Acta Cryst. (1998). D**54**, 1319–1335

Structure of Human Apolactoferrin at 2.0 Å Resolution. Refinement and Analysis of Ligand-Induced Conformational Change

Geoffrey B. Jameson,[a] Bryan F. Anderson,[b] Gillian E. Norris,[b] David H. Thomas[b] and Edward N. Baker[b]*†

[a]*Department of Chemistry, Massey University, Palmerston North, New Zealand, and* [b]*Department of Biochemistry, Massey University, Palmerston North, New Zealand. E-mail: ted.baker@auckland.ac.nz*

(*Received 25 September 1997; accepted 23 March 1998*)

Abstract

The three-dimensional structure of a form of human apolactoferrin, in which one lobe (the N-lobe) has an open conformation and the other lobe (the C-lobe) is closed, has been refined at 2.0 Å resolution. The refinement, by restrained least-squares methods, used synchrotron radiation X-ray diffraction data combined with a lower resolution diffractometer data set. The final refined model (5346 protein atoms from residues 1–691, two Cl^- ions and 363 water molecules) gives a crystallographic *R* factor of 0.201 (R_{free} = 0.286) for all 51305 reflections in the resolution range 10.0–2.0 Å. The conformational change in the N-lobe, which opens up the binding cleft, involves a 54° rotation of the N2 domain relative to the N1 domain. This also results in a small reorientation of the two lobes relative to one another with a further ∼730 Å^2 of surface area being buried as the N2 domain contacts the C-lobe and the inter-lobe helix. These new contacts also involve the C-terminal helix and provide a mechanism through which the conformational and iron-binding status of the N-lobe can be signalled to the C-lobe. Surface-area calculations indicate a fine balance between open and closed forms of lactoferrin, which both have essentially the same solvent-accessible surface. Chloride ions are bound in the anion-binding sites of both lobes, emphasizing the functional significance of these sites. The closed configuration of the C-lobe, attributed in part to weak stabilization by crystal packing interactions, has important implications for lactoferrin dynamics. It shows that a stable closed structure, essentially identical to that of the iron-bound form, can be formed in the absence of iron binding.

1. Introduction

Lactoferrin is a monomeric glycoprotein which is a member of the transferrin family of iron-binding proteins (Brock, 1985; Harris & Aisen, 1989; Baker, 1994). It is found in many mammalian secretions (milk, tears, saliva, mucosal and genital secretions) as well as in white blood cells (Brock, 1985) and has a wide range of reported binding activities. These include, in addition to iron binding, an ability to bind to various kinds of cells (Birgens *et al.*, 1983) and to bind negatively charged species such as DNA (Hutchens *et al.*, 1989), heparin and other glycosaminoglycans (Mann *et al.*, 1994) and lipopolysaccharide (Elass-Rochard *et al.*, 1995). This diversity of interactions accompanies a diversity of proposed functions, which include roles as an antibacterial and antiviral agent, as an antioxidant, as a modulator of the immune and inflammatory responses, as a growth factor and in iron absorption (reviewed by Sanchez *et al.*, 1992); however, many of these proposed functions have yet to be clearly proven.

In its iron-binding role, lactoferrin, like other transferrins, behaves as a classic binding protein. Two Fe^{3+} ions are bound, with high affinity ($K_{app} \simeq 10^{20}$), together with two CO_3^{2-} ions (Baker, 1994). Mechanisms also exist for the release of this tightly bound iron; iron release is stimulated by low pH (Mazurier & Spik, 1980) and, in the case of transferrin at least, by receptor binding (Bali & Aisen, 1991). The apparently conflicting requirements of strong iron binding coupled with an ability to release this tightly bound iron are met by large-scale conformational changes which are of critical importance to the biological activities of these proteins.

The conformational changes that accompany binding and release have been demonstrated by physical and crystallographic studies and are understandable in terms of the three-dimensional structures of the proteins. The structures of lactoferrin, serum transferrin and ovotransferrin are closely similar (Anderson *et al.*, 1987, 1989; Bailey *et al.*, 1988; Kurokawa *et al.*, 1995). In each case the polypeptide chain is folded into two globular lobes representing its N- and C-terminal halves. The two lobes are joined by a short connecting peptide and also make contact *via* a 'cushion' of hydrophobic residues packed between them. Each lobe is further divided into two domains, N1 and N2 for the N-lobe and C1 and C2 for the C-lobe, with the iron-binding site in each lobe being situated deep in the interdomain cleft. Both iron sites have the same construction, with one domain (N1 or C1) providing two ligand residues (Asp and His) and the other domain (N2 or C2) providing two more (two Tyr) together with the bidentate CO_3^{2-} ion. This structure suggested a mechanism by which iron might be

† Present address: School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand.

ISSN 0907-4449

released, involving opening of the binding cleft by a rigid-body domain movement (Baker *et al.*, 1987) which would essentially pull the binding site apart.

The precise nature of this conformational change was demonstrated by our crystallographic analysis of human apolactoferrin at 2.8 Å resolution (Anderson *et al.*, 1990; Norris *et al.*, 1991). Two striking features were apparent: in the N-lobe the binding cleft was opened wide by a rigid-body rotation of the N2 domain relative to N1 of 54°, while the C-lobe, in contrast, was closed even though no metal ion appeared to be bound. Subsequent X-ray solution-scattering measurements have shown that the same kind of conformational change occurs in lactoferrin, serum transferrin and ovotransferrin, but that both lobes are opened in the iron-free forms (Grossmann *et al.*, 1992). The closed C-lobe in our apolactoferrin crystals is assumed to result from crystal-packing considerations, which have selected one possible conformation from several able to coexist in solution. Studies of a second crystal form of apolactoferrin (Baker *et al.*, 1997) have now given crystallographic evidence for structures with both lobes open, but the resolution is limited (3.5 Å).

A number of questions arose from the original apolactoferrin structure analysis which could only be answered by refinement at higher resolution. Are the interactions between the two lobes altered by the opening of the N-lobe? If so, this could give a structural basis to the communication between the two sites that has been shown by mutagenesis (Ward *et al.*, 1996). How are the potential iron ligands and other key functional groups arranged in the open N-lobe? These will be relevant to the mechanisms of iron binding. Are there any conformational differences apparent in the N-lobe that are remote from the immediate environment of the hinge? The nature of the hinge itself and its dependence on a change between two interfaces have been analysed in detail (Gerstein *et al.*, 1993), but further refinement may indicate other sites that adjust when closure occurs. Are there bound ions or solvent molecules that explain why the C-lobe can have a stable closed structure even in the absence of iron binding? To what extent are stabilizing interactions, including those contributed by structural solvent molecules, conserved between open and closed structures?

Here we present details of the refinement of this form of apolactoferrin at 2.0 Å resolution, together with a description of the structure, including bound solvent, and a comparison with the iron-bound form of the protein.

2. Experimental

2.1. *Crystallization*

Crystals were grown as previously described (Norris *et al.*, 1989), from low ionic strength solutions containing 0.05 *M* Tris–HCl pH 7.8, 5%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD) and 5%(*v*/*v*) absolute ethanol, with a protein concentration of 12 mg ml^{-1}. High-quality crystals could only be obtained after enzymatic removal of the carbohydrate, which was shown not to affect the properties of iron binding and release. The crystals were shown to be orthorhombic, cell dimensions a = 152.1, b = 94.6, c = 55.8 Å, space group $P2_12_12_1$, with one molecule of apolactoferrin in the asymmetric unit and a solvent content of ~53% (V_m = 2.61 Å^3 Da^{-1}).

2.2. *Data collection*

The original structure determination of apolactoferrin at 2.8 Å resolution (Anderson *et al.*, 1990; Norris *et al.*, 1991) was based on a set of diffractometer data, collected on an Enraf–Nonius CAD4 diffractometer at room temperature with Cu $K\alpha$ radiation (λ = 1.5418 Å). This diffractometer data set, collected from two crystals, comprised 17783 unique reflections (84% with intensities $I > 2\sigma I$) representing 95% of possible data to 2.8 Å resolution.

For high-resolution refinement, a second data set was collected by oscillation photography at the Synchrotron Radiation Source, Daresbury, UK. Data were collected at room temperature with a wavelength of 0.87 Å. Three crystals, mounted about their a axes, were used to cover a total oscillation range of 92°. This gave 70 film packs, each comprising three films, which were scanned with a Joyce–Loebel Scandig 3 densitometer with a 50 µm sampling grid. The digitized intensities were processed at the Department of Molecular Biology and Biotechnology at the University of Sheffield, UK, using the *MOSFLM* set of programs developed by Wonacott (1980). Both full and partial measurements were retained. The total number of measurements to 1.9 Å resolution was 164907 reflections, which reduced to 52060 reflections. The merging R value for all measurements was 0.086, defined by $R_{\rm merge} = \sum |I_i - \bar{I}| / \sum I_i$, where I_i are the intensity values for individual measurements and $\bar{I}$ are the corresponding mean values.

The final data set used for refinement of the apolactoferrin structure was obtained by merging the diffractometer and oscillation data sets. The addition of the diffractometer data filled in those parts of the diffraction data that were missing from the oscillation data either because of geometrical cut-offs or because of saturation of the film response; these were particularly significant at low resolution. Merging of the two data sets gave a final data set of 53862 independent reflections from a total of 177342 measurements in the resolution range 17–1.9 Å. Of these, 147774 were fully recorded and 29205 partially recorded. The fractional bias of fully recorded reflections relative to partials was −0.012 overall. The overall completeness of the data set

Table 1. *Data collection and processing*

Data collection

	Diffractometer (CAD-4)	Oscillation data (synchrotron)
Number of crystals	1	3
Maximum resolution (Å)	2.8	1.9
Total measurements	19413	164907
Unique reflections	17783	52060
Completeness (%)	95.0	90.2
Reflections with $I > 2\sigma I$	14912	43470
$R_{\rm merge}$	0.090	0.086

Combined data set (2.0 Å)

Resolution range (Å)	*N*ref	Multiplicity	$R_{\rm merge}$	I/σ	Completeness (%)
50.0–4.2	6360	3.0	0.057	116.2	98.8
4.2–3.2	7620	3.7	0.061	67.6	99.9
3.2–2.7	9243	3.5	0.094	26.5	97.9
2.7–2.35	10315	3.3	0.195	10.4	96.0
2.35–2.12	11485	3.2	0.321	5.1	95.4
2.12–2.0	7490	3.0	0.607	2.7	91.1
Overall	52513	3.3	0.083	30.3	96.6

to 2.0 Å resolution was 92%. These and other data-collection statistics are summarized in Table 1.

