Acta Cryst. (1991). B47, 998–1004

Molecular Replacement Solution of the Structure of Apolactoferrin, a Protein Displaying Large-Scale Conformational Change

BY GILLIAN E. NORRIS, BRYAN F. ANDERSON AND EDWARD N. BAKER*

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

(*Received* 13 *May* 1991; *accepted* 9 *July* 1991)

Abstract

The crystal structure of an orthorhombic form of human apolactoferrin (ApoLf) has been determined from 2·8 Å diffractometer data by molecular replacement methods. A variety of search models derived from the diferric lactoferrin structure (Fe_2Lf) were used to obtain a consistent solution to the rotation function. An *R*-factor search gave the correct translational solution and the model was refined by rigid-body least-squares refinement (program *CORELS*). Only three of the four domains were located correctly by this procedure, however; the fourth was finally placed correctly by rotating it manually onto three strands of electron density which were recognized as part of its central β-sheet. The final model, refined by restrained least-squares methods to an *R* factor of 0·214 for data in the resolution range 10·0 to 2·8 Å, shows a large domain movement in the N-terminal half of the molecule (a 54 rotation of domain N2) and smaller domain movements elsewhere, when compared with Fe_2Lf. A feature of the crystal structure is that although the ApoLf and Fe_2Lf unit cells appear very similar, their crystal packing and molecular structures are quite different.

Introduction

Lactoferrin is a member of the transferrin family of iron-binding proteins, which also includes serum transferrin and ovotransferrin (Aisen & Listowsky, 1980; Brock, 1985). These monomeric glycoproteins, of molecular weight ~80 000 dalton, have specific binding sites for two Fe^{3+} ions together with two CO_3^{2-} ions per molecule. The most striking feature of their binding properties is that while iron is bound extremely tightly, with binding constants of the order of 10^{20}, this iron can nevertheless be released. Low-angle X-ray scattering (Kilar & Simon, 1985; Vigh, Cser, Kilar & Simon, 1989) and hydrodynamic measurements (Rossenau-Motreff, Soetewey, Lamote & Peeters, 1971), show that iron binding and release are associated with a significant conformational change, with the molecule becoming markedly more compact when iron is bound.

As multidomain proteins, the transferrins offer a number of possibilities for conformational flexibility. The structure of human diferric lactoferrin (Anderson, Baker, Dodson, Norris, Rumball, Waters & Baker, 1987) showed that the molecule is folded into two globular lobes, one representing the N-terminal half of the molecule (the N-lobe), and the other the C-terminal half (the C-lobe). Each lobe is further subdivided into two domains, N1 and N2 for the N-lobe, C1 and C2 for the C-lobe (Fig. 1). The same four-domain structure has also been found for diferric rabbit serum transferrin (Bailey, Evans, Garratt, Gorinsky, Hasnain, Horsburgh, Jhoti, Lindley, Mydin, Sarra & Watson, 1988) and is presumably common to all transferrins.

Conformational changes during iron binding and release could involve relative movements of the two lobes and/or of the two domains of each lobe. In fact, the location of the iron sites, one in each lobe, buried deep in the cleft separating its two domains, has suggested a mechanism. In this mechanism, relative movements of the domains, about a hinge at the back of the iron site, can open each cleft, thereby facilitating iron release (Baker, Rumball & Anderson, 1987).

The structure analysis of apolactoferrin was undertaken in order to allow a comparison with the structure of diferric lactoferrin, and thereby define the nature of any conformational change. Although the use of the method of molecular replacement for structure solution becomes problematical when large conformational differences exist between search model and target molecule, we were prompted to try this approach because of (i) our failure to prepare suitable heavy-atom derivatives, given a limited supply of crystals, and (ii) an apparent similarity in the unit cells of diferric and apolactoferrin. In the event, the successful use of this method, as described here, shows that it can be successful even where as much as one-quarter of the structure is very substantially changed.

* To whom correspondence should be addressed.

