Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The crystal and molecular structures of diferric porcine and rabbit serum transferrins at resolutions of 2.15 and 2.60 Å, respectively

D. R. Hall,[a]† J. M. Hadden,[a]‡ G. A. Leonard,[b]§ S. Bailey,[a]¶ M. Neu,[a]†† M. Winn[a] and P. F. Lindley[a]*§

[a]CLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England, and [b]Department of Chemistry, University of Manchester, Manchester M13 9PL, England

† Molecular Enzymology Group, Imperial Cancer Research Fund, Clare Hall, South Mimms, Potters Bar EN6 3LD, England.
‡ Astbury Centre for Structural Molecular Biology, School of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, England.
§ European Synchrotron Radiation Facility, 6 Rue Horowitz, F-38043 Grenoble CEDEX, France.
¶ Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mailstop 6-2132, Berkeley, CA 94720, USA.
†† Glaxo-Wellcome Medicines, Research Centre, Gunnels Wood Road, Stevenage, Herts SG1 2NY, England.

Correspondence e-mail: lindley@esrf.fr

Received 9 July 2001
Accepted 15 October 2001

PDB References: pST, 1h76; rST, 1jnf.

The serum transferrins are monomeric proteins with a molecular mass of around 80 kDa and are responsible for the transport of iron in vertebrates. The three-dimensional structures of diferric porcine and rabbit serum transferrin have been refined against X-ray diffraction data extending to 2.15 and 2.60 Å, respectively. Data for both proteins were collected using synchrotron radiation at temperatures of 277 K. The porcine protein crystallizes in the space group $C2$, with unit-cell parameters a = 223.8, b = 44.9, c = 78.9 Å, β = 105.4° with one molecule in the asymmetric unit. The structure was solved by molecular-replacement methods using rabbit serum transferrin as the search model. The structure was refined using *REFMAC*, with a final residual of 13.8% ($R_{\rm free}$ = 18.2% for a 5% data sample) for all data to 2.15 Å. The final model comprises 5254 protein atoms, two Fe^{3+} cations and two CO_3^{2-} anions, one *N*-acetyl glucosamine moiety and 494 water molecules. The rabbit protein crystallizes in space group $P4_32_12$, with unit-cell parameters a = 127.2, c = 144.9 Å and one molecule per asymmetric unit. The structure was solved using the method of multiple isomorphous replacement and refined using *REFMAC* to give a final residual of 18.6% ($R_{\rm free}$ = 22.2% for a 5% data sample) for all data to 2.60 Å. The final model comprises 5216 protein atoms, two Fe^{3+} cations and two CO_3^{2-} anions, a Cl^- anion and 206 solvent molecules; there is no clear indication of the carbohydrate moiety attached to Asn490 (rabbit serum numbering). Both molecules adopt a bilobal structure typical for members of the transferrin family. Each of the structurally homologous lobes contains two dissimilar domains with a single iron-binding site buried within the interdomain cleft. The porcine serum protein lacks an interdomain disulfide bridge close to the connecting peptide between the lobes, but this seems to have little effect on the overall orientation of the lobes. The N-lobes of both proteins possess lysine residues, one from each of the two domains, that lie in close proximity to one another to form the so-called dilysine trigger. The more acid-labile release of iron from serum transferrins than from lactoferrins is discussed.

1. Introduction

The cells of higher organisms have a requirement for iron and to this end nature has devised an elegant transport and uptake system. For transport, the element iron is bound in its ferric (Fe^{3+}) form in a pH-dependent manner by the protein transferrin. Transferrin binds two Fe atoms, each in conjunction with a carbonate anion. At the cell surface (pH 7.4) the iron-loaded transferrin molecule is bound by the transferrin receptor dimer and is internalized *via* receptor-mediated

endocytosis. Iron is released from the transferrin molecule, again in a pH-dependent manner, as the pH of the vesicle is lowered from 7.4 to 5.5 with transferrin remaining bound to the receptor molecule. The iron is then transported out of the vesicle and into the cell and the vesicle returns to the surface of the cell where apo-transferrin is released into the extracellular medium (pH 7.4).

There are three major classes of transferrins: serum transferrins, lactoferrins and ovotransferrins. They are all monomeric glycoproteins with an approximate molecular mass of 80 kDa which bind two ferric ions per molecule with high affinity and have high sequence similarity. The serum transferrins (*x*ST, where *x* represents a species, *e.g.* r for rabbit, p for porcine and h for human) are responsible for the transportation of iron in serum and, as indicated above, delivery of iron to cells *via* receptor-mediated endocytosis (Jandl *et al.*, 1959). Ovotransferrins (where hOT is hen ovotransferrin) are found in avian egg whites and are identical to their serum counterpart except for their attached carbohydrates (Thibodeau *et al.*, 1978), but cannot bind to mammalian receptors. Lactoferrins (where bLF and hLF are bovine and human lactoferrins, respectively) are found in numerous extracellular fluids and in the specific granules of polymorphonuclear lymphocytes and are known to be multifunctional, acting as antimicrobial agents (Arnold *et al.*, 1977), antifungal agents (Singh *et al.*, 1998) and sequence-specific DNA-binding moieties (He & Furmanski, 1995). Sequence homology between the different transferrins is high, with an identity of 60% between hST and hLF, 79% between hST and rST, 71% between hST and pST and 73% between rST and pST. Transferrins also have twofold internal sequence similarity, with 346 residues of the N-terminal and C-terminal halves of hST having 48.6% identity; the equivalent values for pST and rST are 45.5 and 42.9%, respectively (Baldwin, 1993).

