Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a domain-opened mutant (R121D) of the human lactoferrin N-lobe refined from a merohedrally twinned crystal form

Geoffrey B. Jameson,[a] Bryan F. Anderson,[a] Wendy A. Breyer,[a,b] Catherine L. Day,[a]† John W. Tweedie[a] and Edward N. Baker[a,c]*

[a]Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand, [b]Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA, and [c]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

† Current address: Department of Biochemistry, Otago University, PO Box 56, Dunedin, New Zealand.

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 22 January 2002
Accepted 19 March 2002

PDB Reference: lactoferrin, 1l5t, r1l5tsf.

Human lactoferrin is an iron-binding protein with a bilobal structure. Each lobe contains a high-affinity binding site for a single Fe^{3+} ion and an associated CO_3^{2-} ion. Although iron binds very tightly, it can be released at low pH, with an accompanying conformational change in which the two domains move apart. The Arg121Asp (R121D) mutant of the N-lobe half-molecule of human lactoferrin was constructed in order to test whether the Asp121 side chain could substitute for the CO_3^{2-} ion at the iron-binding site. The R121D mutant protein was crystallized in its apo form as it lost iron during crystallization. The crystals were also merohedrally twinned, with a twin fraction close to 0.5. Starting from the initial molecular-replacement solution [Breyer *et al.* (1999), *Acta Cryst.* D**55**, 129–138], the structure has been refined at 3.0 Å resolution to an R factor of 13.9% ($R_{\rm free}$ of 19.9%). Despite the moderate resolution, the high solvent content and non-crystallographic symmetry contributed to electron-density maps of excellent quality. Weakened iron binding by the R121D mutant is explained by occlusion of the anion-binding site by the Asp side chain. The opening of the two domains in the apoR121D structure (a rotation of 54°) closely matches that of the N-lobe in full-length lactoferrin, showing that the extent of the conformational change depends on properties inherent to the N-lobe. Differences in the C-terminal portion of the N-lobe (residues 321–332) for apoR121D relative to the closed wild-type iron-bound structure point to the importance of this region in stabilizing the open form.

1. Introduction

Proteins of the transferrin family play an essential role in iron homeostasis in vertebrates, controlling the levels of free iron in body fluids (Brock, 1985; Aisen & Harris, 1989; Baker, 1994). In mammals, two principal members of this family are involved, serum transferrin (Tf) and lactoferrin (Lf). Serum transferrin has the vital role of transporting iron, as Fe^{3+}, in the bloodstream and delivering it to cells *via* a process of receptor-mediated endocytosis (Klausner *et al.*, 1983; Aisen, 1998). Lactoferrin is not known to have an iron-transport role, but binds iron more strongly than Tf (Aisen & Liebman, 1972) and retains it to lower pH (Mazurier & Spik, 1980), probably for the inhibition of bacterial growth (Bullen *et al.*, 1972) and to limit the iron-catalysed production of damaging free radicals (Baldwin *et al.*, 1984).

Both proteins share very similar structural as well as functional properties (reviewed in Baker, 1994), consistent with

their amino-acid sequence identity of about 60%. In each case, the polypeptide chain is folded into two homologous lobes, representing its N- and C-terminal halves, and each lobe has an iron-binding site located in a deep cleft between two domains. The groups that form the iron-binding sites are also the same, *i.e.* two tyrosine side chains, one aspartic acid and one histidine, together with a carbonate ion that binds to an arginine residue and the N-terminus of an α-helix (Anderson *et al.*, 1989; Baker, 1994). The requirement for an anion, bound synergistically with the Fe^{3+} ion, is a unique feature of transferrins. Iron release can occur at low pH (below pH 6 for Tf and pH 3.5 for Lf) and is associated with a large-scale conformational change in which the two domains that enclose each iron-binding site move wide apart. In the N-lobe of human lactoferrin, this movement involves a rigid-body rotation of 54° of one domain relative to the other (Anderson *et al.*, 1990; Jameson *et al.*, 1998) and similar conformational changes have been defined crystallographically in both the N- and C-lobes of hen ovotransferrin (Kurokawa *et al.*, 1999).

For both lactoferrin and transferrin, their N-lobe half-molecules, Lf_N and Tf_N respectively, can be readily expressed as recombinant proteins (Funk *et al.*, 1990; Day *et al.*, 1992) and offer excellent model systems for testing the roles of particular amino acids by mutagenesis. Having only one iron-binding site, they are free of the complexities of a two-sited protein in which the properties of the two sites overlap. Thus, studies of these half-molecules have probed the effects of mutating the iron ligands (*e.g.* Faber, Bland *et al.*, 1996; Nicholson *et al.*, 1997; MacGillivray *et al.*, 2000) and certain second-shell residues around the iron-binding site (*e.g.* Peterson *et al.*, 2000; Nurizzo *et al.*, 2001).

The carbonate-binding site is an intriguing feature of the structure and function of transferrins. The anion is hydrogen bonded to the N-terminus of an α-helix and to an arginine residue that curls around it to create an anion-binding pocket (Baker, 1994). In the N-lobe of human Lf the arginine residue is Arg121, which sits at the N-terminus of the α-helix. This raises the question as to whether, if an acidic amino acid were substituted for Arg121, the anion-binding requirement might become unnecessary. Acidic side chains frequently 'cap' the N-termini of α-helices, hydrogen bonding to the free NH groups (Baker & Hubbard, 1984), and model building suggested that if Asp were substituted for Arg121 its carboxyl group might simultaneously coordinate the Fe atom. This could create a transferrin that did not release iron at low pH, if protonation of the carbonate is an important factor in iron dissociation as suggested by MacGillivray *et al.* (1998).