0108-7681/91/060998-07$03.00

Experimental

The crystals of human apolactoferrin used for this analysis were prepared from protein whose carbohydrate had been removed enzymatically prior to crystallization (Norris, Baker & Baker, 1989). They grew under similar conditions to those used for human diferric lactoferrin (Baker & Rumball, 1977), *i.e.* dialysis against low ionic strength solutions containing a small amount of added alcohol, and appeared to have a remarkably similar unit cell. Crystals of apolactoferrin were orthorhombic, $a = 152{\cdot}1$, $b = 94{\cdot}6$, $c = 55{\cdot}8$ Å, space group $P2_12_12_1$, while those of diferric lactoferrin were also orthorhombic, $a = 156{\cdot}3$, $b = 97{\cdot}4$, $c = 55{\cdot}85$ Å, space group $P2_12_12_1$. Although no axial length differed by more than 3% between the two forms, the unit-cell volumes differed by only 5%, and the space groups were the same, the intensity distributions did, however, appear quite different.

Data collection was carried out at room temperature, using an Enraf–Nonius CAD-4 diffractometer equipped with a helium-filled diffracted-beam path. Intensities were obtained from a limited step scan (in ω) through each reflection peak, with Gaussian profiles being fitted (Hanson, Watenpaugh, Sieker & Jensen, 1979). With scan widths of 0·3–0·4°, 2000 to 3000 reflections per day could be measured. The stronger reflections (peak height > 2 times background) were used to define peak profiles in different regions of reciprocal space; these were then applied to the data as a whole. Backgrounds, one per reflection, were measured between reciprocal lattice points for short periods of time, then averaged in blocks of reciprocal space. Reflections with negative net intensities, or those that could not be fitted adequately, were omitted. All data were corrected for intensity decay of the crystals, from the intensities of five standard reflections, regularly monitored through each data collection, and were corrected for absorption (North, Phillips & Mathews, 1968).

Two crystals were used for the native data collection, covering the resolution range 20 to 2·8 Å resolution ($\theta = 2{\cdot}2$ to $16{\cdot}2°$). From the total of 19 413 measurements, 947 were omitted as negative intensities or inadequately fitted leaving 18 466 reflections of which 16 355 (84% of total) had intensities $I > 2\sigma_I$. After merging, a final data set of 17 783 unique reflections was obtained (merging R factor, $R_m = 0{\cdot}090$, where $R_m = \Sigma|I - \bar{I}|/\Sigma\bar{I}$).

Attempts were also made to prepare heavy-atom derivatives, but the limited number of crystals then available (less than ten) meant that these studies could not be pursued fully. Data were collected on several crystals soaked in dichloroethylenediamineplatinum(II) ($PtenCl_2$), since this had provided a good derivative in the diferric lactoferrin structure analysis (Anderson, Baker, Norris, Rice & Baker, 1989), but no satisfactory solution of difference Patterson maps could be obtained.

Structure analysis by molecular replacement

Choice of search model(s)

Given the four-domain structure of human lactoferrin (see Fig. 1) and the likelihood that conformational change from the diferric structure would involve relative domain and/or lobe movements, four different search models were used, all based on the diferric lactoferrin structure. These were the N-terminal half-molecule (domains N1 and N2, residues 4–334), the C-terminal half-molecule (domains C1 and C2, residues 345–691), a 'core' fragment comprising domains N1 and C1 (residues 4–90, 252–433 and 596–691), and the whole molecule (residues 4–691). The choice of the 'core' fragment was based on the observation that almost all of the contacts between the two lobes involved domains N1 and C1; thus these seemed likely to form an invariant 'core', with any relative domain movements likely to involve domains N2 and/or C2 which are much less constrained. The whole molecule was still considered a useful possibility because of the similarity of the unit cell to that of diferric lactoferrin, which suggested the conformational change might be very small.

