Acta Cryst. (1995). D**51**, 629–646

Structure of Human Diferric Lactoferrin Refined at 2.2 Å Resolution

BY M. HARIDAS, BRYAN F. ANDERSON AND EDWARD N. BAKER*

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

(*Received* 30 *August* 1994; *accepted* 22 *November* 1994)

Abstract

The three-dimensional structure of the diferric form of human lactoferrin has been refined at 2.2 Å resolution, using synchrotron data combined with a lower resolution (3.2 Å) diffractometer data set. Following restrained least-squares refinement and model rebuilding the final model comprises 5330 protein atoms (691 residues), $2Fe^{3+}$ and $2CO_3^{2-}$ ions, 469 solvent molecules and 98 carbohydrate atoms (eight sugar residues). Root-mean-square deviations from standard geometry are 0.015 Å for bond lengths and 0.038 Å for angle (1–3) distances, and the final crystallographic R factor is 0.179 for all 39 113 reflections in the resolution range 8.0–2.2 Å. A close structural similarity is seen between the two lobes of the molecule, with differences mainly in loops and turns. The two binding sites are extremely similar, the only apparent differences being a slightly more asymmetric bidentate binding of the carbonate ion to the metal, and a slighly longer Fe—O bond to one of the Tyr ligands, in the N-lobe site relative to the C-lobe site. Distinct differences are seen in the interactions made by two cationic groups, Arg210 and Lys546, behind the iron site, and these may influence the stability of the two metal sites. Analysis of interdomain and interlobe interactions shows that these are few in number which is consistent with the known flexibility of the molecule with respect to domain and lobe movements. Internal water molecules are found in discrete sites and in two large clusters (in the two interdomain clefts) and one tightly bound water molecule is present 3.8 Å from the Fe atom in each lobe. The carbohydrate is weakly defined and has been modelled to a limited extent; two sugar residues of the N-lobe glycan and six of the C-lobe glycan. Only one direct protein–carbohydrate contact can be found.

Introduction

Lactoferrin is a member of the family of iron-binding proteins known collectively as the transferrins. These proteins, which also include serum transferrin and ovotransferrin, serve to control the levels of iron (and possibly other elements) in the body fluids of vertebrates, and some invertebrates. As a group, they also have antibacterial and growth-factor activities, while individual proteins have additional specific roles, *e.g.* transferrin as the iron transport protein in serum, and lactoferrin, which is found in white blood cells and secretory fluids (milk, tears, saliva), as a modulator of the immune and inflammatory responses. Numerous reviews of this protein family are available, covering structure and function (Baker & Lindley, 1992; Baker, 1993), physiochemical properties (Harris & Aisen, 1989) and biological roles (Brock, 1985).

A characteristic property of transferrins is their ability to bind, very tightly ($K_{app} \simeq 10^{20}$) but reversibly, two Fe^{3+} ions together with two CO_3^{2-} ions, with a unique synergistic relationship between metal and anion binding (Schlabach & Bates, 1975). Crystallographic studies of human lactoferrin at 3.2 and 2.8 Å (Anderson, Baker, Dodson, Norris, Rumball, Waters & Baker, 1987; Anderson, Baker, Norris, Rice & Baker, 1989) and of rabbit serum transferrin at 3.3 Å resolution (Bailey, Evans, Garratt, Gorinsky, Hasnain, Horsburgh, Jhoti, Lindley, Mydin, Sarra & Watson, 1988), both in their diferric forms, have established the basic features of the transferrin three-dimensional structure. Each protein consists of a single polypeptide chain of 650–700 amino-acid residues, folded into two globular lobes comprising the N-terminal and C-terminal halves of the molecule. The nature and location of the binding sites were also established, one in each lobe, in a deep cleft between two domains, with each Fe^{3+} ion ligated by two Tyr side chains, one His, one Asp and the CO_3^{2-} anion.

Many structural and functional details remain to be clarified by higher resolution analyses, however. The bidentate CO_3^{2-} binding indicated by the 2.8 Å resolution structure refinement of human lactoferrin (Anderson, Baker, Norris, Rice & Baker, 1989) has since been confirmed in crystallographic analyses of the N-terminal half-molecules of rabbit serum transferrin at 2.3 Å resolution (Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990) and human lactoferrin at 2.0 Å resolution (Day, Anderson, Tweedie & Baker, 1993). There are, however, differences between the N- and C-terminal sites which can be detected spectroscopically (Brock, 1985; Harris & Aisen, 1989) and which become more pronounced in metal- and anion-substitution studies (Smith, Anderson, Baker & Baker, 1992; Shongwe, Smith, Ainscough, Baker, Brodie & Baker, 1992). Detailed comparisons of the two sites are necessary if such differences are to be

*To whom correspondence should be addressed.

ISSN 0907-4449

understood in structural terms. This includes the role of solvent, since both NMR (Koenig & Schillinger, 1969) and EXAFS (Hasnain, Evans, Garratt & Lindley, 1987) studies have indicated the presence of water molecules close enough to the binding sites to influence their properties.

Functional differences also exist, both between the N- and C-terminal sites of a given protein and between different members of the transferrin family. The N-terminal site is more acid labile (Evans & Williams, 1978) and also releases iron more rapidly *in vitro* (Kretchmar & Raymond, 1986) while it is the C lobe of transferrin that is specifically acted on by its cellular receptor (Bali & Aisen, 1991). The greater acid stability of lactoferrin relative to serum transferrin, as far as iron release is concerned, may have important implications for their *in vivo* roles and may depend on variations in interdomain or interlobe interactions (Baker & Lindley, 1992). Finally, for neither lactoferrin nor transferrin has the functional role of the attached carbohydrate been elucidated, and the lower resolution analyses gave little indication of its role in the protein structures.

Here we report the crystallographic refinement of the structure of human diferric lactoferrin at the highest resolution attainable from the X-ray data (2.2 Å) together with an analysis of important details of the protein conformation, the metal and anion sites, interlobe and interdomain interactions, solvent structure and as much of the carbohydrate structure as is visible crystallographically.

Experimental

Crystallization and data collection

Human lactoferrin was purified from breast milk, and crystallized in its diferric form, as described previously (Baker & Rumball, 1977). Prior to data collection the crystals were transferred, over a period of a week, from their original mother liquor [0.01 *M* phosphate buffer, pH 7.8, containing 10%(*v*/*v*) ethanol] to a stabilizing mother liquor [0.01 *M* phosphate buffer, pH 7.8, 20%(*v*/*v*) 2-methyl-2,4-pentanediol]; this significantly stabilized the crystals to temperature and radiation damage. These crystals are orthorhombic, $a = 156.26$, $b = 97.40$, $c = 55.85$ Å, space group $P2_12_12_1$, with one molecule in the asymmetric unit.

X-ray diffraction data were collected by oscillation photography at the Synchrotron Radiation Source, Daresbury, England. The wavelength used was 0.88 Å, with a crystal-to-film distance of 123.7 mm. Two crystals were used for the collection of a full data set; the first, mounted about b^* (the longest crystal dimension), was rotated through the full unique 90° in 72 abutting oscillations varying between 2.0 and 0.7°, while the second, mounted about a^* (with b^* parallel to the beam) was used to collect 12 oscillations of 2.0° each and so provide the data missed in the blind region for the first crystal. The 84 film packs (of three films each) were scanned and digitized at 100 μm intervals with an Optronics P1000 rotating-drum microdensitometer. Film images were processed with the *MOSFLM* set of programs developed by Wonacott (1980), with both full and partial reflections being retained.

Scaling and merging made use of the programs *ROTAVATA* and *AGROVATA* of the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Data for each of the two crystals were treated independently to give two data sets, in each of which the partial reflections were added between adjacent films. Some details of these data sets are given in Table 1. Finally these two data sets were combined with the 3.2 Å diffractometer data set collected previously (Anderson, Baker, Norris, Rice & Baker, 1989). The three data sets were merged to give a total of 39 506 independent reflections, of which 31 560 had intensities greater than 3σ, these representing 90.8% of the data theoretically accessible to 2.2 Å resolution (Table 1). The overall merging R value was 0.11. Although the crystals show considerable thermal diffuse scattering and the intensity of the diffraction pattern falls off rapidly beyond 2.5 Å resolution, the proportion of observed data is still reasonable at the nominal cutoff of 2.2 Å resolution (the outermost shell, 2.3–2.2 Å is 80.4% complete, with 51% of the reflections having $I > 3\sigma_I$). A Wilson plot (Wilson, 1949) for the final 2.2 Å data set gave an estimate of 38 Å^2 for the overall temperature factor.

Refinement

The starting model for refinement was that previously described (Anderson, Baker, Norris, Rice & Baker, 1989), which followed refinement against medium-resolution (2.8 Å) data. This included the two Fe^{3+} and two CO_3^{2-} ions, but no solvent. All temperature parameters were re-set to 20 Å^2 prior to refinement. Refinement was against the 39 113 data (*i.e.* all observed data) in the resolution range 8.0–2.2 Å, although data to a maximum spacing of 20 Å were included in all electron-density map calculations.

Initial refinement was with *TNT* Version 2.1 (Tronrud, Ten Eyck & Matthews, 1987). Weights were adjusted so that when the r.m.s. deviation of bond lengths from standard values was 0.02 Å, the equivalent values for interbond angles, planar groups and non-bonded contacts were 3.0°, 0.01 and 0.2 Å, respectively. Refinement then proceeded in a series of rounds; each round consisted of a number of refinement cycles during which the weighting scheme was gradually relaxed until the r.m.s. deviation in bonds approached 0.08 Å, at which point the weights were gradually tightened again to bring the geometrical restraints back to their original values. Each round of loosening and tightening generally comprized 20–25 cycles. These rounds of least-squares

Table 1. *Data collection and processing*

(*a*) Data collection

Diffractometer (CAD-4)		Oscillation data (synchrotron)		
Crystal number	1	Crystal number	1	2
Total measurements*	18840	Rotation axis	b*	a*
Independent reflections	14867	No. of film packs	72	12
Reflections with $I > 2\sigma_I$	13529	Average R_{sym}†	0.045	0.056
R_{merge}‡	0.079	Total measurements*	111743	53974
		Independent reflections	38839	17294
		Reflections with $I > 2\sigma_I$	32274	12254
		R_{merge}‡	0.065	0.072

(*b*) Statistics for combined data*

d_{min} (Å)	Reflections	Completeness (%)	Multiplicity	R_{merge}‡	$I > 3\sigma_I$ (%)
6.60	1778	98.3	2.6	0.092	98.0
4.68	3012	99.4	3.6	0.099	97.9
3.83	3855	99.1	3.6	0.109	97.6
3.32	4376	97.3	4.3	0.118	96.2
2.97	4794	94.3	4.5	0.109	91.3
2.71	5162	92.2	4.6	0.115	84.5
2.51	5418	89.2	4.5	0.125	72.2
2.35	5517	85.1	4.6	0.143	61.6
2.20	5594	80.4	4.7	0.144	50.9
Overall	39506	90.8	4.4	0.110	79.9

* Diffractometer data to d_{min} 3.2 Å, synchrotron data to d_{min} 2.2 Å.

† $R_{sym} = \sum |I - \bar{I}| / \sum I$, summed over all symmetry-related reflections of the same film pack.

‡ $R_{merge} = \sum |I - \bar{I}| / \sum I$.

refinement were interspersed with visual evaluation of the model and rebuilding from $2F_o - F_c$ maps, using the program *FRODO* (Jones, 1978), on an Evans and Sutherland PS330 graphics system.

After the first round of refinement, when the overall R factor had been reduced from 0.325 to 0.301, the two Fe^{3+} ions, two CO_3^{2-} ions and the side chains involved in metal and anion binding, were removed from the model and three further cycles of least-squares refinement were carried out. A subsequent $2F_o - F_c$ electron-density map confirmed the original positions of these moieties. It had been noted during the previous refinement cycles, however, that the Fe-atom position had tended to fluctuate considerably. For all subsequent refinement, therefore, weak restraints ($\sigma = 0.05$ Å) were placed on the iron–ligand bond lengths. The initial target bond lengths were set at 2.0 Å, with these values being periodically updated if consistent movement appeared to be occurring as a result of refinement.

After two further rounds of refinement and rebuilding ($R = 0.254$) a series of 'omit' maps, in which portions of the structure, representing 10% of total atoms, had been removed from the model for the preceding three refinement cycles, were calculated. These were used for a thorough review of the entire protein structure. Most of it appeared well placed within good electron density, but a number of loop conformations were corrected, some side chains were remodelled and three *cis*-prolines were identified, at positions 71, 142 and 628.