Here, we present the structure of the R121D mutant of the human Lf N-lobe half-molecule, refined using data from merohedrally twinned crystals. Although the mutant was set to crystallize in its iron-bound form, it lost iron during crystallization and the structure determined is accordingly that of the apo (iron-free) form. The structure reveals the extent of the conformational change that occurs when iron is released, in comparison with iron-loaded forms of the N-lobe of Lf, and also points to the importance of an α-helix at the C-terminus of the N-lobe in stabilizing the open form.

2. Materials and methods

2.1. Expression, purification and crystallization

The R121D mutation was introduced into the gene for the N-terminal half-molecule of human lactoferrin (residues 1–334) following the methods described for other Lf_N mutants (Faber, Baker *et al.*, 1996; Faber, Bland *et al.*, 1996). The mutated DNA was introduced into the expression vector pNUT (Palmiter *et al.*, 1987), after which the pNUT mutant plasmid was transfected into baby hamster kidney (BHK) cells. Selection of the transformed cells was with 0.5 m*M* methotrexate and expression of the mutant protein was verified by immunoprecipitation and gel electrophoresis. For large-scale production of the protein, the cells were transferred to roller bottles. The mutant protein was secreted into the growth medium from which it was harvested and purified by ion-exchange chromatography as for wild-type Lf_N and other Lf_N mutants (Day *et al.*, 1992; Faber, Baker *et al.*, 1996; Faber, Bland *et al.*, 1996). The final yield of the R121D mutant protein was ~15 mg per litre of growth medium. The presence of iron in the growth medium meant that the protein was obtained with iron already bound. To ensure complete saturation, however, the purified protein was dialysed against 0.05 *M* Tris–HCl pH 8.0 containing 0.1 *M* NaCl and was then titrated with 0.01 *M* ferric nitrilotriacetate solution until no further increase in the absorbance at 450 nm was observed.

For crystallization, a concentrated solution of the protein (50–80 mg ml^{-1}) was dialysed against a solution comprising 0.01 *M* Tris–HCl buffer pH 8.0 containing 12%(*v*/*v*) 2-propanol using 20 µl microdialysis cells (Cambridge Repetition Engineers). Crystals grew over a long period (3–6 months) from an oil phase that separated from the solution. Although the protein was initially in its iron-saturated form, the needle-shaped crystals that separated were colourless, characteristic of the iron-free protein.

2.2. Data collection and crystal characterization

X-ray diffraction data were collected at room temperature on an R-AXIS IIC image-plate detector using Cu $K\alpha$ radiation from a Rigaku RU-200 rotating-anode generator, as described previously (Breyer *et al.*, 1999). The data, extending to approximately 3.0 Å resolution, were processed with *DENZO* (Otwinowski & Minor, 1997) and scaled and merged using programs from the *CCP*4 Suite (Collaborative Computational Project, Number 4, 1994). Final data-collection statistics are given in Table 1. The diffraction pattern exhibited *P*6/*m* symmetry, but careful analysis of the diffraction data showed that the crystals were actually trigonal, space group $P3_1$, and had near-perfect hemihedral twinning (Breyer *et al.*, 1999). The unit-cell parameters were $a = b = 151.3$, $c = 48.6$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. Subsequent molecular-replacement experiments (Breyer *et al.*, 1999) showed that there were two molecules in the asymmetric unit and that the crystals had a very high solvent content of 72% (Matthews coefficient of 4.35 $Å^3\ Da^{-1}$).

Table 1
Data-collection statistics for the crystals of apoR121D.

Unit-cell parameters (Å)	$a = b = 151.3$, $c = 48.6$
Space group	$P3_1$
Wavelength (Å)	1.5418
Resolution range (Å)	30–3.0 (3.11–3.0)
No. of unique reflections	23882
Redundancy	2.8 (2.4)
Completeness (%)	96 (93)
R_{merge} (%)	11.5 (41.8)
$\langle I \rangle / (I)$	6.1 (1.8)

2.3. Structure determination and refinement

The structure was solved by molecular replacement using the two domains (partly truncated) of Lf_N (Day *et al.*, 1993) as search models. The two molecules (four domains) found were assigned to their appropriate twin domains, as described previously (Breyer *et al.*, 1999), to give the starting model for refinement. The model was refined with *SHELXL*97 (Sheldrick, 1997), which uses twinned data for refinement, but detwins the data for calculation of electron-density maps. Refinement, based on F^2, began with values for R and R_{free} of 0.34. Reflections for the test set for R_{free} (Brünger, 1992) were chosen as described in the first footnote to Table 2. The C-terminal helix, which comprises residues 321–332 and was omitted from the molecular-replacement search model, was apparent in electron-density maps ($2F_o - F_c$ and $F_o - F_c$) following the first round of refinement. The aspartate side chain of the mutated residue 121 was also apparent in electron-density maps. In subsequent rebuildings with *TURBO FRODO* (Roussel & Cambillau, 1989), several peptides were flipped and conserved water molecules were added to the model. Tight non-crystallographic symmetry (NCS) restraints were applied to the two molecules; as implemented in *SHELXL*97 (Sheldrick, 1997), equivalent torsion angles (absolute values) and atomic displacement parameters were restrained to similarity. A summary of the target root-mean-square (r.m.s.) deviations for NCS restraints is given in Table 2.

