The Mechanism of Iron Uptake by Transferrins: the Structure of an 18 kDa NII-Domain Fragment from Duck Ovotransferrin at 2.3 Å Resolution

By P. F. Lindley*

SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, Cheshire, England

M. Bajaj and R. W. Evans

Division of Biochemistry, UDMS Guy's Hospital, London SE1 9RT, England

R. C. Garratt†

Department of Crystallography, Birkbeck College, University of London, Malet St, London WC1E 7HX, England

S. S. Hasnain

SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, Cheshire, England

H. Jhoti‡ and P. Kuser

Department of Crystallography, Birkbeck College, University of London, Malet St, London WC1E 7HX, England

M. Neu

SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, Cheshire, England

K. Patel

Division of Biochemistry, UDMS Guy's Hospital, London SE1 9RT, England

R. Sarra

Department of Crystallography, Birkbeck College, University of London, Malet St, London WC1E 7HX, England

R. Strange

SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, Cheshire, England

and A. Walton

Department of Crystallography, Birkbeck College, University of London, Malet St, London WC1E 7HX, England

Abstract

The molecular structure of an iron-containing 18 kDa fragment of duck ovotransferrin, obtained by proteolysis of the intact protein, has been eluci-dated by protein crystallographic techniques at 2.3 Å resolution. This structure supports a mechanism of iron uptake in the intact protein whereby the binding of the synergistic (bi)carbonate anion is followed by binding of the metal with the lobe in the open configuration. These stages are then followed by domain closure in which the aspartic acid residue plays a further key role, by forming an interdomain hydrogen-bond interaction in addition to serving as a ligand to the iron. This essential dual role is highlighted by model building studies on the C-

* Author to whom correspondence should be addressed.
† Present address: Instituto de Física e Química de São Carlos, Departamento de Física e Ciência dos Materiais, Universidade de São Paulo, São Carlos 13569 SP, Brazil.
‡ Present address: Protein Structures Group, Glaxo Group Research, Greenford Rd, Greenford, Middlesex UB6 0HE, England.

terminal lobe of a known human variant. In this variant a mutation of a glycine by an arginine residue enables the aspartic acid to form an ion pair and reduce its effectiveness for both metal binding and domain closure. The X-ray structure of the 18 kDa fragment strongly suggests that the histidine residue present at the iron binding site of the intact protein and arising from the second interdomain connecting strand has been removed during the preparative proteolysis.

1. Introduction

The transferrins constitute an homologous family of monomeric glycoproteins whose primary members are serum transferrins (ST)* in blood plasma, lactoferrin (LF)* found in milk, other secretory fluids and leucocytes, and ovotransferrin (OT)* found in egg white (recent reviews: Huebers & Finch, 1987; Harris & Aisen, 1989; Baker & Lindley, 1992). All the transferrins sequester and solubilize iron, thereby controlling 'available' iron levels with the attendant bacteriostatic and conservation effects. In addition, serum transferrin is the principal iron-transporting protein of vertebrates delivering iron to reticulocytes by means of receptor-mediated endocytosis. Further members of the family have been reported including melanotransferrin (Rose *et al.*, 1986), found as a surface antigen in melanoma cells, but to which no clear functional role has yet been assigned, and an insect transferrin from the tobacco hornworm, *Manduca sexta* (Bartfield & Law, 1990). Melanotransferrin has been shown to bind only one ferric cation (Baker *et al.*, 1992) and sequence homology implies that the binding site for both melanotransferrin and the hornworm protein will almost certainly be in the N-lobe. On the other hand the primary sequence of *Xenopus-Laevis* transferrin (Moskaitis, Pastori & Schoenberg, 1990) indicates that the N-lobe site will probably not bind iron since two key residues involved in iron binding, an aspartic acid and an arginine, are changed to threonine and lysine, respectively. Furthermore, saxiphilin, a saxitoxin binding protein, also appears to show strong sequence identity to the transferrin family (Li & Moczydlowski, 1991). However, alterations at the iron binding sites suggest that, in this case, the iron binding function may have been completely lost and substituted instead by a saxitoxin binding site.

Protein crystallographic studies have shown that the transferrins are folded into two homologous lobes (Anderson *et al.*, 1987; Anderson, Baker, Norris, Rice & Baker, 1989; Bailey *et al.*, 1988),

each of which is comprised of two dissimilar domains with the iron binding sites located deep within the interdomain clefts. Such a structure gives ample scope for molecular flexibility and this appears to be a vital factor in the function of the transferrins. Small-angle X-ray and neutron-scattering measurements (Martel, Kim & Powell, 1980; Kilar & Simon, 1985; Vigh, Cser, Kilar & Simon, 1989; Grossmann, Neu *et al.*, 1992) have long demonstrated that a large conformational change accompanies iron uptake and release and the nature and magnitude of this change has been graphically demonstrated by the X-ray structure of human apolactoferrin (Anderson, Baker, Norris, Rumball & Baker, 1990). In the apolactoferrin structure, domain II in the N-lobe has rotated by some 54° with respect to domain I about a hinge at the rear of the iron binding site, causing the binding cleft to open wide. These observations strongly suggest a 'Venus fly-trap' mechanism for iron uptake in which the anion and metal first bind to domain II of an open lobe, followed by domain closure and complete coordination of the iron. In this paper we present evidence in support of this mechanism by describing the crystal structure of an iron-containing 18 kDa fragment of duck ovotransferrin (18 kDa dOT), corresponding to domain II in the N-terminal lobe of the intact protein. In the intact protein both sites can bind ferric iron as an octahedral high-spin complex involving two tyrosine residues, a histidine, an aspartic acid and a bidentate carbonate anion. The 18 kDa dOT fragment also binds the metal through the two tyrosine residues and the carbonate anion, but lacks the aspartic acid (from domain I); the X-ray studies also indicate that the histidine residue (from the second interdomain strand running from domain II to I) is also absent. The spectroscopic properties of this fragment are significantly different from those of the intact protein, but resemble those produced by the C-lobe of a human transferrin variant (*vide infra*). The fragment thus represents a useful model in investigating the early stages in the mechanism of iron uptake.

In addition, we highlight the crucial role played by the aspartic acid in the domain-closure process. Prior to the structural analysis of human lactoferrin, hLF, (Anderson *et al.*, 1987), biochemical and spectroscopic studies had identified the involvement of tyrosine and histidine residues in iron binding, but the presence of the aspartic acid caused some surprise. This residue plays a key dual role being involved not only in metal binding, but also in an interdomain interaction in the closed lobe iron-loaded conformation. Model-building studies on a known human C-terminal lobe variant and, more recently, small-angle scattering measurements on N-terminal lobe mutants have confirmed the importance of the aspartate residue in the later stages of the uptake

---

* The abbreviations ST, LF and OT are preceded by lower case letters indicating the species as follows: h = human, r = rabbit, c = chicken and d = duck.

mechanism involving domain closure (Grossmann, Mason *et al.*, 1992).

2. Experimental methods and results

(a) Crystallization and data collection

The 18 kDa fragment was prepared by partial digestion of duck ovotransferrin using subtilisin (Evans & Madden, 1984), and crystallized using the hanging-drop technique under the conditions described by Jhoti *et al.* (1988). The crystals are yellow–orange in colour and prismatic in habit with elongation along the c axis, and belong to the trigonal system, space group $P3_1$ with cell dimensions $a = 41.5$ (2), $c = 81.9$ (2) Å, and one molecule per asymmetric unit. X-ray intensity data were measured from one crystal, approximate size $0.5 \times 0.5 \times 2.0$ mm, on the high-intensity wiggler station 9.6, $\lambda = 0.87$ Å, on the Synchrotron Radiation Source (SRS) at SERC Daresbury Laboratory; during the data collection the beam energy was 2 GeV with an average circulating current of some 200 mA. A full set of data was collected in $4°$ oscillation steps over a φ range of $64°$ to a resolution of 2.3 Å; the crystal-to-film distance was 130 mm. The crystal was cooled to 278 K throughout the experiment and translated in the X-ray beam every $24°$ to minimize the effects of radiation damage. The data were scaled and merged to give 6733 unique reflections, 96.3% of the complete data set, from 13822 usable observations with an overall $R_{sym} = 0.044$ for 6535 reflections measured more than once; Agrovata statistics were used for assessment of data quality (Table 1).

