Acta Cryst. (1994). D**50**, 302–316

Structure of Copper- and Oxalate-Substituted Human Lactoferrin at 2.0 Å Resolution*

BY CLYDE A. SMITH, BRYAN F. ANDERSON, HEATHER M. BAKER AND EDWARD N. BAKER†

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

(*Received* 20 *October* 1993; *accepted* 14 *January* 1994)

Abstract

The three-dimensional structure of human dicupric monooxalate lactoferrin, Cu_2oxLf, has been determined to 2.0 Å resolution, using X-ray diffraction data collected by diffractometry to 2.5 Å resolution, and oscillation photography on a synchrotron source to 2.0 Å resolution. Difference electron-density maps calculated between Cu_2oxLf and both dicupric lactoferrin, Cu_2Lf, and diferric lactoferrin, Fe_2Lf, showed that the oxalate had replaced a carbonate in the C-terminal binding site, and that, relative to Cu_2Lf, there were no significant differences in the N-terminal site. The structure was then refined crystallographically by restrained least-squares methods. The final model, in which the r.m.s. deviation in bond distances is 0.017 Å, contains 5314 protein atoms (691 residues), two Cu^{2+} ions, one bicarbonate ion, one oxalate ion, 325 solvent molecules and one sugar residue. The crystallographic *R* factor of 0.193 is for 46 134 reflections in the range 8.0 to 2.0 Å resolution. The oxalate ion is coordinated to copper in a 1,2-bidentate fashion, and the added bulk of the anion results in the rearrangement of the side chains of nearby arginine and tyrosine residues. No other major alterations in the molecule can be observed, the overall protein structure being the same as that for Cu_2Lf and Fe_2Lf.

Introduction

Proteins of the transferrin family, which includes serum transferrin, ovotransferrin and lactoferrin, have the characteristic ability to bind, tightly but reversibly, two Fe^{3+} ions, concomitantly with two CO_3^{2-} anions (Brock, 1985; Harris & Aisen, 1989; Aisen, 1989). Two more distantly related members of the family, melanotransferrin (Rose *et al.*, 1986) and the transferrin from the tobacco hornworm, *Manduca sexta* (Bartfeld & Law, 1990) display similar characteristics to the other transferrins but have been shown to bind only one Fe atom (Baker *et al.*, 1992; Bartfeld & Law, 1990).

Metal- and anion-binding studies have shown that many other metal ions can be bound by transferrins in place of Fe^{3+} and that there are a variety of anions which can replace carbonate yet still promote iron binding. Crystallographic studies on human diferric lactoferrin, Fe_2Lf (Anderson, Baker, Norris, Rice & Baker, 1989), and human dicupric lactoferrin, Cu_2Lf (Smith, Anderson, Baker & Baker, 1992), have demonstrated that the different metal ions bind to the same four protein ligands, and to a carbonate ion. In the case of Fe_2Lf, the two specific sites in the N-terminal and C-terminal lobes are very similar, the ligands and the Fe^{3+} ion occupying roughly equivalent positions, with the carbonate anions coordinated in a bidentate fashion. Bidentate carbonate coordination has also been observed in the structural analyses of the iron-saturated forms of an N-terminal half-molecule fragment of rabbit-serum transferrin (Sarra, Garratt, Gorsinky, Jhoti & Lindley, 1990) and the recombinant N-lobe of human lactoferrin (Day, Anderson, Tweedie & Baker, 1993). In the case of Cu_2Lf, however, the C-terminal site resembles the corresponding site in the diferric molecule, with bidentate carbonate coordination, but the anion in the N-terminal site is monodentate, with the Cu^{2+} ion significantly displaced from the iron position (Smith, Anderson, Baker & Baker, 1992). Thus, metal substitution can cause the small differences between the two sites to become distinctly more pronounced.

The structural parameters governing anion substitution were first investigated by Schlabach & Bates (1975) who concluded that anions should possess a carboxylate group and a second donor group about 6.5 Å away in order to promote synergistic binding; in their model the carboxylate group interacted with a positively charged group on the protein while the proximal second donor group was bound to the iron. Further studies based on electron-spin echo envelope modulation (ESEEM) spectroscopy (Dubach, Gaffney, More, Eaton & Eaton, 1991) has led to the classification of synergistic anions into three categories; carboxylate with a proximal donor group (such as glycolate and salicylate); dicarboxylate with an

* Supported by US National Institutes of Health (HD-20859), Wellcome Trust, Health Research Council of New Zealand and New Zealand Dairy Research Institute.

† To whom correspondence should be addressed (by air mail).

additional donor group (malate and oxaloacetate) and dicarboxylate (oxalate, malonate and maleate). Several modes of binding of these anions have been proposed, all having bidentate coordination of the carboxylate and proximal groups to the metal ion.

We undertook a crystallographic analysis of oxalate-substituted lactoferrin in order to obtain definitive evidence of the mode of coordination of an anion other than carbonate. The complex analysed, dicupric monooxalate monocarbonate lactoferrin, $Cu_2(CO_3)(C_2O_4)$Lf (Cu_2oxLf), was chosen because the use of Cu^{2+} as the associated metal ion gave rise to oxalate binding in only one site and thus also allowed us to determine which site showed a greater affinity for oxalate.

The effects of anion substitution have first been examined through difference electron-density maps, which have previously proved to be extremely sensitive in identifying specific changes in protein structure resulting from the binding or removal of metal ions, the binding of anions or substrates, or the movement of side chains due to mutations (Schneider, Eklund, Cedergren-Zeppezauer & Zeppezauer, 1983; Rees & Lipscomb, 1983; Fothergill & Fersht, 1991; Nar, Messerschmidt, Huber, van de Kamp & Canters, 1992; Lindahl, Liljas, Habash, Harrop & Helliwell, 1992). The resulting structural model has then also been refined at high resolution.

Experimental

Crystallization and data collection

The purification of human apolactoferrin, and the preparation and crystallization of Cu_2oxLf, have been described previously (Smith, Baker & Baker, 1991). The crystals are orthorhombic, $a = 155.8$, $b = 97.1$, $c = 56.2$ Å, space group $P2_12_12_1$, and are isomorphous with those of Fe_2Lf ($a = 156.2$, $b = 97.4$, $c = 55.85$ Å) and Cu_2Lf ($a = 155.9$, $b = 97.0$, $c = 56.0$ Å). Intensity distributions indicated that the substitution of copper for iron and oxalate for carbonate did not result in any major structural changes.

Medium-resolution X-ray diffraction data (to 2.5 Å resolution) were collected with a CAD-4 diffractometer using Cu $K\alpha$ radiation as previously described (Smith, Baker & Baker, 1991). Fig. 1 shows the variation of the proportion of observed reflections (with $I > 2\sigma_I$) as a function of resolution. This shows a marked drop off in the number of observed reflections at about 2.9 Å, with very few observable data beyond 2.6 Å.

A second data set to 2.0 Å resolution was collected, by oscillation photography (Arndt & Wonacott, 1977), at the SRS facility at Daresbury, England, using radiation of wavelength 0.88 Å. Two crystals were used to collect 50 film packs (of three films each) covering 90° of rotation about the a^* axis. The films were digitized with an Optronics P1000 rotating-drum microdensitometer and processed with the *MOSFLM* set of programs developed by Wonacott (1980). The fully and partially recorded reflections were scaled and merged into one data set, with the partial reflections added between adjacent film packs.

Finally, the diffractometer and film data were merged into one file containing 51024 unique reflections. Relevant X-ray data collection statistics may be found in Table 1. The final X-ray data set contains 92% of the total theoretical reflections to 2.0 Å resolution. A Wilson plot (Wilson, 1949) for the final data set is shown in Fig. 2 and gives an estimate of 34 Å^2 for the overall temperature factor. The variation of the proportion of observed reflections with resolution is also shown in Fig. 1. There is a gradual fall off beyond about 2.5 Å, but the proportion (40%) with $I > 2\sigma_I$ is still significant at the nominal cut off of 2.0 Å resolution.

Difference Fourier maps

Cu_2oxLf–Fe_2Lf. A difference Fourier synthesis was calculated using the observed structure amplitudes for Cu_2oxLf and Fe_2Lf and phases calculated from the current Fe_2Lf model (refined at 2.2 Å resolution, $R = 0.179$; Haridas, Anderson & Baker, unpublished), *i.e.* the coefficients were $|F_{Cu_2oxLf(obs)} - k|F_{Fe_2Lf(obs)}|$ and phases $\alpha_{calc(Fe_2Lf)}$.

The highest positive and negative peaks in the difference map were located and displayed on an Evans and Sutherland PS300 graphics system with *FRODO* (Jones, 1978). The largest peaks were in the vicinity of the N-lobe and C-lobe metal-binding sites (shown in Figs. 3*a* and 3*b*). In the N-terminal site (Fig. 3*a*), there is a single pair of positive and negative peaks close to the position occupied by the

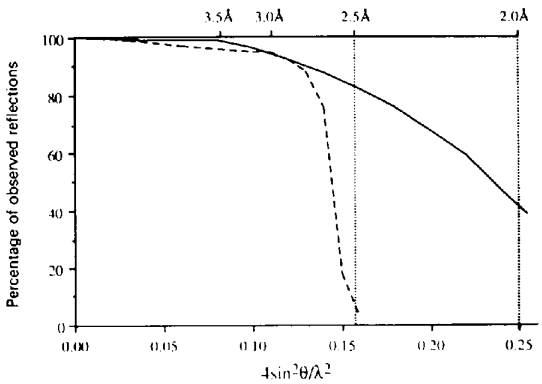


Fig. 1. Percentage of unique reflections with $I > 2\sigma_I$, as a function of $4\sin^2\theta/\lambda^2$ for the diffractometer data set (----) and the final data set obtained by merging the diffractometer and synchrotron data sets (—). The percentage of observed data was calculated in equal concentric shells of $\Delta 4\sin^2\theta/\lambda^2 = 0.02$.