2.4. Other methods

Structure comparisons were performed using *LSQKAB* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994) and hydrogen bonds were identified using the geometrical criteria of Baker & Hubbard (1984). Surface areas and molecular volumes were calculated using *GRASP* (Nicholls *et al.*, 1991), with the side chains of holo $FeLf_N$ (Day *et al.*, 1993) and apoR121D being trimmed, as appropriate, to be identical for both molecules; five side chains that were not defined in electron-density maps for holo $FeLf_N$ were located for R121D. In both molecules, residues 1–4 are disordered and were not included in the calculations. In R121D, residues 322–332 form an α-helix, whereas in holo $FeLf_N$ residues 322–326 form a β-strand and the following residues 328–332 are disordered or highly mobile. In order to compare these C-terminal regions of the two proteins, the volume and surface areas contributed by residues 328–332 in $FeLf_N$ were therefore modelled either as an extended strand or as an α-helix, in each case extending away from the protein in the direction indicated by residue 327; these two representations represent the extremes of molecular compactness. See §3 for further details.

Table 2
Refinement statistics for apoR121D.

Resolution (Å)	20–3.0
R_{cryst} (No. of reflections)	0.139 (22636)
R_{free}† (No. of reflections)	0.199 (795)
Twin fraction	0.445
No. of parameters	19883
No. of restraints	27798
R.m.s.d. bond lengths‡ (target) (Å)	0.005 (0.010)
R.m.s.d. bond-angle distances§ (target) (Å)	0.017 (0.040)
No. of protein atoms	5112
No. of solvent atoms	130
Overall B value (Å^2)	31.1
Mean B value (Å^2)	
Main chain	29.0
Side chain	32.7
Solvent	46.9
Ramachandran plot¶ (%)	
Most favoured regions	82.1
Allowed regions	16.9
Generously allowed regions	0.5
Disallowed regions	0.5

† The data set aside for calculation of R_{free} included, for each randomly selected reflection, its twin-related mate and all symmetry equivalents. ‡ The target r.m.s. deviation from ideal values (Engh & Huber, 1991) corresponds to an effective weighting of that restraint in least-squares calculations. § R.m.s. deviation for bond angles expressed in terms of 1–3 separation. For a standard tetrahedral C1—C2—C3 (bond length 1.54 Å, angle 109.5°), these values correspond to angle of 0.7° (2.6°). ¶ As defined by *PROCHECK* (Laskowski *et al.*, 1993).

3. Results and discussion

3.1. Refinement in the presence of merohedral twinning

Provided that the data are correctly processed and all data to the final cutoff resolution are retained, twinning, especially severe twinning where the two twin domains are nearly equal in volume, can be readily detected by means of the statistic $T = \langle |F| \rangle^2 / \langle |F^2| \rangle$ (Breyer *et al.*, 1999). Evaluated over equal-sized resolution bins, this statistic, in the case of non-centrosymmetric space groups (as is invariably the case for protein structures), has a value of 0.786 for data from a non-twinned crystal and 0.886 for data from a twinned crystal with equal twin fractions of 0.5. For the R121D mutant of half-molecule human lactoferin, the average value of T was 0.866 with an e.s.d. of 0.015 for data divided into ten equal-sized resolution bins (Breyer *et al.*, 1999). This statistic, based on Wilson intensity statistics (Wilson, 1949) and the assumption of randomly distributed atoms, fails when pseudo-symmetry exists, such as pseudo-*I*, pseudo-*C* or pseudo-*F* centring, or when one molecule is related to another by, say, $x + \sim\frac{1}{2}$. In this latter case, the statistic can, however, be usefully applied separately to the strong data (F_{hkl}, where $h = 2n$) and to the weak data (F_{hkl}, where $h = 2n + 1$). No twin law need be specified for this statistic. Alternatively, the data can be analysed assuming various twin laws by the method of Britton

(1972), as implemented by Yeates (1988). Both tests also work well for pseudomerohedral twinning, where the twin law is a symmetry element of a Laue class of higher symmetry and the unit cell is metrically indistinguishable from that of a higher symmetry Laue class. For example, orthorhombic cells with $a = b$ (within experimental error) are predisposed to a twinning mode in which crystal domains are related either by a fourfold rotation parallel to c, or by a twofold rotation about (110), to give, in the limiting case of equal-sized domains, X-ray diffraction data belonging to the tetragonal Laue class, $4/m$ or $4/mmm$, rather than the orthorhombic mmm class.

Although merohedral twinning complicated the structure solution (Breyer *et al.*, 1999), subsequent development and refinement of the structure proceeded smoothly. With $2 \times 332 = 664$ residues, this R121D half-molecule mutant is one of the larger molecules to be elaborated and refined using data from a merohedral twin. Despite the limited resolution of the data, electron-density maps were very clear, probably because of twofold NCS; the water content of the crystals is in excess of 70%, which leads to denser sampling of reciprocal space than in crystals with more usual Matthews coefficients. Several peptides were flipped on the basis of the pattern of positive and negative peaks in difference Fourier maps – the new conformations have been subsequently confirmed in a re-refinement of the parent Lf_N structure (R. D. Kidd and E. N. Baker, personal communication).