Search models included all atoms (main chain and side chain) of the appropriate parts of the diferric structure, taken from a model refined at 2·2 Å resolution (crystallographic R factor 0·191 for a

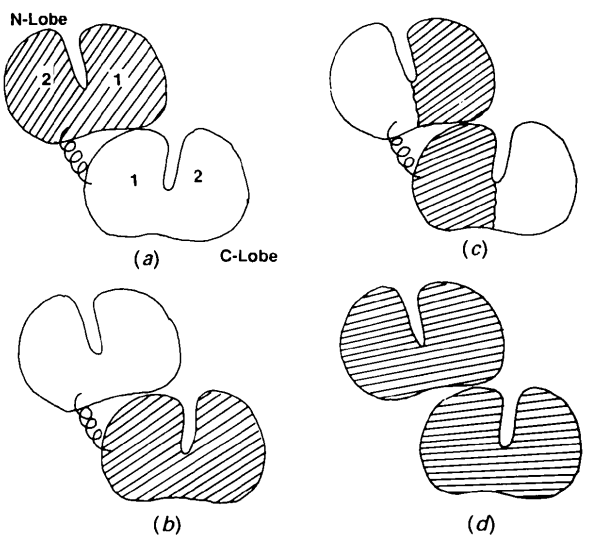

Fig. 1. Schematic diagrams showing search models (shaded portions) derived from the Fe_2Lf structure for rotation function calculations. (*a*) N-lobe, (*b*) C-lobe, (*c*) 'core', and (*d*) whole molecule.

model comprising 5320 protein atoms, 2 Fe^{3+} and 2 CO_3^{2-} ions, and 300 water molecules). The bound ions and solvent molecules were, of course, not included.

Rotation function

Each model was placed in a triclinic orthogonal cell, with edges $a = b = c = 100$ Å and the resulting structure factors were used in rotations against the observed apolactoferrin (ApoLf) data, using the fast rotation function of Crowther (1972). In addition the observed data for diferric lactoferrin (Fe_2Lf) were rotated against the observed ApoLf data. Data within the resolution limits 10 to 3·5 Å were used, without normalization, and all Patterson vectors between 10 and 35 Å considered in the overlap function. For a markedly ellipsoidal molecule such as lactoferrin (approximate dimensions 45 × 45 × 90 Å) this means that the longer vectors along the major axis will not be considered. The overlap function may therefore be less sensitive to small relative movements along this direction (*i.e.* of the two lobes). The results of these rotations are given in Table 1.

The most clear-cut solution was given by the C-lobe (C1 + C2) search model, with a peak of 7·7σ, and the same solution, around (110, 80, −115°) was also obtained from the 'core' (N1 + C1) and whole-molecule models, although at a slightly lower level (5–6σ). The other rotations also gave a peak in approximately the same position, but these were not the highest peak. Indeed the low level (2·1σ) of this peak when the N-lobe was used as search model was the first indication of conformational differences in this part of the molecule. The 'core' was then used in further rotations exploring the effects of variation in the resolution and Patterson sphere ranges. Limiting the vector range to 6·0–22·5 Å reduced the peak-to-background ratio considerably, while increasing the resolution (from 10·0–3·5 to 6·0–3·0 Å) had a similar, but less pronounced, effect. Finally, a finer grid search gave a 'best' peak position of (109, 83, −113·4°).

Translational search

The rotated model was positioned in the apolactoferrin cell using an *R*-factor search, with data in the resolution range 10·0 to 3·0 Å. Initial attempts involved moving the (N1 + C1 + C2) fragment of Fe_2Lf, rotated as above, systematically through the apolactoferrin unit cell. Increments of 1·25 Å were used, with a much finer step size (0·125 Å) used to more accurately define minima. The largest minimum, at (0·06, 0·00, 0·13 Å) and $R = 0{\cdot}400$, was 11·45σ below the mean level of the map; this solution, however, gave a large number of overlaps with symmetry-related molecules. A second translational search, this time using the whole molecule (N1 + N2 + C1 + C2), rotated as before, gave a minimum at (20·28, 2·61, 0·51 Å) and $R = 0{\cdot}423$, which was 13·4σ below the mean level of the map. When this shift vector was applied to the search model, no intermolecular overlaps were found. Moreover, it was found that this solution also corresponded to the second-deepest minimum (6·4σ) in the previous map. This was therefore taken as the correct solution.