Table 1. *Statistics of the data collection*

(*a*) Diffractometer	
No. of crystals used	4
Total reflections measured	30338
Independent reflections	25037
Observed reflections*	21097 (84%)
R_{merge}†	0.062
(*b*) Oscillation photography	
No. of film packs	50
No. of films per pack	3
Average R_{scal}†	0.033
Average R_{sym}†	0.072
R_{merge}†	0.078
Total reflections measured	240281
Independent reflections (non-zero)	49994
(*c*) Final data set (diffractometer + photographic)	
R_{merge}‡	0.077
Independent reflections	51024
Independent reflections to 2.0 Å	50551
Observed reflections to 2.0 Å*	38143 (75%)

* With $I > 2\sigma_I$.

† $R = \sum|I - \bar{I}|/\sum I$; R_{scal} measures the agreement between reflections on successive films of a given pack; R_{sym} measures the agreement between symmetry-related reflections on the same film.

‡ $R = \sum|I - \bar{I}|/\sum I$; the agreement between intensities measured by the diffractometer and by oscillation photography.

Fe atom in Fe_2Lf. These were interpreted in terms of a displacement of the metal ion when copper replaces iron. A similar type of movement of the metal ion has been observed in difference maps calculated between Cu_2Lf and Fe_2Lf (Smith, Baker & Baker, 1991). The lack of difference density at the anion site or at the protein ligands, suggested that the latter were not displaced, and that the anion site was occupied by carbonate, as in Cu_2Lf.

Several large positive and negative peaks in the C-terminal binding site (Fig. 3*b*), indicated that significantly more structural alteration had taken place. A positive feature adjacent to the anion site indicated

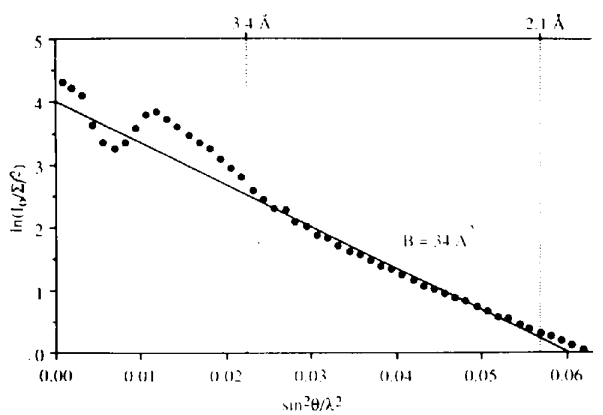


Fig. 2. Wilson plot of $\ln(I_o/\Sigma f^2)$ against $\sin^2\theta/\lambda^2$, with the average observed intensity, I_o, and the sum of the square of the scattering factors, Σf^2, calculated in small concentric shells of $\Delta(\sin^2\theta/\lambda^2) = 0.00125$ Å^{-2}. The gradient, $-2B$, was determined from the best fit line between 3.5 and 2.0 Å and gave an overall B value of 34 Å^2.

that oxalate was bound in place of carbonate, the extra density arising from the two extra atoms of the oxalate ion. A pattern of positive and negative peaks associated with the side chains of Arg465 and Tyr398 pointed to a movement of the Arg465 side chain away from the anion-binding site upon oxalate binding, and a linked movement of the Tyr398 side chain. In addition to these major peaks, positive and negative peaks near the metal ion suggested some movement of the metal. It is not clear whether this movement is related to the substitution of copper for iron or of oxalate for carbonate, although it should be noted that the difference map calculated between Cu_2Lf and Fe_2Lf (Smith, Baker & Baker, 1991) also showed a pair of peaks near the C-lobe metal ion.

Cu_2oxLf–Cu_2Lf. This electron-density map was calculated with coefficients $|F_{Cu_2oxLf\,(obs)} - k|F_{Cu_2Lf|(obs)}|$ and phases derived from the refined coordinates of Cu_2Lf (Smith, Anderson, Baker & Baker, 1992). Once again, there were peaks in the C-terminal binding site indicative of the extra density of an oxalate ion and the accompanying movement of Arg465 and Tyr398, along with a small movement of the Cu atom (Fig. 4). In the N-lobe, however, the largest feature was a small negative peak near the copper site (peak height, 2.8σ). This was not judged to be significant as it was lower than the highest noise peak in the map ($\simeq 3\sigma$). This difference map showed quite clearly that the binding of oxalate in the C-terminal site has little or no effect on the metal-ion geometry in the N-terminal copper-binding site, which remains as in Cu_2Lf.

Refinement

The starting model for refinement was derived from the refined atomic coordinates of dicupric lactoferrin (Smith, Anderson, Baker & Baker, 1992), with the two Cu^{2+} ions, the two carbonate ions, the solvent molecules and the carbohydrate residues removed. The eight metal-binding ligands (four in each binding site) and the anion-binding arginine and threonine residues (Arg121 and 465, Thr117 and 461) were changed to alanine, and all individual temperature factors were given a uniform value of 25 Å^2. This model was initially refined against the medium-resolution diffractometer data for which the initial R factor for data in the resolution range 5.0–2.8 Å (15 607 reflections with $I > 0.5\sigma_I$) was 0.293.

The structure was refined in 'phases' consisting of 20–30 cycles of restrained least-squares refinement, using the *FFT* version of the program *PROLSQ* (Hendrickson & Konnert, 1980), followed by manual model rebuilding with *FRODO* (Jones, 1978). Restraints were placed on bond lengths, angle (1—3) distances, planar groups, chiral volumes, temperature factors, torsion angles and van der Waals con-

tacts. After the first phase (22 cycles), the R factor had dropped to 0.171 and a $2F_o - F_c$ map was then used to add the N-lobe Cu^{2+} ion and the four N-lobe ligand side chains to the model.

Further refinement (25 cycles) with the C-terminal site still empty and no anion in the N-lobe left large pieces of residual density which could be assigned to the bound anions. In the N-lobe (Fig. 5*a*), the density had a bifurcated shape similar to that observed in the N-lobe of dicupric lactoferrin (Smith, Anderson, Baker & Baker, 1992), suggesting monodentate coordination of the carbonate. Such a configuration (with a copper–oxygen distance of 2.0 Å), gave favourable interactions with the surrounding protein structure, with five hydrogen bonds in all. The copper–Tyr92 bond was found to be significantly lengthened relative to the other metal–ligand bonds, again as seen in the N-lobe of Cu_2Lf (Smith, Anderson, Baker & Baker, 1992).

In the C-terminal site, after the placement of the copper ion and the side chains of the four protein ligands, the remaining anion density appeared bilobal in nature (Fig. 5*b*) and too large to be accounted for by a carbonate ion. An oxalate ion, bound in a 1,2-bidentate fashion was found to fit almost perfectly into the available density, making at least five favourable hydrogen bonds. The oxalate anion was fitted with symmetrical coordination to the Cu atom and although only loose restraints (0.05 Å) were placed on the copper–ligand bond distances, this configuration did not change significantly during the course of the refinement.

After four phases of refinement against the medium-resolution diffractometer data, the R factor was 0.170 for data between 5.0 and 2.5 Å resolution (20 064 reflections with $I > 0.5\sigma_I$). At this stage, the higher resolution synchrotron data became available and further refinement was based on the combined (diffractometer and photographic) data set, with a maximum resolution of 2.0 Å. Even at this resolution, however, the metal density was not fully resolved from the ligands. Loose constraints on the metal–ligand bond lengths ($\sigma = 0.05$ Å) were imposed, therefore, with the target values being periodically updated to take into account any consistent shifts in the metal site. The only distance which was not restrained was that of the Cu—O(Tyr92) bond in the N-lobe binding site.

To facilitate the process of solvent-molecule location, the refined diferric and dicupric lactoferrin structures were superimposed upon each other and the solvent molecules which appeared in conserved positions (within 1.0 Å) were added to the Cu_2oxLf model. There were 123 out of 301 Cu_2Lf water moelcules which could be classified as common to both Cu_2Lf and Fe_2Lf, and it was assumed that this 'core' set of solvent molecules might also be common to Cu_2oxLf. All of these solvent molecules were, however, checked against an $(F_o - F_c)$ electron-density map prior to their inclusion in the model. They were retained only if they corresponded to positive peaks in the difference map (at greater than the 3σ level), made reasonable hydrogen bonds to either protein atoms or other water molecules and did not interact with any part of the structure whose conformation was in doubt. A total of 101 of the 'core' solvent molecules were retained in the structure. Additional solvent molecules were fitted into $(F_o - F_c)$ density (contoured at three times the r.m.s. deviation of the map) using the same criteria outlined above. At regular intervals, solvent molecules with high B values or too-close contacts were checked by excluding them from the refinement and calculating $(2F_o - F_c)$ and $(F_o - F_c)$ omit maps to establish their correct position. In addition, some well defined solvent molecules were deliberately omitted and their positions were re-evaluated in a similar way. At a later stage of the structure refinement, $(F_o - F_c)$ electron-density maps were searched systematically by identifying the highest positive peaks and examining these locations on the graphics system. Once again water molecules were added only if they fitted the aforementioned criteria.

Results

The final model

The final atomic coordinate set contains 5314 protein atoms (691 amino-acid residues), 325 solvent molecules, 13 carbohydrate atoms (one *N*-acetylglucosamine residue), two Cu^{2+} ions, one CO_3^{2-} ion and one $C_2O_4^{2-}$ ion. The protein and carbohydrate structures have geometry close to ideal, with r.m.s. deviations of 0.017 and 0.061 Å from standard values of bond lengths and angle (1—3) distances, respectively. After 310 cycles of refinement (13 phases of refinement and model building) the final value of the crystallographic R factor was 0.193 for 46 134 reflections with $I > 1\sigma_I$ between 8.0 and 2.0 Å resolution. Relevant refinement statistics are given in Table 2.*

The average error in atomic coordinates is estimated from a Luzzati plot (Luzzati, 1952) of the R factor as a function of $\sin^2\theta/\lambda^2$, and from a σ_A plot (Read, 1986) of $\ln(\sigma_A)$ as a function of $\sin^2\theta/\lambda^2$, both shown in Fig. 6. In the former case, the mean error in atomic positions can be estimated at about 0.25 Å, while the slope of the σ_A plot indicates an error close

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1LCF and R1LCFSF). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37115). At the request of the authors, the structure factors will remain privileged until 31 December 1994. A list of deposited data is given at the end of this issue.