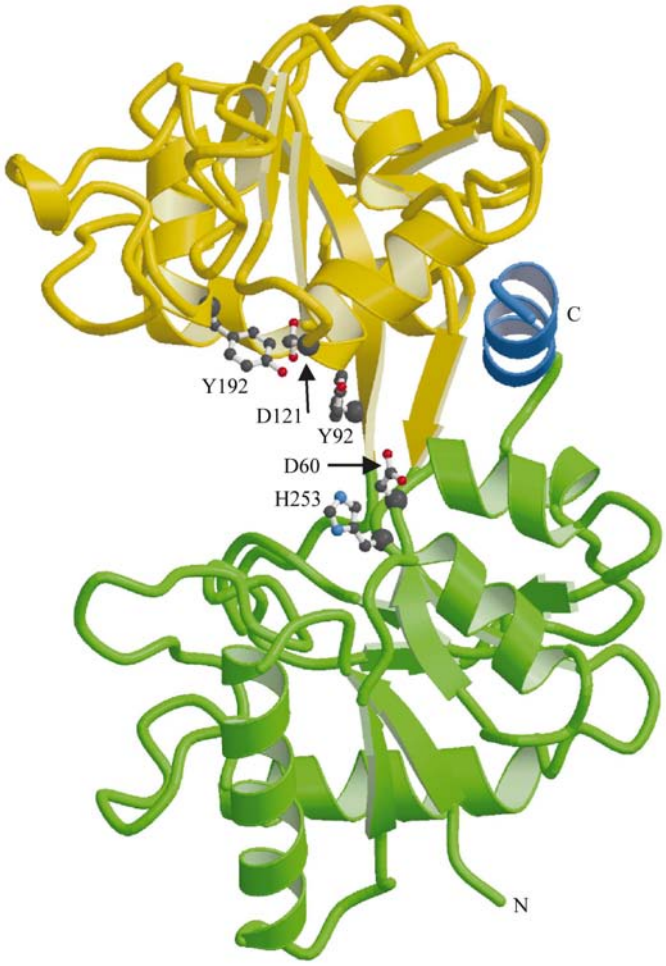

Figure 1
Ribbon diagram of the apoR121D molecule. The N1 domain is shown in green and the N2 domain in gold, with the C-terminal helix 11, residues 321–332, shown in blue. This figure and others were drawn with *MOLSCRIPT* (Kraulis, 1991) and rendered with *Raster3D* (Meritt & Bacon, 1997).

3.2. Molecular structure

The apoR121D protein has an overall molecular structure that is characteristic of the iron-free (apo) form of transferrins (Anderson *et al.*, 1990; Jeffrey *et al.*, 1998; Kurokawa *et al.*, 1999; Mizutani *et al.*, 2000) and indeed of the unliganded forms of the bacterial periplasmic binding proteins, which share very similar folds and functions (Quiocho & Ledvina, 1996). The two α/β domains of apoR121D are widely separated (Fig. 1). The extent of the domain opening in apoR121D compared with the closed iron-bound form of the lactoferrin N-lobe was estimated by first superimposing the N1 domain of intact diferric human Lf (Fe_2Lf; Anderson *et al.*, 1989; Haridas *et al.*, 1995) onto the N1 domain of apoR121D and then determining the rotation required to bring the two N2 domains into coincidence. From this analysis, a domain opening of 53.9° has occurred in apoR121D.

The structure within each domain remains essentially unchanged by the domain movement, emphasizing that domain opening in transferrins is a rigid-body motion (Anderson *et al.*, 1990; Gerstein *et al.*, 1993). When the N1 domains of apoR121D and Fe_2Lf (residues 5–91 plus 252–332, omitting only the disordered N-terminal residues 1–4) are superimposed, 168 C^α atoms can be matched with a root-mean-square (r.m.s.) difference in atomic positions of 0.54 Å. The only significant differences are a peptide flip in the β-turn 301–304, which also occurs in apoLf (Jameson *et al.*, 1998), and a slight shift of the C-terminal helix 321–332 (see later discussion). Superposition of the two N2 domains of apoR121D and Fe_2Lf (residues 92–251) likewise shows very close correspondence; all 160 C^α atoms can be matched with an r.m.s. difference in atomic positions of 0.46 Å.

The substitution of Asp121 for Arg121 in the R121D mutant does not disturb the polypeptide fold. The conformation adopted by the side chain of Asp121 does, however, explain the greatly weakened iron binding of the R121D mutant. Preliminary iron-release studies (Day, 1993) showed that as the pH is lowered this mutant begins to lose iron at a pH value between 7.0 and 7.5, much higher than the wild-type N-lobe half-molecule (~5.0). This weakened iron binding is consistent with the fact that R121D lost all bound iron during crystallization at pH 8.0. We assume that at this pH some of the iron-free form must be present in solution and that this crystallized preferentially over the 3–6 months required for crystallization. In the crystal structure of apoR121D, the carboxylate group of Asp121 is associated with the N-terminus of helix 5 (residues 121–136), with Asp121 $O^{\delta 2}$ being hydrogen bonded (2.6 Å) to the peptide NH of Gly124 as well as to Thr117 $O^{\gamma 1}$ (2.6 Å). In iron-bound transferrins, the helix 5 N-terminus, plus the associated residues Thr117 and Arg121

(Lf N-lobe numbering), bind the CO_3^{2-} ion that is essential for high-affinity iron binding (Baker, 1994) and Thr117 $O^{\gamma 1}$ and Gly124 NH are among the groups that specifically hydrogen bond to it. Likewise, in apo-transferrins this site has a high affinity for other anions, such as Cl^- and SO_4^{2-}, as shown by the crystal structures of apoLf (Jameson *et al.*, 1998) and the apo-oTf N-lobe (Mizutani *et al.*, 2000).