Numerous crystallographic structures of transferrins from the three major classes have been solved (see, for example, Lindley, 2001). However, relatively few precise structural data are available for the serum transferrins, the first solved being the diferric form of rST (Bailey *et al.*, 1988) at a resolution of 3.3 Å. Other partial structures of serum transferrins include the N-terminal half domain from rST (Sarra *et al.*, 1990) and several recombinant N-lobe mutants of hST (MacGillivray *et al.*, 1998; Jeffrey *et al.*, 1998; Bewley *et al.*, 1999; Nurizzo *et al.*, 2001). The structure of the monoferric C-lobe human serum transferrin has also been reported (Zuccola, 1993) at a resolution of 3.3 Å. As a consequence, the differences in iron binding and release between species and classes of serum transferrins are still poorly understood and there is a need for more precise structural data. In order to address this problem, it was decided to re-examine the structure of rabbit serum transferrin using synchrotron X-radiation and to study the structure of porcine serum transferrin (pST). The pST molecule was chosen since the raw material for purification was readily available and, ultimately, the crystals yielded the highest resolution X-ray data for a serum transferrin molecule at our disposal. pST also differs from the rabbit protein both in its disulfide-bridge configurations and the glycosylation sites. This paper reports the structures of diferric rabbit serum transferrin refined to an improved resolution of 2.6 Å and diferric porcine serum transferrin refined against data to 2.15 Å. Except where specifically stated, the residue numbering is that of the individual proteins.

2. Experimental

Both serum transferrins were isolated from their respective serum by fractionation with increasing quantities of ammonium sulfate, followed by purification using ion-exchange chromatography [see Baker *et al.* (1968) for rST].

2.1. Crystallization and data collection

2.1.1. Porcine serum transferrin (pST). Crystallization of pST was achieved in batch mode under paraffin oil using Terazaki plates. A solution of protein, at approximately 120 mg ml^{-1} in water, was mixed with an equal volume of 20%(*w*/*v*) PEG 4000, 0.2 *M* Li_2SO_4 and 0.1 *M* Tris-HCl at pH 8.5. 4 µl of the resulting solution was then introduced into a Terazaki plate that had been pre-filled with oil. All crystallization experiments were set up at 277 K. However, in order to initiate nucleation it was necessary to reduce the temperature at a linear rate of 1 K per day. Crystallization experiments were checked every 24 h and very small plate-like pale red crystals could often be observed within a week. As soon as these small crystals were observed, the temperature was increased by 1 K and maintained. Crystals of dimensions up to 0.5 × 0.2 × 0.05 mm grew within a further 7 d. Unfortunately, the crystals were very fragile, few in number and extremely sensitive to temperature fluctuations. No attempts were made to find cryofreezing conditions and the crystals were mounted in capillaries in a cold room and used for data collection with the minimum time delay.

Diffraction data were collected from two capillary-mounted crystals at station PX9.5, SRS, CLRC Daresbury Laboratory (Table 1). Owing to the goniometer geometry, the closest that the detector could be positioned to the crystal was 310 mm, corresponding to a resolution of 2.05 Å. Data were collected to this resolution at ambient temperature, but close inspection of the data quality in terms of completeness, precision and strength indicated that 2.15 Å was a more realistic limit of reliable data. Autoindexing, integration and scaling of data were performed with the *HKL* suite (Otwinoski & Minor, 1997). Autoindexing revealed the crystals to belong to space group *C*2, with unit-cell parameters a = 223.8, b = 44.9, c = 78.9 Å, β = 105.4°. The space group and unit-cell parameters correspond to a Matthews coefficient (Matthews, 1968) V_M of 2.39 Å^3 Da^{-1} for a monomer in the asymmetric unit, giving a solvent content of 48%. A Wilson plot for the data gave an average thermal coefficient of 22.8 Å^2.

2.1.2. Rabbit serum transferrin (rST). Crystallization of rST was achieved by a batch method (Al-Hilal *et al.*, 1976). Solutions containing in excess of 40 mg ml^{-1} protein in 0.01 *M* Tris–HCl pH 7.2 were adjusted to a pH between 5.3 and 5.6 with acetic or hydrochloric acid and left in a cold room

Table 1
Details of data integration and scaling for pST and rST.