2.3. *Structure refinement*

The starting model for refinement at 2.0 Å resolution was the structure of human deglycosylated lactoferrin described previously at 2.8 Å resolution (Anderson *et al.*, 1990; Norris *et al.*, 1991). This model comprised 5216 protein atoms (from residues 4–691) and one Cl^- ion which had been clearly apparent in the C-lobe anion-binding site, even at 2.8 Å resolution. Some 15 side chains were also missing from the starting model. Initial refinement was with *PROLSQ* (Hendrickson, 1985), using data in the range 5.0–2.0 Å, with restraints on bond lengths, bond angles, planar groups, non-bonded contacts and thermal parameters. A series of rounds of refinement and rebuilding completed the model, but the refinements were rather unstable. Eventually, further refinement was initiated with *SHELXL96* (and *SHELXL97*), where conjugate-gradient least-squares refinement was on F^2 with reflections weighted by their inverse variance as determined from counting statistics and data averaging.

Markedly better electron-density maps permitted (i) rebuilding of a problematic loop at 139–143 with a *trans* conformation for Pro142, which in the metal-bound lactoferrin structures is in a *cis* conformation, (ii) adjustment of several peptides (which generally brought the peptide conformation into more favoured regions of the Ramachandran plot), (iii) rebuilding of the water structure, (iv) correction of side chains of valine, leucine and isoleucine to preferred conformers and (v) rebuilding of the loop at 417–422. Although data in the range 2.12–2.0 Å are rather weak [mean $I/\sigma(I) = 2.7$] electron-density maps calculated with a cut-off at lower resolution (2.2 Å) were not only less well resolved but somewhat noisier. The higher resolution data were therefore retained. The electron densities for two of the remodelled regions are shown in Fig. 1.

Table 2. *Refinement details*

Resolution range (Å)	10–2.0
Number of reflections†	51305
R factor ($R_{\rm free}$)‡	0.201 (0.286)
Number of protein atoms	5346
Number of ions	2 Cl^-
Number of water molecules	363
R.m.s. deviations from standard geometry (target r.m.s.)	
Bond lengths (Å)	0.003 (0.01)
Bond angles (°)§	∼1.7 (3.4)
Torsion angles (°)¶	19.6, 23.2 (−)
Average *B* factors	
Main-chain atoms (Å^2)	48.2
Side-chain atoms (Å^2)	57.1
Water molecules and anions (Å^2)	67.2

† All data used (no σ cutoff). ‡ For 5% of data (2146 reflections); *R* and $R_{\rm free}$ for those data for which $I > 2\sigma I$ are 0.170 and 0.253, respectively. § Estimated from r.m.s. deviation of 0.010 Å for 1–3 distances. ¶ χ_1,χ_2 angles only.

Bond distances were restrained to a target r.m.s. of 0.01 Å from the standard values of Engh & Huber (1991), bond angles as 1–3 distances to 0.02 Å, close contacts to 0.01 Å, planarity of aromatic rings to 0.1 Å, peptide planarity to 0.5 Å, chiral volumes to 0.10 Å^3 and thermal parameters to a difference in *B* of 6 Å^2 between adjacent atoms, except terminal atoms where an r.m.s. difference of 12 Å^2 was the target. In all categories, the r.m.s. difference was less than the target value except for close contacts where it was 0.016 Å. The structure is described by a total of 22138 parameters and restrained by 22087 restraints, including equality restraints with a target r.m.s. of 1.6 Å^2 on thermal parameters of OE1/OE2, OE1/NE2, OD1/OD2, CG1/CG2 (valine), CD1/CD2 (leucine), CD1/CD2 and CE1/CE2 (tyrosine and phenylalanine), CG/CH2 (tryptophan), NH1/NH2 (arginine) CB/CG/CD (proline) and OG1/CG2 (threonine). The value of the minimized function $[\sum w(|F_o|^2 - |F_c|^2)^2]$ is essentially independent of resolution range and magnitude of $|F_c|/|F_c|_{\rm max}$. The final electron-density maps reveal many peaks in plausible positions for water molecules at or below the threshold ratio for selection [peak height/r.m.s. (electron density) of 3.0 for σ_A-weighted difference maps].

3. Results

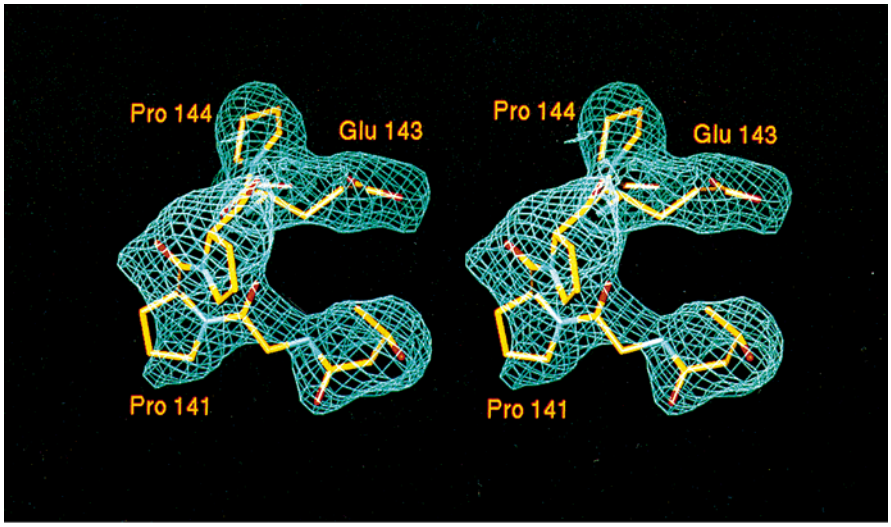
3.1. *Final model*

The final model for this form of apolactoferrin, comprising 5346 protein atoms (from residues 1–691), two Cl^- ions and 363 water molecules, shows good

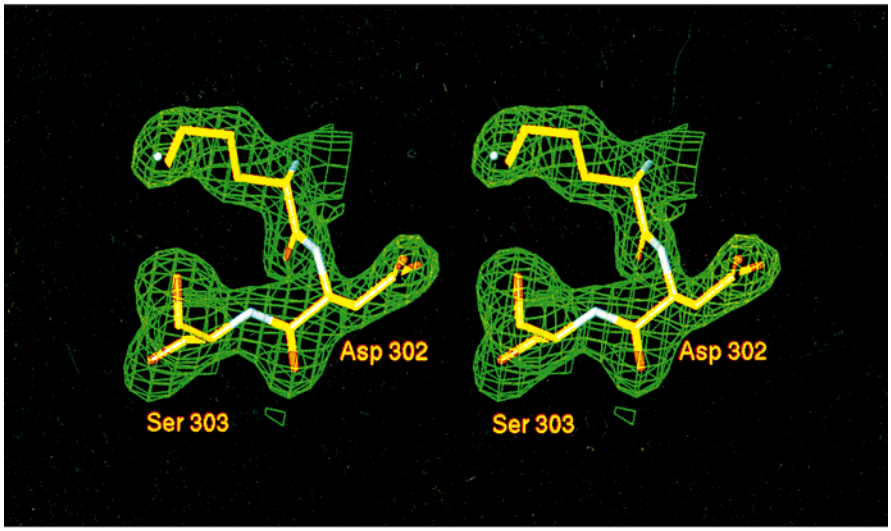
agreement with the measured X-ray data (final R factor 0.201, R_{free} = 0.286) and is tightly restrained to standard geometry. These and other relevant refinement statistics are given in Table 2. Estimates of the mean error in atomic positions have been derived from both Luzzati and σ_A plots (Luzzati, 1952; Read, 1986). These give mean coordinate errors of 0.25 and 0.30 Å, respectively, although these estimates are necessarily imperfect, especially given the wide range of B values in different parts of the molecule, reflecting different degrees of uncertainty (see below).

The difference between R and R_{free} is somewhat larger than we would like but can be attributed to the large number of unrestrained parameters accorded to water molecules and to the mobile loops and side chains whose weak electron density makes only small contributions to the structure factors while requiring the same number of parameters as better defined regions of the structure.

A Ramachandran plot (Fig. 2) of the main-chain conformational angles (Ramakrishnan & Ramachandran, 1965) shows that more than 87% of non-glycine residues lie in the 'most favoured' regions, as defined in the program *PROCHECK* (Laskowski *et al.*, 1993), with a further 12% in the 'additionally allowed' regions. Only two residues appear as outliers, in normally disallowed regions, but these residues, Leu299 and Leu642, are in homologous positions in the two lobes, in each case as the central residue in a γ-turn (Matthews, 1972; Nemethy & Printz, 1972). These two γ-turns, which are located in one wall of the binding cleft of each lobe, appear to be a conserved feature in all transferrins



(*a*)



(*b*)

Fig. 1. Representative sections of the final $2F_o - F_c$ electron-density map. (*a*) The region of the mobile loop 138–143, contoured at 1.1σ. This is one of the poorest regions of electron density. The peptide 141–142 could not be adequately fitted in the *cis* configuration found in the holo form; here it is shown modelled as *trans*, although the fit is still not perfect. (*b*) The well defined region of residues 302–303, where there is a peptide flip in ApoLf compared with Fe_2Lf; contoured at 1.4σ.

(Haridas *et al.*, 1995; Kurokawa *et al.*, 1995). The central residue in each case has (φ, ψ) values around (80, −50°) which are normal for such a structural feature (Baker & Hubbard, 1984). The γ-turns are stabilized by hydrogen bonds from conserved Lys and Glu residues and a conserved water molecule; Lys18 (357) NZ forms a hydrogen bond to 298 O (641 O), Glu15 (354) OE2 to 299 N (642 N) and the water molecule to 299 O (642 O). A number of residues near the boundaries of allowed regions also exist as structurally conserved pairs, *e.g* Trp125 and Trp469 at around (−150, −60°); these are associated with a distortion in helix 5. Finally, the peptide ω angles for the homologous Cys residues 157 and 493 have values of 158 and 149°, respectively, compared with expected values of ±180°. There is no obvious explanation, but there may be stereochemical constraints associated with the close proximity of the disulfide bond 157–173 (493–507) to two others.

The current model includes all residues of the polypeptide chain (residues 1–691). However, two qualifications should be made. Firstly, the first few residues of the polypeptide chain show some heterogeneity. In the original amino-acid sequence analysis of human milk lactoferrin the N-terminal sequence was determined as GRRRRSV (Metz-Boutigue *et al.*, 1984). Recent cDNA sequences confirm this (Powell & Ogden, 1990; Stowell *et al.*, 1991; Ward *et al.*, 1992). On the other hand, a human lactoferrin cDNA derived from the white blood cells of a chronic myeloid leukaemia patient corresponded to an amino-acid sequence beginning GRRRSV, *i.e* with one less Arg residue at the N-terminus; this was the sequence used in our structure determination of diferric human lactoferrin (Anderson *et al.*, 1989). Sequence analyses of lactoferrins from different human milk samples also reveal some heterogeneity in the number of Arg residues (Stowell *et al.*, 1991). In the present work we have modelled the N-terminus as having the sequence GRRRSV, *i.e* with only three Arg residues. The electron density is very weak, however, and we cannot unequivocally distinguish this from the alternative GRRRRSV (with the N-terminal Gly and the first Arg side chain missing).