We conclude that the lowered iron affinity of the R121D mutant arises because the side chain of Asp121 competes for the anion-binding site at the N-terminus of helix 5 and because its negative charge further inhibits carbonate binding. Crystallographic and binding studies of other mutants of Arg121 have shown that iron binding is weakened by destabilization of the anion-binding site (Faber, Baker *et al.*, 1996). Superposition of apoR121D onto Fe_2Lf (Fig. 2) further shows that the carboxylate group of Asp121 is unlikely to be able to substitute functionally for the CO_3^{2-} ion, because although it binds to the same anion-binding site it overlaps poorly with the carbonate position and would not be able to generate the same iron-binding structure. Therefore, we conclude that the Asp121 side chain has to be displaced from the helix 5 N-terminus by CO_3^{2-} before effective iron binding can occur.

3.3. Domain opening in comparison with other transferrins

The domain opening in apoR121D compares remarkably closely with that found to occur in the N-lobe of intact human apolactoferrin (Anderson *et al.*, 1990; Jameson *et al.*, 1998); the domain movement is 53.9° in apoR121D, compared with 54.3° in apoLf. The correspondence is not precise, however. When the open N-lobes of apoR121D and apoLf are superimposed on the basis of their N2 domains, as in Fig. 3, it is apparent that although the extent of domain opening is very similar, there is an additional twist of about 8° in apoR121D compared with apoLf. Thus, when the complete N-lobes of apoR121D and apoLf (residues 6–332) are superimposed, the r.m.s. difference for all 327 C^{α} atom positions is 1.1 Å, whereas the individual domains match with r.m.s. differences of 0.44 Å (N1) and 0.50 Å (N2). Whether this difference results, in apoR121D, from the loss of some small restraining influence of the C-lobe or from altered crystal packing is not clear, but whatever the case it is remarkable that the overall domain movement is so similar in the half-molecule to that in the whole molecule.

The similarity also extends to the N-lobes of other transferrins. In the N-lobe of intact chicken apo-ovotransferrin the domain opening is 53° compared with the closed holo form (Kurokawa *et al.*, 1999) and in the N-lobe half-molecule of chicken apo-ovotransferrin it is just slightly smaller at 49.7° (Mizutani *et al.*, 2000). These differences are small enough to be attributed to slight differences in crystal packing. Likewise, for the N-lobe half-molecule of human transferrin the domain opening for the apo form compared with the holo form is in the range 63–65° for the four independent molecules in the asymmetric unit of the crystal (Jeffrey *et al.*, 1998). The somewhat larger opening for human transferrin arises because it has a slightly more open apo form and a slighly more closed holo form compared with lactoferrin (Jeffrey *et al.*, 1998).

3.4. Importance of the C-terminal helix 321–332

Why are the open forms of the N-lobes of lactoferrin, ovotransferrin and transferrin so remarkably similar? This cannot arise from constraints originating from the C-lobes, since the extent of opening is essentially the same in both full molecules and half-molecules. The domain orientations in the closed holo form are clearly fixed by the Fe^{3+} ion that is covalently bound between them and coordinated to protein ligands that are contributed by both domains, as well as by other hydrogen-bonding and van der Waals interactions. In contrast, in the open forms the hinge simply involves small torsion-angle changes, within easily allowable limits, in the two β-strands that run behind the iron site and there are no restraining interdomain interactions.

The structure of apoR121D, coupled with analyses of the nature of the conformational change (Anderson *et al.*, 1990; Gerstein *et al.*, 1993), offers an explanation for this strong structural conservation of the open forms. The C-terminal helix (helix 11, residues 321–332) plays a key role. In full-length human Lf, this helix remains with the N1 domain during the conformational change and acts as a pivot on which helix 5 (residues 121–136) from the N2 domain