(b) X-ray absorption spectroscopy

The solution sample used was reconstituted from a freeze-dried powder in 0.1 M NaHCO$_3$ at approximately 300 mg ml^{-1}. The solution spectra were recorded on station 8.1 at the SRS, SERC Daresbury Laboratory (Van der Hoek *et al.*, 1986). Data were collected in fluorescence mode as fluorescence excitation spectra (Hasnain, Quinn, Diakun, Wardell & Garner, 1983) using a Canberra 13-element germanium solid-state detector. Monochromatic radiation was produced using a slitless Si(111) double-crystal monochromator. The second crystal was offset from the Bragg angle and the incident intensity at the sample maintained at half maximum in order to minimize harmonic contamination. During the course of the measurements the SRS beam energy was again 2 GeV with an average circulating current of approximately 200 mA. The solution samples were contained in teflon cells and sealed with parafilm. Several scans were recorded each of about 90 min duration and the sample integrity was checked before and after the scans using EPR (elec-

Table 1. *Agrovata statistics for assessment of data quality*

D_{min} (Å)	$\langle I \rangle$	$\sigma(I)$	N_{meas}	N_{ind}	N_{tot}	$I > 3\sigma(I)$	R
9.20	144	6	108	48	55	100.0	0.029
6.88	141	8	258	121	129	99.2	0.030
5.73	135	9	358	178	181	98.3	0.036
5.01	158	12	428	204	218	99.5	0.039
4.51	206	22	446	212	224	99.6	0.036
4.14	191	13	521	246	260	99.2	0.038
3.84	178	11	614	292	297	99.7	0.034
3.60	144	11	645	306	313	98.4	0.040
3.40	137	9	680	316	320	97.8	0.036
3.23	95	7	751	352	355	97.7	0.039
3.08	72	6	802	382	387	95.3	0.043
2.96	49	4	764	366	371	94.1	0.048
2.84	40	4	892	424	426	93.4	0.055
2.74	35	3	855	403	406	90.6	0.056
2.65	31	3	954	447	448	87.7	0.064
2.57	23	3	907	426	443	83.7	0.072
2.49	22	3	865	408	448	85.0	0.081
2.42	19	3	962	451	475	79.8	0.091
2.36	14	2	1037	494	504	76.0	0.111
2.30	13	2	975	459	473	74.8	0.127
Totals			13822	6535	6733	90.1	0.044

tron paramagnetic resonance) measurements. Data reduction and analysis was performed using rapid-curve-wave single- and multiple-scattering theory (Gurman, Binstead & Ross, 1984) in the program *EXCURVE*92 (Binsted, Campbell, Gurman & Stephenson, 1992; Binsted, Strange & Hasnain, 1992) on the raw (*i.e.* unfiltered) k^3-weighted data from extended X-ray absorption fine-structure (EXAFS) analysis.

(c) Molecular replacement

(i) *Search models.* A number of search models were used in the course of the molecular-replacement calculations, based on the N-lobe domain II coordinates of rST (Bailey *et al.*, 1988), the N-lobe half-molecule (Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990), and human apo- and iron-loaded hLF (Anderson *et al.*, 1989, 1990). The search models commenced at residue Ser94 (hST numbering) and were terminated at residue His249. The precise position of the C-terminus on cleavage by subtilisin was difficult to determine, but His249 was chosen since it is the final ligand in the polypeptide chain that binds to the metal in the N-lobe. The search models were used both with and without the ferric and carbonate ions. All the models gave the same molecular-replacement solution, but the final calculations were based on a model derived from domain II of the N-lobe half-molecule of rST, in which the surface loops were deleted and residues with known sequence differences between duck ovotransferrin and the serum transferrin replaced by alanine (55% sequence homology); the *ferric cation and carbonate anion were also omitted.* Unless otherwise stated all the computer programs used are part of the *CCP*4 suite (SERC Daresbury Laboratory, 1979).

(ii) *Cross-rotation function.* The search model, approximate dimensions $34 \times 30 \times 34$ Å, was placed

in a $P1$ orthogonal cell with dimensions $60 \times 60 \times 60$ Å. Structure factors were calculated using the programme GENSFC to a resolution of 2.8 Å and normalized with ECALC; isotropic thermal parameters for all atoms were set at zero. Cross-rotation functions were calculated using ALMN, based on the fast-rotation function of Crowther (1972) as modified by E. Dodson. The resolution range 10.0–2.8 Å was used with the inner and outer Patterson radii for the sphere of integration varying between 3–6 and 17–25 Å, respectively; the optimum results were obtained with radii of 6 and 25 Å. The cross-rotation searches were performed in $2.5°$ steps in all three Eulerian angles with the highest peak located at $\alpha = 95.0$, $\beta = 35.0$ and $\gamma = 302.5°$. The height of the solution peak was 7.05 r.m.s. density compared to only 4.31 r.m.s. density for the second highest peak as indicated in Fig. 1.

(iii) *Translation function.* Translation functions were calculated in both possible space groups, $P3_1$ and $P3_2$ using the T_2 function (Crowther & Blow, 1967) as implemented in the program TFSGEN (Tickle, 1985). The model was rotated according to the rotation angles given above and structure factors generated from the model in the target cell, but with $P1$ symmetry. These structure factors were expanded into partial structure factors for each general position of $P3$ using the CCP4 suite PREPARE–COLLATE–MERGE and combined with the experimentally measured structure factors. Translation-function coefficients were calculated in both possible space groups for input into the fast Fourier transform (Ten Eyck, 1973) using all data between 10.0 and 2.8 Å. In space group $P3_1$ a unique solution was observed with height 11.15 r.m.s. compared to the second highest peak height of only 2.82 r.m.s. (Fig. 2). The translation vector ($T_x = 22.7$, $T_y = 7.21$), was orthogonalized ($T'_x = 11.46$, $T'_y = 7.21$), and applied to the model coordinates for subsequent calculations. Examinations of the unit-cell packing showed no problems with molecular overlap. In space group $P3_2$, the highest and second-highest peaks were only
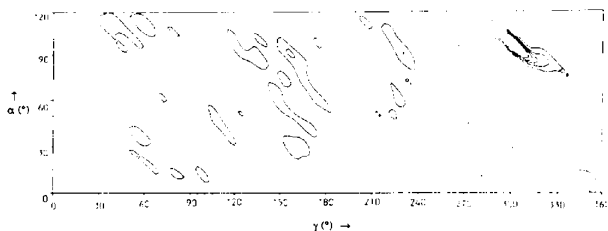
4.58 and 2.96 r.m.s., respectively; neither solution gave acceptable molecular packing and space group $P3_1$ was assumed.

(d) *Refinement*

The overall course of the refinement is summarized in Fig. 3. The molecular-replacement solution was refined using a combination of the least-squares program, RESTRAIN (Driessen, Haneef, Harris, Howlin, Khan & Moss, 1989) and computer-graphics model building. Cycles of refinement were alternated with manual corrections of the model using $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ syntheses with the FRODO software (Jones, 1978) implemented on Evans and Sutherland PS300 and PS390 Picture Systems. The electron-density syntheses were calculated using weighted Fourier coefficients, as described by Read
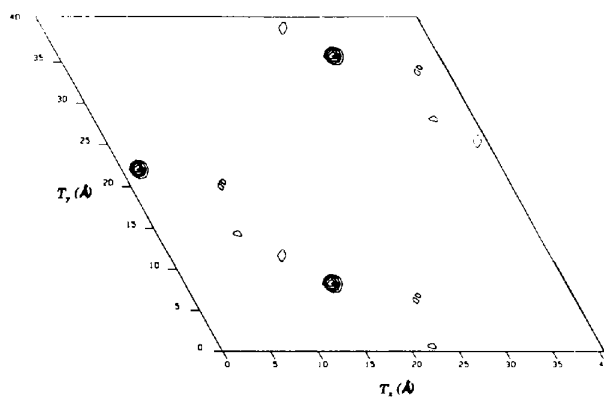

Fig. 2. Translation-function map computed with normalized structure factors between 10.0 and 2.8 Å resolution. Contour levels are 2, 4, 6, 8, 10, 11 and 12 × r.m.s. density. For space group $P3_1$ the translation function is two dimensional with three equivalent origins as shown.
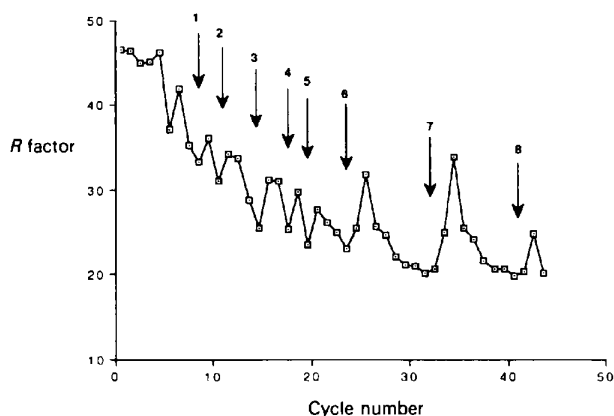

Fig. 3. Variation of the R factor during the course of the refinement. The refinement was carried out in eight stages, each consisting of a model-building session and restrained least-squares or simulated-annealing refinement.