Values in parentheses correspond to the outermost resolution shell: 2.20–2.15 and 2.67–2.60 Å for pST and rST, respectively.

	pST	rST
No. of crystals	2	6
Station at SRS	9.5	9.6
Wavelength (Å)	0.92	0.87
Detector	30 cm MAR IP	30 cm MAR IP
Temperature (K)	277	277
Resolution (Å)	23.31–2.15	48.03–2.60
Space group	*C*2	$P4_32_12$
Unit-cell parameters		
a (Å)	223.8	127.2
b (Å)	44.9	127.2
c (Å)	78.9	144.9
β (°)	105.4	—
Redundancy	3.6 (3.4)	5.6 (2.7)
No. of observations	145143	279069
No. of unique (*hkl*)	41503	36323
Completeness	96.0 (97.4)	97.7 (85.6)
$R_{\rm merge}$	0.073 (0.123)	0.057 (0.248)
$I/\sigma(I)$	15.8 (8.1)	9.9 (2.7)

overnight at a temperature of 277–281 K. Red tetragonal bipyramidal shaped crystals grew within a few days to maximum dimensions approaching 1.0 mm. Specimens with a maximum dimension of some 0.5 mm were mounted in capillaries and used for data collection at station 9.6, SRS, CLRC Daresbury Laboratory as shown in Table 1. After a number of unsuccessful attempts to cryofreeze crystals, data were collected at a temperature of 277 K from six crystals and the data integrated, scaled and merged using *MOSFLM* (Leslie, 1991), *ROTAVATA* and *AGROVATA* (Collaborative Computational Project, Number 4, 1994), respectively. The Matthews coefficient $V_{\rm M}$ was 3.68 Å^3 Da^{-1} assuming that the asymmetric unit contains a monomer, giving a solvent content of some 67%. A Wilson plot for the data gave an average thermal coefficient of 57.8 Å^2.

2.2. Structure solution for pST

The pST structure was solved by molecular replacement. A search model was generated from the coordinates of rST (Bailey *et al.*, 1988), mutating all non-identical residues with respect to pST to alanine and removing all heteroatoms, namely, iron cations, carbonate anions and solvent molecules. Molecular replacement was carried out with *AMoRe* (Castellano *et al.*, 1992; Navaza, 1993) using reflections in the range 10.0–4.0 Å. For completeness, all results for the rotation-function search were used in the translation-function search. Subsequent rigid-body refinement of all results gave an obvious solution where the correlation coefficient was 22.7 (compared with 7.4 for the next nearest solution) and the *R* factor was 49.2 (compared with 54.4 for the closest incorrect solution). The solution (α = 300.04, β = 102.99, γ = 178.83°, T_x = 0.1174, T_y = 0.0000, T_z = 0.0699) was then transformed from Eulerian coordinates to orthogonal coordinates and applied to the model coordinates. These resulting coordinates were then used for refinement.

2.3. Refinement

2.3.1. pST. Initial refinement utilized the least-squares functions of *RESTRAIN* (Driessen *et al.*, 1989), whilst refinement at later stages was performed using the maximum-likelihood functions implemented in *REFMAC* (Murshudov *et al.*, 1997). Temperature factors for the starting model were set to 23 Å^2 and the reflection file was formatted to include a test data set (for free *R* calculations). Model inspection and rebuilding was performed using *O* (Jones *et al.*, 1991) in combination with σ_A-weighted $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron-density maps (Read, 1986).

The molecular-replacement model was initially refined by least squares as six rigid bodies (residues 3–95, 96–244, 245–328, 341–424, 425–579 and 580–676 with pST numbering) delineating the domain organization of rST and using the resolution range 20.0–5.0 Å. This refinement was then extended to include data in the range 4.0–3.0 Å and was followed by a round of positional refinement using data in the range 12.0–2.5 Å to give an *R* factor of 31.4% ($R_{\rm free}$ = 39.3%). The first inspection of electron-density maps and rebuilding of the model were then undertaken. Some sequence disparities between rST and pST were corrected, the linker polypeptide between residues 324 and 339 was deleted and the iron and carbonate of each lobe were placed in well defined density at the binding site at this stage. Further refinement including individual isotropic temperature factors for this amended model, using data between 12.0 and 2.15 Å, gave an *R* factor of 27.4% ($R_{\rm free}$ = 35.7%). Again, map inspection and model rebuilding took place, whereupon loops (residues 31–34, 217–218, 415–420 and 639–642) which were ill defined in the electron-density maps were deleted, the N-lobe chain was extended to residue 327 at the linker peptide and 60 water molecules were placed. (Solvent molecules were only included when they were in chemically acceptable positions, in strong $2|F_o| - |F_c|$ electron density and greater than the 3 r.m.s. level in difference density.) Where no electron density was present for side chains, residues were assigned as alanine. A further round of least-squares isotropic refinement and model building was performed before progressing to maximum-likelihood refinement procedures. The *R* factor from this last stage of least-squares refinement was 23.5%, with an $R_{\rm free}$ of 31.4%.