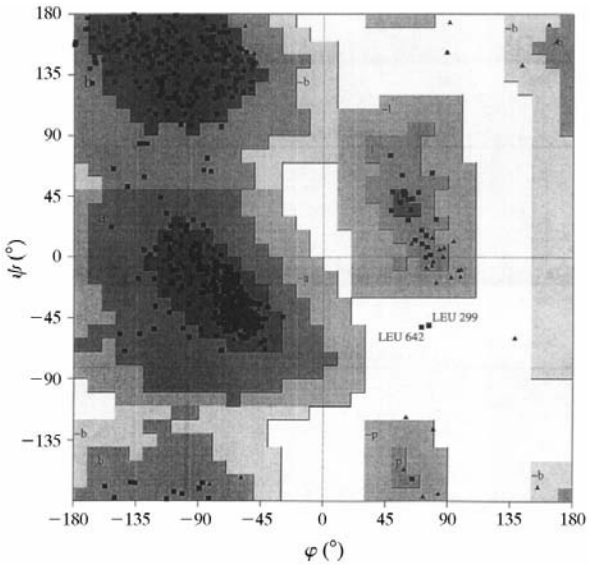


Fig. 2. Ramachandran plot of main-chain torsion angles (φ, ψ) for ApoLf. Glycine residues are indicated by triangles. Figure drawn with *PROCHECK* (Laskowski *et al.*, 1993).

Secondly, some parts of the structure have weak and poorly defined electron density. These include, in addition to residues 1–4, residues 140–142, 219–223 and 417–420, together with the terminal atoms of some 30 of the longer side chains (primarily Arg, Lys, Glu and Gln residues). The electron density of the polypeptide chain is everywhere continuous in $2F_o - F_c$ maps, but only at the level of 1.0σ for the above regions; for these parts of the polypeptide the *B* factors converge at 100–120 Å^2 on refinement and some structural microheterogeneity clearly exists. Side chains for which the terminal atoms have *B* factors greater than 100 Å^2 include Ser5, Gln13, Lys27, Arg28, Arg30, Glu86, Arg87, Lys100, Ser103, Arg120, Asn137, Trp139, Arg210, Asn261, Asp265, Glu276, Lys333, Glu336, Gln479, Glu512, Glu514, Met603, Asn623, Glu660, Gln666, Glu687 and Lys691.

The average *B* factor for all protein atoms is 45.0 Å^2. This relatively high value can be attributed in part to the inclusion of the more poorly ordered parts of the structure, described above, but it also reflects the overall ordering of the crystals; the overall *B* factor derived from a Wilson plot of the measured X-ray data is 40.0 Å^2. The distribution of *B* factors along the polypeptide chain conforms to the expected pattern, with internal regions and secondary structures having lower values (20–45 Å^2) and external loops such as those given above having much higher values.

3.2. *Molecular organization*

The overall structure of this form of apolactoferrin is shown in Fig. 3. The features which distinguish it from the holo form of the protein principally involve the ways in which the four structural domains are reoriented relative to one another. Otherwise the molecular organization does not change from that previously described for diferric lactoferrin (Anderson *et al.*, 1987, 1989). In particular the secondary structural elements are not changed and we use the same notation for helices and β-strands that was used previously.

In the N-terminal lobe a large domain movement (see below) leaves the binding cleft wide open and the location of the hinge between the two domains (Gerstein *et al.*, 1993) allows us to define the domain boundaries more clearly; the N1 domain comprises residues 1–90 and 251–332 and the N2 domain comprises residues 91–250. The C-terminal lobe remains closed, as in the diferric protein, but by analogy with the N-lobe

the C1 domain comprises residues 345–433 and 595–678 and the C2 domain comprises residues 434–594 (for a sequence alignment of N- and C-lobes see Anderson *et al.*, 1989).

Three helices are of particular interest with respect to the domain structure. In the N-lobe, helix 11 (residues 321–332) is closely associated with the N2 domain, yet when N2 moves this helix does not, remaining tied to N1. Thus, we define this helix as part of N1. Helix 12 (residues 334–344) is the 'connecting helix', which is the only covalent connection between the two lobes, and is not formally part of either. Finally, the C-terminal helix (residues 680–691) lies across the surface of the C-lobe and provides much of the non-covalent packing between the lobes. This section of polypeptide is joined to the C2 domain by a disulfide bridge (677–483) just prior to the C-terminal helix and to the C1 domain by a second disulfide bridge (686–405) from within this helix.

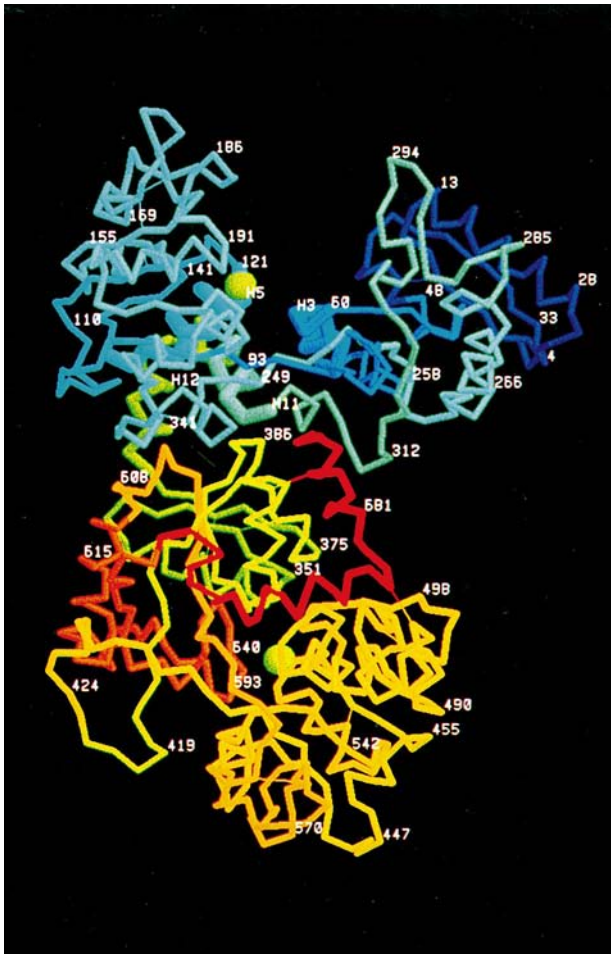


Fig. 3. Cα plot for ApoLf. Residues are colour coded by sequence number, from blue at the N-terminus through to red at the C-terminus. Disulfide bonds are shown by thin dotted lines. Cl^- ions are shown as green spheres. Helices 3, 5, 11 and 12 are shown with thicker lines.

Table 3. *Domain relationships and superpositions*

The diagonal elements show the r.m.s. differences (Å) when the Cα atoms of the equivalent domains of Fe_2Lf and ApoLf are superimposed. The off-diagonal elements show the differences in orientation (°) between pairs of domains in the two structures.

		Fe_2Lf			
		N1	N2	C1	C2
ApoLf	N1	0.39	54.3	9.0	7.9
	N2	—	0.36	55.8	54.5
	C1	—	—	0.37	1.7
	C2	—	—	—	0.41

3.3. *Domain orientations*

The relative orientations of the four domains in this form of apolactoferrin have been compared with the corresponding domains in diferric lactoferrin in a pairwise fashion; the results are summarized in Table 3. In this analysis, in each case a pair of domains is initially superimposed [*e.g.* the N1 domains of human apolactoferrin (ApoLf) and diferric human lactoferrin (Fe_2Lf)] and a further transformation is then applied to bring another pair of domains into correspondence.

When compared in this manner, the major orientational difference is in the N-lobe, where the N2 domain has rotated 54.3° relative to N1, so as to open up the binding cleft (Fig. 4). This confirms the earlier analyses of the N-lobe conformational change (Anderson *et al.*, 1990; Gerstein *et al.*, 1993). In the C-lobe no such conformational change has occurred, the binding cleft is still closed, and the C1 and C2 domains differ by only 1.7° in their orientations in ApoLf relative to Fe_2Lf. This represents a slight twist of one domain relative to the other but no change in closure.

In addition to these movements within the two lobes, there is also a shift in one lobe relative to the other. This is apparent in Fig. 4 and corresponds to a rotation of the N-lobe relative to the C-lobe of about 8–9° (7.9° for the N1/C1 movement and 9.0° for the N1/C2 movement). This has the effect of moving the N1 domain inwards towards the C-lobe in a direction that would lead to closer contacts between the lobes. The C-terminal helix, which is intimately involved in these contacts, is also shifted relative to the rest of the C-lobe by about 1 Å (see Fig. 4).

3.4. *Conformational changes within the domains*

The conformational changes in the apolactoferrin molecule, compared with Fe_2Lf, are essentially rigid-body domain movements. This can be seen from the data in Table 3. For the N1 domain 162 Cα atoms from ApoLf can be superimposed on the corresponding atoms of Fe_2Lf with an r.m.s. deviation of only 0.39 Å. This includes the whole of the N1 domain as defined above including helix 11 and excluding only the flexible N-terminal residues 1–5 and residues 86, 89, 294 and

303. All of the excluded atoms are on flexible loops with high *B* factors, except Ser303 which is at the site of a localized conformational change, described further below. For the N2 domain 155 Cα atoms can be superimposed with an r.m.s. deviation of 0.36 Å, and only residues 139–143 excluded. For the C1 domain the figures are 166 Cα atoms with an r.m.s. deviation of 0.34 Å (residues 418–422 and 622–623 excluded) and for the C2 domain 158 Cα atoms with an r.m.s. deviation of 0.37 Å (residues 480 and 511–512 excluded). The excluded residues are all on loops with high *B* factors, and the overall r.m.s. deviations of 0.3–0.4 Å are comparable with the accuracy of the X-ray analyses, confirming that the domains move essentially as rigid bodies.

When the same superpositions are carried out using all main-chain atoms (N, Cα, C, O) it is clear that the number of local conformational differences is extremely small. In the C-lobe there are none except in the ill-defined loop 418–422 and the r.m.s. deviation when all atoms are used is the same as when Cα only are used. In the N-lobe, there are four peptide flips in the N1 domain (the peptides 85–86, 260–261, 293–294 and 302–303) and a probable change from *cis* to *trans* in the Pro141–Pro142 peptide in the N2 domain; there are also several smaller reorientations of peptide planes.

Differences in the main-chain torsion angles (φ, ψ) along the polypeptide chain, when Fe_2Lf and ApoLf are compared, are shown in Fig. 5. The conformational change at the hinge in the N-lobe, which is responsible for the rigid-body rotation of the N2 domain, has been analysed in detail already (Gerstein *et al.*, 1993); it involves changes in the main-chain torsion angles of residues 90–91 and 249–251 and the details have not been changed significantly by further refinement. Of the other conformational differences noted above, only two seem to be related to the domain movement.

Firstly, when the N2 domain moves, a close contact is made between the rather flexible loop 139–143 and the N-terminus of the connecting helix (334–344), such that in apolactoferrin a new hydrogen bond 139 O···335 N is made (Fig. 6*a*). This may be sufficient to stimulate the *cis–trans* isomerization of the 141–142 peptide, although it should also be noted that it is near the N-lobe glycosylation site (Asn137) and it is possible that deglycosylation of the apo protein could play a part. In the diferric structure this peptide is clearly in the *cis* conformation, whereas in the apo structure the density is less clear but fits best as *trans*.

