., MEYER, T. E. & VAN BEEUMAN, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 6854–6857.
BARTSCH, R. G. (1978). In *The Photosynthetic Bacteria*, edited by R. K. CLAYTON & W. R. SISTROM, pp. 249–279. New York: Plenum Press.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. London: Academic Press.
BRÜNGER, A. T. (1990). *X-PLOR Version* 2.1, Yale Univ., New Haven, CT, USA.
COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.
COWTAN, K. D. & MAIN, P. (1993). *Acta Cryst.* D**49**, 148–157.
CRICK, F. H. C. & MAGDOFF, B. A. (1956). *Acta Cryst.* **9**, 901–908.
CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 554–558.
DOBBS, A. J., ANDERSON, B. F. & BAKER, E. N. (1995). In preparation.
FINZEL, B. C., WEBER, P. C., HARDMAN, K. D. & SALEMME, F. R. (1985). *J. Mol. Biol.* **186**, 627–643.
HENDRICKSON, W. A. (1991). *Science*, **254**, 51–58.
HIGASHI, T. (1990). *J. Appl. Cryst.* **23**, 253–257.
HOWARD, A. J., NIELSEN, C. & XUONG, N. H. (1985). *Methods Enzymol.* **114**, 452–472.
LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.
MALTEMPO, M. M., MOSS, T. H. & CUSANOVICH, M. A. (1974). *Biochim. Biophys. Acta*, **342**, 290–305.
MOORE, M. H., GULBIS, J. M., DODSON, E. J., DEMPLE, B. & MOODY, P. C. E. (1994). *EMBO J.* **13**, 1495–1501.
NAVAZA, J. (1994). *Acta Cryst.* A**50**, 157–163.
NORRIS, G. E., ANDERSON, B. F., BAKER, E. N. & RUMBALL, S. V. (1979). *J. Mol. Biol.* **135**, 309–312.
ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.
VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). *Nature (London)* **303**, 35–40.
WEBER, P. C., HOWARD, A., XUONG, N. H. & SALEMME, F. R. (1981). *J. Mol. Biol.* **153**, 399–424.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
XUONG, N. H., NIELSEN, C., HAMLIN, R. & ANDERSON, D. (1985). *J. Appl. Cryst.* **18**, 342–350.
YASUI, M., HARADA, S., KAI, Y., KASAI, N., KUSUNOKI, M. & MATSUURA, Y. (1992). *J. Biochem.* **111**, 317–324.