Refinement from this point was with the fast Fourier version of *PROLSQ* (Konnert & Hendrickson, 1980), as implemented in the *CCP*4 program suite. All atoms were given individual isotropic B values. Solvent molecules (regarded as water) were added cautiously to the model from $2F_o - F_c$ and $F_o - F_c$ maps; they were only included if they had peak heights exceeding 3σ in $F_o - F_c$ maps, had persisted through several rounds of refinement, were in hydrogen-bonding distance of one or more potential hydrogen-bond partners, and were not close to regions where the protein density was ambiguous. A second major re-evaluation of the model, through 'omit' maps, was carried out after a further five rounds of refinement ($R = 0.185$). It was also found that a number of non-bonded contacts consistently refined to less than 3.0 Å, and a somewhat larger number of hydrogen bonds refined to less than 2.5 Å (though none less than 2.3 Å). These were in well defined regions of the structure and their behaviour was attributed to a lack of sufficient resolution in the X-ray data; the imposition of weak restraints ($\sigma = 0.05$ Å) on 92 such distances rectified this problem.

The full progress of the refinement is summarized in Table 2. Although the R factor was not appreciably reduced in the later stages, the quality of the structure, as judged by geometrical and structural criteria (hydrogen bonding, conformational angles, *etc.*) was significantly improved. In the final stages it was also possible to place some of the more poorly defined parts of the structure, in particular the first few sugar residues of the two carbohydrate chains. Although these did not refine well and apparently make little contribution to the X-ray data, their density did appear consistently in the same place during the refinement, and we concluded that their general positions, at least, could be modelled.

Results

The final model

The final atomic coordinate set contains 5330 protein atoms (from 691 amino-acid residues), two Fe^{3+} ions, two CO_3^{2-} ions, 469 solvent molecules (all modelled as water) and 98 carbohydrate atoms (from four *N*-acetylglucosamine, two fucose and two mannose residues). The protein and carbohydrate structures have geometry close to ideal with r.m.s. deviations of 0.015 and 0.038 Å from standard values of bond lengths and angle (1–3) distances, respectively. The final value of the crystallographic R factor was 0.179 for all 39 113 reflections between 8.0 and 2.2 Å resolution. Refinement statistics are summarized in Table 3.

A Luzzati plot (Luzzati, 1952) of the R factor as a function of resolution gives a value of between 0.25 and 0.30 Å for the average error in atomic coordinates, while a σ_A plot (Read, 1986) gives a similar but slightly higher value of 0.35 Å. Both plots are shown in Fig. 1. Being average maximum values, which assume that the disagreement in observed and calculated structure factors results solely from the coordinates, these overestimate

Table 2. *Course of the refinement*

Round	Atoms	*R* factor	Comments
1	5165	0.325→0.301	Refined with *TNT*, *xyz* refinement only, 5165 protein atoms. Removed Fe^{3+} and CO_3^{2-} ions and ligands for last three cycles.
2	5193	0.304→0.288	Fe^{3+} and CO_3^{2-} ions and ligands replaced. Two NAG residues added. Refined with *TNT*, *xyz* refinement only, 5165 protein atoms, 28 carbohydrate atoms.
3	5254	0.314→0.254	9 side chains and 23 water molecules added. Refined with *TNT*, *B*-factor refinement initiated, 5203 protein atoms, 28 carbohydrate atoms, 23 water molecules.
4	5398	0.295→0.241	Complete rebuild from series of omit maps (blocks of 50 residues removed). Arg residue inserted between 120 and 121. Three *cis* Pro identified. 10 more side chains and 96 waters added. Refined with *TNT*, 5251 protein atoms, 28 carbohydrate atoms, 119 water molecules.
5	5556	0.271→0.201	7 more side chains and 160 waters added, NAG residues removed. 49 residues rebuilt. Refined with *PROLSQ*, 5277 protein atoms, 279 water molecules.
6	5640	0.223→0.183	Complete check against omit maps. 69 residues rebuilt, 14 side chains added, all waters with $B > 60$ Å^2 removed, 51 new waters added. Refined with *PROLSQ*, 5329 protein atoms, 311 water molecules.
7	5637	0.237→0.180	Water model checked. Restraints imposed to correct 74 close contacts. Refined with *PROLSQ*, 5328 protein atoms, 309 water molecules.
8	5910	0.187→0.181	2 N-terminal residues, 7 carbohydrate residues, 201 waters added. 24 waters removed. Refined with *PROLSQ*, 5340 protein atoms, 486 water molecules, 84 carbohydrate atoms.
9	5907	0.182→0.179	1 carbohydrate residue added, water structure reassessed, 17 waters removed. Close contact restraints removed. Refined with *PROLSQ*, 5340 protein atoms, 469 water molecules, 98 carbohydrate atoms.

Table 3. *Refinement statistics*

Resolution limits (Å)	8.0–2.2
Final *R* factor	0.179
No. of reflections used*	39113
No. of protein atoms	5330
No. of solvent molecules	469
Other ions	2 Fe^{3+}, 2 CO_3^{2-}
No. of carbohydrate atoms	98
Average *B* value (protein atoms) (Å^2)	41.3
Average *B* value (all atoms) (Å^2)	43.9
Geometrical deviations†	
Bond lengths (Å)	0.015
Angle distances (Å)	0.038
Planarity (Å)	0.013
Chiral volume (Å^3)	0.223
Non-bonded contacts (Å)	0.205

* Data with $I > 1.0\sigma_I$.
† R.m.s. deviations from standard values.

the error in the well defined parts of the structure such as the two binding sites; we estimate the error in these regions to be of the order of 0.2 Å.

The mean *B* value for all protein atoms is 41.3 Å^2. This relatively high value can be attributed in part to the inclusion of several poorly defined parts of the structure (see below) but it also reflects the overall ordering of the crystals. Nevertheless, the distribution of *B* values along the polypeptide chain follows the expected pattern, with secondary structures and internal regions having lower values and external loops much higher values. This can be clearly seen in Fig. 2. The average *B* value for the atoms of the metal- and anion-binding sites (Fe^{3+} and CO_3^{2-} ions, iron ligands and anion-binding residues, including the helix 5 N terminus; 81 atoms) is relatively low, being 21.8 Å^2 in the N lobe and 20.7 Å^2 in the C lobe. The most mobile parts of the structure, on the other hand, with the highest *B* values (yellow in Fig. 2), are in the N1 domain (the N terminus, the end of helix 1 and several nearby loops) and the C1 domain (the loop 417–423).

Most of the protein structure is represented by well defined electron density (see for example Fig. 3). The fit of the model to the electron density has been quantified by calculating correlation coefficients which provide a residue-by-residue measure of the agreement between model electron density and the final $2F_o - F_c$ electron-density map. Only 2% of residues have main-chain correlation coefficients less than 0.9, and 8% have side-chain correlation coefficients less than 0.9.

The least satisfactory regions of the structure are listed in Table 4. All have high *B* values, and although there is sufficient density to mark their general location, this density is in each case rather weak and fragmented, which is indicative of disorder. Examples are shown in Fig. 4. The N-terminal residues 1–4 project out from the surface of the N1 domain. Modelling their position is made difficult not only by an evident positional disorder but also by a likely chemical heterogeneity; sequencing shows that lactoferrins from different sources can differ in the number of Arg residues at the N terminus, from one to four (Stowell, Rado, Funk & Tweedie, 1991). Residues 293 and 294 are the two central residues of a

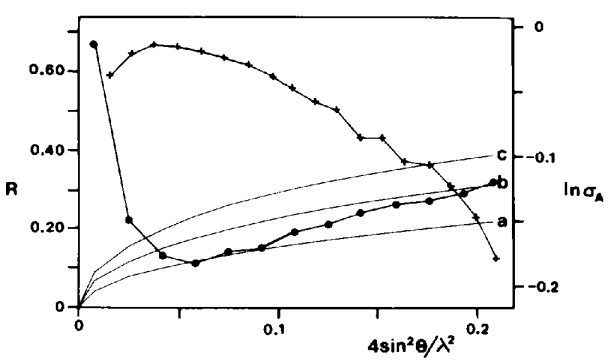


Fig. 1. Plots of the variation of the crystallographic *R* factor (•) and $\ln(\sigma_A)$ (+) with resolution. The theoretical variation of the *R* factor (Luzzati, 1952) for coordinate errors of 0.20, 0.25 and 0.30 Å (curves *a*, *b* and *c*, respectively) is shown. For the σ_A plot (Read, 1986) the gradient of the best fit line between 3.4 and 2.2 Å gives an estimate of 0.35 Å for the r.m.s. coordinate error in the refined model.

Table 4. *Poorly defined regions of the structure*

Residues	Average B value (Å^2)	Comment
1–4	96.5	N terminus. Projects into solvent. May show both spatial and chemical heterogeneity.
293–294	94.1	Central two residues of projecting β-turn. Probably disordered.
417–423	94.0	Large flexible surface loop. No interactions with rest of molecule.
900–901	81.1	Sugar residues of two carbohydrate chains. Little interaction with protein structure. Flexible and probably subject to both spatial and chemical heterogeneity.
910–915	83.8	

β-turn that projects into solution. Several alternative conformations were tried for these residues without any better fit to the density being obtained. The side-chain OH of Ser293 contacts residues 420–422 of a neighbouring molecule but the latter also appear disordered. Residues 417–423 form a large extended surface loop which makes no interactions with the rest of the protein structure and has little evident internal structure. Although density marks the course of this loop (Fig. 4*a*) it is poorly defined and is apparently flexible or disordered. In apo-lactoferrin, which has different crystal packing, this loop takes a rather different course, associated with crystal contacts (Norris, Jameson, Anderson, Thomas & Baker, 1995).

The two carbohydrate chains attached to human lactoferrin are known to be heterogeneous, being represented by at least three chemical structures (Spik, Strecker, Fournet, Bouquelet, Montreuil, Dorland, Halbeek & Vliegenthart, 1982). In the initial MIR-phased electron-density maps some weak carbohydrate density was apparent, but no sugar residues were modelled at that stage. As refinement proceeded this density tended to 'refine away'. At the end, only the use of combined phases, MIR phases to 3.2 Å with calculated (model) phases to 2.2 Å, together with the inclusion of all data with spacings less than 20 Å, gave density with sufficient continuity to be modelled at all. The inclusion of the low-resolution data was particularly important. The terminal *N*-acetylglucosamine residue of the N-lobe glycan chain, and its attached fucose, have quite good density in the final map (Fig. 4*b*) but no further residues of this chain have been modelled. Six sugar residues of the C-lobe glycan chain are included although their electron density (in the combined-phased map) is only continuous at the 0.5σ level; clearly the carbohydrate is subject to spatial disorder or flexibility as well as chemical heterogeneity.

The polypeptide-chain torsion angles are plotted in Fig. 5, in the familiar Ramachandran plot (Ramakrishnan & Ramachandran, 1965), with the various allowed regions defined as by Laskowski, MacArthur, Moss & Thornton (1993) in the program *PROCHECK*. The 'most favoured' regions account for 82.7% of non-glycine residues, while a further 16.6% of residues occupy the 'additional allowed' regions. Only two residues, Leu299 and Leu642, occupy normally disallowed regions; these are the central residues of two γ-turns which are in equivalent positions in the N and C lobes. Their (φ, ψ) values of (82, −50°) and (75, −49°), respectively, are very similar to the characteristic values for classic γ-turns (Matthews, 1972; Nemethy & Printz, 1972), which are around (70, −50°) (Baker & Hubbard, 1984). These two γ-turns are located in part of one wall of the binding cleft of each lobe, and with their highly conserved sequence Leu-Leu-Phe seem likely to be a common feature of transferrin structure (Smith, Anderson, Baker & Baker, 1995). Two other residues, Asn52 and Ser191, occupy the 'generously allowed' regions. Asn52 has the left-handed α configuration, as it is at the C-terminus of a helix with an α_{C2} termination pattern (Baker & Hubbard, 1984). Ser191 has the ε configuration (Sibanda, Blundell & Thornton, 1989), with (φ, ψ) angles (70, −167°); this residue makes an important contribution to the N-lobe binding site, hydrogen bonding with the anion-binding Arg121. The equivalent residue in the C lobe is Gly527, which has the same configuration.

Polypeptide chain folding

The organization of the lactoferrin molecule into two lobes (N lobe and C lobe) and four domains (N1, N2, C1 and C2) has been described previously (Anderson, Baker, Norris, Rice & Baker, 1989), and is shown as a C_α plot in Fig. 6.