Secondly, the flip of the peptide 302–303 (Fig. 6*b*) changes the β-turn 301–304 from a type II configuration in Fe_2Lf to type I in ApoLf. This turn forms part of the interdomain interface in the closed diferric structure and Lys301 forms a salt bridge with Glu216 in the other domain. We suggest that in the closed form residues 301–304 have a more strained configuration, since type II turns strongly favour Gly in position 3 (Wilmot & Thornton, 1988), whereas here residue 3 is Ser. This configuration is presumably a result of the constraints of the closed structure of Fe_2Lf and is stabilized by a hydrogen bond 303 N···289 O. In the open apo structure, however, the constraints are removed, the structure relaxes and the peptide flips to give the more favoured type I turn; the 289–290 peptide is also reoriented, and these residues assume more favourable (φ, ψ) values.

As for the main chain, there are very few side chains that undergo conformational changes that can be linked directly with the domain movement. Two that are of

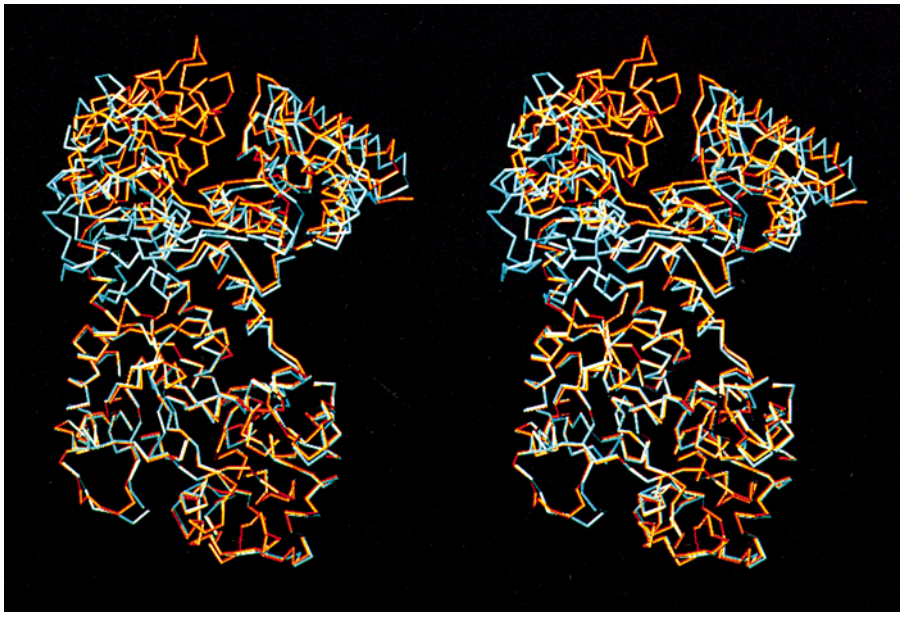


Fig. 4. Superposition of ApoLf (blue) on to Fe_2Lf (orange) showing the differences in domain orientations. The superposition is based on the matching of the two C-lobes.

particular interest, however, are Arg89 and Arg133. Arg89 forms a salt bridge with Glu211 in the opposing (N2) domain, near the hinge at the back of the N-lobe iron site (see Fig. 6*c*). This salt bridge is maintained in both ApoLf and Fe_2Lf, but at the expense of a conformational adjustment of the Arg side chain (Glu211 does not adjust as it is fixed by an intradomain hydrogen bond 211 OE···93 N); the χ_1–χ_4 angles for Arg89 change from (52, −173, 168, 80°) in Fe_2Lf to (−77, −90, −83, −124°) in ApoLf. Arg133 belongs to helix 5 of the N2 domain but hydrogen bonds to the C-terminus of helix 11 (321–332), which is tied to domain N1. When the N2 domain moves Arg133 remains hydrogen bonded to the helix 11 C-terminus, albeit slightly differently, but is forced to adopt a less extended configuration (Figs. 6*d* and 6*e*), with its χ_1–χ_4 angles changed from (−146, −151, 174, 99°) in Fe_2LF to (165, 62, −159, −87°) in ApoLf. This helps to tie the N2 domain, as it moves, to helix 11.

Two other interactions that appear to be of importance to the transition between open and closed states involve the residues Glu66, Arg332 and Tyr324. Glu66 is on helix 3 of the N1 domain and Arg332 is on helix 11; these two residues form a salt bridge in both the diferric and apo forms (Figs. 6*d* and 6*e*). The arginine side chain is in a more extended conformation in the ApoLf structure but the maintenance of this salt bridge may be very important in tying helix 11 to the N1 domain. Tyr324 is also on helix 11. In the closed diferric form it is hydrogen bonded to Thr122 and Asn126 on helix 5 of the N2 domain; as the N2 domain moves by the pivoting of helix 5 on helix 11 (Anderson *et al.*, 1990; Gerstein *et al.*, 1993) as shown in Figs. 6(*d*) and 6(*e*) these hydrogen bonds are broken. In the apo form a water molecule bridges between Tyr324 and Thr122; the change to direct hydrogen bonding between Tyr324 and Thr122 may help lock the closed form together.

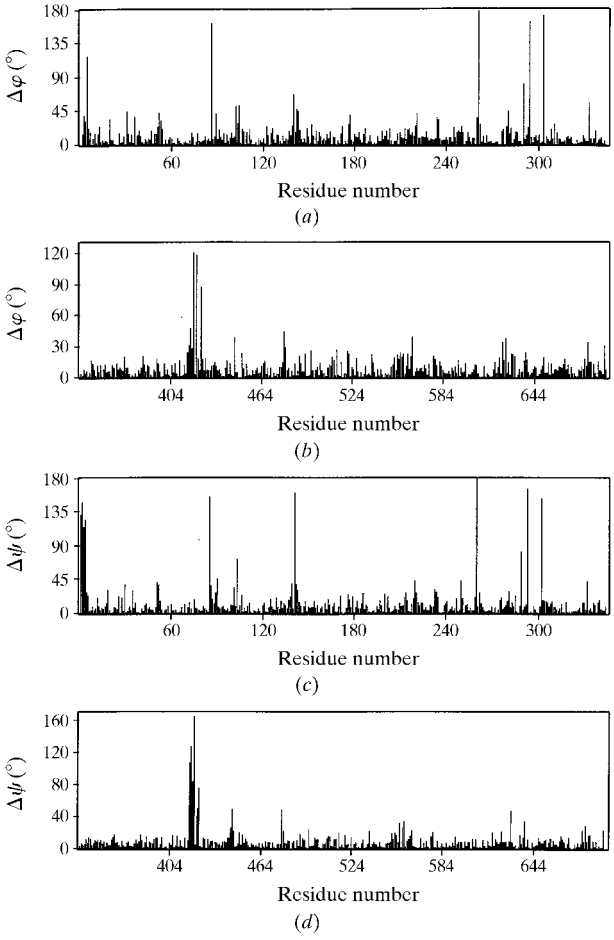


Fig. 5. Differences in main-chain torsion angles ($\Delta\varphi$ and $\Delta\psi$) between ApoLf and Fe_2Lf plotted as a function of sequence number. (*a*) $\Delta\varphi$ for N-lobe; (*b*) $\Delta\varphi$ for C-lobe; (*c*) $\Delta\psi$ for N-lobe; (*d*) $\Delta\psi$ for C-lobe.

The major effect upon side chains of the conformational change in the N-lobe is to change the solvent exposure of side chains in two regions. This phenomenon has been analysed by Gerstein *et al.* (1993), in terms of a change between the open form in which a large interface is exposed (primarily residues of the binding cleft) and the closed form in which a small interface becomes exposed (primarily residues packed between helices 5 and 11). This analysis is not changed by our further refinement of the ApoLf structure; opening of the domains in ApoLf exposes ∼900 Å^2 of solvent-accessible surface in the large interface and buries ∼300 Å^2 in the small interface. Very few side chains in either of these interfaces undergo any change in their conformational angles.

3.5. *N-lobe binding cleft*

The binding cleft contains a number of residues which play important roles either in metal and anion binding or in interdomain interactions in the holo protein. The four iron-binding residues Asp60, Tyr92, Tyr192 and His253 have similar orientations in ApoLf and in Fe_2Lf. On the N1 domain the side chain of Asp60 changes its χ_1 angle from 164 to 67°. It remains hydrogen bonded to the N-terminus of helix 3, acting as an N-cap (Presta & Rose, 1988; Richardson & Richardson, 1988), but moves from hydrogen bonding to 62 NH in Fe_2Lf to 63 NH in ApoLf. Adjacent to Asp60 is His253 with its side chain projecting into the binding cleft; there is no hydrogen-bonded interaction between these residues, however, and in fact they are further apart than in Fe_2Lf as a result of the movement of Asp60. On the N2 domain Tyr92 and Tyr192 are hydrogen bonded together (92 OH···192 OH, 3.15 Å).

Adjacent to the two Tyr residues is Arg210 with its guanidinium group stacked plane-to-plane with the aromatic ring of Tyr92 about 3.5 Å away; again it has essentially the same conformation as it does in Fe_2Lf, although the side-chain *B* factors indicate that it is much less firmly fixed (mean B = 98 Å^2 in ApoLf, 29 Å^2 in