Successive rounds of maximum-likelihood refinement with all data to 2.15 Å using conjugate-gradient matrices or conjugate-direction methods were then interspersed with model inspection and rebuilding. During these later stages of refinement all surface loops were incorporated, with all but the loop 31–34 having well defined $2|F_o| - |F_c|$ electron density at the 1 r.m.s. level. Solvent molecules were assigned on the basis of the previously mentioned criteria. Several attempts were made to model the linker peptide region (residues 334–341) into weak electron density, but no acceptable model could be obtained for this region and it was omitted from the final model. Density was evident for an *N*-acetyl glucosamine (*N*AG) molecule covalently linked to Asn497 N$^{\delta 2}$ and this was incorporated into the model. Throughout the refinement, the

Table 2
Refinement and model geometry statistics for rST and pST.

G factor and Ramachandran analysis were determined by *PROCHECK* (Laskowski *et al.*, 1993).

	pST	rST
Amino acids	676	674
Protein atoms	5254	5216
Solvent atoms	494	206
Other molecules	2 Fe^{3+}, 2 CO_3^{2-}, 1 *N*AG	2 Fe^{3+}, 2 CO_3^{2-}, 1 Cl^-
Resolution limits (Å)	30.0–2.15	30.0–2.6
Working set *R* (observations)	0.136 (38011)	0.185 (34433)
Test set *R* (observations)	0.182 (2001)	0.222 (1718)
Cruickshank's DPI†	0.188	0.323
Average isotropic thermal parameters (Å^2)		
All atoms	25.0	47.9
Protein atoms	23.9	47.9
Main chain	22.0	46.7
Side chain	26.0	49.1
Solvent	36.3	50.4
Residues not found	1–2, 334–341, 688–696	1–2
Residues truncated to alanine or glycine	Gln3, Glu342, Lys345, Glu421	Glu3, Asp257, Asp337, Ser348
Residues in dual conformations	Asp202, Tyr399, Asp444, Asn478, Asp544	None
Distance deviations‡		
Bond distances (Å)	0.010	0.020
Bond angles (°)	1.890	1.942
Planar groups (Å)	0.008	0.008
Chiral volume deviation (Å^3)	0.113	0.121
Overall *G* factor	0.05	−0.13
Ramachandran analysis (%) (No. of residues)		
Favourable	90.3 (543)	85.6 (505)
Additional	9.0 (54)	12.9 (76)
Generous	0.3 (2)	1.0 (6)
Disallowed	0.3 (2)	0.5 (3)

† Cruickshank (1996). ‡ R.m.s. deviations from standard values.

Table 3
Secondary-structure assignment in porcine serum transferrin.

The labelling follows the nomenclature employed by Haridas *et al.* (1995). The α-helices 4, 5*a*, 6*b*, 8*a* and 8*b* in the N-lobe and 3*a*, 6*a*, 8*a*, 8*b*, 8*c* and 12 in the C-lobe are of the 3_{10} type.

N-lobe residues	Nomenclature	C-lobe residues
β-Strands		
4–10	*a*	346–352
35–41	*b*	369–375
59–61	*c*	393–395
76–84	*d*	409–415
91–102	*e*	431–439
114–119	*f*	454–458
154–159	*g*	489–492
205–211	*h*	539–544
226–231	*i*	568–572
235–238	*ia*	575–579
247–259	*j*	588–591
	ja	595–598
304–309	*k*	645–649
α-Helices		
12–29	1	353–366
43–53	2	378–387
62–70	3	396–406
	3*a*	423–425
108–110	4	
127–134	5	463–478
135–137	5*a*	
145–152	6	481–486
166–169	6*a*	499–501
177–179	6*b*	
191–201	7	525–536
212–217	8	545–550
223–225	8*a*	565–567
239–244	8*b*	580–585
	8*c*	600–602
263–278	9	603–617
314–319	10	655–659
321–331	11	661–669
	11*a*	670–673
	12	677–682

geometries of iron sites were unrestrained and H atoms were included as 'riding' atoms in the final stages of the refinement.

2.3.2. rST. Subsequent to partial refinement using *RESTRAIN* and *X-PLOR* (Brünger, 1992), rST was finally refined using the program *REFMAC*, starting essentially with the model from the structure determination at 3.3 Å (Bailey *et al.*, 1988). Initially, residues poorly defined in this medium-resolution model were either omitted or changed to alanine, the iron cations, carbonate anions and solvent molecules were removed and refinement proceeded with working and free (5%) sets of data extended to include all data between 30.0 and 2.6 Å. The initial values for the *R* factor and $R_{\rm free}$ were 28.9 and 30.3%, respectively. Rounds of conjugate-gradient sparse-matrix refinement with bulk-solvent modelling according to the Babinet principle (Tronrud, 1996) were alternated with model building using the *O* program. As far as possible residues were assigned according to the sequence, even when the side-chain density was discontinuous or, as is the case of several surface lysine residues, incomplete at the amino-terminus. No residues were seen with clear dual conformations at 2.6 Å resolution. However, some residues were truncated to alanine as indicated in Table 2 and four residues were mutated as Tyr136Leu, Ser348Gly, Ile498Val and Tyr605Phe to match the electron density rather than the sequence. During the course of the refinement it became clear that the scattering power of a solvent molecule some 3.2 Å from Arg124 $N^{\eta 2}$ was significantly greater than expected. It was therefore assigned on chemical grounds as Cl^-. Such an anion has also been found in the N-lobe half-molecule of rST (R. Sarra, personal communication) and may result from the crystallization conditions. For rST, only one carbohydrate chain is expected at Asn490 (Evans *et al.*, 1988), but there was little evidence for this in the electron density contoured at 1.0 r.m.s. At a contour level of 0.6 r.m.s. two possible conformations became evident, but no attempt was made to insert a glycan moiety into the model. As for pST, the geometries of the iron sites were refined in an unrestrained manner and H atoms were included in the final stages.