The high level of sequence identity between the two halves of the molecule (42%) is reflected in very similar folding of the polypeptide chain. The two lobes are related by an approximate twofold screw axis, the C lobe being superimposable on the N lobe by a rotation of 179.5° and a translation of 24.5 Å, following the transformation matrix,

$$\begin{pmatrix} x_N \\ y_N \\ z_N \end{pmatrix} = \begin{pmatrix} 0.33988 & -0.40748 & -0.84761 \\ -0.39225 & -0.88055 & 0.26603 \\ -0.85477 & 0.24206 & -0.45911 \end{pmatrix} \times \begin{pmatrix} x_C \\ y_C \\ z_C \end{pmatrix} + \begin{pmatrix} -7.727 \\ 46.852 \\ 14.292 \end{pmatrix}$$

When the two lobes are superimposed, a total of 300 residues from each lobe (out of ~ 330) can be matched with an r.m.s. deviation in C_α positions of 1.28 Å. If atoms whose deviation is greater than 3σ are omitted, this falls to an r.m.s. deviation of 1.00 Å for 264 C_α positions. The structurally equivalent regions of the two lobes are listed in Table 5, and the deviations along the course of the polypeptide chain are plotted in Fig. 7. The regions not included are primarily at and around the positions of insertions and deletions (for a sequence alignment, see Anderson, Baker, Norris, Rice & Baker, 1989) together with the final helix of each lobe (helix 12). The latter have quite different orientations; helix 12 of the N lobe

Table 5. *Structurally equivalent regions of the two lobes*

N-lobe residues	C-lobe residues	R.m.s. deviation*
5–27	344–366	1.32
33–69	368–404	1.03
73–83	406–416	0.47
91–102	434–445	0.75
103–133	447–477	0.84
148–163	484–499	1.34
167–175	501–509	0.88
179–218	515–554	0.67
223–259	567–603	0.71
262–280	604–622	1.52
285–291	630–636	0.73
296–310	639–653	0.99
314–332	657–675	0.91

* R.m.s. deviation (Å) for C_α atoms of each section after superposition based on all structurally equivalent regions.

joins the two lobes, whereas helix 12 of the C lobe folds into the interlobe contact area, but is held against the C lobe by a disulfide bridge.

Superposition of the lobes as a whole, however, underestimates the similarity of their folding because the relative orientations of the two domains differ. If the N1 and C1 domains are superimposed the N2 and C2 domains are found to be slightly out of register, and a further rotation of 6.3° is required to bring them into coincidence; the N lobe is more open than the C lobe by this amount. For the individual domains the r.m.s. deviations after superposition are 0.86 Å for N1 and C1 (132 C_α atoms) and 0.75 Å for N2 and C2 (134 C_α atoms).

Secondary structure

Each domain is built around an α/β folding pattern which has a central mixed β-sheet with helices packed against it. The identities and some descriptive features of the various secondary-structure elements are given in Table 6. The topology of the β-sheets has been described previously (Anderson, Baker, Norris, Rice & Baker, 1989) and the nomenclature for the helices and β-strands follows that used for transferrin (Bailey, Evans, Garratt, Gorinsky, Hasnain, Horsburgh, Jhoti, Lindley, Mydin, Sarra & Watson, 1988) as well as lactoferrin.

Refinement at higher resolution has not changed the β-sheet hydrogen-bonding pattern seen at 2.8 Å resolution. Some details of the helices are altered, however. A number of short 3_{10} helices have been identified and are included in Table 6. Residues 105–110 in the N lobe and 449–454 in the C lobe were labelled as helix 4 in the lower resolution analysis; while the course of the chain in each case is roughly helical only one or two 3_{10}-type (1···4) hydrogen bonds are now found. A number of helices have characteristic hydrogen-bonding termination patterns (Table 6), notably the α_{C2} configuration (Baker & Hubbard, 1984) in which a 1···6, 2···5 hydrogen-bonding pattern at the C terminus is associated with the penultimate residue (5) having a left-handed α conformation. The most unusual feature is the wide first turn of helix 5 in each lobe; this helix, which binds the essential carbonate anion at its N terminus, begins with two π-type (1···6) hydrogen bonds (121 O···126 N, 122 O···127 N and 465 O···470 N, 466 O···471 N) before tightening to a classic α-helix. In each lobe, this helix is also characterized by a kink in the middle associated with a proline residue, and by the presence at or close to its C terminus of a glycosylation site (Asn137 and Asn478).

14 turns are found at identical positions, and with the same configurations, in the two lobes, including the two classic γ-turns, 298–300 and 641–643, which appear to be conserved in all transferrins. Many other turns do not match between the two lobes, however, reflecting the fact that the non-homologous regions are principally the connecting loops between secondary structures, where many of the turns are found. There are three turns which incoporate *cis*-proline residues (Pro71, Pro142 and Pro628); in each case the conformation is similar, with

Table 6. *Secondary structures*

Residues	N lobe label*	Type	Residues	C lobe label*	Type
(*a*) Helices†					
12–31	(1)	α_{C2}	351–365	(1)	α_{C1}
41–53	(2)	α_{C2}	376–388	(2)	α_{C2}
60–69	(3)	α_{C1}	395–407	(3)	α_{C2}
105–108	(4)	3_{10}	449–453	(4)	3_{10}
121–136	(5)	π, α_{C1}	465–481	(5)	π, α
144–153	(6)	α_N	484–488	(6)	3_{10}
166–170	(6*a*)	3_{10}	501–505	(6*a*)	3_{10}
190–203	(7)	α_{C2}	526–539	(7)	α_{C2}
212–218	(8)	α	547–554	(8)	α_N, α_{C1}
223–227	(8*a*)	3_{10}	567–571	(8*a*)	3_{10}
241–245	(8*b*)	3_{10}	585–589	(8*b*)	3_{10}
263–279	(9)	α	605–621	(9)	α
315–321	(10)	α	658–664	(10)	α
321–332	(11)	α_{C1}	664–678	(11)	α_{C1}
334–344	(12)	α	680–691	(12)	α_{C1}
(*b*) β-Strands‡					
5–10	(*a*)		345–350	(*a*)	
33–39	(*b*)		368–374	(*b*)	
54–59	(*c*)	*b* 55	389–394	(*c*)	*b* 390
75–82	(*d*)	*b* 77	408–415	(*d*)	*b* 410
90–100	(*e*)		435–443	(*e*)	
112–116	(*f*)		456–460	(*f*)	
153–158	(*g*)	*b* 154	489–494	(*g*)	*b* 490
204–211	(*h*)	*b* 205	540–547	(*h*)	*b* 541
226–231	(*i*)		570–575	(*i*)	
246–258	(*j*)	*b* 247	590–602	(*j*)	*b* 591
304–310	(*k*)	*b* 305	647–653	(*k*)	*b* 648
(*c*) Turns§ (matching between lobes)					
109–112		II	453–456		II
118–121		II	462–465		II
158–161		II	494–497		II
162–166		α	498–501		I
170–173		I	504–507		I
179–182		II	515–518		II
184–187		III	520–523		III
187–190		I	523–526		I
(210–213)		III	546–549		III
231–234		I	575–578		I
238–241		I	582–585		III
287–290		I	(632–635)		I
298–300		γ (82, −50)	641–643		γ (75, −49)
301–304		II	644–647		I

Table 6 (*cont.*)

N lobe Residues	label*	Type	C lobe Residues	label*	Type
(*d*) Turns (non-matching)					
69–72		VIa(*cis*-Pro)	362–365		I
84–87		III	365–368		III′
140–143		VIa(*cis*-Pro)	422–424		γ (−68, 64)
177–180		III	(424–427)		I
220–223		I	442–445		III
279–282		Irregular	444–446		γ (−88, 53)
283–286		III	445–448		I
292–295		I	483–485		γ (−85, 72)
311–314		I	510–513		I
			554–557		III′
			560–563		I
			563–566		I
			602–605		III
			621–624		I
			624–627		I
			626–629		VIa(*cis*-Pro)
			636–639		II′
			653–656		I′
			655–657		γ (−81, 75)

* Helices and β-strands labelled as in Anderson, Baker, Norris, Rice & Baker (1989).

† Helices are defined as extending from the first hydrogen-bonded residue to the last. The label α_N denotes a helix with 1–4 hydrogen bonding at its N terminus, α_{C1} a helix which tightens to 1–4 hydrogen bonding towards its C terminus and α_{C2} a helix with a 1–6, 2–5 hydrogen-bonding pattern at its C terminus (Baker & Hubbard, 1984).

‡ β-strands defined by their hydrogen-bonding pattern. All have β-type (φ, ψ) angles except for those labelled *b*, which have a β-bulge conformation, approximately (−100, −45).

§ β-turns labelled according to Venkatachalam (1968) and Wilmot & Thornton (1988). All have 1–4 hydrogen bonds except for those given in parentheses, for which the distances are slightly greater than 3.5 Å. γ-turns are identified by their 1–3 hydrogen bonding (Baker & Hubbard, 1984); for these the (φ, ψ) angles of the central residue are also given.

Pro at position 3, (φ, ψ) angles approximately (−70, 150°) at position 2 and (−80, 0°) at position 3, and both 1···4 and 1···5 hydrogen bonds. The two conserved γ-turns have the quite rare 'classic' γ-turn configuration (Nemethy & Printz, 1972; Matthews, 1972) with the central residue having normally unfavourable (φ, ψ) angles, around (75, −50°), but with excellent hydrogen-bond geometry (O···H = 1.91 and 1.89 Å, N—H···O = 152° and 155°). Other γ-turns, however, have the more common configuration in which the central residue has (φ, ψ) angles around (−90, 70°) (Baker & Hubbard, 1984).

Overall the β-sheets are the most highly conserved secondary-structural elements between the two lobes; equivalent sheets have essentially identical hydrogen bonding and conformation. This reflects the fact that they form the core of each lobe. Helices are more variable. Some match closely, even to having the same termination patterns, but others differ where insertions or deletions lengthen or shorten them at the protein surface. Turns are the most variable elements betwen the two lobes, with only about half being conserved; this clearly reflects their surface location and their frequent association with insertions and deletions.

Table 7. *Geometry of the iron sites*

N lobe Atoms*	Bond length (Å)	C lobe Atoms†	Bond length (Å)
Fe—O_{60}	2.00	Fe—O_{395}	2.08
Fe—O_{92}	2.28	Fe—O_{435}	1.92
Fe—O_{192}	1.92	Fe—O_{526}	2.10
Fe—N_{253}	2.21	Fe—N_{597}	2.13
Fe—O_1	2.08	Fe—O_1	2.13
Fe—O_2	2.22	Fe—O_2	2.06

N lobe Atoms*	Bond angle (°)	C lobe Atoms†	Bond angle (°)
O_{60}—Fe—O_{92}	90	O_{395}—Fe—O_{435}	87
O_{60}—Fe—O_{192}	161	O_{395}—Fe—O_{526}	166
O_{60}—Fe—N_{253}	86	O_{395}—Fe—N_{597}	80
O_{60}—Fe—O_1	91	O_{395}—Fe—O_1	95
O_{60}—Fe—O_2	84	O_{395}—Fe—O_2	87
O_{92}—Fe—O_{192}	106	O_{435}—Fe—O_{526}	102
O_{92}—Fe—N_{253}	94	O_{435}—Fe—N_{597}	80
O_{92}—Fe—O_1	95	O_{435}—Fe—O_1	100
O_{92}—Fe—O_2	155	O_{435}—Fe—O_2	161
O_{192}—Fe—N_{253}	83	O_{526}—Fe—N_{597}	91
O_{192}—Fe—O_1	97	O_{526}—Fe—O_1	95
O_{192}—Fe—O_2	85	O_{526}—Fe—O_2	88
N_{253}—Fe—O_1	170	N_{597}—Fe—O_1	174
N_{253}—Fe—O_2	109	N_{597}—Fe—O_2	116
O_1—Fe—O_2	61	O_1—Fe—O_2	64

* Ligand atoms Asp60 OD1, Tyr92 OH, Tyr192 OH, His253 NE2, carbonate atoms O1, O2.

† Ligand atoms Asp395 OD1, Tyr435 OH, Tyr526 OH, His597 NE2, carbonate atoms O1, O2.

Metal and anion sites

Metal and anion binding are highly interdependent. Higher resolution refinement has confirmed that the carbonate ion is indeed coordinated to the metal ion in bidentate mode. This interaction is shown in Fig. 8, while details of the metal geometry are given in Table 7, and the hydrogen bonds that hold the anion to the protein are listed in Table 8.

The binding sites in the two lobes are remarkably similar, at least within the accuracy of the present X-ray analysis. Superposition of the two sites (Fe^{3+} and CO_3^{2-} ions, the iron ligands and the anion-binding residues, including the N-terminal residues of helix 5, a total of 81 atoms) gives an r.m.s. deviation of only 0.31 Å. For bond lengths and angles the r.m.s. deviations between the two sites are also small, 0.15 Å and 6.5°, respectively. The metal–ligand bond lengths, which were not restrained in the later stages of refinement, are all close to 2.0 Å, as expected for Fe^{III} coordination, and as indicated by EXAFS studies (Garratt, Evans, Hasnain & Lindley, 1986). Nevertheless, two differences between the sites may be significant, even though they are at the error level. In the N lobe the bond to Tyr92 consistently refined to beyond 2.2 Å, in marked contrast to the equivalent bond in the C lobe (to Tyr435); it is the Tyr92 bond that also lengthens to 2.7 Å when Cu^{II} is bound (Smith, Anderson, Baker & Baker, 1992). Secondly, although the carbonate coordination is almost exactly symmetrical in both sites, it is less symmetric in the N-lobe site; again it is in this site that the carbonate

coordination becomes much less symmetric when Cu^{II} is bound. The small chelate bite angle of ~63° means that the geometry of the metal ion is distorted from true octahedral symmetry. One carbonate O atom is *trans* to the His ligand, with an angle of O—Fe—N(His) of ~170°, while the other is *trans* to the first Tyr ligand with an angle O—Fe—O(Tyr) of ~155°.