Table 2. *Refinement statistics*

Resolution limits (Å)	8.0 2.0
Final R factor	0.193
No. of reflections used*	46134
No. of protein atoms	5314
No. of solvent molecules	325
Other ions	2 Cu^{2+}, 1 $C_2O_4^{2-}$, 1 CO_3^{2-}
No. of sugar residues	1 NAG
Average B value (protein atoms) (Å^2)	40.1
Average B value (all atoms) (Å^2)	41.2
R.m.s. shift in position (final cycle) (Å)	0.004
R.m.s. shift in atomic B (final cycle) (Å^2)	0.51
Geometrical deviations†	
Bond lengths (Å)	0.017 (0.019)
Angle distances (Å)	0.061 (0.050)
Planarity (Å)	0.014 (0.020)
Chiral volume (Å^3)	0.188 (0.15)
Non-bonded contacts (Å)	0.204 (0.50)

* *PROLSQ* refinement against 2.0 Å resolution film + diffractometer data, using data with $I > 1.0\sigma_I$.

† Values given are the final r.m.s. deviations from standard values for these parameters with target values in parentheses.

to 0.37 Å. As these are only mean coordinate errors, the true error is probably somewhat less than this in well defined regions of the structure (the binding site for example) but substantially more in some poorly defined parts of the molecule such as surface loops.

Conformational angles

A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the main-chain conformational torsion angles, φ and ψ (Fig. 7), shows that about 83% lie within the core region, with 98% lying within the allowed regions of conformational space. The exceptions are mostly in ill defined surface regions of the structure, namely external loops including residues 87 and 283, as well as the N-terminus, residues 1–5, where the poor definition of the electron density makes interpretation of the correct folding of the polypeptide chain very difficult. Plots of average main-chain or side-chain atomic temperature factors as a function of residue number indicate quite clearly the location of these portions of the structure and although every effort was made to determine their correct conformation with the use of omit maps, they remain poorly defined in relation to the rest of the structure.

Two residues, Leu299 and Leu642, have clear, well resolved electron density, but still have (φ, ψ) angles outside the normally occupied regions. Residues Leu299 and Leu642 are the central residues in two γ-turns, similar to those observed in thermolysin (Matthews, 1972). Such turns are generally characterized by main-chain conformational angles $\varphi = 70°$, $\psi = -60°$ (Baker & Hubbard, 1984) for the central residue; the values for Leu299 and Leu642 are (68, $-29°$) and (60, $-36°$), respectively. The residues to either side (298 and 300, 641 and 643) interact *via* 1—3 hydrogen bonds (Baker & Hubbard, 1984). These γ-turns are in equivalent positions in the N- and C-lobes of the protein, occurring as one of a series of loops in a portion of the polypeptide chain flanking the binding cleft. The sequence in this region is highly conserved in most members of the transferrin family and analysis of the atomic coordinates from the structures of diferric rabbit-serum transferrin (Bailey *et al.*, 1988) and the N-terminal lobe of rabbit-serum transferrin (Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990), indicate similar turns at identical positions in the polypeptide chain, with (φ, ψ) angles near (70, $-50°$).

Analysis of protein structures from the Brookhaven Protein Data Bank (Morris, MacArthur, Hutchinson & Thornton, 1992) shows that there are a small number of non-glycinyl residues which adopt an ε configuration (initially described in Sibanda, Blundell & Thornton, 1989). In Cu_2oxLf, Ser191, with main-chain torsion angles (63, $-174°$), has this conformation which has been noted in a number of proteins, including Arg133 in the loop 130–137 in alcohol dehydrogenase (Eklund *et al.*, 1976), an aspartate (Asp174) in the protease tonin (Fujinaga & James, 1987) and the active serine (Ser144) in the fungal triglyceride lipase from *Rhizomucor miehei* and related enzymes (Derewenda & Derewenda, 1991). The residue in the C-lobe of Cu_2oxLf which corresponds in position to Ser191 is a glycine (Gly527), which has very similar (φ, ψ) angles (66, $-170°$). In all the other transferrins whose sequences are known, there are glycine residues at this position in *both* lobes and there is a strong sequence conservation around it. This suggests that the conformation taken by Ser191 is a consequence of a highly conserved folding pattern in this region which immediately precedes one of the metal ligands (Tyr192 or Tyr528).

Polypeptide chain and domain structure

The polypeptide chain folding of Cu_2oxLf is closely similar to that observed in Cu_2Lf and Fe_2Lf. This was first shown by the difference maps calculated between Cu_2oxLf and the latter structures, which suggested that the only significant differences were in the vicinity of the metal- and anion-binding sites. This conclusion is further supported by superposition of the three structures; all three have the same domain structure, secondary structure and general organization as described for Fe_2Lf (Anderson, Baker, Norris, Rice & Baker, 1989).

Superposition of Cu_2oxLf on to Fe_2Lf using all 691 C_α positions gives an overall r.m.s. deviation in C_α positions of 0.65 Å. This figure is exaggerated, however, by the large contributions from the first five residues at the N-terminus which have an average deviation in C_α positions of almost 5 Å. The overall r.m.s. value drops to 0.39 Å when these five residues

are excluded, and to 0.29 Å (Table 3) when regions of the structure with B values greater than 80 Å^2 are also excluded (24 residues). The largest deviations in C_α positions between Cu_2oxLf and Fe_2Lf are at the N-terminus (as noted above) and in external loop regions near residues 86, 220, 283 and 420. The internal parts of the structure near the metal-binding sites have deviations in the order of 0.1–0.3 Å.

Superpositions of the individual lobes and domains of Cu_2oxLf on to those of Fe_2Lf and Cu_2Lf (Table 3) confirm the similarities in their folding. The N-lobe of Cu_2oxLf seems to fit slightly less well to its counterpart in Fe_2Lf (r.m.s. deviation 0.28 Å) than does the C-lobe (r.m.s. deviation 0.23 Å). This may be either because the substitution of Cu^{2+} for Fe^{3+} has induced changes in the extent of closure of the N-lobe, or because of the greater flexibility of portions of the N-lobe compared with corresponding regions in the C-lobe. The latter is also indicated by the observation that the average main-chain temperature factors in the N-lobe are somewhat larger than those in the C-lobe.

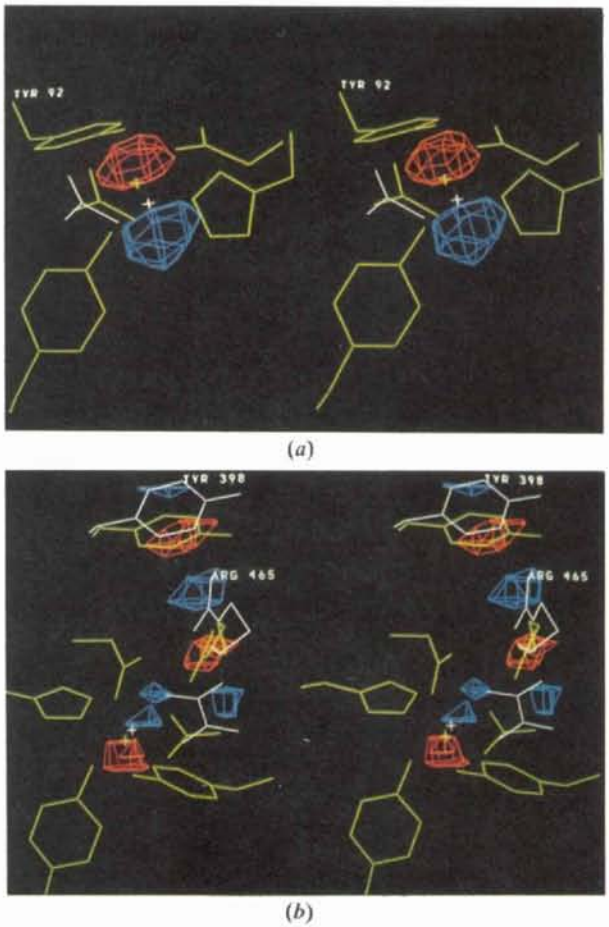


Fig. 3. Stereoviews of the $||F_{Cu_2oxLf}|_{(obs)} - k|F_{Fe_2Lf}|_{(obs)}|$ difference electron-density map contoured at 5.5 times the r.m.s. deviation of the map, showing the main positive (blue) and negative (red) peaks. In (*a*) the peaks in the vicinity of the N-terminal metal-binding site are shown, with atomic positions from the Fe_2Lf model shown in yellow and the position of the copper atom and the carbonate anion in the refined Cu_2oxLf model in white. In (*b*) the C-terminal metal-binding site is indicated, the Fe_2Lf model once again shown in yellow, with the Cu atom, oxalate anion and the side chains of Arg465 and Tyr398 in the Cu_2oxLf model given in white.