2.4. Quality of models

2.4.1. pST. The final conjugate-gradient matrix refinement was performed on a model incorporating residues 3–333 and 342–687, two iron cations, two carbonate anions, one *N*AG molecule and 494 waters. A final *R* factor of 13.8% ($R_{\rm free}$ = 18.2%) for all reflections was observed (see Table 2 for summary of refinement and geometry statistics). A mean

temperature factor of 23.9 Å^2 for all protein atoms is close to the value obtained from the Wilson plot, with the values being lower in regions of secondary structure and higher at external loops (Fig. 1). A Ramachandran analysis as implemented in *PROCHECK* (Laskowski *et al.*, 1993) shows that two leucine residues, Leu298 and Leu639, are found in disallowed regions. Both occupy similar sites within each lobe, being found in a classic γ-turn. Fig. 2 illustrates these residues as well as showing representative electron density for this model.

The electron density is contiguous at the 1 r.m.s. level for the main chain except for the linking peptide of residues 334–341; those residues assigned as the start and end of the linker peptide have higher than average temperature factors. It is assumed that this is a very mobile region of the model. For five residues there was notable $|F_o| - |F_c|$ electron density into which two side-chain conformations could be built and refined successfully. Table 2 lists these side chains in addition to those residues truncated to alanine owing to the lack of interpretable density at the 1 r.m.s. level. Three *cis* prolines were identified at positions 73, 141 and 144.

2.4.2. rST. The final model comprises residues 3–673, 5227 protein atoms, two Fe^{3+} cations and two CO_3^{2-} anions, a Cl^- anion and 224 solvent molecules; there is no clear indication of the carbohydrate moiety attached to Asn490. A final *R* factor of 18.6% (R_{free} = 22.2%) for all reflections was observed as shown in Table 2. The mean temperature factor for all protein atoms is 47.4 Å^2. A *PROCHECK* analysis indicates that three residues are in disallowed regions of the Ramachandran plot. In a similar manner to pST, two of these, Leu294 and Leu627, are involved in the γ-turn in each lobe. The third, Asn106, is involved in a 2.65 Å intermolecular hydrogen bond with Asp13 (symmetry operation: $y - \frac{1}{2}, -x + \frac{1}{2}, z + \frac{1}{4}$) and lies in well defined density. The need for this intermolecular interaction also appears to have an effect on the loop Asn12–His14 that has a different conformation to that found in pST, causing the φ/ψ values for Asn12 to fall close to the *p* region of the Ramachandran plot. The connecting peptide, residues 332–338, is a 'best efforts' attempt to match the electron density that is almost continuous at the 1 r.m.s. level of $2|F_o| - |F_c|$ electron density. It is highly probable that this region occupies more than one conformation in the crystal and is flexible in solution, *c.f.* pST. In general, the C-lobe appears to be slightly more poorly defined than the N-lobe; the average isotropic thermal coefficients of the protein atoms is 54.2 Å^2 for the C-lobe and 40.3 Å^2 for the N-lobe.

3. Results and discussion

3.1. Overall molecular structure

The three-dimensional arrangements of all the members of the transferrin family whose structures have so far been determined are fundamentally the same. Since the pST structure has been determined to a relatively high resolution, it will be used to illustrate the major features. Thus, diferric pST is a bilobal molecule with overall dimensions of 90 × 65 × 45 Å (Fig. 3). The sequence identity between the two lobes is 45.5% and this similarity is carried over into its structure, where an overlay of the C-lobe onto the N-lobe has an r.m.s. deviation of 1.27 Å for 293 C$^\alpha$ atoms. For rST, the sequence identity is 42.9% and 290 C$^\alpha$ atoms superpose with an r.m.s. deviation of 1.31 Å. For both serum proteins, a superposition of the lobes requires a rotation of approximately 167° followed by a translation of 23 Å.

Each lobe of the transferrin structure is split into two dissimilar domains (NI and NII, CI and CII in Fig. 3) with a deep cleft in between where the Fe^{3+} cation is bound in conjunction with the synergistic carbonate anion. In pST, domain NI is composed of residues 3–94 and 250–333 and domain NII comprises residues 95–249. Likewise in the C-lobe, CI comprises residues 342–430 and 592–687 and CII residues 431–591. Thus, in both lobes the similar NII and CII domains are formed from a continuous polypeptide separating the two regions that make up the similar domains NI and CI. Two antiparallel β-strands connect the two domains of each lobe and form the hinge region about which the domain II pivots to give the open-lobe form of the apo protein.