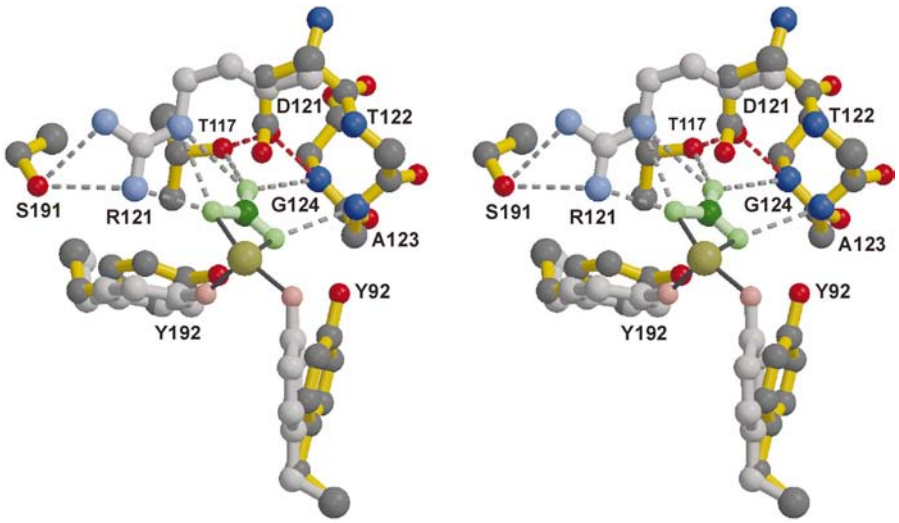

Figure 2
Stereoview of the mutation site of apoR121D (full colour) compared with $FeLf_N$ (lighter colours). The hydrogen bonds from the N-terminus of helix 5 and Thr117 $O^{\gamma 1}$ to the Asp121 carboxylate group are shown as red dashed lines. The hydrogen bonds involving the carbonate group (green) in the wild-type protein from Arg121, Thr117 and the N-terminus of helix 5 are shown as dark grey dashed lines. For clarity, the side chains of Thr122 and Ala123 are omitted. The superimposed structures show that the Asp121 side chain is poorly positioned to bind to the iron centre directly and interferes with CO_3^{2-} binding and, by association, iron binding, thus explaining the greatly weakened iron binding in the R121D mutant.

rotates (Anderson *et al.*, 1990). Furthermore, the open form is stabilized by a 'small interface', primarily between residues from helices 5 and 11, that is enhanced in the open form relative to the closed form (Gerstein *et al.*, 1993).

Strikingly, crystallographic analyses of the closed holo forms of both wild-type and mutant human Lf half-molecules have shown that the C-terminal region beyond residue 320 is almost invariably disordered. In wild-type Lf_N (Day *et al.*, 1993) the helix 321–332 is unwound such that residues 322–326 form a β-strand and the chain is disordered beyond Thr327. In the Lf_N mutants D63S (Faber, Bland *et al.*, 1996), R121E and R121S (Faber, Baker *et al.*, 1996), H253M (Nicholson *et al.*, 1997) and R210K (Peterson *et al.*, 2000), the residues beyond Gly321 are either completely or largely disordered. In contrast, in the open apoR121D structure, residues 321–332 form a well ordered α-helix just as in intact human lactoferrin; the *B* factors for the main-chain atoms of these residues average 26.8 Å^2. Significantly, the interactions between helix 11 and helix 5 in apoR121D are exactly as in apoLf and when these structures are superimposed on the basis of their N2

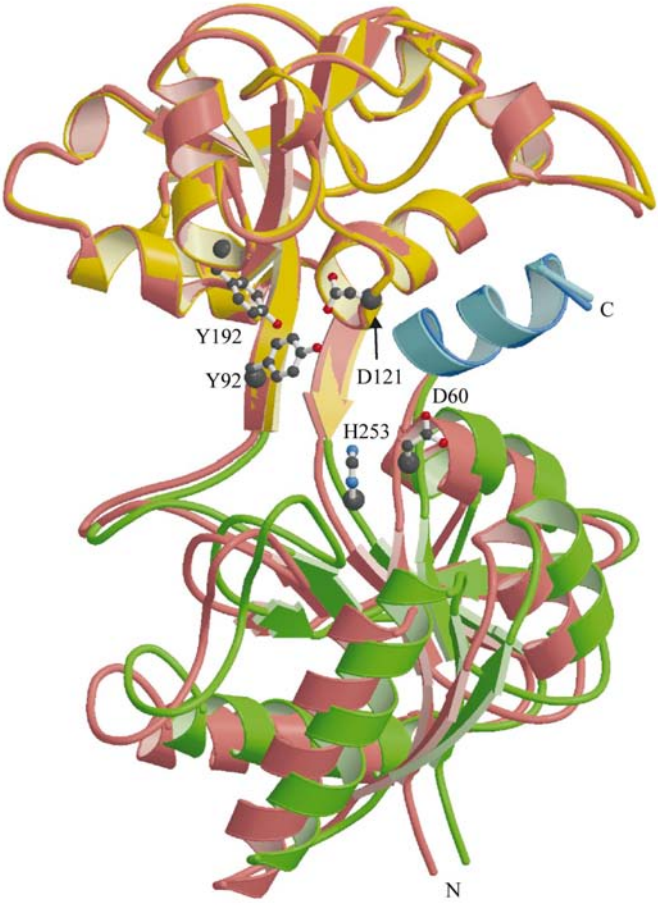

Figure 3
Overlay of the apoR121D and the apoLf N-lobe structures after superposition on the basis of their N2 domains. The apoR121D structure is in green (N1 domain) and gold (N2 domain); the apoLf N-lobe structure is in pink. Selected active-site residues of the R121D mutant are shown as ball-and-stick. Note how closely the two C-terminal helices of R121D and apoLf superimpose.