The anion binds to a positively charged site on domain 2 of each lobe. This is formed by the N terminus of helix 5, residues 121–124 (465–468) and the side chain of Arg121 (465), as well as the loop joining β-strand *f* to helix 5 which contributes Thr117 (461) to the site. These residues contribute to an exquisite fit of the anion between metal ion and protein in which the full hydrogen-bonding potential of the carbonate ion is utilized (Fig. 8 and Table 8). Each of the two coordinated carbonate O atoms accepts a hydrogen bond through its remaining lone pair (O1 from a peptide NH, O2 a bifurcated hydrogen bond from the NH1 and NE H atoms of the Arg residue), and the non-coordinated O atom, O3, accepts two hydrogen bonds, through its two lone pairs. The hydrogen bonds all have highly favourable geometry. The C lobe appears to provide a slightly tighter anion site, the hydrogen bonds being on average approximately 0.1 Å shorter.

Differences between the two sites are found outside the immediate coordination sphere. Apart from differences in interdomain interactions, which are described separately, two differences are particularly noteworthy. Firstly, the 'essential' (anion-binding) arginine side chain differs in its interactions (Fig. 9). In the N lobe Arg121 hydrogen bonds through NH1 to the hydroxyl group of Ser 191, but this interaction is not possible in the C lobe where the equivalent residue is a glycine (Gly527), as it is in all other transferrins. In the C lobe, on the other hand, NH1 of the equivalent arginine (Arg465) is instead hydrogen bonded to the phenolic O atom of Tyr526 (Phe190 in the N-lobe). The different positions of these two hydrogen-bonding partners for the arginine may be responsible for the small variations in the two anion sites and certainly influence the different ways in which the sites respond to anion substitution (Smith, Anderson, Baker & Baker, 1995; Baker, Anderson, Smith & Baker, 1995).

Secondly, a characteristic feature of all transferrins is the presence of a second cationic group at the 'back' of the iron site near the two Tyr ligands, at positions 210 in the N lobe and 546 in the C lobe (lactoferrin numbering). In the N lobe of human lactoferrin Arg210 makes a complex set of interactions (Fig. 10), involving hydrogen bonds from NH2 to the phenolic O atoms of Tyr192 (an iron ligand) and Tyr82 (a nearby non-ligated residue); two amino–aromatic hydrogen bonds (Burley & Petsko, 1986; Levitt & Perutz, 1988), from NE to the ring of Tyr192 (3.59 Å) and from NH1 to the ring of Tyr82 (3.33 Å); and a fifth hydrogen bond from NH1 to a tightly bound water, W113. In both of the amino–aromatic hydrogen bonds the H atoms point directly towards the centres of the aromatic rings. The guanidinium group also packs plane-to-plane with the other Tyr ligand, Tyr92, 3.7 Å away; this, too, is a common packing mode for Arg side chains with aromatic rings

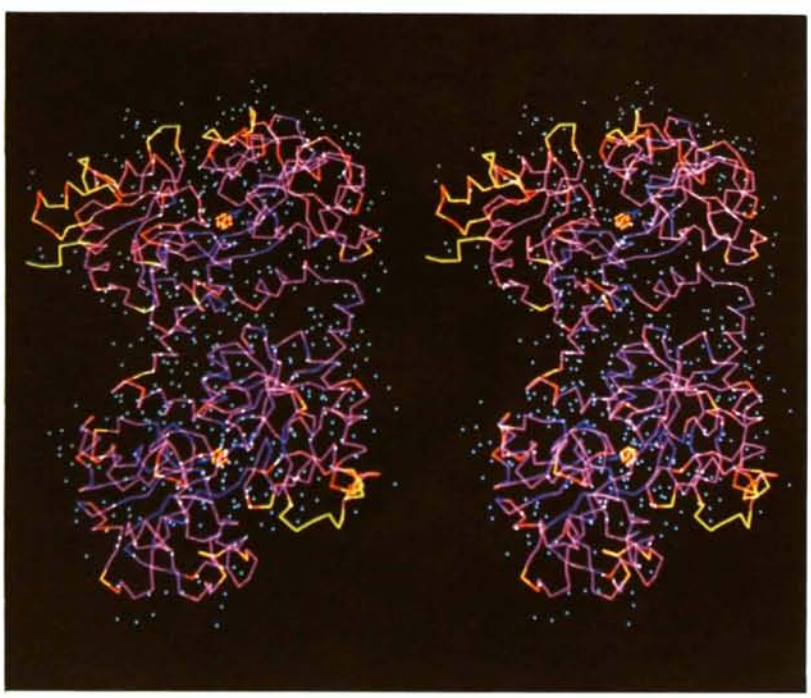


Fig. 2. Stereo C_α diagram in which residues are coloured according to their main-chain *B* values (blue <20 Å^2, purple 20–45 Å^2, orange 40–65 Å^2, yellow >65 Å^2). The distribution of solvent molecules in the final refined model is also shown, and the two Fe atoms are shown as large orange spheres.

(Flocco & Mowbray, 1994) and must help stabilize the iron site.

In the C lobe, however, the equivalent residue is Lys546, which makes no such interactions, being hydrogen bonded only to water molecules; one of the latter, *W4*, is within 0.2 Å of the position filled by 210 NH2 in the N lobe and makes very similar hydrogen bonds, while another, W317, is close to the position of 210 NH1 and fills its role (Fig. 11). Thus, Lys546 makes only indirect interactions with the protein.

One tightly bound water molecule is found close to the Fe atom in each lobe; in the N lobe, water W1 is hydrogen bonded to the ligand Tyr92 OH (2.4 Å), Gly61 N (2.9 Å), Thr122 OG1 (2.9 Å) and Pro251 O (3.0 Å) and is only 3.8 Å from the Fe atom, while in the C lobe water W3 hydrogen bonds to Tyr435 OH (2.5 Å), Gly396 N (3.0 Å), Thr466 OG1 (2.6 Å) and Pro595 O (2.8 Å) and is 3.9 Å from the Fe atom. No other water molecules are within 5 Å of the N-lobe Fe atom, although in the C lobe there is a second such water, W4, which is the one that occupies the site equivalent to 210 NH2 and which is 4.2 Å from the Fe atom. At a greater distance from each Fe atom (beyond 6 Å) are several bound waters near the back of the site, and a much larger number in the interdomain cavity (see later).

Interdomain interactions

The binding cleft of each lobe is predominantly polar in nature and is formed by the inner surfaces of its two domains. The wall provided by domain 1 comprises the N termini of helices 1, 2 and 3 together with a section of polypeptide (294–303 in the N lobe, 638–646 in the C lobe) that is made up of series of turns, including the conserved γ-turn, 298–300, and whose conformation is stabilized by hydrogen bonds from a number of highly conserved side chains (Fig. 12). The wall provided by domain 2 comprises the N termini of helices 5 and 7, one side of helix 8 and a loop (183–186 in the N lobe, 519–522 in the C lobe) at the mouth of the cleft.

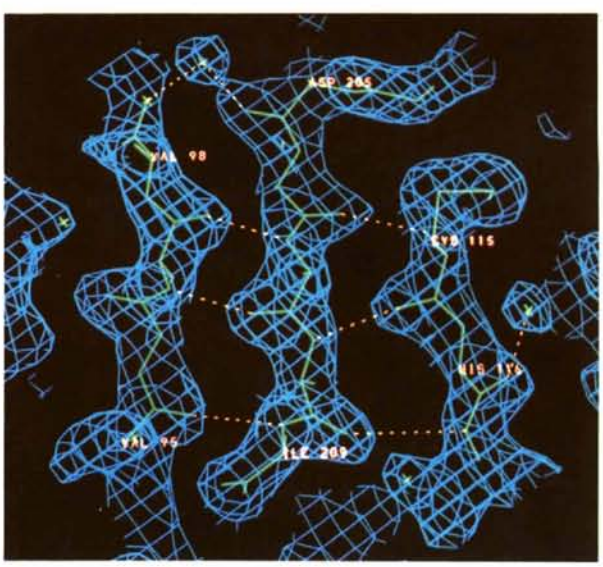
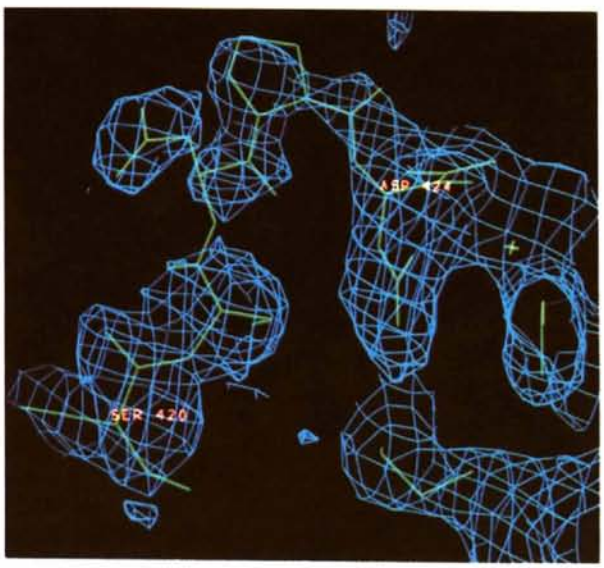
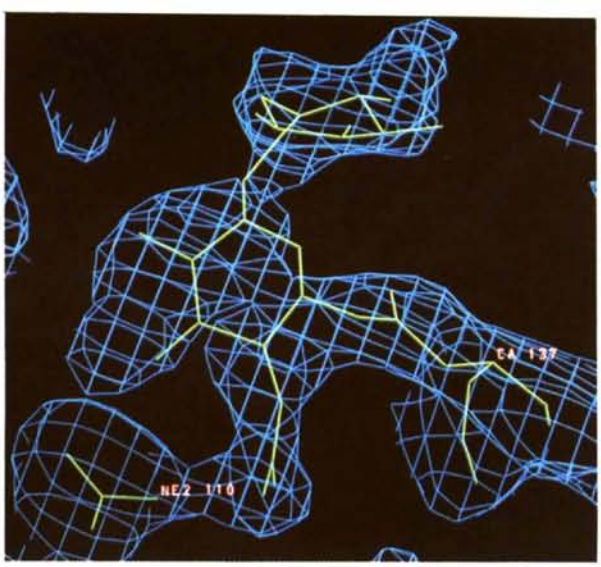


Fig. 3. Representative electron density from a well defined region of the protein structure, the central β-sheet of the N2 domain. The final $2F_o - F_c$ map is contoured at a level of 1.5σ, and hydrogen bonds are shown with broken lines.



(*a*)



(*b*)

Fig. 4. Electron density from two of the most poorly defined parts of the structure, with B values in the range 80 to 90 Å^2. The final $2F_o - F_c$ map is contoured at a level of 0.75σ. In (*a*) part of the external loop 417–423 is shown and in (*b*) the two sugar residues that have been located from the N-lobe carbohydrate chain, NAG 900 and FUC 901, are shown. Note the interaction of the *N*-acetyl group of NAG 900 with the side chain of Gln110.

Table 8. *Anion hydrogen-bonding distances* (Å) *and angles* (°)

Hydrogen bond	O···N	N—H···O	C—O···H	Hydrogen bond	O···N	N—H···O	C—O···H
O1···N(123)	2.83	160	110	O1···N(467)	2.59	157	122
O2···NE(121)	2.87	135	97	O2···NE(465)	2.65	139	97
O2···NH2(121)	2.85	132	111	O2···NH2(465)	2.88	128	126
O3···N(124)	3.07	141	130	O3···N(468)	2.94	153	125
O3···OG1(117)*	2.59	115	150	O3···OG1(461)*	2.53	112	149

* Distance is O···O, angles are C_β—OG1···O and C—O···OG1.

The most striking feature of the binding cleft is the relative lack of direct protein–protein interactions between the two domains (Table 9). The only interactions common to both lobes are those made by the iron ligand Asp60 (Asp395 in the C lobe). As well as binding to the Fe atom, its carboxyl group is placed between two helix N-termini, helix 3 in domain 1 and helix 5 in domain 2, such that its non-coordinating carboxylate O atom receives hydrogen bonds from the peptide N atoms of Gly62 (397) on helix 3 and Thr122 (466) on helix 5 (Fig. 13). Since helix N-termini carry a partial positive charge (Hol, van Duijnen & Berendsen, 1978) this must provide a strong interdomain interaction.

In the N-lobe two salt bridges cross between the two walls of the cleft, Lys296···Asp217 and Lys301···Glu216. The former cannot be present in serum transferrins or bovine lactoferrin, where residue 217 is changed to Asn; the latter is not found in rabbit transferrin because of a conformational difference (Baker & Lindley, 1992). Two other salt bridges in the N lobe,

Table 9. *Interdomain interactions*

Only direct protein–protein hydrogen bonds are listed.