Fig. 1. Section, $\beta = 35°$, of the cross-rotation function map (modified N-lobe rST model against 18 kDa-fragment data). For this map the parameters used were: resolution range 10.0–2.8 Å, Patterson vectors between 6.0 and 25.0 Å, step size $2.5°$, normalized structure factors used as coefficients. Contour levels are 2.5, 3.0, 4.0, 5.0 and 6.0 × r.m.s. density.

(1986), employing the program *SIMWT* (I. J. Tickle, unpublished work). Towards the end of the refinement the *X-PLOR* program (Brünger, 1988) was used in an attempt to clarify a number of ill defined surface loops.

Stage 1. Prior to commencing the refinement the *R* factor was 46% confirming the plausibility of the molecular-replacement solution. Initially the model was refined as five rigid bodies (residues 94–136, 146–162, 170–176, 186–216 and 218–249 corresponding to well defined regions of secondary structure), using the resolution range 10.0–5.0 Å, and then extending the range through 4.0 to 3.0 Å. This was followed by constrained/restrained refinement and increasing the resolution range to 2.3 Å with all the reflection data incorporated. Restrained refinement of the atomic coordinates, followed by individual isotropic temperature-factor refinement gave an *R* factor of 34.7% with a correlation coefficient of 80.4%. It is important to note that the Fe^{3+} cation and carbonate anion were not included in this initial model, but electron-density syntheses at the end of this stage clearly indicated the presence of the metal ion in a position identical to that found in the structures of intact transferrins.

Stage 2. The metal cation was inserted and residues 245–249 inclusive were deleted because of lack of electron density. Further refinement as described in stage 1, without the low-resolution rigid-body stages, led to a slight increase in the *R* factor to 35.5%, but gave a significant improvement in the electron-density maps. The anion was observed to lie in a similar position to that found in an intact transferrin molecule. The C-terminus was rebuilt up to residue 248, since there was no electron density for His249 and the chain appeared to point away from the metal binding site.

Stages 3 *and* 4. During these stages, nearly all the sequence corrections could be incorporated. There was little evidence for an arginine side chain at residue 248 (no electron density and high isotropic thermal parameters) and a recheck of the cDNA sequence, Fig. 4 (Bajaj, Faik & Evans, 1992), of the intact protein confirmed that this residue should indeed be an alanine. Solvent molecules were also included under the criteria that (*a*) they were in chemically acceptable positions, (*b*) they lay in strong density in $2|F_o| - |F_c|$ maps, and (*c*) they were at, or above, the 3σ r.m.s. level in difference electron-density syntheses. The *R* factor, 24.7% at the end of stage 3, reduced to 22.8% at the end of stage 4. Overall the electron density was well defined, but the loop at 135–146 (17 residues including 5 insertions), a highly variable region in transferrin sequences, was particularly fragmented and difficult to interpret. In addition, it was not possible to unambiguously determine the nature of the remaining ligand(s) at

the iron binding site, although the location of the cation on the surface of the fragment suggested that they could be solvent related. For the next stages of the refinement two water molecules were included at about 2.5 and 2.7 Å, respectively, from the ferric ion to give an octahedral coordination; the nearer molecule occupied a very similar position to that occupied by Asp63 (rST numbering) in the intact protein.
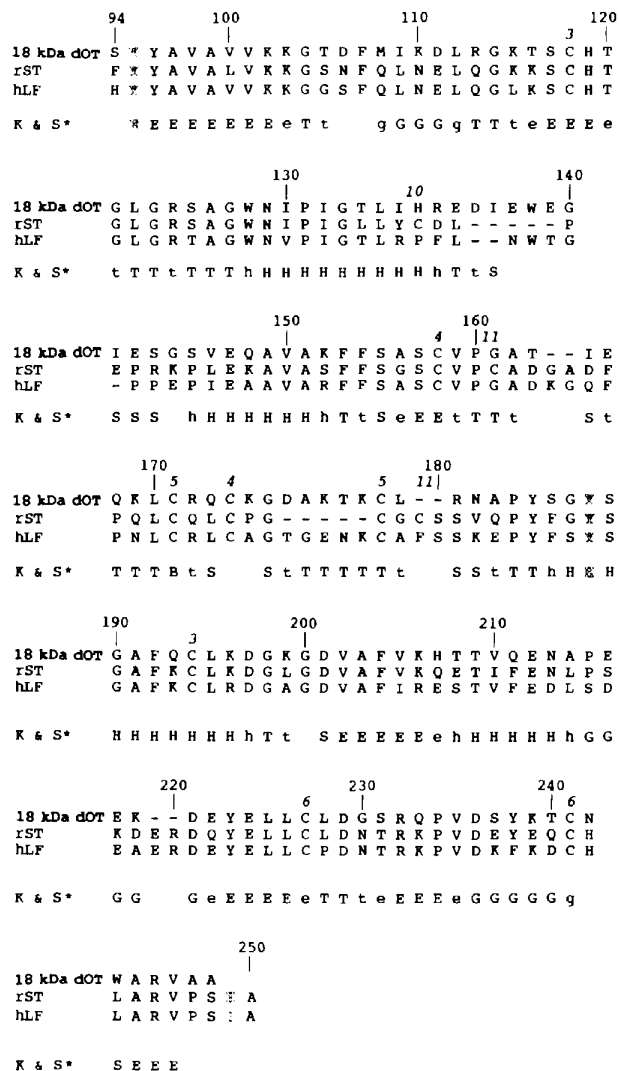

Fig. 4. Primary sequence information for the 18 kDa fragment of dOT and a comparison with the equivalent domain II in rST and hLF. Residues directly bound to iron are outlined; disulfide bridges are indicated by italics above the sequence. The major secondary elements of the 18 kDa dOT fragment are indicated beneath the primary sequences. The secondary structure assignments of the 18 kDa dOT fragment are given by the algorithm of Kabsch & Sander (1983). H = α-helix, E = β-strand, G = 3_{10} helix, T = turn and S = bend. 18 kDa dOT, Bajaj *et al.* (1992); rST, MacGillivray (1989); hLF, Metz-Boutigue *et al.* (1984).

Stages 5, 6 *and* 7. Refinement was continued with the simulated-annealing program *X-PLOR* (Brünger, Krukowski & Eriksson, 1990), and for each stage strain in the starting model was relieved by some 200 cycles of conjugate-gradient energy minimization. In all stages a slow-cooling protocol was adopted, between 3000 and 300 K for stages 5 and 7, and 2000 and 300 K for stage 6; in each case 25 K temperature decrements were used with 50 steps at 0.5 fs giving overall cooling time intervals of 2.7 ps for stages 5 and 7 and 1.7 ps for stage 6. However, in stages 6 and 7 the cooling protocol was preceded by a heating stage with an overall heating time of 1.0 ps (0.001 ps steps). Further conjugate-gradient minimization, 80 cycles, was followed in stage 6 and 7 by three alternate steps of positional and individual isotropic temperature-factor refinement with 40 cycles per step. The resulting *R* factor was 19.6% and the correctness of the structure was checked by a systematic set of omit syntheses, whereby 16 residues were omitted, cycles of restrained refinement performed and an omit electron-density map computed. With the exception of the 135–146 loop region where there was little

improvement, electron density was observed consistent with the omitted residues.