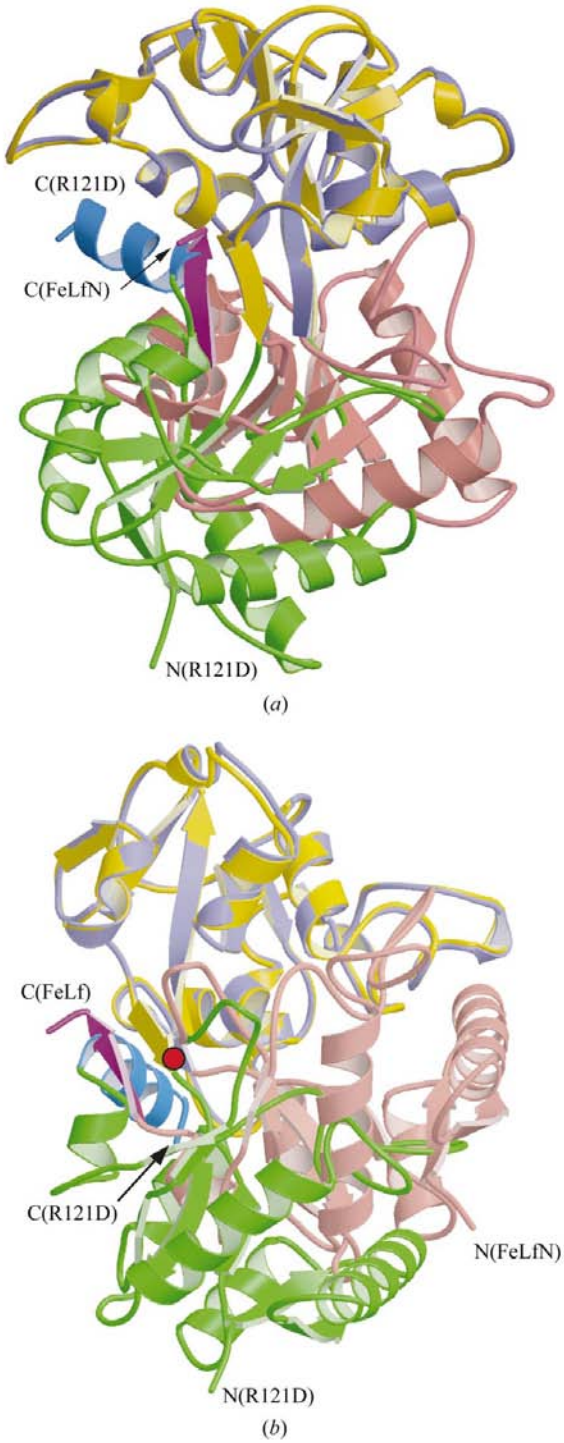

Figure 4
Overlay of the apoR121D and $FeLf_N$ structures after superposition on the basis of their N2 domains. The apoR121D structure is in green (N1 domain) and gold (N2 domain). The corresponding colouring scheme for $FeLf_N$ is pink (N1 domain) and purple (N2 domain). The C-terminal helix (residues 321–332) of apoR121D is shown in blue. The C-terminal strand (residues 321–327) of $FeLf_N$ is shown in magenta. (*a*) View highlighting the β-strands that link the N1 and N2 domains and the C-terminal strand of $FeLf_N$. Relative to the orientation of the molecules in Fig. 3, the molecules are rotated by 180° about an axis running vertically in the plane of the page. (*b*) View down the axis of rotation between open and closed forms of Lf. The position of this axis is marked by a red-filled circle.

domains, helix 11 in each of the two structures is also found to match perfectly (Fig. 3). Thus, a defining feature of these two apo structures is the preservation of this helix 5/helix 11 interface.

We propose that the ordering of the C-terminal α-helix in apoR121D arises because of the conserved interface between this helix and helix 5 (and the body of the N2 domain) that comes into effect when the domains open. This is nicely consistent with the 'two-interface' analysis of Gerstein *et al.* (1993), who showed that the conformational change in the N-lobe of lactoferrin (and by implication other transferrins) could be seen as a shift between two close-packed interfaces. The closed form is stabilized by interactions across a large interface, primarily the binding cleft. In the open form, the interactions across the large interface are much reduced, but interactions at the small interface, primarily involving the C-terminal helix and the N2 domain, come into effect. Thus, although the domain opening of the apoR121D structure differs from that of apoLf in detail, there is a remarkable similarity in the extent of the conformational change when the N-lobes of transferrins, both from whole molecules and half-molecules, are compared.

An intriguing question that remains is why the C-terminal region (residues 322–332) is invariably found to be poorly ordered and non-helical for the closed holo forms of Lf_N and its mutants. The answer, we propose, may lie in conformational entropy and surface free-energy effects. A highly significant feature of apo and holo full-length lactoferrin is that despite their very different conformations, open and closed, the solvent-exposed surface areas are essentially identical (Jameson *et al.*, 1998). For the N-lobe half-molecule of lactoferrin, Lf_N, trimmed to residues 4–321, the open apo form (represented by apoR121D) has a surface area of 12 180 Å^2, whereas the closed holo form has a substantially smaller surface area of 11 760 Å^2. In this calculation, all side chains were trimmed to the same length, as described in §2.4. When the full molecule, residues 4–332, is considered, the surface areas of open and closed forms become nearly equal as a result of the very different conformations of the C-terminal region, residues 321–332. In the open form, represented by apoR121D, residues 322–332 are observed to form an α-helix. For the closed form, represented by $FeLf_N$, residues 322–326 form a third β-strand alongside the canonical two-stranded β-sheet that links the N1 and N2 domains (Fig. 4). To avoid contact with Pro142, the polypeptide chain then makes a 90° turn at Ala327 to head off into solution. Modelling residues 328–332, either as an extended strand or as an α-helix (see §2.4), increases the surface area to 12 430 Å^2 (when modelled as a strand) or 12 310 Å^2 (when modelled as a helix). In both cases, the surface area of closed form is now very similar to the value of 12 310 Å^2 that is calculated for the open form. Thus, relative to the closed form, the ordered C-terminal helix of the open form serves to reduce the solvent-exposed surface area. On the other hand, the disordering of the C-terminal region, part of which can associate with the β-strands that link the N1 and N2 domains, provides conformational entropy and generates a similar surface area in the closed form; both factors stabilize the closed holo structure. Conformational entropy (Stone, 2001) has been identified as stabilizing the binding of small-molecule pheromones to major urinary protein (MUP) of mouse (Zidek *et al.*, 1999) through *increased* flexibility of the protein main chain on ligand binding, as observed by NMR. Conformational entropy associated with highly flexible loops also appears to stabilize a highly invaginated solvent-exposed hydrophobic pocket in bovine β-lactoglobulin (Jameson *et al.*, 2002).