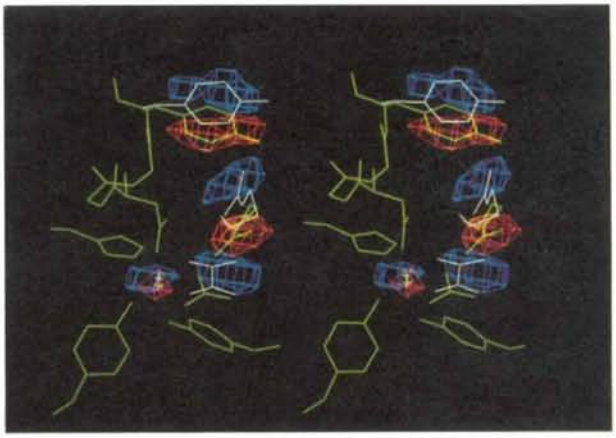


Fig. 4. Stereoview of the $||F_{Cu_2oxLf}|_{(obs)} - k|F_{Cu_2Lf}|_{(obs)}|$ difference electron-density map in the vicinity of the C-terminal metal-binding site, contoured at six times the r.m.s. deviation of the map, showing the main positive (blue) and negative (red) peaks. The Cu_2Lf model is indicated in green, and the Cu atom, oxalate anion and the side chains of Arg465 and Tyr398 in the Cu_2oxLf model are given in white.

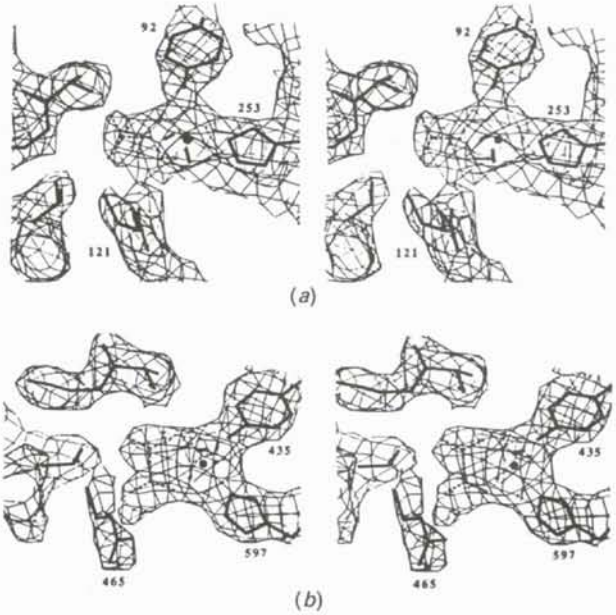


Fig. 5. Stereoviews of the $2F_o - F_c$ electron-density map of Cu_2oxLf, calculated at the time the anions (dashed lines) were first built into the model ($R = 0.177$ for data between 5.0 and 2.6 Å). In (*a*) the N-terminal metal-binding site is shown, indicating the positions of two of the metal-binding residues (Tyr92 and His253) and the anion-binding Arg121 (solid lines). In (*b*), the corresponding view of the C-terminal site is shown with the positions of Tyr435, His597 and Arg465 indicated.

Comparisons of the individual domains show the extent to which the closure of the domains over the respective metal and anion sites varies between Cu_2oxLf, Cu_2Lf and Fe_2Lf. If the N2 domains of Cu_2oxLf and Fe_2Lf are superimposed, for example, they show an r.m.s. deviation of 0.20 Å for all C_α atoms. Furthermore, a rotation of 0.9° is then required to bring the two N1 domains into coincidence, implying a difference in closure of the two N-lobes of 0.9°. A similar figure of 0.6° was found when the N-lobes of Cu_2Lf and Fe_2Lf were compared in this way (Smith, Anderson, Baker & Baker, 1992). Similar figures for the superposition and relative orientations of all the individual domains of Cu_2oxLf are summarized in Table 3.

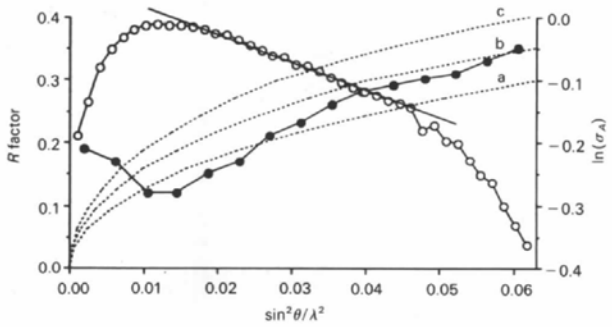


Fig. 6. Plots of the variation of the crystallographic R factor (●) and $\ln(\sigma_A)$ (○) with resolution. The theoretical variations of the R factor (Luzzati, 1952) for coordinate errors, Δr, indicated by the dashed curves a, b and c, of 0.20, 0.25 and 0.30 Å, respectively, are shown. The gradient of the σ_A plot is equal to $\pi^3(|\overline{\Delta r}|^2)$, where $|\overline{\Delta r}|^2$ is the square of the mean error in the atomic coordinates. The slope of the best-fit line between 3.4 and 2.2 Å gives an estimate of the r.m.s. coordinate error in the refined model.

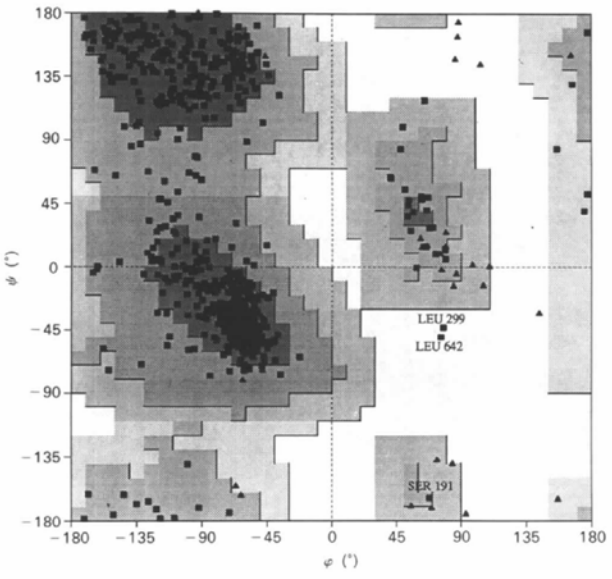


Fig. 7. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the main-chain torsion angles (φ and ψ) for the final refined model of Cu_2oxLf. The plot was calculated with the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). Out of 600 non-glycine and non-proline residues, 496 (83%) had torsion angles in the core regions of the plot, with a further 91 (15%) occupying the allowed regions.

Table 3. *Structural comparisons and domain relationships*

(*a*) Superposition of Cu_2oxLf onto Fe_2Lf*

	R.m.s. deviation (Å)	
Whole molecule	0.29 (C_α)	0.32 (all main chain)
N-lobe onto N-lobe	0.28	0.30
C-lobe onto C-lobe	0.23	0.26
Secondary-structure elements only†	0.25	0.27

(*b*) Domain relationships between Cu_2oxLf and Fe_2Lf‡

		Fe_2Lf N1	N2	C1	C2
	N1	0.27 Å	0.9°	0.5°	0.2°
Cu_2oxLf	N2	–	0.20 Å	0.3°	1.1°
	C1	–	–	0.22 Å	0.8°
	C2	–	–	–	0.25 Å

(*c*) Domain relationships between Cu_2oxLf and Cu_2Lf‡

		Cu_2Lf N1	N2	C1	C2
	N1	0.22 Å	0.2°	0.2°	0.3°
Cu_2oxLf	N2	–	0.18 Å	0.0°	0.6°
	C1	–	–	0.20 Å	0.5°
	C2	–	–	–	0.19 Å

* Omitting residues with $B > 80$ Å^2 (1–4, 85–87, 220–223, 280–285, 417–423).

† Defined as in Anderson *et al.* (1987).

‡ The diagonal elements give the r.m.s. deviation in C_α positions when equivalent domains of the two structures are superimposed. The off-diagonal elements show the relative rotations between pairs of domains.

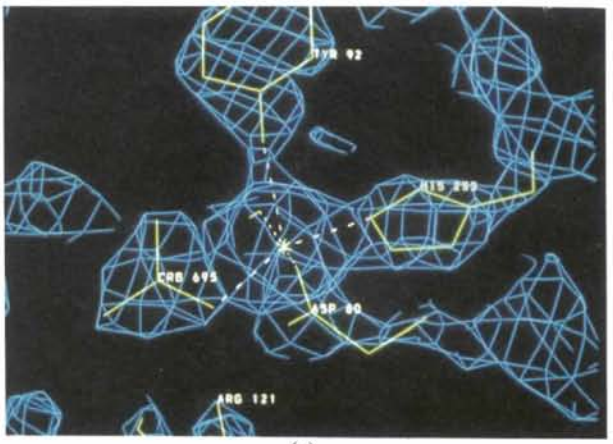
The metal and anion sites

In the final model for Cu_2oxLf, the copper geometries in the two sites are clearly different. The N-site is five-coordinate with an approximately square pyramidal geometry, as is the N-site of Cu_2Lf (Smith, Anderson, Baker & Baker, 1992), with a long apical bond of 2.7 Å to Tyr92, and the Cu atom displaced 0.1 Å out of the plane of the other four ligands, towards the apical ligand. The C-site, on the other hand, is an almost regular octahedron with five of the six bonds between 1.9 and 2.1 Å (the bond to Tyr528 is lengthened to ~2.5 Å). Figs. 8(*a*) and 8(*b*) show the final $(2F_o - F_c)$ electron density in the N-lobe and C-lobe metal-binding sites, and relevant bond lengths and angles in the two copper sites are listed in Table 4.