The assignment of secondary structure and the three-dimensional arrangement for diferric pST is shown in Table 3 and Fig. 3. It can be seen that each domain has a central mixed β-sheet flanked on either side by α-helices. The folding of the domains is an approximate β/α repeat. Several short 3_{10}-helices (Table 3) are observed and are also found in hLF (Haridas *et al.*, 1995). There are two classic γ-turns at 297–

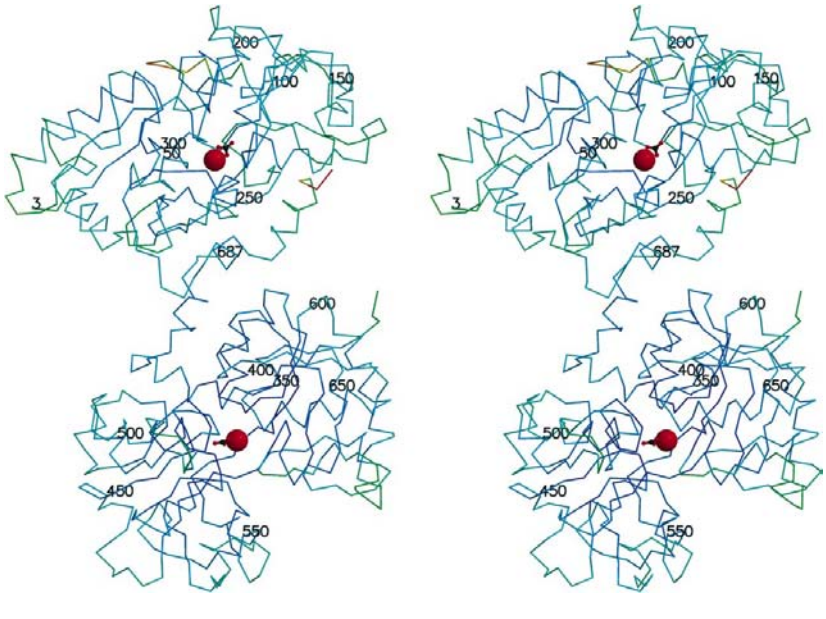

Figure 1
Stereoview C$^\alpha$ trace of pST with the temperature factors of C$^\alpha$ atoms displayed in a colour ramp from blue (5 Å^2) to red (100 Å^2). The iron cations are shown as red CPK spheres and the carbonates as ball-and-stick representations.

Table 4
The geometry of iron coordination in the N- and C-lobe binding sites of pST and rST.

The numbering scheme refers to the pST structure. The solvent molecule closest to the cation is also included in the distances.

Bond lengths (Å).

	N-lobe		C-lobe	
Atoms–residue	pST	rST	pST	rST
Fe—Asp $O^{\delta1}$ (62/396)†	2.05	2.1	2.02	2.4
Fe—Tyr1 OH (94/431)†	2.02	1.8	1.88	1.7
Fe—Tyr2 OH (192/526)†	1.82	1.8	1.87	1.9
Fe—His N^{ε} (253/594)†	2.20	2.1	2.19	2.1
Fe—CO_3^{2-} O_1	2.09	1.9	2.06	1.9
Fe—CO_3^{2-} O_2	2.19	2.0	2.08	2.2
Fe—H_2O	3.80	3.8	3.88	3.7

Bond angles (°).

	N-lobe		C-lobe	
	pST	rST	pST	rST
Asp $O^{\delta1}$—Fe—Tyr1 OH	87	90	92	77
Asp $O^{\delta1}$—Fe—Tyr2 OH	175	168	167	165
Asp $O^{\delta1}$—Fe—His $N^{\varepsilon2}$	88	89	84	84
Asp $O^{\delta1}$—Fe—CO_3^{2-} O_1	85	89	89	83
Asp O^{δ}—Fe—CO_3^{2-} O_2	86	78	83	97
Tyr1 OH—Fe—Tyr2 OH	98	91	99	102
Tyr1 OH—Fe—His $N^{\varepsilon2}$	96	96	95	93
Tyr1 OH—Fe—CO_3^{2-} O_1	96	91	93	87
Tyr1 OH—Fe—CO_3^{2-} O_2	156	156	157	155
Tyr2 OH—Fe—His $N^{\varepsilon2}$	90	79	88	81
Tyr2 OH—Fe—CO_3^{2-} O_1	96	102	97	113
Tyr2 OH—Fe—CO_3^{2-} O_2	90	104	90	89
His $N^{\varepsilon2}$—Fe—CO_3^{2-} O_1	166	173	170	166
His $N^{\varepsilon2}$—Fe—CO_3^{2-} O_2	107	104	107	111
CO_3^{2-} O_1—Fe—CO_3^{2-} O_2	61	69	65	68

† N/C-lobe residue number.

299 and 638–640 (Fig. 2) that are conserved in all transferrins. The comparable turns are found at 298–300 and 641–642 in hLF and 293–295 and 626–628 in rST. As is commonly found in other transferrin structures, *cis*-proline residues are a feature and the N-lobe of pST contains three of them at 73, 141 and 144. The structure of rST has two comparable *cis*-prolines at 74 and 145 (equivalent to 73 and 144 in pST), whilst hLF has them at 71 (73 in pST), 142 (144 in pST) and at 629.