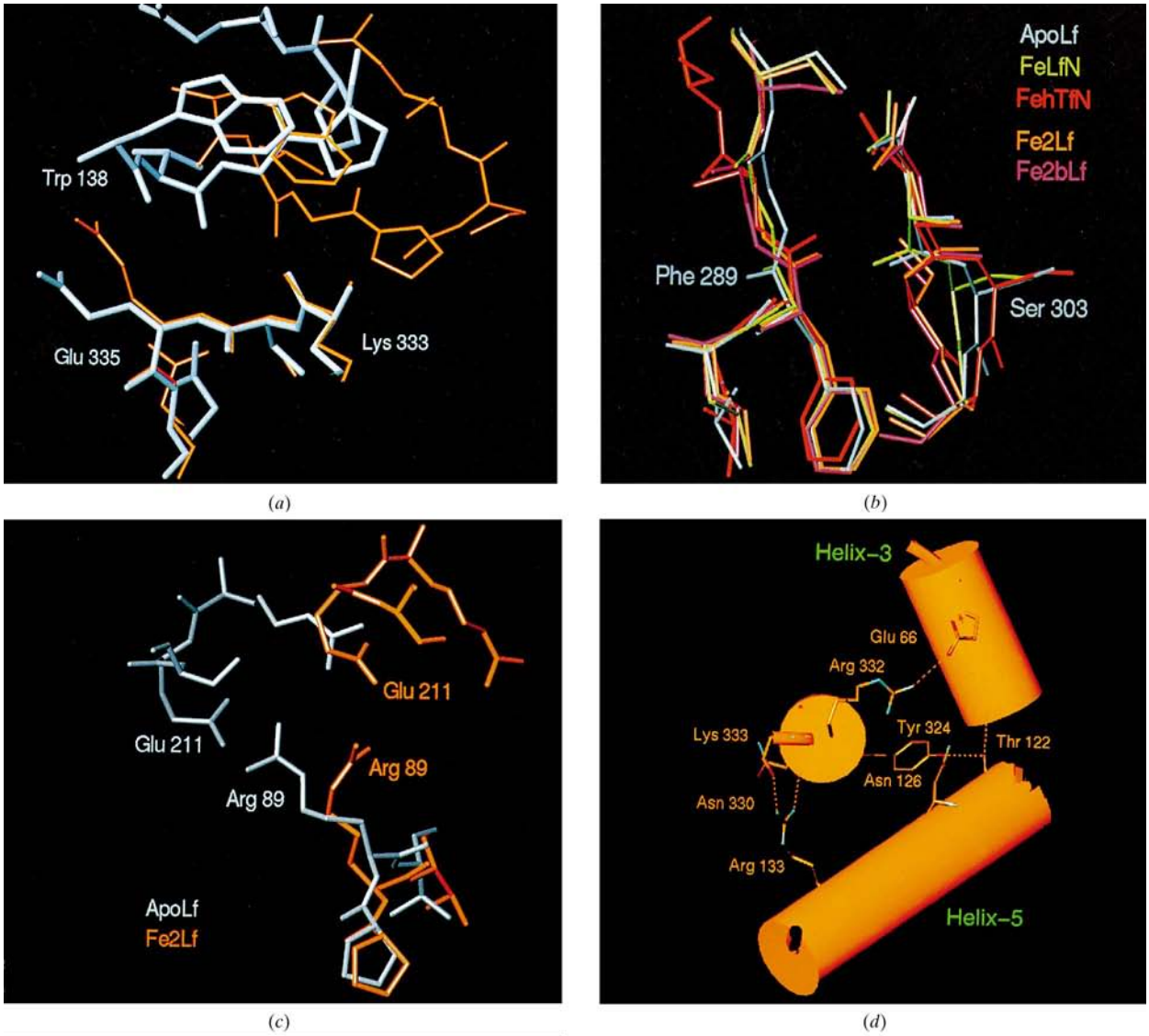
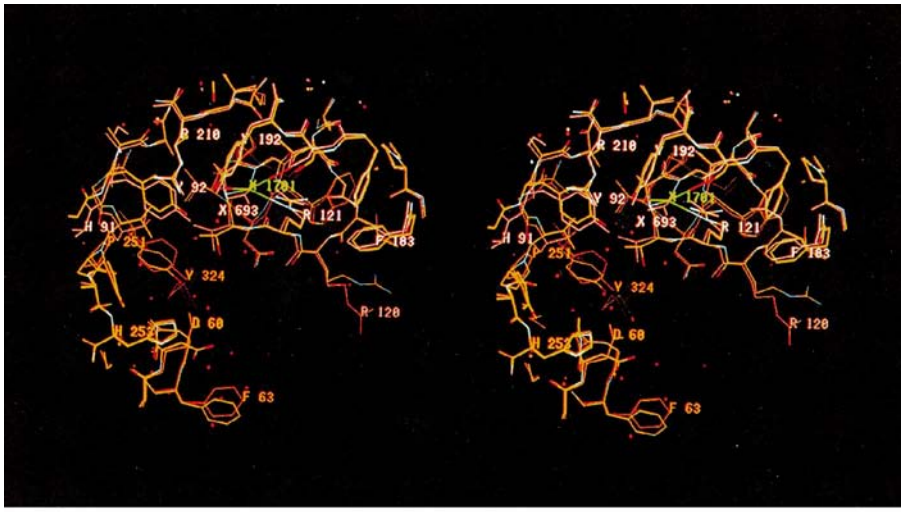


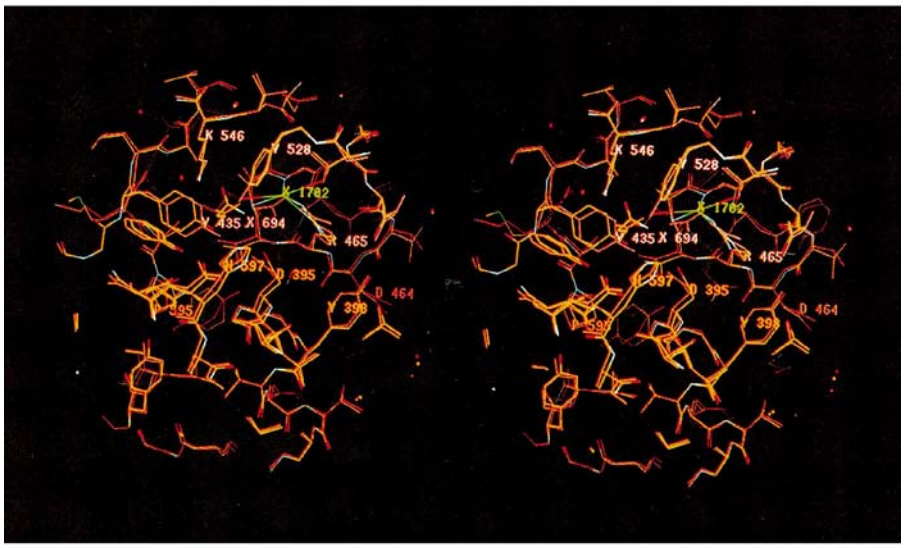
Fig. 6. (*a*) Conformational changes that result from movement of the N1 and N2 domains, with ApoLf in blue and Fe_2Lf in orange. (*a*) Movement of the loop 137–143 to contact residues 334–337 at the N-terminus of the connecting helix in ApoLf. (*b*) Conformational differences in the 302–303 peptide and the neighbouring region 287−291. Here the N1 domains of ApoLf (blue), human Fe_2Lf (orange), bovine Fe_2Lf (magenta) and the ferric N-terminal half-molecules of human transferrin (red) and human lactoferrin (green) are superimposed. (*c*) Conformational changes in the side chain of Arg89 that maintain the Arg89–Glu211 salt bridge in ApoLf (blue) and Fe_2Lf (orange). (*d*) Side-chain interactions linking helix 3, helix 11 and helix 5 in Fe_2Lf. (*e*) Helix movements and changes in side-chain hydrogen bonding in the transition between Fe_2Lf (orange helices, yellow side chains) and ApoLf (blue helices, blue side chains).

Fe_2Lf). On the other side of the two Tyr residues is Arg121 and the N-terminus of helix 5, which comprise the CO_3^{2-} binding site on the N2 domain in Fe_2Lf (Fig. 7*a*). These residues are again arranged as in Fe_2Lf, but instead of CO_3^{2-} a possible Cl^- ion occupies this site. We identify this as Cl^- because (i) its density is spherical, (ii) when modelled as Cl^- it has a *B* factor comparable with the surrounding structure and (iii) the distances to potential hydrogen-bond donors are all in the range 3.2–3.6 Å: Thr117 OG1, 3.39 Å; Arg121 NE, 3.51 Å; Ala123 N, 3.55 Å; Tyr192 OH, 3.17 Å.

In addition to the iron ligands, a number of side chains that project into the binding cleft are involved in specific interactions in the closed holo structure. They include Arg121 (at the anion-binding site), Tyr82, Thr122, Arg210, Glu216, Asp217, Lys296 and Lys301 (all involved in interdomain hydrogen bonds). Others that make water-bridged interdomain interactions or van der Waals contacts are Ser12, Phe63, Phe183, Ser185, Ser193, Gln295 and Asp297. Two general observations can be made. Firstly, almost all of these side chains have the same conformations in ApoLf as in Fe_2Lf, making the point that this is a largely preformed interface, in which the side chains have favourable conformations in the apo form that fit together upon domain closure (Gerstein *et al.*, 1993). Only Ser185, Lys296 and Lys301 have significantly different orientations, the latter two because they become reoriented to form interdomain salt bridges in the holo protein (Lys296···Asp217 and Lys301···Glu216). Secondly, the *B* factors are consid-



(*a*)



(*b*)

Fig. 7. Stereoviews of the two binding clefts of ApoLf. (*a*) The open N-lobe cleft showing its preformed nature. (*b*) The closed C-lobe cleft. In both diagrams the ApoLf structure is taken as the reference and is coloured by atom type. Superimposed separately on it are the N1(C1) domains (orange) and the N2 (C2) domains (mauve) taken from the Fe_2Lf structure. The two Cl^- ions (X1701 and X1702) are green, with hydrogen bonds shown by dotted lines. The Fe^{3+} and CO_3^{2-} ions of Fe_2Lf are denoted X693 (X694).

erably higher in the open cleft of the apo protein. The mean *B* factor for the side chains of the 15 residues listed above is 59 Å^2 in ApoLf and 43 Å^2 in Fe_2Lf (both ApoLf and Fe_2Lf have similar overall *B* factors). The mean *B* factor for the corresponding residues in the closed, but iron-free, C-lobe binding cleft of ApoLf is lower (48 Å^2) and more similar to the mean value for the C-lobe of Fe_2Lf (36 Å^2).

3.6. *C-lobe binding cleft*

The closed C-lobe of this form of ApoLf has an almost identical structure to that of Fe_2Lf. Although no metal ion is bound, the four side chains that act as iron ligands, Asp395, Tyr435, Tyr528 and His597 are hardly rearranged at all; the r.m.s. difference in atomic positions is only 0.38 Å when these residues, together with residues 465–468 which define the anion site, are superimposed on their equivalents on the holo protein. In these orientations the OH groups of Tyr435 and Tyr528 can form a weak hydrogen bond (3.17 Å) but there is otherwise no change. In almost all other respects the structure within the binding cleft is the same as in Fe_2Lf. All the C-lobe interdomain interactions found in Fe_2Lf are preserved in ApoLf. This includes the hydrogen bonds made by Asp395 OD1 with Gly397 N, Thr466 N and Thr466 OG1, even though no iron is bound. Even the water molecules bound in the interdomain cleft are conserved; 20 out of 23 found in the binding cleft of the Fe_2Lf C-lobe are also present in ApoLf, within 0.5 Å of the Fe_2Lf positions, and a further two are within 1.0 Å.

As in the N-lobe, a presumed chloride ion is bound in the C-lobe anion-binding site (Fig. 7*b*). It is represented by a spherical peak of electron density, and when modelled as a water molecule its *B* factor reduces to an anomalously low value; as a Cl^- ion its *B* factor is 30 Å^2, comparable with the surrounding atoms. This Cl^- ion is also situated 3.0–3.6 Å from a number of potential hydrogen-bond donors, Thr461 OG1 (2.95 Å), Arg465 NE (3.41 Å), Arg465 NH2 (3.60 Å), Gly468 N (3.38 Å) and Tyr528 OH (3.61 Å). We note that Cl^- was present in the crystallization medium, supporting this assignment.