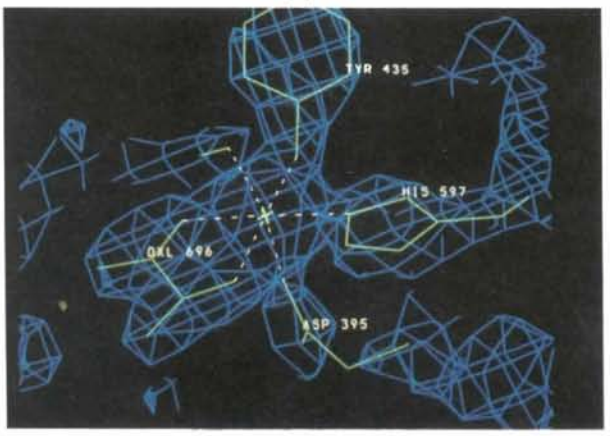
The hydrogen-bonding pattern around the N-terminal carbonate is similar to those observed in Fe_2Lf (Anderson, Baker, Norris, Rice & Baker, 1989), Cu_2Lf (Smith, Anderson, Baker & Baker, 1992) and the N-lobe of rabbit transferrin (Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990), and is shown in Fig. 9(*a*). Five hydrogen bonds are formed, two with Arg121 (involving the N_ε and $N_{\eta 2}$ atoms of the guanidinium group), two with main-chain N

atoms on the N-terminus of helix 5 (residues 123 and 124) and one with Thr117 ($O_{\gamma 1}$). The C-terminal oxalate anion is involved in six hydrogen bonds (Fig. 9*b*), two with the guanidinium group of Arg465 (one each with N_{ε} and $N_{\eta 2}$), two with the main-chain N atoms of residues 467 and 468 (at the N-terminus of helix 5), and two with $O_{\gamma 1}$ of Thr461. A list of the hydrogen bonds associated with the carbonate and oxalate ions is given in Table 5.

The effects of metal and anion substitution can be shown by superposition of the binding sites of Cu_2oxLf on those of Fe_2Lf and Cu_2Lf, using the polypeptide structure to define the superposition. These are shown in Figs. 10(*a*) and 10(*b*) for the N-lobe and C-lobe, respectively. In the N-lobes of these three structures the principal movement is that of the metal ion; in Cu_2oxLf the Cu^{2+} ion has moved



(*a*)



(*b*)

Fig. 8. The final $2F_o - F_c$ electron density in the two metal-binding sites. The final model of Cu_2oxLf is superimposed in green. In (*a*) the N-terminal site is shown. The Cu atom, the monodentate carbonate ion and three of the four metal-binding ligands (Asp60, Tyr92 and His253) are shown, the remaining ligand (Tyr192) being below the plane of the page. In (*b*) the bidentate oxalate ion, the Cu atom and three of the C-lobe binding ligands, Asp395, His597 and Tyr435 are shown. The fourth ligand, Tyr528, projects into the plane of the page.

Table 4. *Bond lengths* (Å) *and bond angles* (°) *in the copper-coordination sphere*

	N-lobe site	C-lobe site
Bond*		
Cu—O60(395)	2.0	2.1
Cu—O92(435)	2.7	2.0
Cu—O192(528)	2.0	2.5
Cu—N253(597)	2.0	2.1
Cu—O_1(anion)	–	2.0
Cu—O_2(anion)	2.0	1.9
Angle*		
O60(395)—Cu—O92(435)	77	84
O60(395)—Cu—O192(528)	167	173
O60(395)—Cu—N253(597)	96	90
O60(395)—Cu—O_1(anion)	–	85
O60(395)—Cu—O_2(anion)	86	85
O92(435)—Cu—O192(528)	94	103
O92(435)—Cu—N253(597)	86	84
O92(435)—Cu—O_1(anion)	–	97
O92(435)—Cu—O_2(anion)	116	170
O192(528)—Cu—N253(597)	92	92
O192(528)—Cu—O_1(anion)	–	92
O192(528)—Cu—O_2(anion)	91	88
N253(597)—Cu—O_1(anion)	–	175
N253(597)—Cu—O_2(anion)	157	97
O_1(anion)—Cu—O_2(anion)	–	81

* Residue numbers in parentheses are those for the C-lobe site.

Table 5. *Hydrogen-bonding interactions between the carbonate and oxalate ions and the surrounding protein in* Cu_2oxLf

Hydrogen bonds	O···H (Å)	C—O···H (°)	N—H···O (°)	O···N (Å)	H—N···O (°)
N-lobe					
O_1—N(123)	1.69	111	155	2.63	16
O_2—N_{ε}(121)	2.00	94	123	2.68	39
O_2—$N_{\eta 2}$(121)	2.00	108	132	2.77	32
O_3—N(124)	2.09	124	129	2.83	35
O_3—$O_{\gamma 1}$(117)	–	153*	116*	2.44*	–
C-lobe					
O_1—N(467)	2.06	112	157	3.01	16
O_3—N(468)	1.87	135	147	2.76	22
O_3—$O_{\gamma 1}$(461)	–	126*	117*	2.49*	–
O_4—$O_{\gamma 1}$(461)	–	120*	91*	2.77*	–
O_4—N_{ε}(465)	1.97	92	140	2.81	27
O_4—$N_{\eta 2}$(465)	1.91	83	132	2.69	32
O_4—O_{w4}(721)†	–	101‡	–	2.53‡	–

* These values related to the C_1—O_3···$O_{\gamma 1}$ and C_B—$O_{\gamma 1}$···O_3 angles and the O_3···$O_{\gamma 1}$ distance.

† Atom 721 O_{w4} is a water molecule. The residue number and atom name are as they appear in the final coordinate file. This water molecule is labelled W1 in Fig. 12.

‡ These values relate to the C1—O_4···O_{w4} angle and the O_4···O_{w4} distance.

0.8 Å from the Fe^{3+} position, while in Cu_2Lf the movement appears slightly greater at 1.0 Å. Associated with this movement of Cu^{2+} from the Fe^{3+} position are small consistent movements of His253 (0.5–0.7 Å) and Tyr92 (0.3–0.4 Å) in the two Cu^{2+}-substituted structures, compared with Fe_2Lf. The carbonate ion also undergoes small rotations about O(3). The anion movement appears smaller (8°) in Cu_2oxLf than in Cu_2Lf (20°), though whether this is significant in terms of the accuracy of fitting such anions is hard to tell. The combined effect of

this rotation, however, together with the movement of the metal ion and the slight opening of the N-lobe, is to give monodentate coordination of the anion in both copper-substituted structures and bidentate coordination in the iron structure.

In the C-lobe of Cu_2oxLf, the Cu atom has shifted ~ 0.2 Å relative to the metal in Cu_2Lf and ~ 0.3 Å relative to that in Fe_2Lf, while the four protein ligands occupy the same positions to within about 0.1–0.2 Å. The most evident difference between the C-lobe metal-binding sites in these structures is clearly related to the substitution of the carbonate anion by an oxalate. The incorporation of the larger oxalate anion results in the displacement of the side chain of Arg465 (the average movement in the four atoms of the guanidinium group, N_ε, C_ζ, $N_{\eta 1}$ and $N_{\eta 2}$, is in the order of 2.7 Å) and brings the $N_{\eta 1}$ atom within hydrogen-bonding distance of the O_η atom of Tyr526 (2.9 Å). Fig. 11 shows the relative movement of Arg465 and the new hydrogen-bonding interaction with Tyr526, resulting from the substitution of carbonate by oxalate.

The movement of the bulky arginine side chain also causes the movement of the side chain of Tyr398, which projects into the binding cleft from the opposing (C1) domain, by ~ 1.5 Å. No hydrogen-bonding interaction is involved, as the O_η–$N_{\eta 1}$ distance is over 5.5 Å, and the observed movement must be caused by van der Waals interactions between the aromatic ring and the approaching arginine. The side chain of Tyr398 interacts only with solvent molecules in all four lactoferrin structures (two in Fe_2Lf, apoLf and Cu_2Lf, and three in Cu_2oxLf), and is, therefore, relatively easy to displace to facilitate the binding of larger anions.

The substitution of oxalate for carbonate also leads to some rearrangement of bound solvent molecules in the binding cleft of Cu_2oxLf, relative to Cu_2Lf (Fig. 12). The position vacated by the phenolic O atom of Tyr398 is occupied by a solvent molecule (~ 1 Å from the O_η position in Cu_2Lf) unique to Cu_2oxLf; it is hydrogen bonded to the $N_{\eta 1}$ atom of Arg465. In addition, there is another unique water molecule resulting from the anion substitution, near the position occupied by the guanidinium group of Arg465 in Cu_2Lf. It is hydrogen bonded to the $N_{\eta 2}$ (Arg465) in Cu_2oxLf and the solvent network, and is also directly hydrogen bonded to atom O_4 of the oxalate anion (Fig. 12). The majority of the other solvent molecules near residues 398, 465 and 526 are in conserved positions relative to Cu_2Lf, including molecules which are hydrogen bonded to the O_η atoms of Tyr398 and Tyr526 in both structures.