N lobe	C lobe
Asp60 OD2···Thr122 N	Asp395 OD2···Thr466 N
Asp60 OD2···Thr122 OG1	Asp395 OD2···Thr466 OG1
Glu66 OE1···Arg332 NH2	Glu352 OE2···Asn522 ND2
Tyr82 OH···Arg210 NH2	Ser375 O···Asn520 ND2
Tyr82 ring···Arg210 NH1	Thr638 O···Thr521 OG1
Arg89 NH2···Glu211 OE2	
Lys296 NZ···Asp217 OD2	
Lys301 NZ···Glu216 OE2	

Glu66···Arg332 and Arg89···Glu211 are positioned such that they do not appear likely to constrain domain movements significantly. The only other direct interdomain interactions in the N lobe involve Arg210, behind the iron site (described above and depicted in Fig. 10) and hydrophobic contacts between Pro42 and Phe183 at the mouth of the cleft.

In the C lobe there are no interdomain salt bridges, and there are just three hydrogen bonds, all involving the loop 519–522, at the mouth of the cleft. The residues equivalent to Glu216 and Asp217 are both amides (Gln552 and Asn553). Lys301 is replaced by Asn644, whose shorter side chain only interacts with the other domain indirectly, through solvent molecules. Likewise, Arg210 is replaced by Lys546, which does not reach Tyr415 (the equivalent of Tyr82) and again interacts through a water bridge.

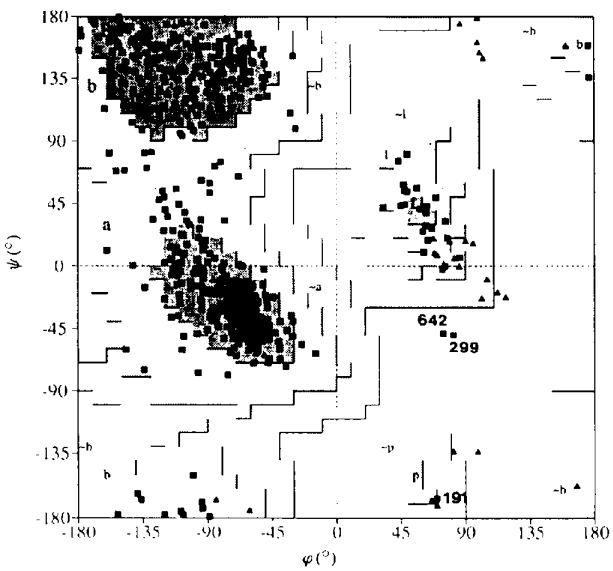


Fig. 5. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the main-chain torsion angles (φ and ψ) for the final refined model. The plot was calculated with the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). Glycine residues are identified by triangles, non-glycine residues by squares. Residues 191, 299 and 642 are discussed in the text.

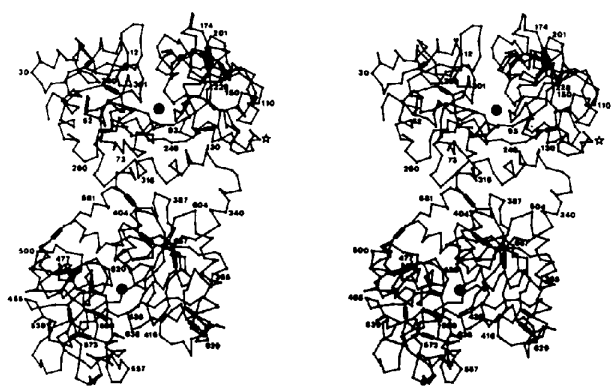


Fig. 6. Stereo C_α plot of the diferric lactoferrin molecule. The N lobe is above, the C lobe below, and the domains are labelled N1, N2, C1 and C2. The Fe atoms are shown with filled circles and the two glycosylation sites are marked with stars. Disulfide bridges are shown with solid bars.

The predominant feature of the interdomain cleft is the large number of well ordered solvent molecules, of which 23 have been identified in the N lobe and 28 in the C lobe. The majority of these (18 in the N lobe, 21 in the C lobe) occupy the large interdomain cavity that stretches between the iron site and the mouth of the cleft. Most of the interactions between the domains are, therefore, mediated by solvent. In the N lobe we have noted nine single-water bridges between protein groups on the two domains, and in the C lobe 13 such bridges, while there are many more bridges involving multiple waters.

Interlobe interactions

The only covalent connection between the two lobes is the three-turn helix, residues 334–344. This helix is on the outside, and rather displaced from, the main area of contact between the lobes, which involves, on the N lobe, the loop 311–314 together with helices 10 (315–321) and 11 (321–332), and on the C lobe two further helices, numbers 2 (376–388) and 12 (the C-terminal helix 680–691).

Again there are very few direct interactions (Table 10). Two salt bridges are found, Asp315–Lys386 and Arg313–Asp379, together with one possible hydrogen bond from the hydroxyl of Ser322 to the carbonyl O atom of Leu385. The main feature of the interlobe contact area is the presence of a number of hydrophobic sidechains, contributed by both lobes. These comprise Pro311, Ile314, Leu318 and Phe325 on the N lobe and Leu385, Ala685, Phe688 and Leu689 on the C lobe; although there are few close contacts between these side chains, solvent is excluded, and they appear to provide a hydrophobic 'cushion' between the lobes.

Table 10. *Interlobe interactions*

Hydrogen bonds and salt bridges listed if the distance is less than 3.5 Å. Non-bonded contacts listed if the distance is less than 4.0 Å. Water bridges only listed where a single water is involved.

Direct	Water bridged
Arg313 NH2···Asp379 OD2	Arg313 NH1···*W49*···Lys386 NZ
Asp315 OD2···Lys386 NZ	Arg313 NH2···*W50*···Asp379 O
Ser322 OG···Leu385 O	Arg313 N···*W51*···Pro681 O
Pro311 CD···Phe688 ring	Thr90 OG1···*W111*···Phe688 O
Ile314 CD1···Ala685 CB	His91 ND1···*W27*···Lys691 OE
Ile314 CD1···Leu689 CD2	Arg249 NH2···*W170*···Lys691 O
Gly317 CA···Lys386 CG	
Leu318 CD2···Leu385 CD1	
Leu318 O···Leu689 CD2	
Phe325 ring···Lys386 O	

Outside this hydrophobic region there are, in addition to the salt bridges mentioned above, a small number of hydrogen-bonding contacts mediated by water molecules. The most significant of these may be the water bridges between the polypeptide chain C-terminus (the end of helix 680–691) and residues Thr90, His91 and Arg249. The latter are all at or near the hinge which mediates domain opening in the N lobe (Gerstein, Anderson, Norris, Baker, Lesk & Chothia, 1993) and recent comparisons show that when domain opening occurs, in apo-lactoferrin, these waters are displaced to allow direct contact with the C-terminus.

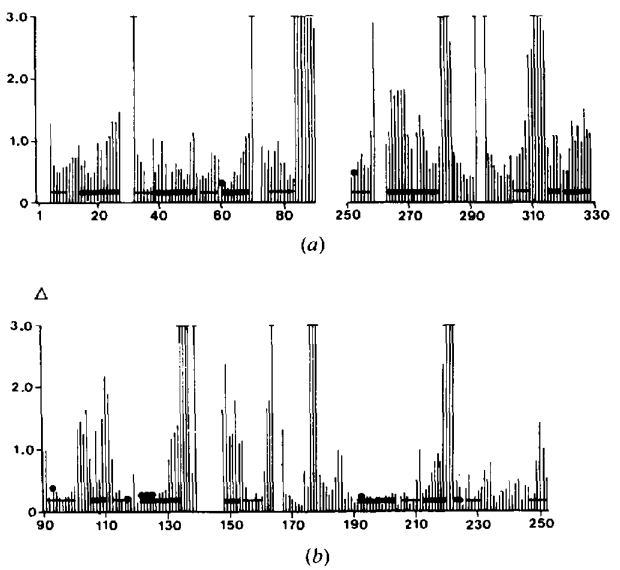


Fig. 7. Plot showing structural deviations between the N lobe and the C lobe after superposition, as a function of residue number. Separate superpositions have been carried out for (*a*) domain C1 on N1 and (*b*) domain C2 on N2, using main-chain atoms only. For each residue the deviation is the average for N, C_α, C and O atoms. Sites of insertions and deletions have been omitted, and secondary structures are indicated with thick lines (helices) and thin lines (β-strands).

Internal polar side chains

A small number of internal polar side chains make important contributions to the protein folding (Table 11). These residues, many of them charged, make multiple hydrogen-bonding interactions, often utilizing their full hydrogen-bonding potential. They tend to lie just inside the protein surface, out of contact with the external water, particularly in 'loopy' regions of the structure, where they have the role of stabilizing loops of the polypeptide chain, tying them together or linking them to secondary-structure elements (*e.g.* Fig. 14).

Many of these residues are conserved, between both lobes and through almost all transferrins. Lys18 (357), on helix 1, forms three hydrogen bonds to a 'loopy' section of polypeptide that forms part of one wall of the interdomain cleft. This same section is also hydrogen bonded by Asp302 (645) in the wall; by Glu15 (354) on helix 1, which further links it to β-strand *a*; and by Glu80 (413) and Tyr82 (415), both on β-strand *d*. The latter residues continue the hydrogen-bonding network to β-strand *c* through Thr58 (Ser393) and β-strand *j* through His253 (597). These multiple hydrogen-bonding interactions clearly help to define a large part of the wall of the binding cleft, which also includes the conserved γ-turn 298–300 (641–643).

Table 11. *Internal hydrogen-bonded side chains*

Residues in bold are invariant in all transferrin sequences to date; italic bold indicates almost complete invariance (all but one or two species).

N lobe			C lobe		
Residue		Hydrogen bonded to	Residue		Hydrogen bonded to
(*a*) Matching in both lobes					
Glu15	OE1	11 N, *W8*	***Glu354***	OE1	350 N, *W5*
	QE2	299 N, *W17*		OE2	642 N, *W12*, *W218*
Lys18	NZ	288 O, 290 O, 298 O, *W318*	**Lys357**	NZ	633 O, 635 O, 641 O *W370*
Asp55	OD1	7 N, 263 NZ	**Asp390**	OD1	346 N, 341 NH1
	OD2	258 NH2		OD2	602 NH2, 341 NH2
Glu80	OE1	301 N, *W2*	**Glu413**	OE1	393 OG, 644 N, *W10*
	OE2	82 OH, *W2*		OE2	415 OH, 644 ND2
Tyr82	OH	80 OE2, 253 N, 210 NH2	***Tyr415***	OH	413 OE2, 597 N, *W4*
Ser114	OG	116 NE2, 156 OG	**Ser458**	OG	460 NE2, 492 OG
His116	ND1	124 O	**His460**	ND1	468 O
	NE2	114 OG		NE2	458 OG
Glu187	OE1	174 N, 189 N	**Glu523**	OE1	508 N, 525 N
	OE2	175 N, 181 N		OE2	509 N, 517 N
Arg236	NE	228 OE2	**Arg580**	NE	*W451*
	NH1	104 O, *W31*		NH1	445 O, 448 O
	NH2	99 O, 228 OE2		NH2	443 O, *W451*
Arg258	NE	49 O	**Arg602**	NE	604 OD2
	NH1	52 OD1, 261 O		NH1	384 O, 389 O
	NH2	54 O, 55 OD2		NH2	390 OD2, 604 OD1
Lys263	NZ	5 O, 55 OD1	***Lys605***	NZ	342 O, 344 O, *W94*
Asp302	OD1	291 OG, 297 N	**Asp645**	OD1	636 OG, 638 N, 640 N
	OD2	291 OG, *W318*		OD2	636 OG, *W370*
Tyr319	OH	252 O, 252 OG	***Tyr662***	OH	596 O
Tyr324	OH	122 OG1, 126 ND2	***Tyr667***	OH	466 OG1, 470 ND2
(*b*) Non-matching					
Arg75	NE	264 OE2	Arg430	NE	*W270*
	NH1	311 O, 314 O		NH1	424 OD1, 648 OE1
	NH2	310 O		NH2	648 OE2
Arg133	NE	*W32*	***Arg524***	NE	509 O
	NH1	330 O, *W39*		NH1	511 OE1
	NH2	333 O		NH2	508 O, 510 O
Arg224	NE	218 O	***Asp604***	OD1	602 NH2, *W93*
	NH2	218 O, 219 O		OD2	602 NE

Asp55 (390), totally conserved in both lobes, is on β-strand *c*, from where it hydrogen bonds to β-strand *a* and also forms two salt bridges; in the N lobe these are to Arg258 and Lys263 on the loop joining β-strand *j* to helix 9, while in the C lobe they are to Arg602 (the equivalent of Arg258) and Arg341 on the inter–lobe connecting helix. Ser114 (488) and His116 (490), totally conserved in all transferrins, are on β-strand *f*, from where they hydrogen bond to β-strand *g* and helix 5, respectively. Glu187 (523) is completely buried yet has no basic group to ion pair with; instead, each of its carboxylate O atoms receives two hydrogen bonds from peptide NH groups and in so doing tie a 'loopy' region of domain 2 together. Arg236 (580) links the loop connecting β-strands *i* and *j* with the loop joining β-strands *e* and *f*, through hydrogen bonds to peptide O atoms, while Arg 258 (602) and Lys 263 (605), both on the loop connecting β-strand *j* to helix 9, hydrogen bond to β-strand *a*, helix 2 and β-strand *c*. Finally, two tyrosines on helices 10 and 11, Tyr319 (662) and Tyr324 (667) help to join these helices, which run across the back of the iron site, to the main body of each lobe.