Table 1. *Rotation function results*

Search model	Highest peak (α, β, γ) (°)	Height	Correct peak (α, β, γ) (°)	Height
Whole molecule	(110, 80, −115)	5·9σ	(110, 80, −115)	5·9σ
N-lobe	(20, 84, 70)	3·4σ	(105, 80, −90)	2·1σ
C-lobe	(110, 80, −115)	7·7σ	(110, 80, −115)	7·7σ
Core (N1 + C1)	(110, 85, −112·5)	5·0σ	(110, 85, −112·5)	5·0σ
Fe_2Lf data	(0, 0, 15)	2·5σ	(110, 80, −112·5)	1·6σ

Initial refinement and model correction

The rotated and translated whole molecule was next subjected to a series of rigid-body refinements, using the program *CORELS* (Sussman, 1985), in order to optimize the positioning of the molecule, and to try to account for any relative domain or lobe movements in the apolactoferrin structure. The model was refined first as a single rigid body (the whole molecule), then two (the two lobes), four (the four domains) and finally fourteen rigid groups (fourteen coherent structural elements) (see Table 2). At the same time the resolution limits of the data used were gradually increased, from 10·0–8·0 to 6·0–4·0 Å.

At the end of this procedure the crystallographic *R* factor appeared reasonable ($R = 0{\cdot}398$ for data in the range 10·0–2·8 Å resolution). However, the correlation coefficients never exceeded ~0·5, and a $2F_o - F_c$ electron density map showed good, well-fitted density for the three domains N1, C1 and C2, but weak, broken, uninterpretable density for the fourth domain, N2. It also appeared that the rigid-body refinement had moved N2 *closer* to N1, *i.e.* in a direction opposite from that expected. An attempt was made to improve the phasing by carrying out 20 cycles of conventional restrained least-squares refinement (allowing the geometry to relax considerably before re-tightening) using the program *TNT* (Tronrud, Ten Eyck & Matthews, 1987). This was followed by solvent flattening (Wang, 1985) using as masks both the whole molecule and the (N1 + C1 + C2) fragment.

Electron density maps with coefficients $3F_o - 2F_c$ were then calculated, based on the (N1 + C1 + C2) fragment, with and without a phase contribution from the solvent flattening. These maps showed little density which could be attributed to the N2 domain.

Table 2. *Rigid-body refinement*

Number of groups	Resolution range (Å)	Number of cycles	*R* factor*	Correlation coefficient†
(i) Initial model				
1[a]	10·0–8·0	6	0·462→0·456	
2[b]	10·0 8·0	5	0·456→0·440	
4[c]	10·0–8·0	6	0·440→0·395	0·484→0·530
	9·0 7·0	6	0·431→0·429	0·420→0·468
	8·0–6·0	9	0·450→0·430	0·378→0·427
	7·0–5·0	8	0·442→0·425	0·391→0·439
14[d]	7·0 5·0	6	0·425→0·409	0·464→0·479
	6·0 4·0	4	0·418→0·411	0·466→0·485
	10·0 4·0	1	0·417	0·491
(ii) After correctly placing N2 domain				
2[b]	10·0 8·0	2	0·328→0·307	0·677→0·725
4[c]	10·0 8·0	4	0·307→0·263	0·725→0·795
	10·0 6·0	4	0·325→0·314	0·721→0·743
7[e]	7·0 5·0	4	0·346→0·319	0·683→0·727

Notes: (*a*) Whole molecule: residues (4–691). (*b*) Two lobes: (4–333), (334–691). (*c*) Four domains: (4 90 + 252–333), (91–251), (334–433 + 596–691), (434–595). (*d*) Fourteen structural units: (4 55), (56–84 + 252–320), (85–93), (94–159 + 190 230 + 321–332), (160–189), (231–251), (333–344), (345–387), (388–416 + 595 663), (417–436), (437–496 + 527–575 + 664–677), (497–526), (576–594), (678–691). (*e*) Four domains, with N2 split into four parts: (4–90 + 252–333), (91–160), (161–190), (191–231), (232–251), (334–433 + 596–691), (434–595).

* *R* factor defined as $R = \sum| F_o - F_c |/\sum|F_o|$.