The simulated-annealing refinement moved the solvent molecules at the iron binding site further away from the metal to greater than 3.0 Å, but this was probably due to the incompleteness of the model in this region. A further synthesis and refinement omitted the Fe^{3+} ion and all atoms within a radius

(a)

(b)

Fig. 6. (a) Wilson plot for the 18 kDa fragment. The equation of the least-squares straight line is given by $y = 5.695 - 7.254x$ and the correlation coefficient is 98.7%. The overall value for $U_{iso} = 0.46$ Å^2. (b) Plots of the average U_{iso} values (Å^2) for main-chain and side-chain atoms of the 18 kDa dOT fragment. The residues which have side chains with high values of U_{iso} are given by their single amino-acid code and number. These diagrams were produced with software kindly provided by Morris *et al.* (1992).
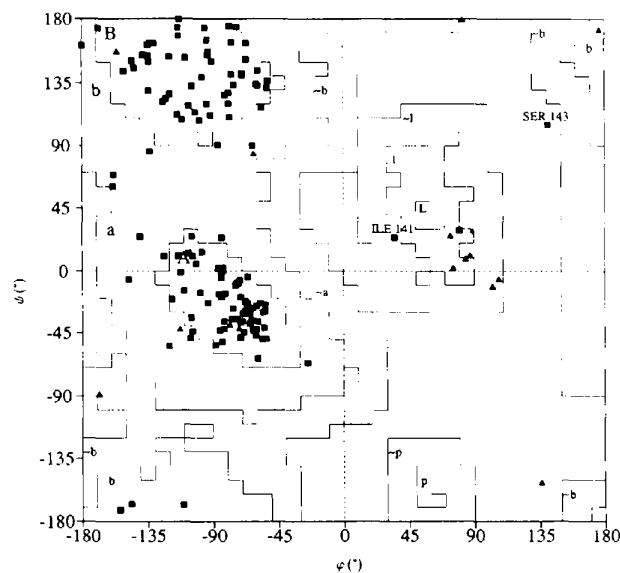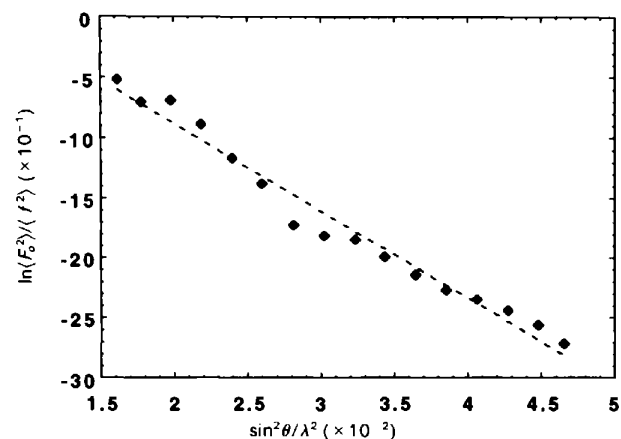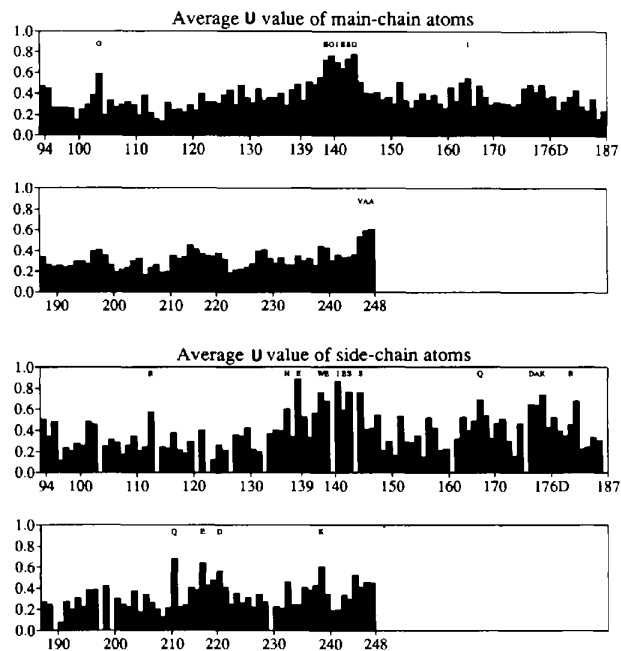

Fig. 5. Analysis of the secondary structure: a Ramachandran plot of the backbone torsion angles of the refined coordinates of the 18 kDa dOT fragment. The 15 glycine residues are marked by closed triangles (▲). The φ and ψ torsion angles were not restrained during the refinement. This diagram was produced with software kindly provided by Morris *et al.* (1992). The residue statistics are as follows:

Residues	Number	%
In total	159	
In most-favoured regions (A, B, L)	116	84.7
In additionally allowed regions (a, b, l, p)	19	13.9
Remainder	2	1.4

Both remaining residues Ile141 and Ser143 belong to the ill defined loop region, residues 135–146.

of 7.0 Å of the cation. This synthesis clearly showed density corresponding to the ferric ion, the carbonate anion and the two tyrosine residues. Additional density was again observed in the vicinity of the two solvent molecules (*i.e.* at the aspartic acid site in the intact protein), but it was still not possible to unequivocally define the nature of this density.

Stage 8, *final refinement using RESTRAIN.* The final stages of the refinement were undertaken with the *RESTRAIN* program to give an *R* factor of 19.5% for all data between 7.0 and 2.3 Å resolution and correlation coefficient 94.8%; further details are given in Table 2. The Ramachandran plot (Ramachandran & Sasisekharan, 1968), Fig. 5, is an indication of the overall correctness of the structure,

Fig. 7. A ribbon diagram of the 18 kDa dOT fragment, drawn with software written by Evans (1992). The location of the synergistic carbonate anion and the ferric cation at the N-terminus of the helix formed by residues 124–134, helix 5, is clearly shown together with tyrosine residues 95 and 188. The remaining ligand(s) to the iron have been omitted, but the exposure of the cation to solvent is apparent. The poorly defined loop region 135–146 (17 residues including 5 insertions) is at the top right-hand corner of the diagram, immediately following the 124–134 helix.
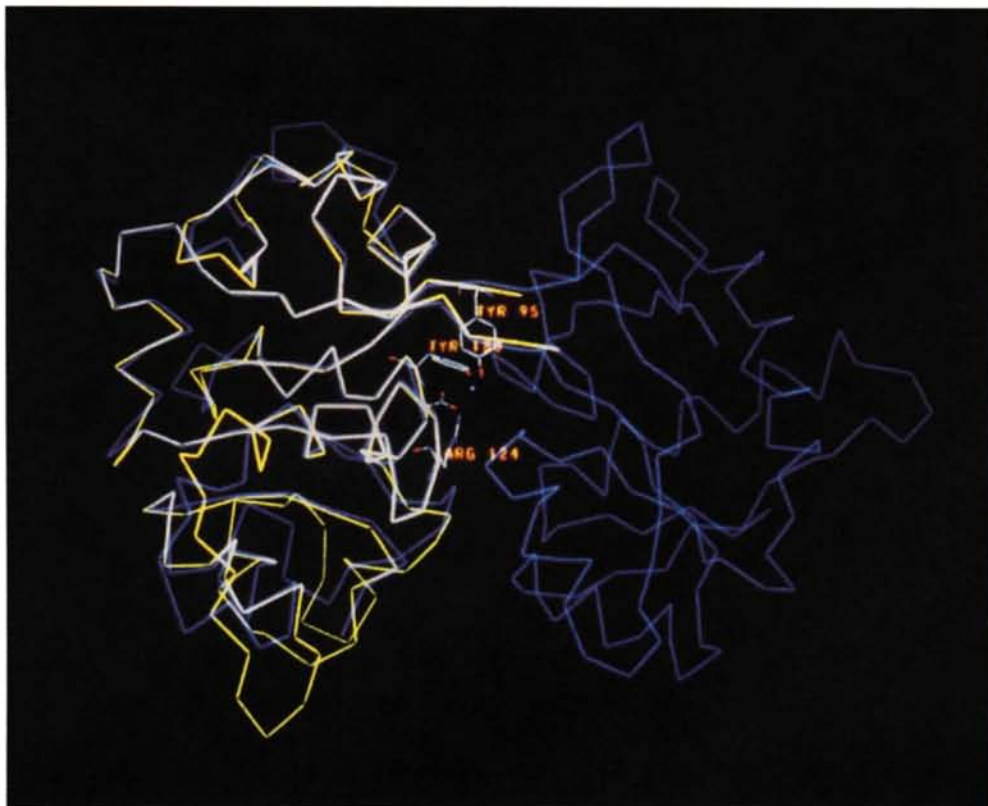

Fig. 8. Superposition of the C$^\alpha$ backbones of the 18 kDa dOT fragment (yellow) and the N-lobe half-molecule of rST (blue).

showing that 84.7% of residues lie in sterically favoured regions and 13.9% in additionally allowed regions; the φ and ψ torsion angles were not restrained during the initial and final stages of the refinement. Luzzati (1952) and Read (1986) plots enable the average positional error of the structure to be estimated as 0.30 Å, but the well defined regions of secondary structure should be better than this (< 0.2 Å) and the surface loops substantially worse.