Temperature factors

The average main-chain temperature factors for Cu_2oxLf as a function of residue number for the N-

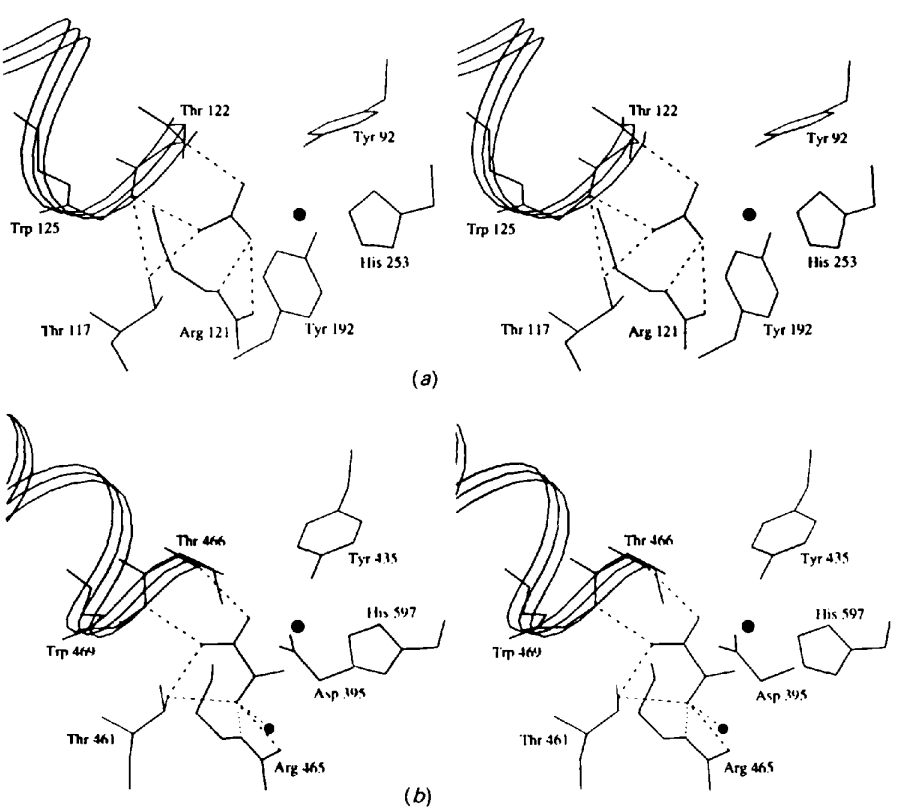
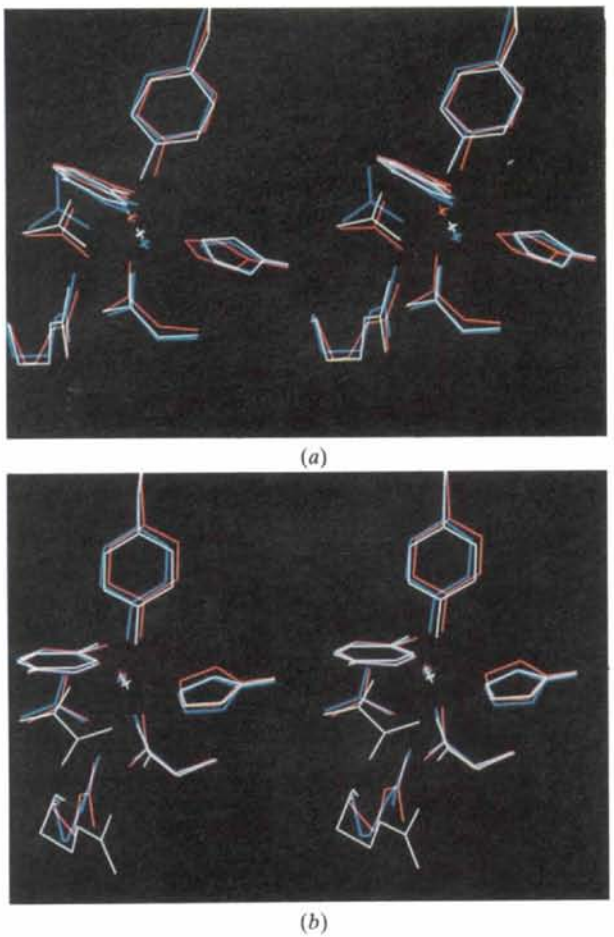


Fig. 9. Stereoviews of the hydrogen-bonding interactions involving (*a*) the N-terminal lobe carbonate ion and (*b*) the C-terminal lobe oxalate ion. The Cu atoms, shown as filled circles, are at the centre of the figures. The N-terminus of helix 5, residues 122–125 in the N-lobe (466–469 in the C-lobe), is shown as a ribbon at the left. In the C-lobe, the solvent molecule unique to Cu_2oxLf, hydrogen bonded to the oxalate and Arg465, is indicated by a filled circle at bottom centre.

and C-lobes are shown in Figs. 13(*a*) and 13(*b*), respectively. The average main-chain *B* value of 38.4 Å^2 (see Table 2) is consistent with the weakness of the high-resolution data and with the overall *B* estimated from the Wilson plot (~34 Å^2) (Wilson, 1949).

In both lobes, the fluctuation of *B* values along the polypeptide chain can be correlated with the patterns of secondary structure. The residues with low *B* values correspond to the β-strands, the helices and the areas associated with the metal- and anion-binding ligands (as indicated in Figs. 13*a* and 13*b*). Most of the residues in these parts of the chain are internal and are stabilized by a large number of hydrogen-bonding and other interactions with protein atoms and ordered water molecules. The exceptions in the N-lobe are the initial three or four residues of β-strand *a*, where it connects to the poorly ordered N-terminal residues, the majority of helix 1 which is solvent exposed at one end of the N1 domain, and helix 8*a*, a short piece of 3_{10} helix which also lies on the molecular surface [the nomenclature of the helices and β-strands is as given by Anderson, Baker, Norris, Rice & Baker (1989)]. In the C-lobe, the secondary structure is quite well defined except for the last few residues of the C-terminal helix, number 12.

Table 6 gives the average main-chain and side-chain atomic temperature factors for the whole molecule (excluding solvent and the glycan), individual lobes and the four domains. These results suggest that the N1 domain contains regions which are somewhat less ordered than the corresponding areas in the C1 domain, and, although it is not correct to generalize that the whole of the N1 domain is more mobile, a significant portion of it appears to be more flexible than the remainder. Fig. 13(*a*) would indicate



(*a*)

(*b*)

Fig. 10. Superpositions of Cu_2oxLf (white) onto Fe_2Lf (red) and Cu_2Lf (blue). In (*a*) the N-terminal binding sites are shown, while in (*b*) the C-terminal sites are presented. The C_α atoms in the helices and β-sheets of the N2 and C2 domains were used to define the superposition.

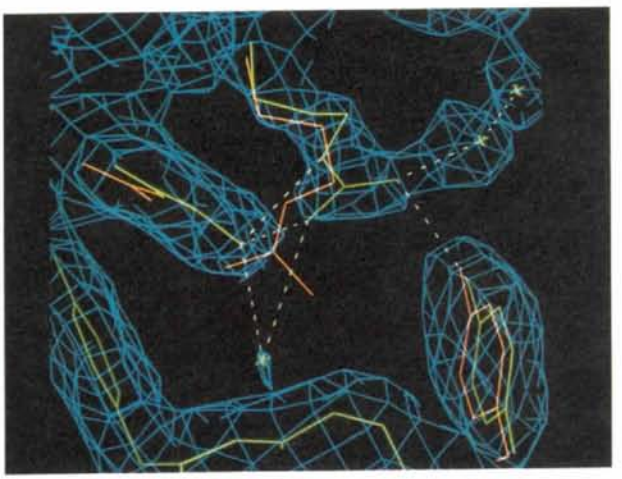


Fig. 11. The superposition of Arg465 (centre) and Tyr526 (lower right) in the C-terminal lobe binding site of Cu_2oxLf (green) onto the corresponding site in Cu_2Lf (red). The carbonate and oxalate ions are shown edge-on at the top left. The electron density shown corresponds to the final model of Cu_2oxLf.

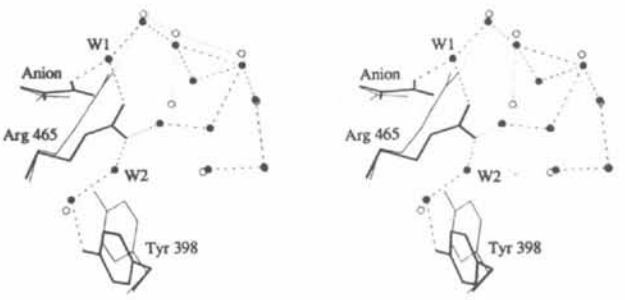


Fig. 12. Superposition of the solvent network in the C-lobe cleft of Cu_2oxLf (thick lines, closed circles, dashed hydrogen bonds) onto the C-lobe cleft of Cu_2Lf (thin lines, open circles, dotted hydrogen bonds). Two solvent molecules occupying positions in Cu_2oxLf near where the side chain of Arg465 and the O_η of Tyr398 are located in Cu_2Lf are labelled *W*1 and *W*2, respectively. The C_α atoms of the α-helices and β-strands were used in the least-squares superposition.

that it is the initial 50 or so residues which contribute most to the high B values and that another region in the second half of this domain (residues 280–300) also has B values significantly greater than the average. However, in the C1 domain, although the region corresponding to the latter (residues 620–640) also has high B values, the initial 50 residues have values less than or equal to the average. The reason for these differences is clear from analysis of the model, in that while the first 50 residues of the N-lobe are located on the exterior of the molecule, the same area in the C-lobe lies close to the N-lobe, in the inter-lobe contact area, and exposure to the external solvent is reduced.

The regions with high B values (>70 Å^2) are mostly external loops connecting the elements of secondary structure. In these parts, the majority of the hydrogen-bonding interactions are with solvent and consequently the protein structure is flexible and not well defined.

Solvent structure

A total of 325 solvent molecules are included in the final model of Cu_2oxLf with an average B value of 54 Å^2. Of these water molecules, 267 can be classified as belonging to the first hydration sphere, in that they make at least one hydrogen bond with a protein atom. The remainder (58) interact only with other solvent molecules. In all, the 'first sphere' solvent molecules make a total of 715 interactions to both protein and other solvent atoms, an average of 2.7 hydrogen bonds per water molecule. Although this is significantly higher than the average of 1.7 hydrogen bonds reported by Baker & Hubbard (1984) from a selection of refined protein structures taken from the Protein Data Bank, that value was only for interactions with protein atoms. The average number of hydrogen bonds to the 'second sphere' water is 1.8. There is a direct correlation between the number of hydrogen bonds to solvent molecules and the average B value, as shown in Table 7(*a*), in that molecules which make only 1 or 2 interactions have high average B values, while those which interact with four or five other atoms have significantly lower average B values. This has previously been observed in other well refined protein structures (Baker, 1988, and references therein) and it is clear that the more hydrogen-bonding partners they have, the more constrained the water molecules are, and the lower is their temperature factor.