The residues that do not match between the two lobes also tend to be less well conserved through the transferrin family. Two deserve particular mention. Arg75, on β-strand *d*, is hydrogen bonded to three peptide O atoms of the loop 310–314 which is an important contributor to the contacts between the two lobes; the equivalent region in the C lobe, 653–657, has no such role, and a somewhat different conformation, and perhaps this is why the Arg is not present in the C lobe. Arg133 extends from the C terminus of helix 5 in the N lobe to hydrogen bond to the C-terminus of helix 11, just prior to the inter-lobe connecting helix. This Arg residue is only found in lactoferrins, in the N lobe, possibly linked to the fact that only lactoferrins have a helical connecting peptide between lobes. It is also analogous in position to disulfide bridge 10, which joins helix 5 to the connecting peptide in some serum transferrins.

Carbohydrate structure

The two carbohydrate chains are both *N*-linked, to Asn137 in the N lobe and Asn478 in the C lobe. These two attachment sites are in roughly equivalent locations on the two lobes, at or near the C terminus of helix 5; Asn 478 is in the final turn of helix 5 while Asn137 is one residue beyond it.

Of the N-lobe carbohydrate chain only two sugar residues, the terminal *N*-acetylglucosamine (NAG) and its attached fucose, are visible (Fig. 4*b*). These make only one direct contact with the protein structure, a hydrogen bond between the acetyl group of the NAG sugar and the side chain amide N atom of Gln110. The same interaction is seen in a complex formed between a lectin and an N2 domain fragment of human lactoferrin (Bourne, Mazurier, Legrand, Rougé, Montreuil, Spik & Cambillau, 1994); in this structure, too, this is the only dirct contact between the carbohydrate and the protein part of the lactoferrin structure. Several water molecules appear to link the fucose to the protein but these have high *B* values, and cannot be regarded as tightly bound.

Although more residues of the C-lobe carbohydrate have been modelled, as reasonably continuous (though weak) density can be seen, these make no direct contacts with the protein. The acetyl group of the terminal *N*-acetyl-glucosamine is linked by a water molecule, W261, to the peptide O atom of residue 475, but this is the only apparent water bridge. The residues included in our model are shown in Fig. 15. All have been modelled in chair conformations except for mannose 3, which fits better as the boat conformer; we stress, however, that the density is poor, the carbohydrate clearly suffers from disorder, and only its approximate position, not detailed conformation, is reliable. This is not surprising given the paucity of interactions with the protein to which it is attached and the lack of any intermolecular (crystal) contacts which might stabilize it.

Solvent structure

The crystals used for this analysis were equilibrated in a mother liquor containing 20%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD). Many of the solvent sites identified during refinement are clearly occupied by water molecules, as judged by the size and shape of the peaks, and their environment. This is especially true of the well ordered sites. Some peaks on the outside of the molecule are more diffuse and could represent MPD molecules or phosphate ions rather than water. In particular, several spherical peaks that have strong density and low *B* values, but no hydrogen-bond partners within 3.5 Å of the peak centre, may be phosphate ions. At the present resolution, however, and without compelling evidence from their environments, they cannot unequivocally be distinguished from disordered water; accordingly all solvent molecules are treated as water.

Apart from the few strong peaks mentioned above, all sites that do not make at least one potential hydrogen-bond contact either with protein or with another solvent molecule have been eliminated from the final model. A total of 469 water molecules remain; for a protein of 691 residues this is not unreasonable. Although to a first approximation the modelled water forms a surface monolayer (76% of waters make at least one contact with the protein) the distribution is not regular (See Fig. 2). Well defined sites are not seen around the more mobile regions of the protein structure, such as helix 1 on the N lobe and certain surface loops. On the other hand clusters of water molecules are found in other regions where they are trapped between elements of the protein structure. The water molecules in the interdomain cavity of each lobe (see below) are classic examples.

A total of 42 water sites are wholly internal, 20 within the N lobe and 22 within the C lobe. These water molecules have zero solvent accessibilities, as calculated by the method of Lee & Richards (1971), they tend to be located in discrete sites with space for only one or two waters, and they make a large number of hydrogen-bonded contacts to the surrounding protein structure. On average each one of these internal waters makes 3.7 hydrogen bonds (3.3 of these with protein atoms), implying that their hydrogen-bonding potential is almost fully utilized and that they make a significant contribution to the stability of the protein structure. The most common location for such sites is between, or inside, loops of the polypeptide at the protein surface, or at the points where β-strands diverge away from one another; these waters probably help define the surface structure. Some, as noted earlier, are buried at or behind the iron site, or help mediate interdomain interactions.

In addition to the 42 essentially discrete sites two large internal clusters are found, within the interdomain cavity of each lobe. This cavity is filled with a network of ordered water molecules, 18 in the N-lobe cavity and 21 in the C-lobe cavity. These waters are highly inter-connected, with an average of 3.2 hydrogen bonds each, 1.8 of these being with other waters. The majority of them are anchored to protein groups around the lining of each cavity but a few, in the centre, are hydrogen bonded only to other waters. The water molecules in these two large cavities have higher *B* values than those in the discrete internal (average 43 Å^2 compared with 32 Å^2), implying that they are more mobile, but there is still a considerable degree of correspondence between the water positions in the two cavities, with ten in each cavity matching to within 1.0 Å (Table 12). The anion-binding arginine side chain (Arg121 in the N lobe, Arg465 in the C lobe) projects into the cavity, and it is this that allows it to move to accept larger anions in the

Table 12. *Solvent sites conserved between lobes*

Only those with deviations < 1.0 Å listed for internal and interdomain sites and < 0.7 Å for others.

Solvent	N lobe Hydrogen-bond partners	Solvent	C lobe Hydrogen-bond partners	Deviation (Å)
(*a*) Internal				
*W*19	116O, 189O	*W*21	460N, 460O, 493O, 525O	0.01
*W*20	95O, 123O	*W*22	438N, 438O, 467O	0.30
*W*354	105NE2, 106N, 231O, 235O	*W*86	450N, 575N, 575O, 579O	0.30
*W*42	213OG1, 216OE1, 301NZ, *W*48	*W*243	549OG1, 552NE2, *W*99, *W*228	0.37
*W*312	96N, 230O, 245O	*W*71	439N, 574O, 589O	0.37
*W*1	61N, 92OH, 122OG1, 251O	*W*3	396N, 435OH, 466OG1, 595O	0.38
*W*2	253ND1, 80OE1, *W*18, *W*195	*W*10	597ND1, 413OE1, *W*11	0.40
*W*9	8NE1, 58N, 58OG1, 299O, *W*8	*W*6	347NE1, 393N, 642O, *W*14	0.63
*W*318	18NZ, 288O, 290N, 300O, 302OD2	*W*370	357NZ, 636OG, 641O, 643O, 645OD2	0.68
(*b*) Interdomain cavity				
*W*17	15OE1, *W*18, *W*193	*W*12	354OE1, *W*14, *W*62	0.27
*W*8	9O, 15OE2, *W*9, *W*193	*W*5	348O, 354OE2, *W*6, *W*14	0.30
*W*314	297OD2, *W*18, *W*194, *W*366	*W*101	644OD1, *W*11, *W*83, *W*208	0.56
*W*192	*W*16, *W*191, *W*193	*W*15	377OG1, *W*13, *W*14, *W*62	0.61
*W*366	*W*194, *W*314, *W*415	*W*83	465NH1, *W*62, *W*101, *W*208, *W*421	0.68
*W*415	15OE1, *W*366	*W*218	354OE1, *W*62, *W*274, *W*421	0.70
*W*18	297O, *W*2, *W*314	*W*10	640O, *W*101	0.79
*W*313	191OG, 193N, 297OD2, *W*48	*W*7	529N, 640ND2, *W*208, *W*401	0.82
*W*193	58O, *W*8, *W*17, *W*192	*W*14	393O, *W*5, *W*6, *W*12	0.96
*W*191	*W*190, *W*192	*W*62	526OH, *W*12, *W*15, *W*83, *W*218	0.99
(*c*) Others				
*W*331	167O, 171N, 180NZ	*W*229	499O, 501O, 505N	0.20
*W*34	202O, 205OD2, *W*332	*W*84	538O, 541OD2	0.29
*W*30	99N, 205O, *W*36	*W*72	442N, 442NH2, 445OD2, 541O	0.37
*W*25	319O, *W*322	*W*403	662O, *W*133	0.38
*W*310	64O, 68N, 74O	*W*339	399O, 403N	0.46
*W*164	125NE1, 146OE2	*W*74	469NE1	0.51
*W*247	287O, 290O, *W*178	*W*226	632O, 635O	0.52
*W*324	113NZ, 203O	*W*234	457NZ, 539O, *W*231	0.57
*W*289	330ND2, *W*58, *W*380	*W*187	669O, 673ND2, *W*449	0.58

binding site (Smith, Anderson, Baker & Baker, 1995), by displacement of solvent.

Some 72 water molecules (*i.e.* 36 pairs) are found that match between the two lobes, at least as closely as does the protein structure itself (within 1.0 Å). A selection of these are listed in Table 12, to give an idea of the degree of correspondence and the sites they occupy. The superposition of water sites was made more difficult by the different closures of the two lobes, which meant that an 'average' superposition, based on the whole lobe, gave an inadequate measure of how well the sites agreed. A more correct superposition was achieved by a domain-by-domain comparison; the C2 domain was superimposed on the N2 domain, using 134 C_α positions, and their associated waters were compared, and a similar, separate comparison was done for the N1 and C1 domains and their waters. Those waters that are located between the domains were included in both cases and showed similar agreement. About half of the internal water sites (18 out of 42) are found as matching pairs in the two lobes; where they do not agree this is largely because of sequence differences. Likewise 20 of the 40 waters in the interdomain cavities also appear as matching pairs (within 1.0 Å). Most of the other sites that match are in crevices in the protein surface, attached to main-chain atoms, or to conserved side-chain atoms. The former probably help to stabilize the protein, but some of the latter are probably conserved more because certain conserved side-chains offer favourable binding sites (*e.g. W*74 and *W*164, bound to a Trp side chain).

Intermolecular contacts

Very few intermolecular protein–protein contacts are found in the diferric lactoferrin crystals. Those direct contacts that do exist are listed in Table 13. Most are hydrogen-bonded interactions, involving the side chains of Lys99, Arg272, Glu276, Ser293 and Ser308 in the N lobe, Glu336 in the connecting helix, and Asp 422, Lys454, Lys581, Glu585, Ser588, Arg622, Asp629, Lys630 and Glu684 in the C lobe. Charged side chains predominate, and each molecule makes eight salt bridges with neighbouring molecules (Lys99–Glu585, Glu276–Arg622, Glu336–Lys454 and Lys630–Glu684, together with their four equivalents); as in the case of the recombinant half-molecule of lactoferrin (Day, Anderson, Tweedie & Baker, 1993), for these crystals, grown at low ionic strength, ionic interactions may be a significant determinant of the crystal packing.

The main contact regions are, in the N lobe, the C terminus of helix 7 (residues 201–202), the flank of helix 9 (272–276), and the loop joining β-strand *k* to helix 10 (308–311); and in the C lobe the poorly ordered loop joining β-strand *d* to β-strand *e* (417–424), the loop joining β-strand *i* to β-strand *j* (581–588) and the loops following helix 9 (622–629). The C lobe makes slightly more contacts with neighbouring molecules than does the N lobe (16, compared with 13) and there is just a single intermolecular contact involving the carbohydrate; the mannose residue 4 of the C lobe glycan contacts Ala70 on the N lobe of a neighbouring molecule and this may contribute to the slightly better ordering of the C-lobe carbohydrate chain.