3.7. *Interactions between the lobes*

In the diferric form of human lactoferrin, the interactions between the two lobes involve two components, the connecting helix (334–344), which provides the only covalent connection between them, and a cluster of residues that form a packed hydrophobic interface some distance from the connecting helix. These are contributed by a loop (311–314), helix 10 (315–321) and helix 11 (321–332), all from the N1 domain, and by helix 2 (376–388) from the C1 domain and the C-terminal helix (680–691) which packs against the N1 domain (Haridas *et al.*, 1995). The interactions include two salt bridges but are otherwise hydrophobic, burying ∼1200 Å^2 of solvent-accessible surface, ∼640 Å^2 between the N1 domain and the C-terminal helix and ∼540 Å^2 between N1 and the rest of the C1 domain.

The analysis of domain orientations (above) shows that the domain movement in the N-lobe does have an impact on the C-lobe, in that the relative orientations of the two lobes are changed by 8–9°. In particular the relative orientations of the N1 and C1 domains, which contribute almost all of the interactions in Fe_2Lf are changed by 7.9°. How does this occur? Here we consider the ways in which the N-lobe conformational change impacts on the C-lobe.

The residues that pack between the N1 and C1 domains (including those of the C-terminal helix) maintain essentially the same interactions in Fe_2Lf and ApoLf. None of the side chains changes its conformation, the two salt bridges Arg313···Asp379 and Asp315···Lys386 are still present, and the buried surface area between the N1 and C1 domains is essentially unchanged in ApoLf. Thus, the N1/C1 interface is preserved. In part this is because the interactions are mostly hydrophobic and somewhat diffuse; also the C-terminal helix, which provides most of the C-lobe residues in the interface, moves by 1–2 Å in response to the N-lobe movement.

The major factor in the readjustment between the two lobes is the movement of the N2 domain. When the N-lobe opens, new contacts are made with two parts of the C-lobe. Firstly, residues 232–234 from the N2 domain come into contact with residues 603–604 of the C1 domain; no specific interactions are made but ∼260 Å^2 of solvent-accessible surface is buried. Secondly, residues 243–249 move into much closer contact with residues 689–691 at the end of the C-terminal helix, burying a further 310 Å^2 of surface area. This increase in contact between the two lobes can be clearly seen in Fig. 8. Several specific interactions are formed, the side chains of Lys243 and Arg249 hydrogen bonding to terminal CO groups of the helix (243 NZ···690 O, 2.82 Å; 249 NE···689 O, 3.25 Å). The interactions between these basic side chains and the negatively charged helix C-terminus must be highly favourable. Thr91, adjacent to the hinge, is also brought close enough to the helix C-terminus to hydrogen bond weakly to it (91 OG1···691 O, 3.32 Å). Finally, there are also changes in the helix 334–344 which links the two lobes. The N-terminus of this helix is displaced by the movement of the N2 domain (Figs. 3 and 6*d*), causing this end of the helix to change the angle at which it leaves the N-lobe by 25°. At the other end of the connecting helix a similar adjustment occurs. This change, at the connection between the connecting helix and the first β-strand of the C-lobe, is about 17°, but in an opposite sense, giving a net effect of 8°.

3.8. *Solvent-accessible surface*

The domain and lobe movements, with consequent changes in interdomain and interlobe interactions between ApoLf and Fe_2Lf, lead to changes in solvent accessibility for many parts of the structure. Absolute solvent accessibility and changes in accessibility, residue by residue, are plotted in Fig. 9.

A quite unexpected feature, however, is the extent to which the total solvent-accessible surface area of apolactoferrin matches that of diferric lactoferrin; in fact the overall molecular surface area of ApoLf is 260 $Å^2$ less than that of Fe_2Lf in spite of its apparently more open structure. The opening of the N-lobe binding cleft in ApoLf increases the net solvent-accessible surface of the N-lobe by itself by 790 $Å^2$. However, this is more

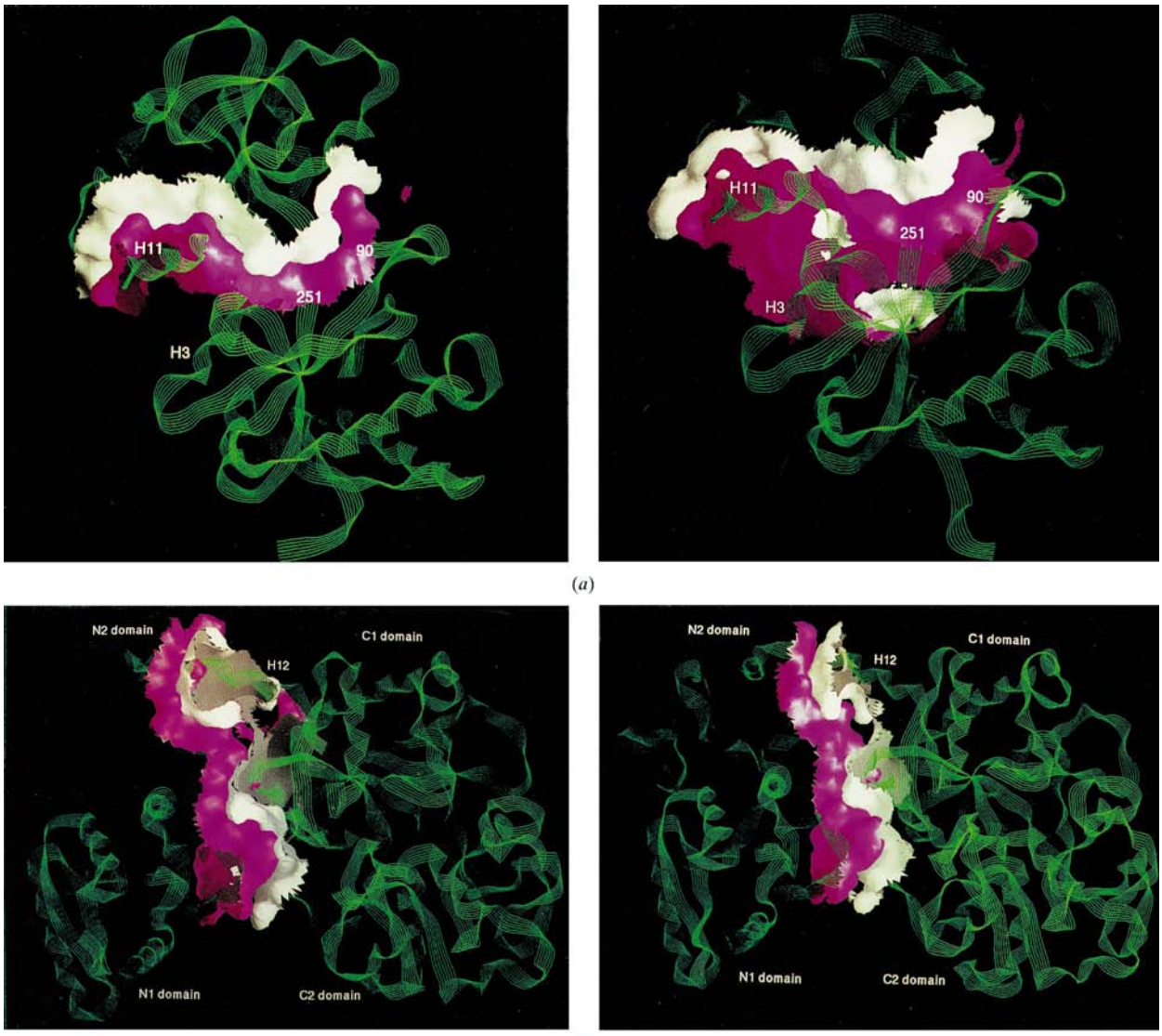


Fig. 8. Changes in the solvent-accessible surface (*a*) for the N1/N2 interface in the lactoferrin N-lobe and (*b*) for the N-lobe/C-lobe interface. In each case the open ApoLf structure is on the left and the closed Fe_2Lf structure is on the right. Surface area buried between N1 and N2 domains is much larger in the closed form (*a*, right panel). On the other hand, the surface area buried between the N- and C-lobes is greater in the open form (*b*, left panel). In (*a*) the N2 domain surface is white and the N1 domain surface is magenta; only the portions of the surface that come into contact (within 4.5 Å) are shown. The much larger contact surface for the closed form (*a*, right) arises because the N2 domain (upper) rotates over the N1 domain (lower) about an axis running left to right across the page. In (*b*) the N-lobe surface is magenta and the C-lobe surface is white. The greater contact area between the lobes in the open form (left) arises because of the movement of the N2 domain (upper, left) up against the connecting helix (H12) and the C-lobe. The rotation axis is approximately about an axis running top to bottom across the page. Figure prepared with *GRASP* (Nicholls *et al.*, 1993).

Table 4. *Solvent-accessible surface comparisons for ApoLf and Fe_2Lf*

The tabulated values are solvent-accessible surface areas (in Å^2) calculated using the program *ASC* (Eisenhaber & Argos, 1993), with a 1.4 Å probe radius. In separating the contributions of individual structural elements there are small overlaps, but these amount at most to ~50 Å^2. Values for ApoLf are tabulated above those for Fe_2Lf. The absolute surface areas for the isolated domains are given on the diagonals and have been corrected for truncated side chains of residues 86, 416 and 637 in Fe_2Lf. The surface areas buried between domains (and the connecting and C-terminal helices) are given as off-diagonal values. The total surface areas are 28428 and 28693 2 Å^2 for the complete ApoLf and Fe_2Lf molecules, respectively.

	N1	N2	C1	C2	Connecting helix	C-terminal helix
N1 (1–90, 250–332)	9327	1373	542	0	132	685
	9273	2161	543	0	178	686
N2 (91–249)		7401	256	0	445	365
		7468	0	0	281	53
C1 (345–433, 595–678)			8787	2522	465	752
			8788	2387	461	732
C2 (434–594)				7513	0	36
				7543	0	2

than offset by the burial of other surfaces when the N2 domain moves. As summarized in Table 4 and illustrated in Fig. 8, new contacts bury a further 260 Å^2 between the N2 and C1 domains: 160 Å^2 between the N2 domain and the connecting helix (334–344) and 310 Å^2 between the N2 domain and the C-terminal helix (680–691). Elsewhere, other changes also occur, as a result of the orientational difference in the C1 and C2 domains, and of changes in the mobile loop 416–422 of the C1 domain, which in ApoLf packs closer to the rest of the C1 domain than in Fe_2Lf.

3.9. *Solvent molecules*

Solvent molecules (all assumed to be water, except for the two Cl^- ions noted earlier) were located in the ApoLf structure without any reference to the Fe_2Lf structure. Comparisons of these water-molecule positions in the two structures are made more difficult by the domain movement in the N-lobe of ApoLf. Therefore we carried out this comparison in three parts, superimposing (i) the N1 domains and their associated solvent, (ii) the N2 domains and their associated solvent and (iii) the entire C-lobe and associated solvent. Solvent molecules between domains or between lobes were included in each comparison.