3.2. Iron-binding sites

The iron-binding sites in each lobe are very similar and comprise a distorted octahedral coordination involving four protein ligands, namely two tyrosines, a histidine, an aspartate and a bidentate synergistic carbonate anion. For pST, the protein ligands are Tyr94 and Tyr192 in the N-lobe (Tyr431 and Tyr526 in the C-lobe), His253 in the N-lobe (His594 in the C-lobe) and Asp62 in the N-lobe (Asp396 in the C-lobe). A water molecule found at approximately 4 Å from the metal and therefore not directly bound to it takes part in a hydrogen-bonded network including a bridge between the hydroxyl O atom of Tyr94 (Tyr431) and the carboxyl group of Asp62 (Asp396). Thus, the aspartate residues not only bind to iron, but also make hydrogen bonds with residues at the N-termini of α-helices 3 and 5. These are inter-domain interactions and are observed in all other iron-loaded transferrin structures. Stereoviews of the N- and C-lobe iron-binding sites in pST are shown in Fig. 4, whereas Table 4 lists the respective distances and angles. The resolution of pST (2.15 Å) provides a model of a diferric serum transferrin for which the iron coordination distances can be interpreted reliably, since the metal–ligand distances were unrestrained during refinement.

The iron–ligand bond distances and angles at both the N- and C-lobes of pST are comparable. This contrasts with hLF where the Fe—OH (Tyr92) distance is 2.28 Å, but the Fe—OH (Tyr435) distance is 1.92 Å. In the case of rST these distances are similar: 1.77 Å in the N-lobe and 1.70 Å in the C-lobe. The longer bond length in the N-lobe of hLF, taken in conjunction with the overall greater asymmetry of the metal site, has been suggested to be important for the N-lobes predilection to release iron more readily than the C-lobe (Haridas *et al.*, 1995). The largest difference in iron ligands for the serum transferrins occurs for the C-lobe aspartate ligand in rST, which is considerably longer at 2.40 Å than that observed in any other transferrin (Table 4). It has been shown that this ligand is not essential for iron binding (Faber *et al.*, 1996), but that it does contribute significantly to stability owing to the formation of interdomain hydrogen bonds.

Table 5
The ligands for carbonate in each lobe and their bond distances for pST and rST.

Carbonate atom–ligand	N-lobe bond lengths (Å)/residue number		C-lobe bond lengths (Å)/residue number	
	pST	rST	pST	rST
O1—Ala N	2.8/125	3.0/126	2.9/464	3.0/457
O1—Fe	2.09	1.9	2.08	1.9
O2—Arg N^{ε}	2.8/123	2.8/124	2.9/462	2.8/455
O2—Arg $N^{\eta1}$	2.9/123	2.9/124	2.8/462	2.8/455
O2—Fe	2.19	2.0	2.06	2.2
O3—Thr $O^{\gamma1}$	2.7/119	2.6/120	2.7/458	2.9/451
O3—Gly N	2.9/126	3.2/127	3.3/465	3.1/458
O3—AlaN	3.3/125	—	—	3.1/457

3.3. Binding of the synergistic carbonate anion

The carbonate binds between the iron and a positively charged site on domain II of each lobe. The +1 positive charge exerted by an arginine side chain (Arg123, Arg462) and the macro-dipole of the N-terminus of helix 5 almost match the charge of the carbonate. Table 5 shows that for pST the carbonate in each lobe is hydrogen bonded in a similar manner and all ligands are from domain II. Each carbonate O atom binding to iron also forms hydrogen bonds, with the amide of an alanine in one case and a bifurcated bond from N^{ε} and $N^{\eta1}$ of an arginine in the other, in a manner comparable to hLF. In conjunction with the contributions from a glycine amide and the dipole of helix 5, the third carbonate O atom hydrogen bonds with a threonine (Thr119, Thr458) from β-strand *f* in each lobe. The carbonate ligands are similar but not identical to those of hLF. Haridas *et al.* (1995) have

suggested that for hLF carbonate binding at the C-lobe is slightly tighter since the ligand distances are on average 0.1 Å shorter than in the N-lobe and that this explains at least in part the more facile release of iron from the N-lobe. The environments of the carbonate anions in the rST structure are essentially the same as those for pST.