† Correlation coefficient defined as $[(\sum F_o F_c) - (\sum F_o)(\sum F_c)]/[\sum F_o^2 - (\sum F_o)^2][\sum F_c^2 - (\sum F_c)^2]^{1/2}$.

However, two short stretches of density did extend from residues 91 and 251, which were the cutoff points between domains N1 and N2. These stretches of density were taken as being the two interdomain connecting peptides, extending into domain N2, and were fitted as residues 91–96 and 247–251. Adjacent to these two was a third short strand of density (see Fig. 2) which we interpreted as residues 207–211; in the Fe_2Lf structure these three polypeptide strands form part of the central β-sheet of domain N2. The entire N2 domain from the Fe_2Lf structure, residues 91–251, was then fitted as a rigid group on the C_α positions of these 16 residues. This new position for the N2 domain, which was rotated about 50° from the previous position, gave rise to no unacceptable intermolecular contacts.

Rigid-body refinement using *CORELS* was then used again, on this revised model, to try to achieve optimum placement of the four domains. The results (Table 2) strongly suggested that the new model was indeed correct, with *R* values reduced from around 0·4 to around 0·3, and correlation coefficients increased to about 0·7–0·8. Domain N2 was moved by ~6° in this refinement but the other three domains were hardly changed (maximum rotation 0·4°). Further refinement in which the N2 domain was divided into four blocks also produced only small changes (1–2°).

Restrained least-squares refinement

Further refinement was by restrained least-squares refinement using the program *TNT* (Tronrud, Ten Eyck & Matthews, 1987), using all data in the resolution range 10·0 to 2·8 Å for which $I > 0{\cdot}5\sigma_I$ (17 686 reflections). In the initial six cycles geometric restraints were relaxed, in order to allow relatively large atomic movements; the *R* value dropped from 0·332 to 0·262 while the r.m.s. deviation of bond lengths from standard values increased to 0·13 Å. After a further 14 cycles, in which the geometry was gradually tightened and *R* increased to 0·280, a series of $2F_o - F_c$ 'omit' maps were calculated. Blocks of ~70 residues at a time were omitted from the model and three cycles of least-squares refinement carried out (to reduce bias in the calculated structure factors) prior to the calculation of each map; between

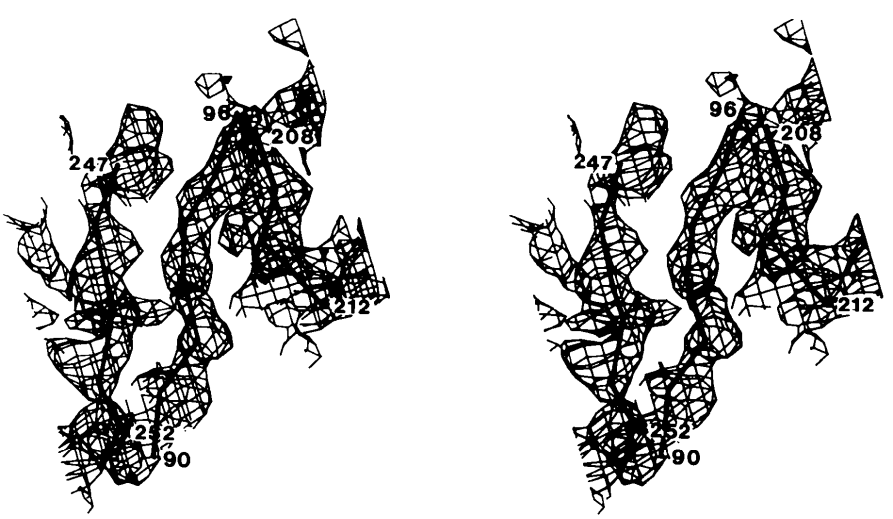

Fig. 2. Portion of the $3F_o - 2F_c$ electron density map calculated from a model comprising domains N1, C1 and C2, with phases further improved by solvent flattening. The three strands of electron density shown are the only clear density for the N2 domain; they were fitted as three strands of the N2 domain β-sheet and used to position the whole domain (treated as a rigid body). The α-carbon trace is shown for each strand.

them, ten such 'omit' maps covered the entire structure.