For the N1 domain 30 water molecules are found in ApoLf within 1.0 Å of corresponding water molecules in Fe2Lf. Of these, five are truly internal and are on average within 0.4 Å of the Fe_2Lf positions. The others are from around the periphery of the domain, in surface crevices, mostly bound to main-chain C═O or NH groups. The opening of the interdomain cleft leaves very few solvent molecules in this region that correspond with those in Fe_2Lf; clearly the closing of the domains in the holo protein traps and stabilizes a number of bound water molecules. The same is true of the N2 domain. All of the seven internal water molecules are retained (mean difference in position 0.41 Å) but in total only 13 water molecules associated with the ApoLf N2 domain correspond within 1.0 Å to bound water molecules in Fe_2Lf.

Not surprisingly, there is much greater correspondence in the closed C-lobe. All of the 19 internal water molecules present in the C-lobe of Fe_2Lf are also found in ApoLf, with a mean positional difference of only 0.44 Å. Likewise, almost all (22 out of 23) of the water molecules within the binding cleft of Fe_2Lf are also present in ApoLf with a mean difference of 0.39 Å. This provides an independent validation of the sites of these bound water molecules. It also demonstrates that the solvent molecules bound within the interdomain cleft are not influenced by whether iron is bound or not, so long as the same domain closure is maintained. Elsewhere around the surface of the C-lobe a further 41 water molecules correspond within 1.0 Å of those in Fe_2Lf. In all, about one third of the water molecules in ApoLf have counterparts in Fe_2Lf within this distance criterion.

3.10. *Crystal-packing contacts*

Crystal-packing contacts between the molecules of ApoLf are relatively few in number (Table 5). In the N-lobe, only nine atoms make contacts less than 3.5 Å with neighbouring molecules (seven involving atoms from the N1 domain and two from the N2 domain). This makes it unlikely that the crystal packing has a significant effect on the extent or nature of the conformational change in the N-lobe. The contacts made by the N-lobe mostly involve rather flexible side chains projecting from the surface of the N1 domain (Arg27, Arg39, Glu51, Arg53), together with a projecting loop 177–180 from the N2 domain.

In the C-lobe the numbers of contacts are greater, but still small; 21 atoms in total are involved in crystal-packing interactions, 18 from the C2 domain and three from C1. It is probably significant that most of the contacts involve the C2 domain, since these seem most likely to account for the closed structure of the C-lobe.

Most involve side chains, however, and none of the contact regions are extensive. All of the contacts made by the C-lobe involve surface-loop regions rather than secondary structural elements; on the C1 domain the very poorly defined loop carrying Asp422 and on the C2 domain two main regions: the adjacent βe–α4 and α7–βh loops (445–448 and 539–541) and a region comprising a series of loops centred on residues 559–564 following helix 8 (see Anderson *et al.*, 1989; Haridas *et al.*, 1995, for full descriptions of the polypeptide-chain conformation).

4. Discussion

Refinement of the apolactoferrin structure has confirmed the essential features of the earlier analyses (Anderson *et al.*, 1990; Gerstein *et al.*, 1993). The large conformational change in the N-lobe occurs as a rigid-body movement of the N2 domain, in which it rotates 54° relative to N1. It involves a very localized hinge, located at Thr90 and Pro251 in two antiparallel β-strands that pass behind the iron site. It is mediated by a 'see-saw' motion between the two interfaces; in the transition from closed to open structure a large interface (the binding cleft) is opened up and a small interface (primarily between helices 5 and 11) is closed (Gerstein *et al.*, 1993).

Refinement has, however, revealed a number of features that were not apparent at lower resolution. Firstly, there are some local conformational adjustments in both N-lobe domains as a result of the domain rotation. The most intriguing of these is the 'flip' of the peptide 302–303 in the N1 domain interface of the binding cleft. This converts the β-turn 301–304 from a less favourable type II conformation in the closed diferric form to a more favourable type I conformation in the open ApoLf. The change seems to correlate with the extent of domain closure in lactoferrin. The turn is type II in Fe_2Lf and in lactoferrin mutants with similar or greater domain closure (Faber, Baker *et al.*, 1996; Faber, Bland *et al.*, 1996) but relaxes to type I in more open structures such as the wild-type recombinant N-lobe half-molecule (Day *et al.*, 1993) and the present ApoLf structure. It is also type I in both serum transferrin (Bailey *et al.*, 1988) and ovotransferrin (Kurokawa *et al.*, 1995). Thus, it is assumed to result from some greater constraints only present in the closed diferric lactoferrin N-lobe.

The other local changes primarily involve side chains whose conformations adjust to accommodate the domain movement; for example, Arg89 adjusts to maintain its salt bridge with Glu211 near the hinge and Arg133 changes its conformation to retain its hydrogen bonding with the C-terminus of helix 11. Both are

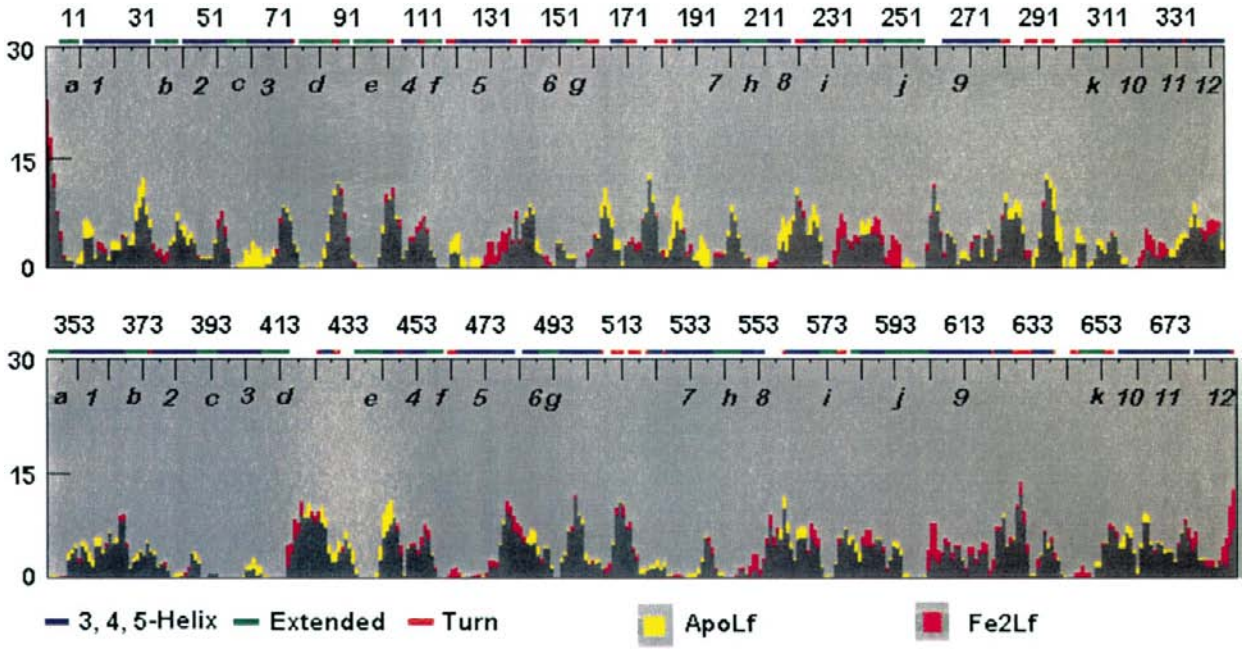


Fig. 9. Plot of solvent accessibility as a function of residue number, comparing ApoLf with Fe_2Lf. Yellow shows where solvent exposure is increased in ApoLf, magenta shows where it is increased in Fe_2Lf and grey shows where it is unchanged. The yellow patches around residues 60, 120, 190, 215 and 300 form the surface of the binding cleft between N1 and N2 domains. The magenta patches around residues 130, 230–250 and 320 represent regions that are exposed in the 'small interface' and between the lobes in Fe_2Lf. The magenta patches in the lower panel relate to decreased contacts between the lobes in Fe_2Lf, relative to ApoLf. Bars under residue numbers identify regions of helix (blue), sheet (green) and turn (red), labelled as in Anderson *et al.* (1989). Figure prepared with *WPDB* (Shindyalov & Bourne, 1995).

Table 5. *Crystal-packing contacts*

Residue	Contact	Distance (Å)†
Arg27 NH2	Asp629 OD1^{i}	2.50
Arg39 NH1	Asp422 OD1^{i}	2.89
Arg39 NH2	Asp422 OD2^{i}	2.98
Glu51 OE2	Gly177 N^{ii}	2.88
Arg53 NE	Asp422 OD2^{i}	3.08
Pro71 O	Glu178 N^{ii}	3.22
Ser293 OG	Arg500 O^{iii}	3.34
Ser293 OG	Asn502 N^{iii}	3.44
Ser293 OG	Asn502 ND2^{iii}	2.78
Asp445 OD2	Ala561 N^{iv}	2.88
Asp445 OD2	Glu560 CA^{iv}	3.17
Thr446 OG1	Arg524 NH1^{iv}	2.99
Thr446 OG1	Arg533 NH2^{iv}	3.48
Ser447 OG	Ala561 N^{iv}	3.17
Ser447 OG	Trp562 N^{iv}	3.01
Gly455 O	Lys564 NZ^{iv}	3.10
Asp541 OD1	Asn559 ND2^{iv}	3.19
Arg580 NE	Glu511 OE2^{iv}	3.01

† Hydrogen bonds and non-bonded contacts less than 3.5 Å.
Symmetry-related positions: (i), $\frac{1}{2} - x, -y, -\frac{1}{2} + z$, (ii), $-x, \frac{1}{2} + y, -\frac{3}{2} - z$, (iii), $-x, -\frac{1}{2} + y, -\frac{1}{2} - z$, (iv), $\frac{1}{2} - x, 1 - y, \frac{1}{2} + z$.

presumably low-energy transformations but contribute to limiting the extent of the domain movement. The *cis–trans* isomerization of the Pro141–Pro142 peptide occurs in a highly flexible loop, apparently because the N2 domain rotation brings this loop up against the N-terminus of the connecting helix.

The most important new information concerns the light this structure throws on cooperativity in transferrins. The two iron sites are ~40 Å apart in the holo form (Anderson *et al.*, 1987) and evidence for cooperativity of binding has been contradictory (Brock, 1985; Harris & Aisen, 1989; Lin *et al.*, 1991). Calorimetric studies, however, show that in ovotransferrin iron binding in one lobe is signalled to the other; binding in the N-lobe reduces binding in the C-lobe, whereas binding the C-lobe enhances binding in the N-lobe (Lin *et al.*, 1991). This is assumed to occur through interactions between the two lobes. Likewise, in lactoferrin, mutagenesis of the C-lobe Tyr residues to knock out iron binding destabilizes binding in the N-lobe (Ward *et al.*, 1996), as does the complete removal of the C-lobe, as in the N-lobe half-molecule of lactoferrin (Day *et al.*, 1992).