The overall average value of the isotropic thermal parameter, 0.41 Å^2, compares favourably with that of 0.46 Å^2 predicted from a Wilson plot (Wilson, 1942), Fig. 6(*a*). Fig. 6(*b*) shows the average values of U_{iso} for the main-chain and side-chain atoms as a function of residue; the maxima in general correspond to the poorly defined loop regions. Over 100 solvent molecules were incorporated in the final model according to the criteria indicated in
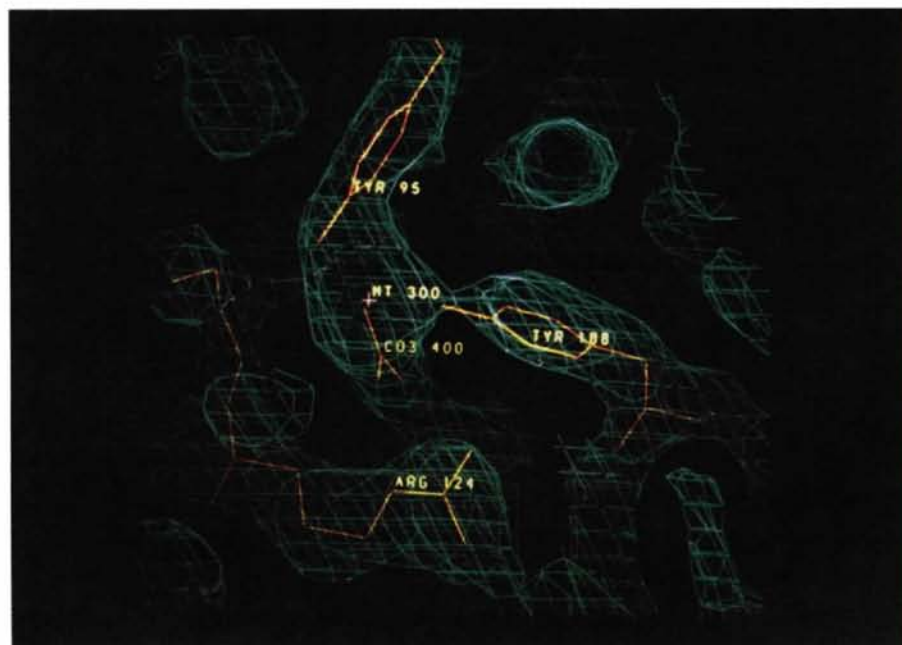
(*a*)
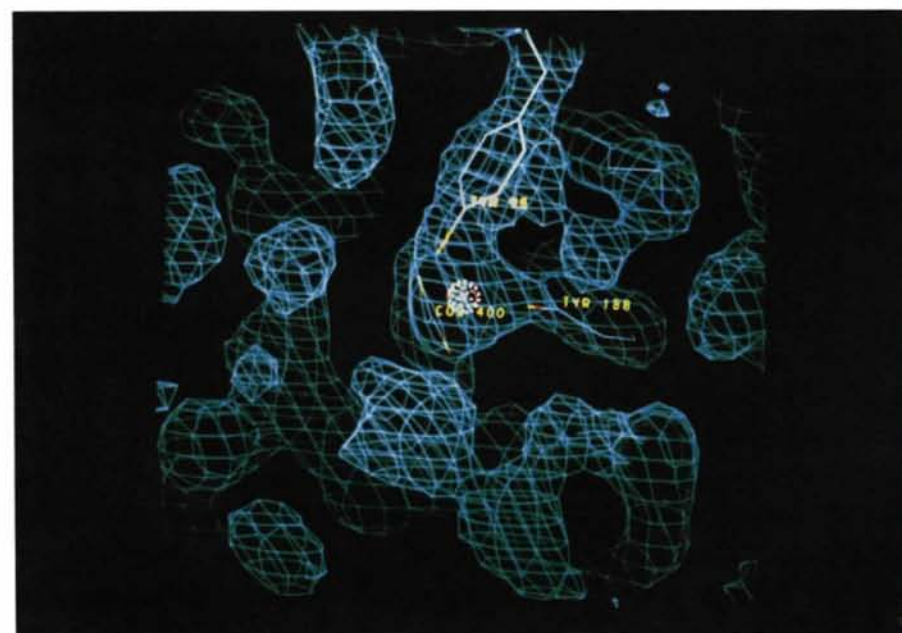
(*b*)

Fig. 9. The electron density at the iron binding site, contour level at 1 r.m.s. (*a*) Density for the ferric cation, the carbonate anion and the two tyrosine residues is clearly defined. (*b*) The remaining density at the iron binding site is shown in the lower foreground beneath the iron cation and Tyr188. On the left-hand side of the figure is the main-chain density for residues Arg124 and Ser125; the side chain of Arg124 extends behind the additional density, see (*a*). The tripeptide Asp.Gly.Gly gives a reasonable fit to the additional density, but an unambiguous interpretation must await the collection of higher resolution data.

refinement stages 4 and 5; Table 3 gives an analysis of the solvent molecules in terms of their isotropic thermal parameters.

3. Results and discussion

(a) Overall comparison with the rST NII domain

The structure of the 18 kDa N-terminal domain fragment of duck ovotransferrin is represented schematically as a ribbon diagram in Fig. 7. It adopts a very similar conformation to domain NII observed in both the intact rST protein (Bailey *et al.*, 1988) and the N-lobe half-molecule (Sarra *et al.*, 1990) as would be predicted by the relative ease with which the structure solution was found by molecular-replacement methods.* Thus, some 75 C^{α} atoms representing well defined secondary structure (out of 179) can be superimposed, Fig. 8, with an r.m.s. deviation of 0.60 Å (maximum displacement of 1.41 Å); the fit is clearly very good in the vicinity of the iron site and elsewhere the major variations occur in the surface-loop regions. However, although the N-terminus at residue Ser94 is well defined, the C-terminus appears to extend only to residue Ala248 and there is no indication of the presence of His249.

(b) The iron binding site

Fig. 9 shows the electron density at the iron binding site in the 18 kDa fragment of dOT. This density can be clearly interpreted in terms of two tyrosine residues and a carbonate anion distributed about the ferric cation in an almost identical manner to that found in the intact lobe. The r.m.s. fit is 0.47 Å for 77 atoms common to the 18 kDa dOT fragment and the N-terminal lobe half-molecule of rST (Sarra *et al.*, 1990) (77 common atoms: Fe^{3+}, CO_3^{2-}, Tyr95, Tyr188, Thr120 and residues 121–127 inclusive), with a maximum deviation of 1.22 Å for the C^{β} of Arg124. The carbonate anion is bidentate to the cation and also forms an intricate network of hydrogen bonds with Arg124, Thr120 and the main-chain amino groups of residues Ala126 and Gly127. The carbonate anion and the two tyrosine residues provide four ligands to the ferric cation; Asp63 and His249 which complete the octahedral coordination in the intact lobe are not present. If the iron environment is to remain similar to that in the intact lobe,

---

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1OVB, R1OVBSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37073 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 31 December 1993.