Manual rebuilding used the program *FRODO* (Jones, 1978) on an Evans and Sutherland PS330 interactive graphics system. An immediate indication of the correctness of the model was given by the appearance of good density for several side chains which had not been included in the original search models because of disorder in the Fe_2Lf structure. These included two side chains in the N2 domain (Lys 237 and Lys 241). Other side chains required some reorientation and particular attention was paid to residues at the joins between groups used for the rigid-body refinement, as some disortion was apparent in these regions.

Further cycles of restrained least-squares refinement then reduced the R value to 0·213 for a model with acceptable geometry (r.m.s. deviation from standard bond lengths 0·020 Å). Individual temperature factors were refined for the first time after ten cycles, when R was 0·249, followed by a further 20 cycles of refinement, in which geometrical restraints were relaxed and then tightened again. The refined model comprised 5231 atoms, representing residues 3–418 and 424–691; no density could be seen for residues 1–2, and only weak discontinuous density for 419–423 which form a poorly defined external loop in the Fe_2Lf structure also.

Discussion

The final refined structure deviates considerably from the Fe_2Lf structure used for the various search models in the molecular replacement analysis (see Fig. 3). The main differences are in the relative domain orientations, summarized in Table 3. Superpositions of each of the individual domains on to the corresponding domains of the Fe_2Lf stucture show that there is very little difference *within* domains, the r.m.s. deviation in C_α positions ranging from 0·43 Å for the C1 domain to 0·65 Å for the N2 domain (Table 3). Although these figures may change somewhat as the refinement is continued at higher resolution, it is clear that the major movements in going from the holo- to the Apo-Lf structure are rigid-body domain rotations.

The largest movement is the 54° rotation of the N2 domain, which causes the binding cleft in the N-terminal half of the molecule to be opened wide. This suggests a mechanism for metal binding in which first the anion (CO_3^{2-}) and then the metal ion (Fe^{3+}) bind to one domain while the molecule is in this 'open' conformation, followed by a conformational change in which the cleft is closed, by domain rotation, to complete the metal-ion coordination (Anderson, Baker, Norris, Rumball & Baker, 1990). The least movement is seen in the C-terminal half of the molecule, in which the cleft is still closed, even though no metal is bound. In fact the relative movement of domains C1 and C2 is only 1·3° (Table 3). The possible structural and functional implications of this have been discussed elsewhere (Anderson, Baker, Norris, Rumball & Baker, 1990).

Given the large conformational change in the N-lobe, involving rotational of the N2 domain, it is not surprising that the N-lobe search model gave a

Fig. 3. C_α stereo plots of structures of apolactoferrin (above) and diferric lactoferrin (below), showing the conformational transition from an open N-lobe in ApoLf to a closed N-lobe in Fe_2Lf.

Table 3. *Domain relationships between* Fe_2Lf *and ApoLf*

The diagonal elements (N1 *versus* N1 *etc.*) give the r.m.s. deviation in C_α positions when the corresponding domains in Fe_2Lf and ApoLf are individually superimposed. The off-diagonal elements show the relative angular rotations between pairs of domains in the two structures. These were computed, for example, by superimposing ApoLf on Fe_2Lf using the N1 domains to define the transformation matrix, and then seeing what further transformation was required to superimpose the N2 domains.

		Fe_2Lf			
		N1	N2	C1	C2
	N1	0·52 Å	53·3°	8·1°	7·7°
	N2		0·65 Å	54·6°	53·7°
ApoLf	C1			0·43 Å	1·0°
	C2				0·44 Å

poor solution to the rotation function. It is apparent, in retrospect, that this search model did give peaks for both the N1 and N2 domain orientations; the latter was in fact the highest in the map but was ignored because it was inconsistent with the solutions using other search models. On the other hand the C-lobe, even though it represents only half the structure, gave a good clear solution because its structure is essentially unchanged. Similarly the 'core' search model (N1 + C1) gave a good solution even though there is a small readjustment between the two lobes which results in a change of 8° in the relative orientations of N1 and C1 domains (Table 3). The success of the 'core' as a search model suggests that a correct solution to the rotation function could have been obtained even if both lobes had shown the 'open' conformation (*i.e.* with both N2 and C2 undergoing a large-scale rotation). It also emphasizes the value, when using molecular replacement on a multidomain protein structure, of careful examination of the structure when choosing a search model. (Alternatively one can choose a variety of search models, as we did, and seek a solution common to all).