Compared with half-molecule forms, there is a different combination of stabilizing interfaces in full-length lactoferrin, which include contributions from the C-lobe, that lead to conservation of surface area between the open and closed states. For Lf_N we make the counter-intuitive prediction that modifications of the C-terminal region (residues 322–332) that increase helical propensity, without affecting the helix 5/helix 11 interface of the open form, will destabilize the closed conformation of Lf_N species.

This work was supported by the US National Institute of Child Health and Human Development (grant number HD-20859 to ENB) and by the award of a Fulbright Scholarship to WAB.

References

Aisen, P. (1998). *Metal Ions Biol. Syst.* **35**, 585–631.

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

Aisen, P. & Liebman, A. (1972). *Biochim. Biophys. Acta*, **257**, 314–323.

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.

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

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

Baldwin, D. A., Jenny, R. R. & Aisen, P. (1984). *J. Biol. Chem.* **259**, 13391–13394.

Breyer, W. A., Kingston, R. L., Anderson, B. F. & Baker, E. N. (1999). *Acta Cryst.* D**55**, 129–138.

Britton, D. (1972). *Acta Cryst.* A**28**, 296–297.

Brock, J. (1985). *Metalloproteins*, Part 2, *Metals with Non-Redox Roles*, edited by P. Harrison, pp. 183–262. London: MacMillan Press.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.

Bullen, J. J., Rogers, H. J. & Leigh, L. (1972). *Brit. Med. J.* **1**, 69–75.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Day, C. L. (1993). PhD thesis, Massey University, New Zealand.

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.

Engh, R. A. & 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.

Funk, W. D., MacGillivray, R. T. A., Mason, A. B., Brown, S. A. & Woodworth, R. C. (1990). *Biochemistry*, **29**, 1654–1660.

Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M. & Chothia, C. (1993). *J. Mol. Biol.* **234**, 357–372.
Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–646.
Jameson, G. B., Adams, J. J. & Creamer, L. K. (2002). *Intl Dairy J.* **12**, 319–329.
Jameson, G. B., Anderson, B. F., Norris, G. E., Thomas, D. H. & Baker, E. N. (1998). *Acta Cryst.* D**54**, 1319–1335.
Jeffrey, P. D., Bewley, M. C., MacGillivray, R. T. A., Mason, A. B., Woodworth, R. C. & Baker, E. N. (1998). *Biochemistry*, **37**, 13978–13986.
Klausner, R. D., Ashwell, J. V., Van Renswoude, J. B., Harford, J. & Bridges, K. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 2263–2267.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Kurokawa, H., Dewan, J. C., Mikami, B., Sacchettini, J. C. & Hirose, M. (1999). *J. Biol. Chem.* **274**, 28445–28452.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
MacGillivray, R. T. A., Bewley, M. C., Smith, C. A., He, Q.-Y., Mason, A. B., Woodworth, R. C. & Baker, E. N. (2000). *Biochemistry*, **39**, 1211–1216.
MacGillivray, R. T. A., Moore, S. A., Chen, J., Anderson, B. F., Baker, H., Luo, Y., Bewley, M., Smith, C. A., Murphy, M. E., Wang, Y., Mason, A. B., Woodworth, R. C. & Baker, E. (1998). *Biochemistry*, **37**, 7919–7928.
Mazurier, J. & Spik, G. (1980). *Biochim. Biophys. Acta*, **629**, 399–408.
Meritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Mizutani, K., Yamashita, H., Mikami, B. & Hirose, M. (2000). *Biochemistry*, **39**, 3258–3265.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Nicholson, H., Anderson, B. F., Bland, T., Shewry, S. C., Tweedie, J. W. & Baker, E. N. (1997). *Biochemistry*, **36**, 341–346.
Nurizzo, D., Baker, H. M., He, Q.-Y., MacGillivray, R. T. A., Mason, A. B., Woodworth, R. C. & Baker, E. N. (2001). *Biochemistry*, **40**, 1616–1623.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Palmiter, R. D., Behringer, R. R., Quaife, C. J., Maxwell, F., Maxwell, I. H. & Brinster, R. L. (1987). *Cell*, **50**, 435–443.
Peterson, N. A., Anderson, B. F., Jameson, G. B., Tweedie, J. W. & Baker, E. N. (2000). *Biochemistry*, **39**, 6625–6633.
Quiocho, F. A. & Ledvina, P. S. (1996). *Mol. Microbiol.* **20**, 17–25.
Roussel, A. & Cambillau, C. (1989). *Silicon Graphics Geometry Partners Directory*, pp. 72–78. Mountain View, California: Silicon Graphics.
Sheldrick, G. M. (1997). *SHELXL*97. University of Göttingen, Germany.
Stone, M. J. (2001). *Acc. Chem. Res.* **34**, 339–345.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Yeates, T. O. (1988). *Acta Cryst.* A**44**, 142–144.
Zidek, L., Novotny, M. V. & Stone, M. J. (1999). *Nature Struct. Biol.* **6**, 1118–1121.