Table 13. *Intermolecular contacts*

	Contact*	Distance (Å)
Ala70 O	MAN913 O4^{III}	3.02
Lys99 NZ	Glu585 OE1^{I}	3.13
Asp201 O	Ser588 OG^{I}	2.41
Gly202 O	Ser588 OG^{I}	3.02
Gly202 O	Lys581 NZ^{I}	2.79
Arg272 NH1	Gly624 O^{II}	3.06
Arg272 NH1	Cys627 O^{II}	2.79
Glu276 OE1	Arg622 NH2^{II}	3.02
Glu276 OE2	Arg622 NH2^{II}	3.07
Ser293 OG	Ser420 O^{I}	3.21
Ser293 OG	Asp422 OD2^{I}	2.86
Arg309 N	Asp629 OD2^{II}	3.37
Ser308 OG	Asp629 OD1^{II}	3.18
Glu336 OE1	Lys454 NZ^{II}	2.76
Ala340 CB	Thr480 O^{III}	2.76
Lys630 NZ	Glu684 OE2^{III}	3.17

* Hydrogen bonds less than 3.4 Å, non-bonded contacts less than 3.5 Å. Symmetry-related positions: (I), $-\frac{1}{2}+x, \frac{1}{2}-y, -z$; (II), $-x, -\frac{1}{2}+y, \frac{1}{2}-z$; (III), $-x, \frac{1}{2}+y, \frac{1}{2}-z$.

The small number of interactions at each contact point shows that the molecules are not tightly packed, and that the crystal packing is not likely to have disturbed the protein structure significantly. The reasonable geometry of the contacts, which are not restrained by the refinement methods, also helps validate the quality of the modelling of the structure

Discussion

Refinement has confirmed most of the details of the diferric lactoferrin structure seen at lower resolution (Anderson, Baker, Norris, Rice & Baker, 1989). In particular, the carbonate ion, which was only tentatively placed at 2.8 Å resolution, is clearly bound in bidentate fashion to the Fe atom of each site, while also making an exquisite set of hydrogen-bonded interactions with the anion site on domain 2 (N2 or C2) of each lobe. Its structural role is to neutralize the positive charge presented by the 'essential' arginine side chain (Arg121 or 465) and the helix 5 N terminus, and fill two *cis* coordination positions on the Fe atom.

Refinement also confirms the lack of any water (or hydroxide) ligand on the Fe atom. The closest water molecule to the metal ion is the one that is hydrogen bonded to the 'first' Tyr ligand of each site (Tyr92 in the N lobe, Tyr435 in the C lobe). This water molecule, which is ~3.8 Å from the Fe atom, is tetrahedrally surrounded by four protein atoms from different parts of the structure, each making a strong hydrogen bond (Fig. 16). It has been found to be present in every transferrin structure so far reported, including the iron-bound N-lobe

half-molecules of human lactoferrin (Day, Anderson, Tweedie & Baker, 1993), rabbit transferrin (Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990) and chicken ovotransferrin (Dewan, Mikami, Hirose & Sacchetini, 1993) and both lobes of diferric lactoferrin (this work) and rabbit transferrin (P. F. Lindley, personal communication). It seems most likely that it is this water molecule that was detected by Koenig & Schillinger (1969) in NMR relaxation studies.

The two iron sites are found to be virtually identical in terms of the immediate coordination environment of each metal ion. The spectroscopic differences that are found between the N- and C-terminal sites when iron is bound are very small and probably arise from very small structural differences similar to those observed for lactoferrin; here the differences appear confined to a slightly longer bond to the first Tyr ligand in the N-lobe site, compared with the C-lobe site, and a slightly greater asymmetry in the anion coordination in the N-lobe site. The differences are so small, in relation to the likely error in the analysis, that they would not be taken as significant were it not for the fact that differences that are similar, but larger, are found when Cu^{2+} is substituted for Fe^{3+} and when oxalate is substituted for carbonate; metal or anion substitution seems to enhance intrinsic differences between the two sites. When Cu^{2+} is substituted for Fe^{3+} in human lactoferrin the N-lobe carbonate anion becomes monodentate, and the bond to Tyr92 is lengthened to 2.7 Å, whereas the C-lobe carbonate remains bidentate and the bond to Tyr435 remains at ~2.0 Å (Smith, Anderson, Baker & Baker, 1992). Likewise when oxalate is substituted for carbonate the N-lobe anion is distinctly asymmetrically bound whereas that in the C lobe remains symmetric (Baker, Anderson, Smith & Baker, 1995).

Differences in the iron-release characteristics of lactoferrins relative to transferrins and of the N-lobe site relative to the C-lobe site have been attributed to differences in the interactions made by two residues at the 'back' of the iron site (Baker & Lindley, 1992; Dewan, Mikami, Hirose & Sacchetini, 1993). In the N lobe of human lactoferrin these are Arg210 and Lys301; in the C lobe they are Lys546 and Asn644. In rabbit transferrin and chicken ovotransferrin both residues in the N lobe are Lys, and their close, hydrogen-bonded interaction has led to them being described as a 'di-lysine trigger' (Dewan, Mikami, Hirose & Sacchetini, 1993) which could be responsible for the lesser acid stability of the N-lobe site of these proteins, relative to lactoferrin and to the C-lobe site.

In the N-lobe site of human lactoferrin Arg210 and Lys301 are not in contact. Arg210, which belongs to the N2 domain, is involved in a network of interactions with the two Tyr ligands and with a third Tyr belonging to the N1 domain; these interactions certainly seem likely to enhance the stability of the site and help account for the much lower pH of iron release from this site in lactoferrin, 3.0–4.0, compared with 5.0–6.0 in transferrin. The situation is, however, more complex than just the substitution of Arg for Lys at position 210 (lactoferrin numbering). Firstly, bovine lactoferrin has Lys in this position and does not show markedly different iron-release properties. Secondly the other residue involved, Lys301, is involved in a salt bridge with Glu216 and this must add further stability. This salt bridge is not found in rabbit transferrin or chicken ovotransferrin even though both residues are conserved. The differences between lactoferrin and transferrin in this region appear to be associated with a 'flip' of the peptide 302–303 (Baker & Lindley, 1992), leading to a different set of interdomain interactions with Glu216 and making the turn 301–304 type II in lactoferrin but type I in transferrin.

In the C lobe of lactoferrin the equivalent residues are Lys546 and Asn644. These residues are not in contact, and neither of them makes any direct interaction with protein atoms of the opposing domain; all the interdomain interactions in this region are mediated by water molecules. Although it will be interesting to compare the refined structure of a transferrin C lobe when this becomes available, a tentative conclusion at this stage is that there are no obvious acid-sensitive interdomain contacts in the C lobe; the differences between the pH-mediated iron-release properties of lactoferrins and transferrins are primarily properties of the N lobe.

The paucity of direct protein–protein contacts between the domains is consistent with the requirement for reversible binding by transferrins. Once the iron site is disrupted, and with it the strong interdomain hydrogen bonds made by the Asp ligand, there are few other interdomain interactions to hold the binding cleft in its closed state. At the 'lips' of the interdomain cleft the N lobe has only hydrophobic non-bonded contacts, and the C lobe three hydrogen bonds. Most of the other interdomain interactions are mediated by water molecules, apart from the two salt bridges in the N lobe of lactoferrin, Lys296–Asp217 and Lys301–Glu216, and the interactions involving Arg210.

Interactions between the N-terminal and C-terminal lobes of lactoferrin show a distinct lack of specificity. Apart from the two salt bridges, Arg313–Asp379 and Asp315–Lys386, and one hydrogen bond, there are only hydrophobic interactions and water-mediated contacts. If the same is true for other transferrins then it is not surprising that different relative orientations of the N and C lobes are found between human lactoferrin and rabbit transferrin (Baker & Lindley, 1992) and between human and bovine lactoferrin (Baker, Anderson, Baker, Haridas, Jameson, Norris, Rumball & Smith, 1991). The relative lobe orientations could be altered by receptor interactions or by structural changes (such as opening of the binding cleft) in one or other lobe; this could provide a means of signalling between lobes. The relative lobe orientations could also be influenced by factors as weak as crystal packing forces.

The area between the lobes has also been implicated in a recent suggestion that residues on the surface of the N lobe are in a similar juxtaposition to the catalytically important residues of ribonuclease (Sharada Devi, Das & Pandit, 1994). This follows earlier reports (Furmanski, Li, Fortuna, Swamy & Das, 1989) of an isoform of lactoferrin that has ribonuclease activity but does not bind iron. The residues in question, His91 and His246, face into a distinct cleft between the lobes, adjacent to the interlobe contact area. The presence of other positively charged amino acids in this region, notably Lys243 and Arg249, that could assist in binding of negatively charged species such as nucleic acids, makes this an intriguing suggestion; it is at least conceivable that isoforms of lactoferrin could exist in which the relative lobe orientations are altered to generate or modulate such an activity.

The lack of defined three-dimensional structure in the two carbohydrate chains is consistent with the absence of interactions between protein and carbohydrate; apart from the covalent linkages to the Asn residues, the N-lobe glycan chain makes only one direct contact with the protein (involving the terminal *N*-acetylglucosamine) and the C-lobe glycan none. There have been suggestions that the carbohydrate on lactoferrin influences iron binding, with the ability to bind iron apparently reduced by deglycosylation (Legrand, Mazurier, Colavizza, Montreuil & Spik, 1990). Iron binding also apparently alters the reactivity of the carbohydrate with lectins (Li & Furmanski, 1993). It is difficult to see a structural basis for such effects, given the lack of protein–glycan interactions either in our structure, or in the lactoferrin–

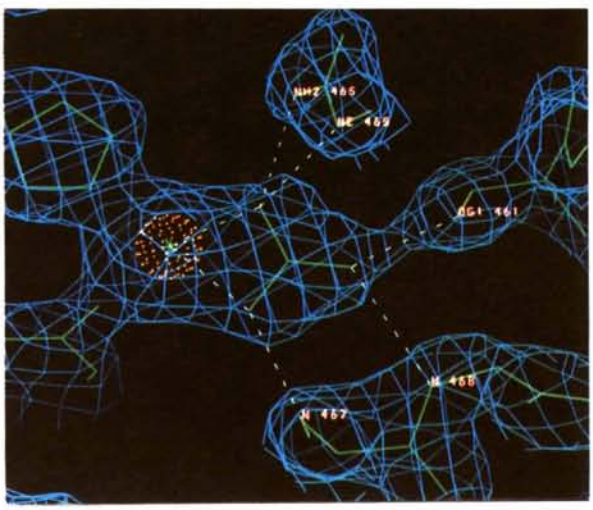


Fig. 8. Electron density for the carbonate ion showing its bidentate coordination to the metal ion (red sphere) and hydrogen bonding with the surrounding protein structure. Density for the C-lobe site, taken from the final $2F_o - F_c$ map, contoured at a level of 1.5σ.

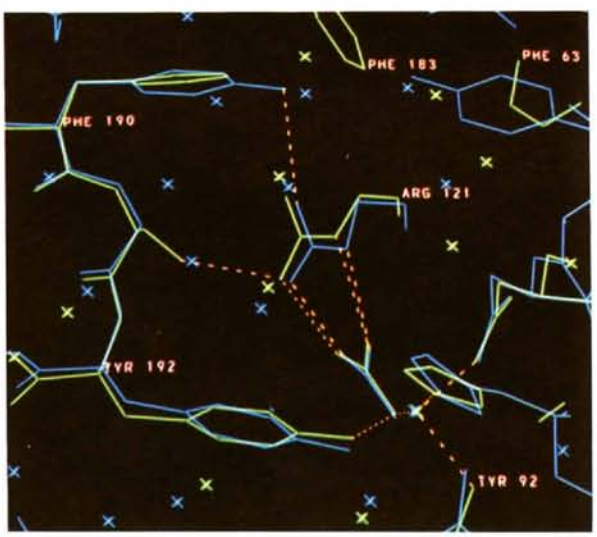


Fig. 9. Superposition of the 'essential' Arg residues from the two anion-binding sites, Arg121 (yellow) and Arg465 (blue), showing their different interactions with Ser191 (N lobe) and Tyr526 (C lobe).

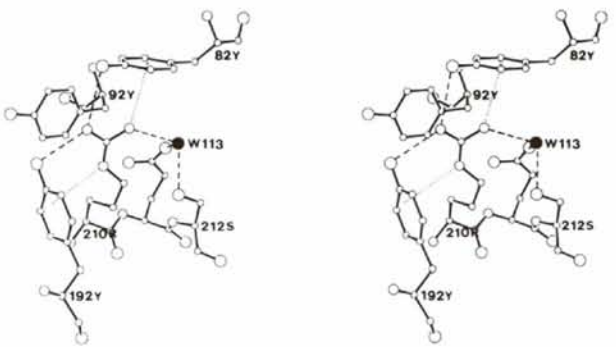


Fig. 10. Stereo diagram showing the interactions made by Arg210 at the back of the N-lobe iron site. Hydrogen bonds are shown, and include both amino-aromatic (dotted lines) and conventional hydrogen bonds (dashed lines). The guanidinium group also packs plane-to-plane with one of the iron ligands, Tyr92.