An interesting sidelight to our selection of the 'core' as the best search model was an independent observation (R. C. Garratt and P. F. Lindley, personal communication), from comparison of the diferric lactoferrin (Anderson, Baker, Dodson, Norris, Rumball, Waters & Baker, 1987) and transferrin (Bailey, Evans, Garratt, Gorinsky, Hasnain, Horsburgh, Jhoti, Lindley, Mydin, Sarra & Watson, 1988) structures, that relative domain displacements occur between the two molecules, but that the domains which are displaced *least* are N1 and C1, again implying that they form a structural 'core'. A similar situation also occurs in the dimeric liver alcohol dehydrogenase structure. The ligand-induced conformational change for this enzyme involves rigid-body movements of its two catalytic domains relative to the two coenzyme-binding domains (Eklund, Samama, Wallen, Branden, Akeson & Jones, 1981), with the latter forming a similar core to the two-subunit four-domain structure.

It is not clear why the whole molecule (including the incorrectly placed N2 domain) gave a better solution to the translational search than the three-domain (N1 + C1 + C2) model; in retrospect, however, the same correct solution was given by both, albeit as only the second-lowest minimum in the latter case. *CORELS* refinement of the original solution (with N2 wrongly placed) did achieve some improvement, since small movements of N1, C1 and C2 domains did occur, and these then hardly moved at all in subsequent refinement after N2 had been re-positioned. On the other hand, even with a simple model of only four rigid groups (the four domains) and low-resolution data (10–8 Å) the radius of convergence was insufficient to move N2 in the right direction. The best indication that the initial structure was wrong in some fundamental way was given by the low value for the correlation coefficient, which remained less than 0·5; as soon as N2 was corrected this increased to 0·7–0·8.

A curious feature of the apolactoferrin crystal structure is the similarity of the unit-cell dimensions to those of the diferric lactoferrin crystals (and with the same space group) even though the crystal packing is quite different. The packing in the two cases is illustrated schematically in Fig. 4. The molecular

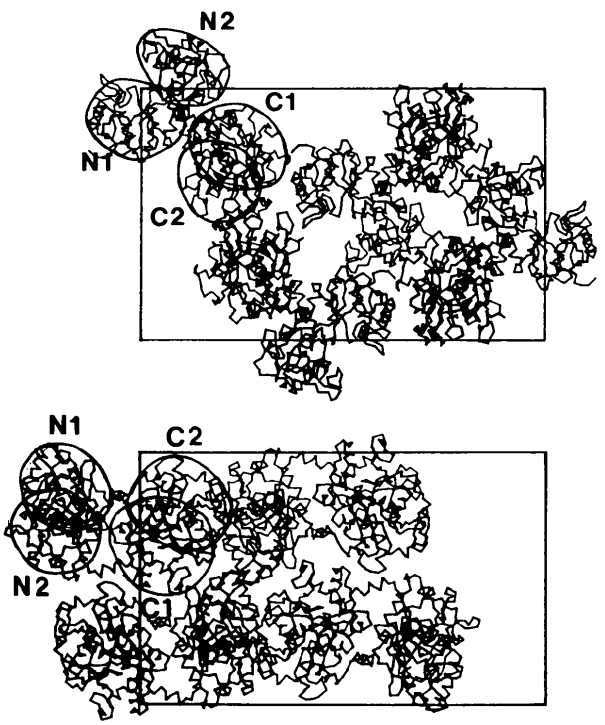

Fig. 4. Schematic diagrams of molecular packing in unit cells of apolactoferrin (above) and diferric lactoferrin (below) ([001] projection). In each case one molecule is outlined to identify the four domains.

orientation in the apolactoferrin cell is related to that in the diferric lactoferrin cell by a rotation of 90° about the *a* axis, 20° about *b*, followed by 20° about *c*. The similarity in unit cells seems to be largely coincidental, although in the *c* direction at least it may be attributed to the ellipsoidal shape of the molecules; a rotation of ~90° about *a* does not change the *c* dimensions of the molecule.