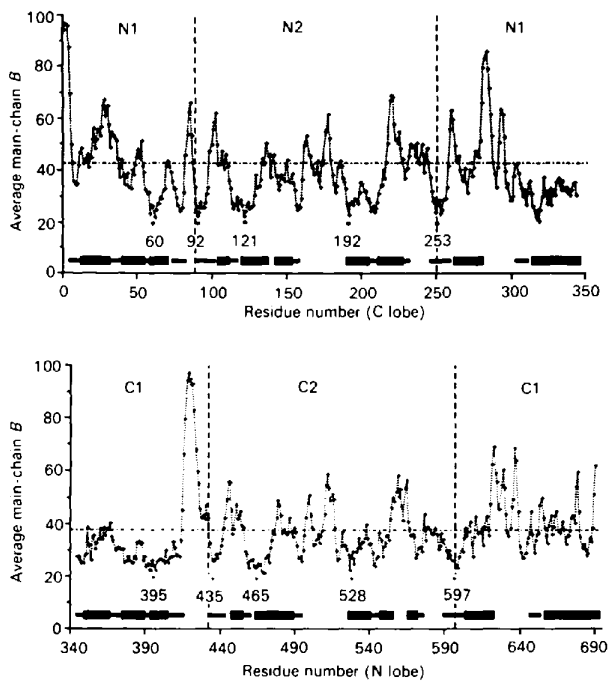


Fig. 13. Plots of the average main-chain B values (Å^2) against residue number for (*a*) the N-terminal lobe (1–344) and (*b*) the C-terminal lobe (345–691). The positions of the helices and β-strands are indicated by ▬ and —, respectively. The asterisks (*) show the location of the four metal-binding ligands in both lobes, along with the anion-binding Arg121 (N-lobe) and Arg465 (C-lobe).

Comparison of the final solvent model of Cu_2oxLf with that of Fe_2Lf indicates that there are 138 water molecules which can be classified as being in conserved positions, in that they are within 1.0 Å of each other. In addition, there are a further 36 water molecules in Cu_2oxLf which lie between 1.0 and 2.0 Å of water molecules in Fe_2Lf. Analysis of the individual temperature factors of these 174 'common' solvent molecules shows a good correlation between their B values, the number of hydrogen bonds associated with each molecule, and the difference in position of the pairs of solvent molecules in the two structures. These results are summarized in Tables 7(*b*) and 7(*c*). The solvent molecules whose positions differ most are generally those with the highest B values and the least number of hydrogen-bonding partners; these tend to be either at the surface or second sphere solvent. Only three of the common solvent molecules are in crystal contact regions; although the crystals are isomorphous the contacts are few in number and there is little well defined solvent in these regions.

Of the water molecules present in the interdomain clefts, about 75% are part of the core solvent model. The pattern of hydrogen bonding in the cleft of the N-lobe is virtually identical to that observed in Cu_2Lf and Fe_2Lf, while the solvent structure in the C-lobe site varies principally because of the oxalate substitution. Several water molecules appear to play a significant part in the hydrogen-bonding interactions in the two binding sites. The active-site histidines (His253 and His597) are both bonded to the copper through $N_{\varepsilon 2}$ and hydrogen bonded to a water molecule through $N_{\delta 1}$. These water molecules in turn

Table 6. *Average main-chain and side-chain temperature factors* (Å^2) *for* Cu_2oxLf

	Overall	N-lobe	C-lobe	N1	N2	C1	C2
Average main-chain *B**	38.4	40.0	36.9	41.8	38.1	38.2	35.4
Average side-chain *B*	43.7	43.9	40.3	45.7	41.7	41.5	38.7

* Averages are for protein atoms only. Residues 1–333 make up the N-lobe and residues 345–691 the C-lobe. The residues making up the four domains are as follows: N1 1–90 and 252–333; N2 91–251; C1 345–433 and 596–691; C2 434–595.

Table 7. *Analysis of solvent molecules in* Cu_2oxLf

(*a*) All solvent molecules – number of hydrogen bonds

Hydrogen bonds	0	1	2	3	4	5
No. of water molecules	19	62	85	81	53	25
Average *B* value (Å^2)	66	63	59	52	43	45

(*b*) Common solvent between Cu_2oxLf and Fe_2Lf* – number of hydrogen bonds

Hydrogen bonds	0	1	2	3	4	5
No. of water molecules	6	30	35	48	43	13
Average difference in position (Å)	1.01	0.89	0.79	0.62	0.58	0.47
Average *B* value (Å^2)	59	61	53	46	40	40

(*c*) Common solvent between Cu_2oxLf and Fe_2Lf – difference in position

	Distance between equivalent water molecules (Å)			
	0.0–0.5	0.5–1.0	1.0–1.5	1.5–2.0
No. of common solvent	79	59	19	17
Average *B* (Å^2)	41	51	56	64
Total hydrogen bonds	251	148	38	40
Average bonds/water molecule†	3.2	2.5	2.0	2.4
No. of first sphere water	75	51	13	11
Total bonds to protein atoms	164	87	18	19
Total bonds to solvent molecules	87	61	20	21

* Solvent molecules found to have a difference in position in the two structures ≤ 2.0 Å following least-squares superposition of the structures.

† Based on the total number of water molecules in each range, although 6 water molecules out of 174 do not appear to interact with any protein or solvent atoms.

interact with the interdomain solvent network *via* two or three other solvent molecules and are also hydrogen bonded to Glu80 ($O_{\epsilon 1}$) and Glu413 ($O_{\epsilon 1}$), respectively. Comparison of the Cu_2oxLf model with Fe_2Lf and Cu_2Lf indicates that these water molecules are in conserved positions (r.m.s. deviation 0.36 Å).

Another solvent molecule common to both sites and to all three structures is hydrogen bonded to the phenolic O atom of Tyr92 (Tyr453 in the C-lobe). This makes identical contacts in both lobes, namely to $O_{\delta 1}$ of Asp60 (395), $O_{\gamma 1}$ of Thr122 (466) and the carbonyl O atom of Pro251 (595), in addition to the O_{η} atom of Tyr92 (435), and their positions are closely conserved between Cu_2oxLf, and Cu_2Lf and Fe_2Lf (r.m.s. deviation 0.31 Å).

Discussion

Three important results come from the determination and refinement of the structure of Cu_2oxLf, and its comparison with those of Fe_2Lf and Cu_2Lf. Firstly, Cu^{2+}, as a bound metal ion, and oxalate, as its synergistically bound anion, are able to induce the same domain closure, from the 'open' apo form, as occurs with the physiologically important Fe^{3+} and carbonate ions (Anderson, Baker, Norris, Rumball & Baker, 1990). Secondly, oxalate is shown to coordinate to the metal in a 1,2-bidentate mode, as predicted by some spectroscopic studies, and thirdly oxalate is shown to bind preferentially in the C-lobe site of lactoferrin, when Cu^{2+} is the associated metal ion, rather than the N-lobe site, as had been expected from comparison with serum transferrin (Zweier & Aisen, 1977; Baker *et al.*, 1990).

The superpositions of Cu_2oxLf on to Fe_2Lf and Cu_2Lf show that within each of the four domains (N1, N2, C1 and C2) the polypeptide-chain conformation is essentially identical in all three structures. Moreover, the relative orientations of the domains are hardly changed by substitution of a different metal ion (Cu^{2+} for Fe^{3+}) or anion (oxalate for carbonate). In the N-lobe of Cu_2oxLf the relative orientation of the N1 and N2 domains is changed by 0.9° compared with Fe_2Lf, the cleft being more open by this extent in Cu_2oxLf. This is a very small difference, but it may be real given that in Cu_2Lf the N-lobe also appears slightly more open, by 0.6°, than in Fe_2Lf (Smith, Anderson, Baker & Baker, 1992). Superposition of the N-lobe of Cu_2oxLf on to that of Cu_2Lf shows very close correspondence, with a difference in domain orientations of only 0.2°, implying that substitution of oxalate in the C-lobe causes no disturbance to the N-lobe. In the C-lobe, the domains are again slightly more open in Cu_2oxLf than in Fe_2Lf, but the difference is equally small (only 0.7° between the orientations of the C1 and C2

domains in the two structures). The largest difference between any of the three structures is the 1.1° difference in the relative orientations of the N2 and C2 domains in Cu_2oxLf and Fe_2Lf; although still very small this is consistent with the view that these are the domains which move, relative to N1 and C1 which represent a more fixed 'core' to the molecule (Norris, Anderson & Baker, 1991).

The 'interlocking sites' model of anion binding in transferrins (Schlabach & Bates, 1975) was far sighted in predicting that synergistic anions should bind both to the metal ion and to a positively charged group on the protein but is now shown to be incorrect in the detail of anion binding. Both carboxylate groups of the oxalate ion interact with positively charged groups on the protein, one with the helix 5 N-terminus and the other with the arginine side chain (Arg465 in the C-lobe). Both also coordinate to the metal ion. This 1,2-bidentate metal coordination is as predicted from electron-spin echo envelope modulation (ESEEM) studies (Dubach, Gaffney, More, Eaton & Eaton, 1991) and suggests that the predictions of similar bidentate coordination for other anions will prove to be correct. The observed differences in chemical shifts for the two carboxylate groups, seen in NMR studies of oxalate-substituted transferrins (Bertini *et al.*, 1986), must be due to their different environments with respect to their interactions with the protein.

The 1,2-bidentate coordination of the oxalate ion is also consistent with the coordination seen in small molecule copper-oxalato species such as the $[Cu(C_2O_4)_2(H_2O)_2]^{2-}$ ion (Gleizes, Maury & Galy, 1980) and $[Cu(bipy)(C_2O_4)(OH_2)].2H_2O$ and $[Cu(bipy)(C_2O_4)].2H_2O$ (Fitzgerald *et al.*, 1982). In the first two complexes, the oxalate is symmetrically coordinated to the copper (Cu—O bond lengths between 1.93 and 1.97 Å), while the latter complex has the oxalate ion bridging between two Cu atoms, with asymmetric bidentate 1,2-coordination (the Cu—O bond lengths are 1.99 and 2.32 Å). Monodentate and 1,1-bidentate coordination can occur, the latter more common for higher dicarboxylates (Oldham, 1987). Given the more favourable chelate bite angle of the 1,2-bidentate configuration (80–90°, compared with ~60° for the 1,1-bidentate mode), and the limited space available in the binding cleft, it seems probable that other dicarboxylate anions such as malonate and maleate will bind similarly to oxalate.