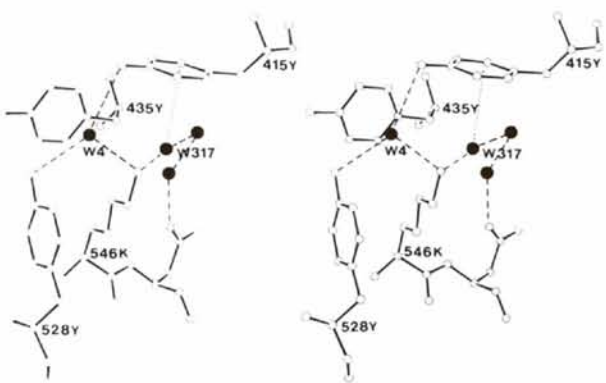


Fig. 11. Stereo diagram showing the environment of Lys546 at the back of the C-lobe iron site. The orientation is the same as for Fig. 11 so that the interactions made by Lys546 can be compared with those for Arg210. Note how water *W*317 occupies the position of Arg210 NH1, and *W*4 the position of Arg210 NH2.

lectin complex recently described (Bourne, Mazurier, Legrand, Rougé, Montreuil, Spik & Cambillau, 1994). The observations may have more to do with the ways the experiments were carried out, since our own binding studies using purified deglycosylated lactoferrin showed no difference in iron-binding ability when compared with the native protein (Baker, Norris & Baker, unpublished).

The characteristics of the molecule as a whole are that each of the four domains shows a strong structural integrity, enhanced by a number of highly conserved side chains which maintain the structures of loop regions, often holding them to the secondary-structural core of the domain. The interactions between domains, on the other hand, are weak and relatively non-specific, consistent with the requirements of a protein whose function demands flexibility and in which domains move as rigid bodies.

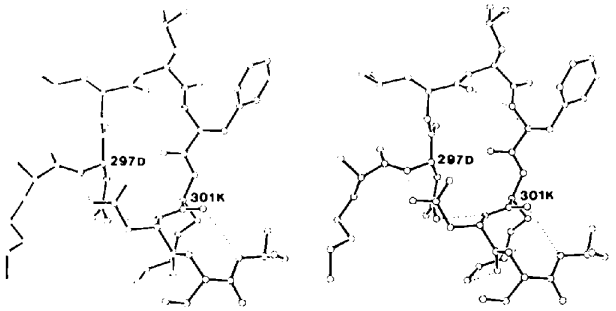


Fig. 12. Stereo diagram showing residues 296–304 which form part of one wall of the binding cleft and are highly conserved in transferrins. Note the γ-turn 298–300 and hydrogen bonds from two Asp side chains, 297 and 302, which help stabilize the polypeptide conformation.

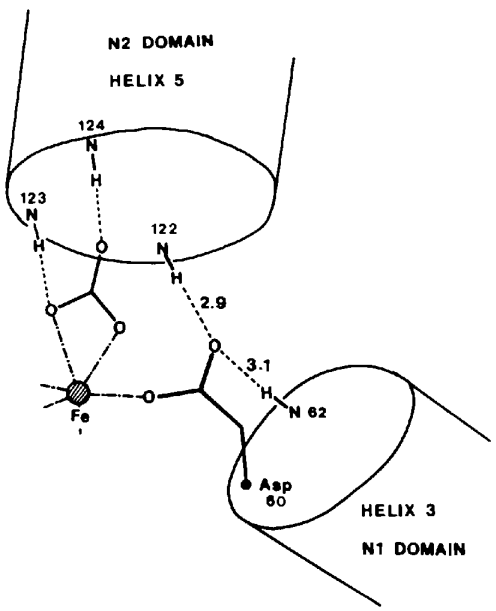


Fig. 13. Schematic diagram showing the role of the Asp ligand in simultaneously binding to iron and hydrogen bonding between two helix N termini. Distances and residue numbers are for the N-lobe site.

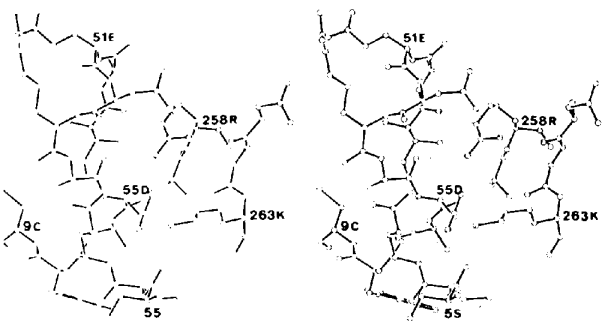


Fig. 14. An example of the role played by conserved polar side chains in tying together different secondary-structural elements through extensive hydrogen-bond networks. Asp55, on β-strand *c* of the N lobe, hydrogen bonds to 7 N (β-strand *a*), Lys263 NZ (helix 9) and Arg258 (β-strand *j*), while Arg258 also hydrogen bonds to 49O, 54O and Asp52 (helix 2) and 261O (loop between β-strand *j* and helix 9).

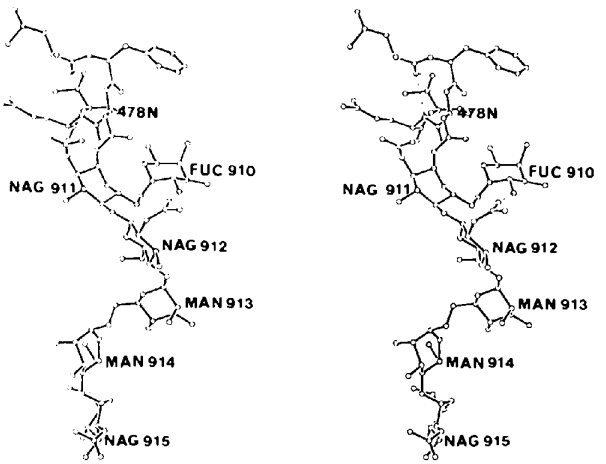


Fig. 15. Stereo diagram showing the structure of the six residues of the C-lobe carbohydrate chain that have been modelled.

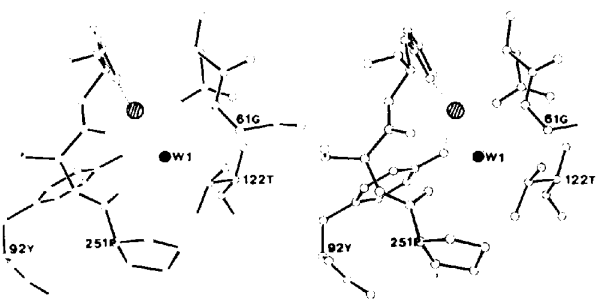


Fig. 16. Water molecule *W1*, which is tightly bound close (3.8 Å) to the N-lobe iron site, and which may be the bound water molecule detected by spectroscopic studies (see text). A water molecule occupies the equivalent site in the C lobe and these appear to be a conserved feature of transferrins.

We gratefully acknowledge support from the US National Institutes of Health (grant HD-20859), the Health Research Council of New Zealand and the New Zealand Dairy Research Institute. ENB also receives research support as an International Research Scholar of the Howard Hughes Medical Institute. Thanks are also due to colleagues at Massey University for their encouragement and help and to Professor Peter Lindley and members of the Birkbeck College transferrin group for helpful dicussions and generous sharing of unpublished data.*

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1LFG). Free copies may be obtained through The Managing Editor, International Union of Crytallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0194).

References

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.

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. (1993). *Perspectives on Bioinorg. Chem.* **2**, 161–205.

BAKER, E. N., ANDERSON, B. F., BAKER, H. M., HARIDAS, M., JAMESON, G. B., NORRIS, G. E., RUMBALL, S. V. & SMITH, C. A. (1991). *Int. J. Biol. Macromol.* **13**, 122–129.

BAKER, H. M., ANDERSON, B. F., SMITH, C. A. & BAKER, E. N. (1995). In preparation.

BAKER, E. N. & HUBBARD, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

BAKER, E. N. & LINDLEY, P. F. (1992). *J. Inorg. Biochem.* **47**, 147–160.

BAKER, E. N. & RUMBALL, S. V. (1977). *J. Mol. Biol.* **111**, 207–210.

BALI, P. K. & AISEN, P. (1991). *Biochemistry*, **30**, 9947–9952.

BOURNE, Y., MAZURIER, J., LEGRAND, D., ROUGÉ, P., MONTREUIL, J., SPIK, G. & CAMBILLAU, C. (1994). *Structure*, **2**, 209–219.

BROCK, J. H. (1985). *Metalloproteins*, Vol. 2, edited by P. M. HARRISON, pp. 183–262. London: McMillan.

BURLEY, S. K. & PETSKO, G. A. (1986). *FEBS Lett.* **203**, 139–143.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4 (1994). *Acta Cryst.* D**50**, 760–763.

DAY, C. L., ANDERSON, B. F., TWEEDIE, J. W. & BAKER, E. N. (1993). *J. Mol. Biol.* **232**, 1084–1100.

DEWAN, J. C., MIKAMI, B., HIROSE, M. & SACCHETINI, J. C. (1993). *Biochemistry*, **32**, 11963–11968.

EVANS, R. W. & WILLIAMS, J. (1978). *Biochem. J.* **173**, 543–552.

FLOCCO, M. M. & MOWBRAY, S. L. (1994). *J. Mol. Biol.* **235**, 709–717.

FURMANSKI, P., LI, Z. P., FORTUNA, M. B., SWAMY, V. B. & DAS, M. R. (1989). *J. Exp. Med.* **170**, 415–429.

GARRATT, R. C., EVANS, R. W., HASNAIN, S. S. & LINDLEY, P. F. (1986). *Biochem. J.* **233**, 479–484.

GERNSTEIN, M., ANDERSON, B. F., NORRIS, G. E., BAKER, E. N., LESK, A. M. & CHOTHIA, C. (1993). *J. Mol. Biol.* **234**, 357–372.

HARRIS, D. C. & AISEN, P. (1989). *Iron Carriers and Iron Proteins*, edited by T. M. LOEHR, pp. 241–351. New York: VCH.

HASNAIN, S. S., EVANS, R. W., GARRATT, R. C. & LINDLEY, P. F. (1987). *Biochem. J.* **247**, 369–375.

HOL, W. G. J., VAN DUIJNEN, P. T. & BERENDSEN, H. J. C. (1978). *Nature (London)*, **273**, 443–446.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KOENIG, S. H. & SCHILLINGER, W. E. (1969). *J. Biol. Chem.* **244**, 6520–6526.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

KRETCHMAR, S. A. & RAYMOND, K. N. (1986). *J. Am. Chem. Soc.* **108**, 6212–6218.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LEE, B. & RICHARDS, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

LEGRAND, D., MAZURIER, J., COLAVIZZA, D., MONTREUIL, J. & SPIK, G. (1990). *Biochem. J.* **266**, 575–581.

LEVITT, M. & PERUTZ, M. F. (1988). *J. Mol. Biol.* **201**, 751–754.

LI, Y. & FURMANSKI, P. (1993). *Biochem. Biophys Res. Commun.* **196**, 686–691.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MATTHEWS, B. W. (1972). *Macromolecules*, **5**, 818–819.

NEMETHY, G. & PRINTZ, M. P. (1972). *Macromolecules*, **5**, 755–758.

NORRIS, G. E., JAMESON, G. B., ANDERSON, B. F., THOMAS, D. H. & BAKER, E. N. (1995). In preparation.

RAMAKRISHNAN, G. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

SARRA, R., GARRATT, R., GORINSKY, B., JHOTI, H. & LINDLEY, P. (1990). *Acta Cryst.* B**46**, 763–771.

SCHLABACH, M. R. & BATES, G. W. (1975). *J. Biol. Chem.* **250**, 2182–2188.

SHARADA DEVI, A., DAS, M. R. & PANDIT, M. W. (1994). *Biochim. Biophys. Acta*, **1205**, 275–281.

SHONGWE, M. S., SMITH, C. A., AINSCOUGH, E. W., BAKER, H. M., BRODIE, A. M. & BAKER, E. N. (1992). *Biochemistry*, **31**, 4451–4458.

SIBANDA, B. L., BLUNDELL, T. L. & THORNTON, J. M. (1989). *J. Mol. Biol.* **206**, 759–777.

SMITH, C. A., ANDERSON, B. F., BAKER, H. M. & BAKER, E. N. (1992). *Biochemistry*, **31**, 4527–4533.

SMITH, C. A., ANDERSON, B. F., BAKER, H. M. & BAKER, E. N. (1995). *Acta Cryst.* D**51**, 302–316.

SPIK, G., STRECKER, G., FOURNET, B., BOUQUELET, S., MONTREUIL, J., DORLAND, L., HALBEEK, H. VAN & VLIEGENTHART, J. F. G. (1989). *Eur. J. Biochem.* **121**, 413–419.

STOWELL, K. M., RADO, T. A., FUNK, W. D. & TWEEDIE, J. W. (1991). *Biochem. J.* **276**, 349–355.

TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

VENKATACHALAM, C. M. (1968). *Biopolymers*, **6**, 1425–1436.

WILMOT, C. M. & THORNTON, J. M. (1988). *J. Mol. Biol.* **203**, 221–232.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

WONACOTT, A. J. (1980). *A Suite of Programs for On-line Evaluation and Analysis of Integrated Intensities on Small Angle Rotation/Oscillation Photographs*, Cambridge, England.