Finally, although molecular replacement has been successful in this structure analysis, a recent elaboration of the method, in which molecular dynamics calculations are used to give flexibility to a search model (Brünger, 1990) may make problems of this type much more amenable to solution.*

We wish to thank Dr P. M. Colman for helpful discussions on both the molecular replacement procedures and the rigid-body refinement, and Mrs E. J. Dodson for helpful advice on the molecular replacement methods and the *CCP*4 program package. We are grateful to Dr P. F. Lindley and members of the Birkbeck College transferrin group for help, encouragement and provision of unpublished results, and we gratefully acknowledge financial support from the US National Institutes of Health (grant HD-20859), the Medical Research Council of New Zealand and the New Zealand Dairy Research Institute.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory, and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37051 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

AISEN, P. & LISTOWSKY, I. (1980). *Annu. Rev. Biochem.* **49**, 357–393.
ANDERSON, B. F., BAKER, H. M., DODSON, E. J., NORRIS, G. E., RUMBALL, S. V., WATERS, J. M. & BAKER, E. N. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 1769–1773.
ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RICE, D. W. & BAKER, E. N. (1989). *J. Mol. Biol.* **209**, 711–734.
ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RUMBALL, S. V. & BAKER, E. N. (1990). *Nature (London)*, **344**, 784–787.
BAILEY, S., EVANS, R. W., GARRATT, R. C., GORINSKY, B., HASNAIN, S., HORSBURGH, C., JHOTI, H., LINDLEY, P. F., MYDIN, A., SARRA, R. & WATSON, J. L. (1988). *Biochemistry*, **27**, 5804–5812.
BAKER, E. N. & RUMBALL, S. V. (1977). *J. Mol. Biol.* **111**, 207–210.
BAKER, E. N., RUMBALL, S. V. & ANDERSON, B. F. (1987). *Trends Biochem. Sci.* **12**, 350–353.
BROCK, J. H. (1985). *Metalloproteins*, Part II, edited by P. HARRISON, pp. 183–262. London: MacMillan.
BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46–57.
CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178, New York: Gordon and Breach.
EKLUND, H., SAMAMA, J.-P., WALLEN, L., BRANDEN, C.-I., AKESON, A. & JONES, T. A. (1981). *J. Mol. Biol.* **146**, 561–587.
HANSON, J. C., WATENPAUGH, K. D., SIEKER, L. & JENSEN, L. H. (1979). *Acta Cryst.* A**35**, 616–621.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KILAR, F. & SIMON, I. (1985). *Biophys. J.* **48**, 799–802.
NORRIS, G. E., BAKER, H. M. & BAKER, E. N. (1989). *J. Mol. Biol.* **209**, 329–331.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
ROSSENAU-MOTREFF, M. Y. F., SOETEWEY, R., LAMOTE, R. & PEETERS, H. (1971). *Biopolymers*, **10**, 1039–1048.
SUSSMAN, J. L. (1985). *Methods Enzymol.* **115**, 271–303.
TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
VIGH, R., CSER, L., KILAR, F. & SIMON, I. (1989). *Arch. Biochem. Biophys.* **275**, 181–184.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Acta Cryst. (1991). B**47**, 1004–1010

Polymorphism of L-Glutamic Acid: Decoding the α–β Phase Relationship *via* Graph-Set Analysis

BY JOEL BERNSTEIN

Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva 84105, *Israel*

(*Received* 6 *February* 1991; *accepted* 30 *July* 1991)

Abstract

Graph-set analysis is employed to study the similarities and differences in the hydrogen-bonding patterns of the two polymorphs of L-glutamic acid. There are considerable similarities between the two structures, and only the higher-order graph sets reveal the differences. These are used to gain insight into the crystallization chemistry of the two forms as well as the structural relationship between them and the significantly different molecular conformations found in them.

0108-7681/91/061004-07$03.00