Table 2. *Refinement parameters, including target and actual r.m.s. deviations from target values after the last cycle of refinement, and analysis of R factor versus resolution*

(a) Refinement parameters

	18 kDa dOT
Resolution range (Å)	7.00–2.30
No. of reflections (no σ cut-off)	6555
No. of restraints	3193
No. of protein atoms	1218
No. of Fe³⁺ cations	1
No. of (bi)carbonate anions	1
No. of water molecules	106
Total No. of parameters refined*	5342
Observation/parameter ratio	1.82
R factor†	19.5
Weighted R factor‡	19.8

Stereochemistry	Target r.m.s.	R.m.s.
Distances <2.12 Å (bond lengths)	0.025	0.008
Distances, 2.12–2.62 Å (≈ 3.3° for bond angles)	0.050	0.011
Distances >2.62 Å (C^{α}—C^{α} virtual bonds)	0.070	0.016
Peptide planarity (Å)	0.020	0.003
All other planes (Å)	0.0125	0.001
Chirality	0.025	0.013

(b) Analysis of R factor *versus* resolution

Resolution range (Å)	R factor	No. of reflections
7.00–3.96	14.4	1059
3.96–3.24	14.8	1096
3.24–2.86	20.0	1111
2.86–2.62	24.1	1102
2.62–2.44	28.5	1089
2.44–2.30	33.8	1098

* Overall scale and temperature factor, three positional, and one isotropic thermal parameters per atom.

† R factor $= (\sum ||F_o| - |F_c||)/\sum |F_o|$.

‡ Weighted R factor $= [(\sum w|F_o - F_c|^2)/\sum w|F_o|^2]^{1/2}$ where all data were given the same weight, $w = 2.5 \times 10^{-5}$. An analysis of the weighting scheme showed approximately constant averages of $\sum (w_i F_o - F_c|^2)$ in terms of batches of increasing $|F_o|$ and $\sin\theta/\lambda$.

Table 3. *Distribution of U_{iso} for solvent molecules*

No. of waters	U_{iso} range (Å²)
2	0.11–0.20
9	0.21–0.30
15	0.31–0.40
20	0.41–0.50
27	0.51–0.60
15	0.61–0.70
10	0.71–0.80
8	0.81–0.90

any remaining ligand(s) must not only satisfy the charge imbalance due to the absence of the aspartic acid (the histidine will be neutral in the intact protein), but also any geometrical requirements. In the early stages of the refinement, electron-density syntheses appeared to indicate the possibility of a solvent molecule liganded to the iron at a distance of 2.0 Å, almost exactly in the position which would be occupied by the carboxyl O atom of the aspartate; a second less well defined solvent molecule at a longer distance could also be discerned near the histidine position. As the refinement progressed it became clear that this model was an oversimplification and that the residual density could accommodate a larger moiety. Many different models were explored including networks of water molecules, 2,4-methylpentane-diol used in the crystallization medium (and various

isomers), various anions such as (bi)carbonate and phosphate, and fragments of peptide chain which may have arisen from the proteolytic preparation. In particular, the possibility that the density corresponded to His249 was investigated, but it proved impossible to fit in the side chain to the density and simultaneously link this residue to Ala248 which lay in well defined density remote from the iron binding site. Attempts were made to fit a variety of tri- and tetrapeptides with all models having a carboxylate group binding to the ferric cation and maximizing the hydrogen bonding between the peptide and residues and solvent molecules in the vicinity of the iron binding site. One of the more feasible of these models is the Asp.Gly.Gly tripeptide from the sequence of the intact dOT molecule. None of the models investigated refined satisfactorily either in terms of electron density or acceptable isotropic temperature factors. However, the density can be better explained in terms of a peptide fragment rather than solvent structure and the possibility of more than one pep-

tide occupying statistically disordered positions cannot be discounted. It appears that the elucidation of the correct model must await the collection of even higher quality and higher resolution diffraction data, possibly using cryogenic techniques.

The X-ray absorption spectra of the 18 kDa dOT NII fragment appear to be consistent with the X-ray structural information. Simulations based only on a four-coordinate model of the iron with two tyrosine residues and a bidentate carbonate anion do not fully explain either the spectrum or its transform and the preferred model is a five-coordinate model achieved with the addition of a low-Z ligand at around 2.0 Å, Fig. 10(a). Using this model a difference spectrum, Fig. 10(b), produces a maximum in the transform at around 3.5–3.8 Å in close agreement with the X-ray density. Insertion of several low-Z atoms at around this distance certainly improves the agreement between the observed and simulated absorption spectra and their transforms, but does not enable an unambiguous model to be chosen.
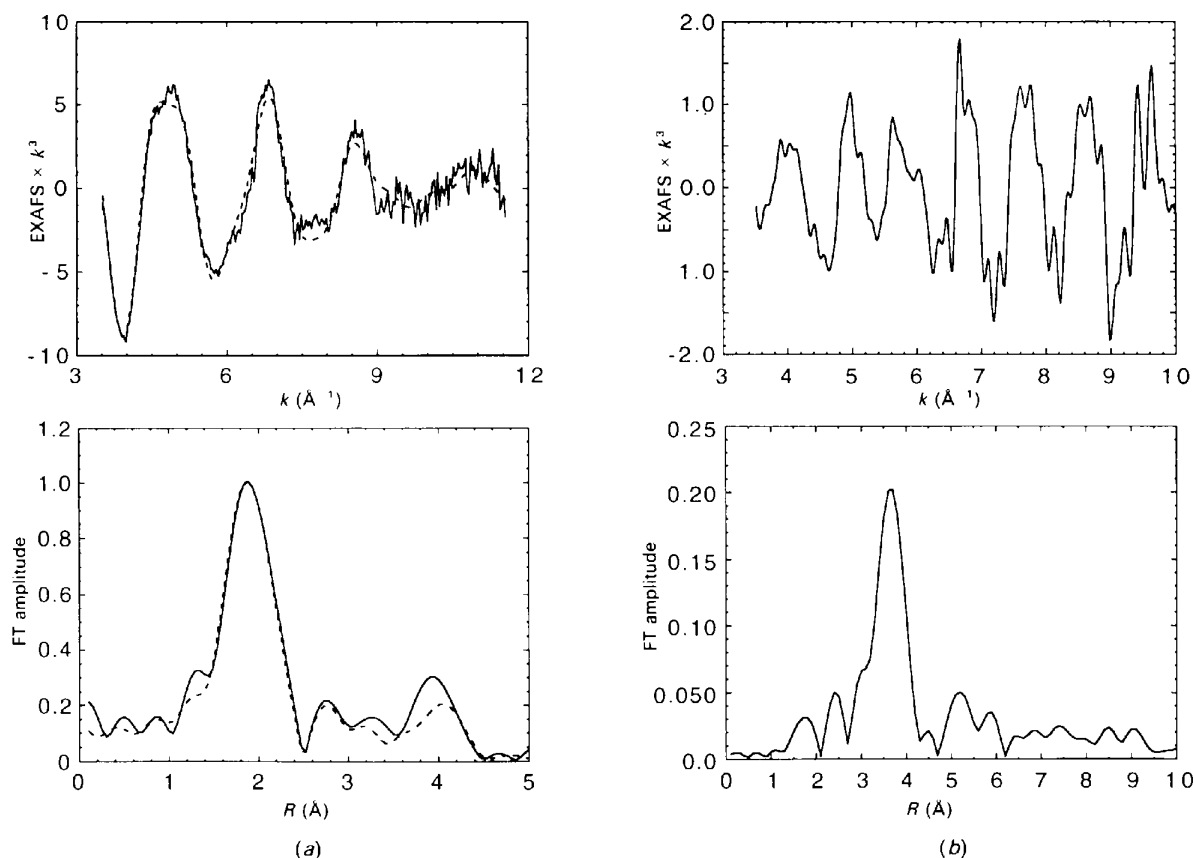

Fig. 10. Extended X-ray absorption fine-structure analysis of the 18 kDa fragment. (a) The Fe K-edge solution and its transform (lower panel). The dotted lines represent a simulation of the spectra and transform using a model based on the ferric cation, carbonate anion, two tyrosine residues, and a further low-Z ligand (assumed to be oxygen), all at about 2.0 Å from the iron. The fit is relatively poor at high values of k, $k = 6$–9 Å^{-1}. (b) Difference EXAFS spectrum and transform after the removal of the simulation described in (a). The peak at 3.7 Å can be ascribed to several low-Z ligands in the vicinity of the iron; see proposed model in text.