The present apolactoferrin structure shows that the interactions between the lobes are indeed altered by the conformational change in the N-lobe. The movement of the N2 domain when the binding cleft opens brings it up against the C-lobe. This has two effects: changing the relative orientations of the two lobes by 8–9° and introducing new contacts between them. These new contacts are significant because they bury a further ~610 Å^2 of surface area between the lobes, increasing the buried surface by ~50%; two specific interactions, between Lys243 and Arg249 on the N2 domain and the end of the C-terminal helix are also added. It is this C-terminal helix (680–691) which is the major participant in the increased interactions. Significantly it occupies a critical position on the C-lobe, being attached to both the C1 and C2 domains through two disulfide bridges. Thus, it is entirely plausible that movement of this helix (as seen in the present structure) could perturb iron binding by altering the ease of domain movement between C1 and C2. Likewise, domain movements in the C-lobe, as a result of iron binding, could move the C-terminal helix and so signal to the N-lobe.

The precise nature of signalling between the lobes appears to vary between different transferrins (Lin *et al.*, 1994). This is understandable in terms of the three-dimensional structures determined to date. In lactoferrin, serum transferrin and ovotransferrin the relative orientations of the N- and C-lobes differ by up to 15° (Kurokawa *et al.*, 1995). This in turn produces some differences in interlobe interactions; in diferric ovotransferrin, for example, residues at the N-lobe hinge, Thr90 and Ser91, together with the nearby Lys308, make direct contact with the C-terminal helix, whereas in lactoferrin there is no direct interaction in the diferric form. The importance for function is that these results indicate a mechanism through which iron-dependent changes in interlobe interactions can produce changes in the overall structure and that these changes can be species specific. This has clear implications for receptor binding,

The closed C-lobe in the present structure is virtually identical to that in the diferric form of lactoferrin. The only conformational changes are in several very flexible external loops, and even the water molecules bound in the interdomain cleft, in other internal sites and in well defined surface sites are retained between the two structures. There is no metal ion or other species bound in place of iron to explain the closed structure. Rather, a stable closed structure can be formed even in the absence of iron binding. This parallels the situation found in the related family of periplasmic binding proteins (Quiocho, 1990), where closed but ligand-free structures have been found for both the arabinose and glucose/galactose binding proteins (Quiocho, 1991; Flocco & Mowbray, 1994). The implication is that for the ligand-free proteins there is a dynamic equilibrium between open and closed forms in solution, with little energy difference between them. Although lattice interactions (weak and few in number) help stabilize the closed structure in the crystals, these closed structures must be capable of existing, even if only transiently, in solution.

We anticipate that an apo structure with both lobes open is the dominant species in the solution. This would be consistent with solution-scattering studies (Grossmann *et al.*, 1992) and our preliminary X-ray studies on the second crystal form of apolactoferrin in which both lobes are open (Baker *et al.*, 1997). In solution, this 'all-open' structure must, however, be able to sample closed

structures from time to time. Iron binding is then to the open form, initially to the two Tyr ligands and the carbonate ion which cluster closely together on the N2 (or C2) domain. This would be as seen in the N2 domain fragment of ovotransferrin (Lindley *et al.*, 1993). When the closed structure is sampled, through the natural dynamics of the molecule, the iron acts as a lock which fastens the two domains together when it binds to the other two ligands. We should emphasize, however, that the precise nature of the open form(s) of apolactoferrin in solution may vary under different solution conditions and attempts to model radii of gyration at low pH, using various domain orientations, have not been fully successful (Mecklenburg *et al.*, 1997).

The concept of balance between open and closed forms is strengthened by the calculations of solvent-accessible surface in ApoLf and Fe_2Lf. The fact that the large increase in surface area of the N-lobe when its binding cleft opens is offset by an almost equal burial of other surfaces has profound implications. First, the entropy difference between the closed Fe_2Lf structure and the 'one open, one closed' ApoLf structure must be small, as their surface areas are virtually identical. Second, the numbers of hydrogen bonds also remain very similar, as few interdomain hydrogen bonds are lost on domain opening. Thus, the enthalpy change should also be similar, implying only a small free-energy difference between the two structures. This balance does not fully carry over to the single-lobe proteins, however, because in these there is no second lobe to offset the increase in solvent-accessible surface. We would expect, therefore, that the molecular mechanism controlling the relative free energies of open and closed half molecules, and their kinetics of iron binding and release need not be the same as for the whole molecule.

Finally, we note that chloride ions are bound in the anion-binding sites of both lobes, just as carbonate is in the holo protein. Chloride was a component of the crystallization medium, explaining its presence. More importantly, this demonstrates the strong affinity of lactoferrin (and other transferrins) for anions. It again emphasizes the link with the bacterial periplasmic binding proteins, which have essentially the same fold as each lobe of lactoferrin and the same mechanism of binding, and amongst which are sulfate- and phosphate-binding proteins that possess anion-binding sites very closely analogous to the anion-binding site in each lobe of lactoferrin (Baker *et al.*, 1987).

We gratefully acknowledge support from the US National Institutes of Health (grant HD-20859), the Wellcome Trust and the Health Research Council of New Zealand. ENB also gratefully acknowledges research support from the Howard Hughes Medical Institute through an International Research Scholarship. Thanks are also due to colleagues at Massey University for their encouragement and help, and to Drs P. F. Lindley and F. A. Quiocho for generous sharing of unpublished data.

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.

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. (1994). *Adv. Inorg. Chem.* **41**, 389–463.

Baker, E. N., Anderson, B. F., Baker, H. M., Faber, H. R., Smith, C. A. & Sutherland-Smith, A. J. (1997). *Lactoferrin: Interactions and Biological Functions*, edited by T. W. Hutchens & B. Lonnerdal, pp. 177–191. New Jersey: Humana Press.

Baker, E. N., Anderson, B. F. & Rumball, S. V. (1987). *Trends Biochem. Sci.* **12**, 350–353.

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

Bali, P. K. & Aisen, P. (1991). *Biochemistry*, **30**, 9947–9953.

Birgens, H. S., Hansen, N. E., Karle, H. & Ostergard-Kristensen, L. (1983). *Br. J. Haematol.* **54**, 383–391.

Brock, J. H. (1985). *Metalloproteins*, Vol. 2, edited by P. M. Harrison, pp. 183–262. London: Macmillan.

Day, C. L., Anderson, B. F., Tweedie, J. W. & Baker, E. N. (1993). *J. Mol. Biol.* **232**, 1084–1100.

Day, C. L., Stowell, K. M., Baker, E. N. & Tweedie, J. W. (1992). *J. Biol. Chem.* **267**, 13857–13862.

Eisenhaber, F. & Argos, P. (1993). *J. Comput. Chem.* **14**, 1272–1280.

Elass-Rochard, E., Roseanu, A., Legrand, D., Trif, M., Salmon, V., Motas, C., Montreuil, J. & Spik, G. (1995). *Biochem. J.* **312**, 839–845.

Engh, R. H. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Faber, H. R., Baker, C. J., Day, C. L., Tweedie, J. W. & Baker, E. N. (1996). *Biochemistry*, **35**, 14473–14479.

Faber, H. R., Bland, T., Day, C. L., Norris, G. E., Tweedie, J. W. & Baker, E. N. (1996). *J. Mol. Biol.* **256**, 352–363.

Flocco, M. M. & Mowbray, S. L. (1994). *J. Biol. Chem.* **269**, 8931–8936.

Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M. & Chothia, C. (1993). *J. Mol. Biol.* **234**, 357–372.

Grossmann, J. G., Neu, M., Pantos, E., Schwab, F. J., Evans, R. W., Townes-Andrews, E., Lindley, P. F., Appel, H., Thies, W.-G. & Hasnain, S. S. (1992). *J. Mol. Biol.* **255**, 811–819.

Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–646.

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

Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.

Hutchens, T. W., Maguson, Y. S. & Yip, T.-T. (1989). *Pediatr. Res.* **26**, 618–622.

Kurokawa, H., Mikami, B. & Hirose, M. (1995). *J. Mol. Biol.* **254**, 196–207.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lin, L.-N., Mason, A. B., Woodworth, R. C. & Brandts, J. F. (1991). *Biochemistry*, **30**, 11660–11669.
Lin, L.-N., Mason, A. B., Woodworth, R. C. & Brandts, J. F. (1994). *Biochemistry*, **33**, 1881–1888.
Lindley, P. F., Bajaj, M., Evans, R. W., Garrett, R. C., Hasnain, S. S., Jhoti, H., Kuser P., , Neu, M., Patel, K., Sarra, R., Strange, R. & Walton, A. (1993). *Acta Cryst.* D**49**, 292–304.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Mann, D. M., Romm, E. & Migliorini, M. (1994). *J. Biol. Chem.* **269**, 23661–23667.
Matthews, B. W. (1972). *Macromolecules*, **5**, 818–819.
Mazurier, J. & Spik, G. (1980). *Biochim. Biophys. Acta*, **629**, 399–408.
Mecklenburg, S. L., Donohue, R. J. & Olah, G. A. (1997). *J. Mol. Biol.* **270,** 739–750.
Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. & Jolles, P. (1984). *Eur. J. Biochem.* **145**, 659–676.
Nemethy, G. & Printz, M. P. (1972). *Macromolecules*, **5**, 755–758.
Nicholls, A., Bharadwaj, R. & Honig, B. (1993). *Biophys. J.* **64**, 166.
Norris, G. E., Anderson, B. F. & Baker, E. N. (1991). *Acta Cryst.* B**47**, 998–1004.
Norris, G. E., Baker, H. M. & Baker, E. N. (1989). *J. Mol. Biol.* **209**, 329–331.
Powell, M. J. & Ogden, J. E. (1990). *Nucleic Acids Res.* **18**, 4013.
Presta, L. G. & Rose, G. D. (1988). *Science*, **240**, 1632–1641.
Quiocho, F. A. (1990). *Philos. Trans. R. Soc. London Ser. B*, **326**, 341–351.
Quiocho, F. A. (1991). *Curr. Opin. Struct. Biol.* **1**, 922–933.
Ramakrishnan, G. & Ramachandran, G. A. (1965). *Biophys. J.* **5**, 909–933.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Richardson, J. S. & Richardson, D. C. (1988). *Science*, **240**, 1648–1652.
Sanchez, L., Calvo, M. & Brock, J. H. (1992). *Arch. Dis. Child.* **67**, 657–661.
Shindyalov, I. N. & Bourne, P. E. (1995). *J. Appl. Cryst.* **28**, 847–852.
Stowell, K. M., Rado, T. A., Funk, W. D. & Tweedie, J. W. (1991). *Biochem. J.* **276**, 349–355.
Ward, P. P., Lo, J.-Y., Duke, M., May, G. S., Headon, D. R. & Conneely, O. M. (1992). *Biotechnology*, **10**, 784–789.
Ward, P. P., Zhou, X. & Conneely, O. M. (1996). *J. Biol. Chem.* **271**, 12790–12794.
Wilmot, C. M. & Thornton, J. M. (1988). *J. Mol. Biol.* **203**, 221–232.
Wonacott, A. J. (1980). *MOSFLM. A Suite of Programs for the On-Line Evaluation and Analysis of Integrated Intensities on Small Angle Rotation/Oscillation Photographs*. Cambridge, England.