The most surprising aspect of the structure of Cu_2oxLf was the discovery that oxalate had bound preferentially in the C-lobe binding site. Although chicken ovotransferrin apparently also binds oxalate more strongly in its C-terminal site, when Cu^{2+} is the associated metal ion (Zweier, 1981), human serum transferrin binds oxalate only in its N-terminal site with Cu^{2+} (Zweier & Aisen, 1977) and in human lactoferrin the N-terminal anion site appears to have rather more space available for a larger anion (Baker *et al.*, 1990).

The C-lobe anion pocket clearly has the ability to accept a larger ion. The Arg465 side chain, interacting only with the anion and a network of water molecules, is pushed aside to accommodate the oxalate. The solvent structure near the arginine is disrupted and the only other movement appears to be in the position of Tyr398, which is pushed away by the arginine side chain. The movement of the arginine also brings the $N_{\eta 1}$ atom into hydrogen-bonding distance of O_η of Tyr526, thus stabilizing its new conformation.

The preferential binding of oxalate in the C-site raises an intriguing question. If there is more room in the N-lobe binding cleft, sufficient to allow the metal ion to move by ~1 Å and the carbonate to reorient itself and still maintain a favourable hydrogen-bonding network, why does an oxalate ion not bind preferentially in this site under the conditions of the substitution experiment? (Note that these are substitution experiments, implying that it is the comparison of the relative stabilities of the carbonate and oxalate structures which matters).

Two factors may be important. Firstly, the extra space in the N-lobe site allows the Cu^{2+} ion to take up its more favoured five-coordinate, square-pyramidal geometry, through small movements of the metal, carbonate and other ligands, but without displacement of the arginine. Substitution of oxalate would probably enforce a less favourable octahedral geometry, assuming 1,2-bidentate coordination (in the C-lobe oxalate substitution would replace one octahedral geometry by another). Secondly, substitution of a larger anion in either lobe requires displacement of the arginine, and the interactions which are made by this residue may be critical. In the N-lobe of lactoferrin, with carbonate as the anion, Arg121 is stabilized by hydrogen bonds to the anion, the side chain of Ser191, and a water molecule. Displacement could disrupt these interactions without producing new ones. In the C-lobe, however, Arg465 is hydrogen bonded only to the anion and a water molecule when carbonate is bound; displacement by oxalate, however, leads to a new stabilizing hydrogen bond to Tyr526.

When considering other transferrins it must be environmental effects of this nature which account for binding differences. In serum transferrin the residue equivalent to Ser191 is glycine, which cannot hydrogen bond with the arginine. Thus, the latter may be more easily displaced and so allow oxalate to be readily substituted in the N-lobe. In the C-lobe of diferric rabbit serum transferrin, with carbonate as the anion, the binding-site arginine, Arg465, is already hydrogen bonded to Tyr512 (the equivalent of Tyr526); disruption of this interaction when arginine is moved could inhibit oxalate substitution

and explain why oxalate will only displace carbonate from the N-lobe site. Differences in domain closure may also play a part, but are more difficult to quantify.

In the case of chicken ovotransferrin, the copper geometry appears to be different at the two sites as judged by their different spectroscopic parameters ($\lambda_{max} \simeq 440$ and 430 nm for N-terminal and C-terminal monocupric complexes, respectively) (Yamamura, Hagiwara, Nakazato & Satake, 1984). These values appear to be the reverse of those for lactoferrin, for which the cupric complex of the N-terminal half molecule has a λ_{max} of 424 nm (Day, Stowell, Baker & Tweedie, 1992) compared with a λ_{max} of 434 nm for the averaged sites of Cu_2Lf (Ainscough, Brodie & Plowman, 1979). Despite this, oxalate substitution appears to occur preferentially in the C-lobe for both proteins, emphasizing that details of the three-dimensional structure are ultimately required to interpret such results.

We gratefully acknowledge the award of a Graduate Assistantship and an Assistant Lectureship (to CAS) by Massey University, and the encouragement and interest shown by Drs Andrew Brodie and Eric Ainscough, in addition to many useful discussions. Thanks are also due to Professor Peter Lindley and members of the Birkbeck College transferrin group for their helpful discussions and generous sharing of unpublished data. This work is funded by US National Institutes of Health (HD-20859), the Wellcome Trust (UK), the Health Research Council of New Zealand and New Zealand Dairy Research Institute. ENB also receives research support as an International Research Scholar of the Howard Hughes Medical Institute.

References

AINSCOUGH, E. W., BRODIE, A. M. & PLOWMAN, J. E. (1979). *Inorg. Chim. Acta*, **33**, 149–153.

AISEN, P. (1989). *Iron Carriers and Iron Proteins*, edited by T. LOEHR, pp. 355–371. New York: VCH Publishers.

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.

ARNDT, U. W. & WONACOTT, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

BAILEY, S., EVANS, R. W., GARRATT, R. C., GORINSKY, B., HASNAIN, S. S., HORSBURGH, C., JHOTI, H., LINDLEY, P. F., MYDIN, A., SARRA, R. & WATSON, J. L. (1988). *Biochemistry*, **27**, 5804–5812.

BAKER, E. N. (1988). *J. Mol. Biol.* **203**, 1071–1095.

BAKER, E. N., ANDERSON, B. F., BAKER, H. M., HARIDAS, M., NORRIS, G. E., RUMBALL, S. V. & SMITH, C. A. (1990). *Pure Appl. Chem.* **62**, 1067–1070.

BAKER, E. N., BAKER, H. M., SMITH, C. A., STEBBINS, M. R., KAHN, M., HELLSTRÖM, K. E. & HELLSTRÖM, I. (1992). *FEBS Lett.* **298**, 215–219.

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

BARTFELD, N. S. & LAW, J. H. (1990). *J. Biol. Chem.* **265**, 21684–21691.

BERTINI, I., LUCHINAT, C., MESSORI, L., SCOZZAFAVA, A., PELLACANI, G. C. & SOLA, M. (1986). *Inorg. Chem.* **25**, 1782–1786.

BROCK, J. H. (1985). *Metalloproteins*, Part II, *Metal Proteins with Non-Redox Roles*, edited by P. 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.

DEREWENDA, Z. S. & DEREWENDA, U. (1991). *Biochem. Cell Biol.* **69**, 842–851.

DUBACH, J., GAFFNEY, B. J., MORE, K., EATON, G. R. & EATON, S. S. (1991). *Biophys. J.* **59**, 1091–1100.

EKLUND, H., NORDSTROM, B., ZEPPEZAUER, E., SODERLUND, G., OHLSSON, I., BOIWE, T., SODERBERG, B.-O., TAIPA, O., BRÄNDÉN, C.-I. & AKESON, A. (1976). *J. Mol. Biol.* **102**, 27–59.

FITZGERALD, W., FOLEY, J., MCSWEENEY, D., RAY, N., SHEAHAN, D., TYAGI, S., HATHAWAY, B. & O'BRIEN, P. (1982). *J. Chem. Soc. Dalton Trans.* pp. 1117–1121.

FOTHERGILL, M. D. & FERSHT, A. R. (1991). *Biochemistry*, **30**, 5157–5164.

FUJINAGA, M. & JAMES, M. N. G. (1987). *J. Mol. Biol.* **195**, 373–396.

GLEIZES, A., MAURY, F. & GALY, J. (1980). *Inorg. Chem.* **19**, 2074–2078.

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

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Biomolecular Structure, Function, Conformation and Evolution*, Vol. 1, edited by R. SRINIVASAN, pp. 43–57. Oxford: Pergamon Press.

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

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

LINDAHL, M., LILJAS, A., HABASH, J., HARROP, S. & HELLIWELL, J. R. (1992). *Acta Cryst.* B**48**, 281–285.

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

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

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992). *Proteins*, **12**, 345–364.

NAR, H., MESSERSCHMIDT, A., HUBER, R., VAN DE KAMP, M. & CANTERS, G. W. (1992). *FEBS Lett.* **306**, 119–124.

NORRIS, G. E., ANDERSON, B. F. & BAKER, E. N. (1991). *Acta Cryst.* B**47**, 998–1004.

OLDHAM, C. (1987). *Comprehensive Coordination Chemistry*, Vol. 2, edited by G. WILKINSON, R. D. GILLARD & J. A. MCCLEVERTY, pp. 435–459. Oxford: Pergamon Press.

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

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

REES, D. C. & LIPSCOMB, W. N. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 7151–7154.

ROSE, T. M., PLOWMAN, G. D., TEPLOW, D. B., DREYER, W. J., HELLSTRÖM, K. E. & BROWN, J. P. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 1261–1265.

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

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

SCHNEIDER, G., EKLUND, H., CEDERGREN-ZEPPEZAUER, E. & ZEPPEZAUER, M. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 5289–5293.

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

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

SMITH, C. A., BAKER, H. M. & BAKER, E. N. (1991). *J. Mol. Biol.* **219**, 155–159.

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

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

YAMAMURA, T., HAGIWARA, S., NAKAZATO, K. & SATAKE, K. (1984). *Biochem. Biophys. Res. Commun.* **119**, 298–304.

ZWEIER, J. L. (1981). *J. Biol. Chem.* **255**, 2782–2789.

ZWEIER, J. L. & AISEN, P. (1977). *J. Biol. Chem.* **252**, 6090–6096.