(c) Ser125, Lys206 and Glu212

In the intact iron-loaded transferrin lobe Ser125 and Glu212 are involved in interdomain contacts which keep the lobe in the closed conformation. Ser125 forms an interdomain hydrogen bond with Asp63 (also one of the iron binding residues) in the intact lobe, but in the present structure it interacts with two solvent molecules, at 2.7 and 3.1 Å respectively. Glu212 is a key residue with respect to interdomain contacts in the N-lobes of the transferrins. In rST it forms a hydrogen bond with Ser298, but in hLF (Glu216 in hLF numbering) this is replaced by a salt bridge with Lys296 (302). This difference between the two proteins may well account for the lower acid stability of the N-lobe of the serum proteins with respect to the lactoferrins (Baker & Lindley, 1992; Sarra, 1992). In the 18 kDa dOT fragment Glu212 hydrogen bonds to the main-chain amino groups of Asp107 and Phe106 in an adjacent symmetry-related molecule ($O^\epsilon 1 \cdots N = 3.3$ and 3.0 Å, respectively), and to two water molecules, both at 2.7 Å.

In the rST Lys206 is directed to within 3.2 Å of Lys296 from domain I and this pair of lysine residues, close to the iron site in the interdomain hinge region, may present a secondary anion binding which could account for the stimulation of iron release by some non-synergistic anions. In the 18 kDa dOT NII fragment, Lys206 forms an ion pair with Asp106 from a symmetry-related molecule ($N^\epsilon \cdots O^\delta$ distances of 2.6 and 3.4 Å) and also hydrogen bonds to a water molecule at 2.7 Å.

(d) Mechanism of iron uptake

The 18 kDa dOT NII-domain structure is clearly consistent with an iron uptake mechanism, shown schematically in Fig. 11, in which the initial stages are binding of the synergistic anion, followed by the cation, to domain II with the transferrin lobe in the opened conformation. The positive macro-dipole at the N-terminus of helix 5 [see Bailey et al. (1988) for helix numbering], which also includes Arg124, is orientated towards the metal site and is hardly an ideal environment to receive a ferric cation (Fig. 7). However, this configuration readily explains the requirement for the synergistic carbonate anion which effectively neutralizes the positive charges and confers site specificity for binding ferric ions. In the intact protein the cation charge is then balanced by the two tyrosinate residues, Tyr188 on the N-terminus of helix 7 and Tyr95 on the first interdomain-connecting strand, before the domain-closure stage introduces the aspartic acid Asp63 (Asp392 in the C-terminal lobe) on the N-terminus of helix 3 in domain I. The environment of the ferric cation is completed by the neutral His249 residue

sited on the second interdomain-connecting strand as the peptide chain traverses back from domain II to domain I. This mechanism is consistent with experiments on iron-binding kinetics which show a conformational change to be the rate-limiting step in iron binding (Cowart, Kojima & Bates, 1982; Cowart, Swope, Loh, Chasteen & Bates, 1986). The EPR spectrum of the intermediate differs from those of the N- and C-terminal half-molecules of cOT (Evans & Madden, 1984), and its resemblence to that of the 18 kDa fragment (Evans & Patel, 1992) lends further credence to the mechanism proposed. Clearly, in the later stages of the uptake mechanism, the aspartate plays an important dual role since not only does one of the carboxylate oxygen atoms bind to the iron, but the second also forms an interdomain hydrogen-bond through the O^γ of residue Ser125 (residue 125 is a semi-conserved serine or threonine in the transferrins). Indeed, small-angle X-ray solution scattering measurements of N-lobe hST mutant in which the aspartate is changed to serine, indicate that domain closure is incomplete (Grossmann, Mason et al., 1992).
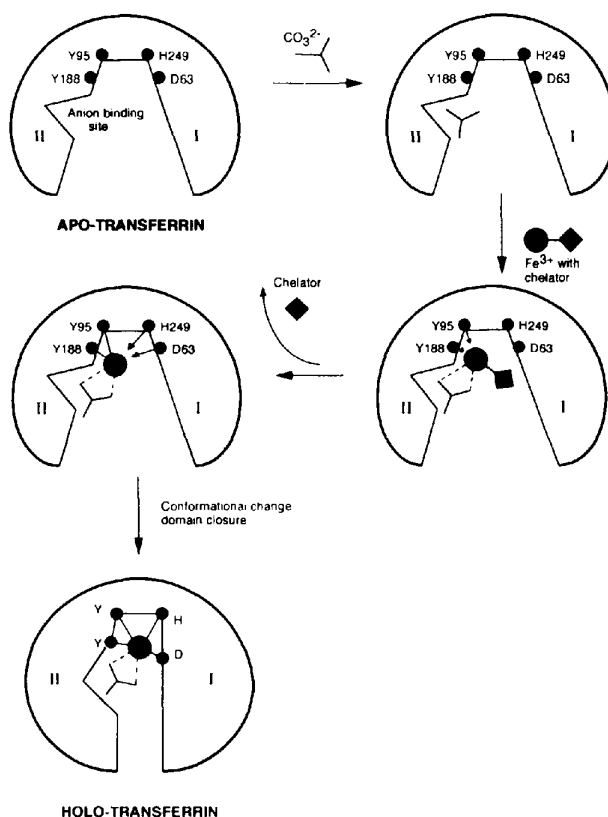

Fig. 11. A schematic representation of the mechanism of iron uptake by a transferrin lobe after Baker et al. (1991). The two tyrosine residues involved in iron binding, Tyr95 and Tyr188, are not shown.

(*e*) *Variant of human transferrin*

A variant of human serum transferrin has been identified which possesses a normal N-lobe, but a C-lobe in which a glycine, two residues after the aspartic acid (Asp392) involved in iron binding, is substituted by an arginine (Evans *et al.*, 1988). This modified C-lobe is capable of binding iron, but has several unusual spectroscopic properties (Evans, Williams & Moreton, 1982). In particular, both its visible and ESR spectra resemble that of the 18 kDa dOT fragment in which the aspartic acid is absent. Further recent small-angle X-ray solution scattering measurements indicate that the C-lobe may not be fully closed (Grossmann, 1992).

In the native protein, residues Asp392, Gly393, Gly394 and Phe395 form a type I turn (Schulz & Schirmer, 1979), as do the homologous residues in the N-terminal lobe. One of the carboxylate oxygen atoms of the Asp is bound to the iron, whereas the other forms hydrogen bonds to Thr457 and the main-chain NH of Gly394. However, Wilmot & Thornton (1988) have shown that type I turns with Asp at position *i* and Arg at position *i* + 2 are relatively favoured because of specific side-chain interactions. A feasible model for the human variant C-lobe is therefore a type I turn in which the Asp carboxylate group also forms an ion pair with the guanidinium moiety of the substitute Arg394 as shown in Fig. 12. Such a structure would severely reduce the propensity of the Asp either to bind to the iron, thereby giving rise to spectroscopic properties similar to the 18 kDa dOT fragment, or to form an interdomain hydrogen bond, leaving the domains in an open configuration in agreement with the solution scattering measurements. The variant also shows a tenfold reduction in binding constant for the receptor complex compared with the native transferrin (Young *et al.*, 1984), suggesting that receptor recognition involves the iron-loaded closed-domain structure.

(*f*) *Future studies*

In the transferrins the polypeptide chain is folded into two homologous lobes, each consisting of two dissimilar domains; domain II is a contiguous portion of the chain, whereas I is comprised of two non-contiguous regions. Domain II fragments have been isolated from both the N-lobes of dOT and hLF, but so far no domain II fragment from a C-lobe has been reported. This may be due to the presence of a semi-conserved disulfide bridge in the C-lobe (number 7) which spans the two domains and may well have an influence on the difference in iron binding between the two lobes. As this bridge can be selectively reduced in hen ovotransferrin (Williams, Moreton & Goodearl, 1985), it may be possible to isolate a domain II fragment from proteolysis of the partially reduced protein.

Of the four protein ligands at the iron binding site in the transferrins, the two tyrosine residues (95 and 188 in the N-lobe) play a fundamental role in the initial stages of iron binding, whereas Asp63 appears to facilitate domain closure as well as providing a negatively charged ligand to the iron. The role of His249 in the mechanism of iron binding is less well defined, although its presence as a neutral ligand completing the octahedral environment of the iron is chemically sensible. However, site-directed mutagenesis techniques now enable these ideas to be tested further and a number of experiments on the N-lobe of hST are in progress whereby the tyrosine and histidine residues are being systematically changed and interchanged. These experiments will, hopefully, support the mechanism of iron uptake suggested by the structures of apolactoferrin, the iron-loaded transferrins and the 18 kDa dOT fragment.
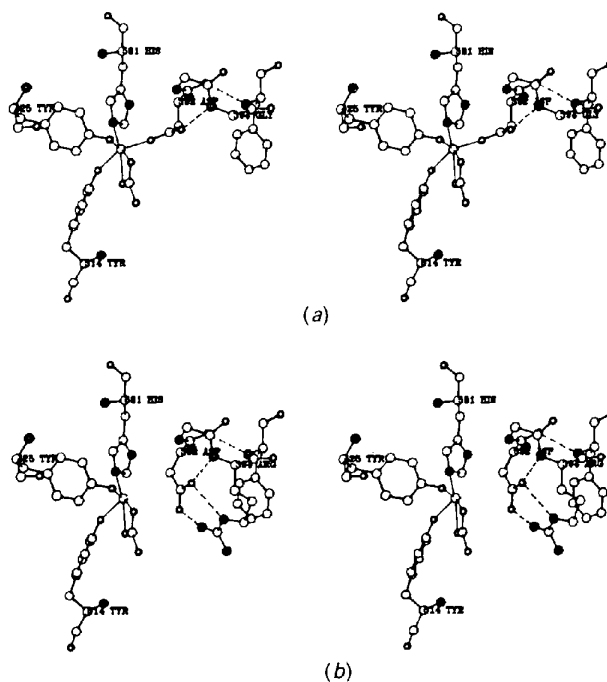
(*a*)

(*b*)

Fig. 12. The hypothetical C-terminal iron binding site in the human variant transferrin. (*a*) A stereoview of the C-lobe iron binding site in the native serum protein. (*b*) A stereoview of the C-lobe iron binding site in the variant. Arg394 forms a twin-contact ion pair with Asp392 drawing it away from the iron. Asp392 retains its hydrogen bond with the amide of residue 394 and the type I turn remains intact.

References

ANDERSON, B. F., BAKER, H. M., DODSON, E. J., NORRIS, G. E., RUMBALL, S. V., WATERS, J. M. & BAKER, E. N. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 1768–1774.

ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RICE, D. W. & BAKER, E. N. (1989). *J. Mol. Biol.* **209**, 711–734.

ANDERSON, B. F., BAKER, H. M., NORRIS, G. E., RUMBALL, S. V. & BAKER, E. N. (1990). *Nature (London)*, **344**, 784–787.

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

BAJAJ, M., FAIK, P. & EVANS, R. W. (1992). In preparation.

BAKER, E. N., ANDERSON, B. F., BAKER, H. M., HARIDAS, M., JAMESON, G. B., NORRIS, G. E., RUMBALL, S. V. & SMITH, C. A. (1991). *Int. J. Biol. Macromol.* **13**, 121–129.

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

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

BINSTED, N., CAMPBELL, J. W., GURMAN, S. J. & STEPHENSON, P. C. (1992). *EXCURV*92. SERC Daresbury Laboratory, Warrington WA4 4AD, England.

BINSTED, N., STRANGE, R. W. & HASNAIN, S. S. (1992). *Biochemistry*. In the press.

BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

BRÜNGER, A. T., KRUKOWSKI, A. & ERIKSSON, J. (1990). *Acta Cryst.* A**46**, 585–593.

COWART, R., KOJIMA, N. & BATES, G. (1982). *J. Biol. Chem.* **257**, 7560–7565.

COWART, R., SWOPE, S., LOH, T. T., CHASTEEN, N. D. & BATES, G. W. (1986). *J. Biol. Chem.* **261**, 4607–4614.

CROWTHER, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. ROSSMAN, pp. 173–178. New York: Gordon & Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

DRIESSEN, H. P. C., HANEEF, I., HARRIS, G. W., HOWLIN, B., KHAN, G. & MOSS, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.

EVANS, R. W. & MADDEN, A. D. (1984). *Biochem. Soc. Trans.* **12**, 661–662.

EVANS, R. W., MEILAK, A., AITKEN, A., PATEL, K. J., WONG, C., GARRATT, R. C. & CHITNAVIS, B. (1988). *Biochem. Soc. Trans.* **16**, 834–835.

EVANS, R. W. & PATEL, K. (1992). Unpublished results.

EVANS, R. W., WILLIAMS, J. & MORETON, K. (1982). *Biochem. J.* **201**, 19–26.

EVANS, S. V. (1992). *J. Mol. Graphics*. In the press.

GROSSMANN, J. G. (1992). Unpublished results.

GROSSMANN, J. G., MASON, A. B., WOODWORTH, R. C., NEU, M., LINDLEY, P. F. & HASNAIN, S. S. (1992). In preparation.

GROSSMANN, J. G., NEU, M., PANTOS, E., SCHWAB, F. J., EVANS, R. W., TOWNES-ANDREWS, E., LINDLEY, P. F., APPEL, H., THEIS, W.-G. & HASNAIN, S. S. (1992). *J. Mol. Biol.* **225**, 811–819.

GURMAN, S. J., BINSTEAD, N. & ROSS, I. (1984). *J. Solid State Phys.* C**17**, 143–151.

HARRIS, D. C. & AISEN, P. (1989). *Physical Biochemistry of the Transferrins*. In *Physical Bioinorganic Chemistry*, Vol. 5, *Iron Carriers and Iron Proteins*, edited by T. LOEHR, pp. 239–351. New York: VCH Publishers.

HASNAIN, S. S., QUINN, P. D., DIAKUN, G. P., WARDELL, E. M. & GARNER, C. D. (1983). *J. Phys. E*, **17**, 40–43.

HUEBERS, H. A. & FINCH, C. A. (1987). *Physiol. Rev.* **67**, 520–582.

JHOTI, H., GORINSKY, B., GARRATT, R. C., LINDLEY, P. F., WALTON, A. R. & EVANS, R. W. (1988). *J. Mol. Biol.* **200**, 423–425.

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

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KILAR, F. & SIMON, I. (1985). *Biophys. J.* **48**, 799–802.

LI, Y. & MOCZYDLOWSKI, E. (1991). *J. Biol. Chem.* **266**, 15481–15487.

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

MACGILLIVRAY, R. T. A. (1989). Personal communication.

MARTEL, P., KIM, S. M. & POWELL, B. M. (1980). *Biophys. J.* **31**, 371–380.

METZ-BOUTIGUE, M.-H., JOLLES, J., MAZURIER, J., SCHOENTGEN, F., LEGRAND, D., SPIK, G., MONTREUIL, J. & JOLLES, P. (1984). *Eur. J. Biochem.* **145**, 659–676.

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

MOSKAITIS, J. E., PASTORI, R. L. & SCHOENBERG, D. R. (1990). *Nucleic Acids Res.* **18**, 6135.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

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

ROSE, T. M., PLOWMAN, G. D., TEPLOW, D. B., DREYER, W. J., HELLSTROM, K. E. & BROWN, J. P. (1986). *Biochemistry*, **83**, 1261–1265.

SARRA, R. (1992). PhD thesis, Univ. of London, England.

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

SCHULTZ, G. E. & SCHIRMER, R. H. (1979). *Principles of Protein Structure*. New York: Springer-Verlag.

SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, England.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

TICKLE, I. J. (1985). *Molecular Replacement*, edited by P. A. MACHIN, pp. 22–26. SERC Daresbury Laboratory, Warrington WA4 4AD, England.

VAN DER HOEK, M. J., WERNER, W., VAN ZYLEN, P., DOBSON, B. R., HASNAIN, S. S., WORGAN, J. S. & LUICKX, G. (1986). *Nucl. Instrum. Methods Phys. Res.* A**246**, 380–384.

VIGH, R., CSER, L., KILAR, F. & SIMON, I. (1989). *Arch. Biochem. Biophys.* **275**, 181–184.

WILLIAMS, J., MORETON, K. & GOODEARL, A. D. J. (1985). *Biochem. J.* **228**, 661–665.

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

WILSON, A. J. C. (1942). *Nature (London)*, **150**, 152.

YOUNG, S. P., BOMFORD, A., MADDEN, A. D., GARRATT, R. C., WILLIAMS, R. & EVANS, R. W. (1984). *Br. J. Haematol.* **56**, 581–587.