In pST the carbonate-binding arginine in each lobe differs in the other hydrogen bonds it forms. In the N-lobe the $N^{\eta1}$ and $N^{\eta2}$ atoms of Arg123 form hydrogen bonds to a common solvent molecule, W340 (2.9 and 3.4 Å, respectively). $N^{\eta2}$ has a hydrogen bond to another solvent molecule, W314 (3.3 Å), in a water-filled pocket that is found above the iron-binding site for all transferrins. These solvent molecules only have interactions with other solvent molecules. In the C-lobe a different pattern of hydrogen bonding is observed. Thus, Arg462 $N^{\eta1}$ hydrogen bonds to a solvent molecule, W166 (3.1 Å), which has extensive interactions involving not only other solvent molecules but also the main-chain amide of Tyr526 (3.1 Å) and Asp637 $O^{\delta1}$ (2.8 Å). As a consequence, an indirect bond is

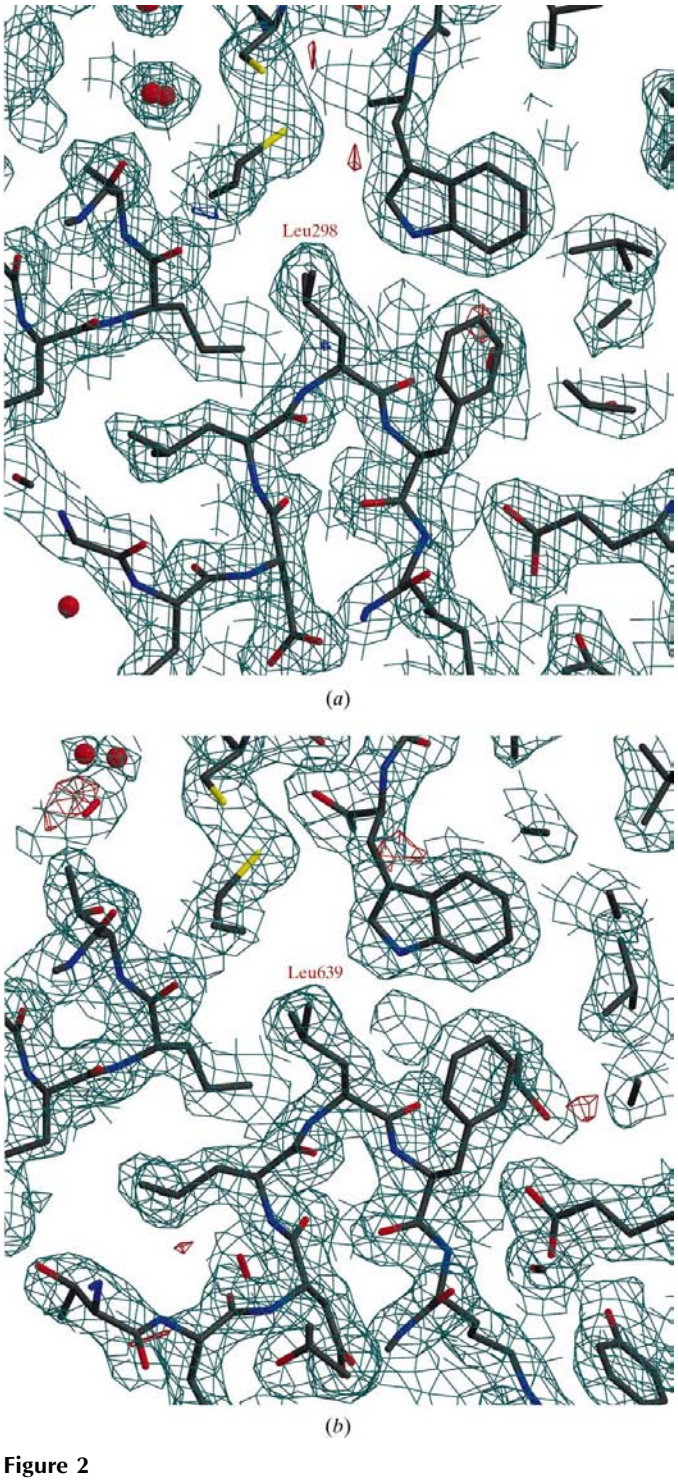

Figure 2
Representative sections of the model of pST illustrating the two residues found in disallowed regions of a Ramachandran plot. These are in classic γ-turns: (*a*) Leu298 found in the N-lobe and (*b*) Leu639 found in the C-lobe. The $2|F_o| - |F_c|$ map (green) is contoured at 1 r.m.s., whilst $|F_o| - |F_c|$ maps are contoured at +3 r.m.s. (blue) and −3 r.m.s. (red). The electron density was converted for *MOLSCRIPT* from *O* plots using *PLT2MOL* (C. S. Bond, personal communication).

Figure 3
The model of pST showing the bilobal structure and domain arrangement whereby NI is equivalent to CI (red and blue secondary structure) and NII is equivalent to CII (green and cyan secondary structure). The liganded iron and carbonate of both lobes is shown. The *N*-acetyl glucosamine (*N*AG) that is covalently linked to Asn497 is shown in stick representation.

made from domain CII to domain CI *via* this solvent molecule. The $N^{\eta 2}$ of the C-lobe arginine also hydrogen bonds to a solvent molecule, W165 (3.1 Å), but this water, in a similar manner to the N-lobe, only has strong interactions with other solvent molecules near the iron-binding site. However, it also has a weak hydrogen bond to Tyr524 OH (3.4 Å) and this is a same-domain interaction not present in the N-lobe. The water-mediated domain–domain interaction of the arginine in the C-lobe may be another factor underlying stronger iron binding in the C-lobe compared with the N-lobe.

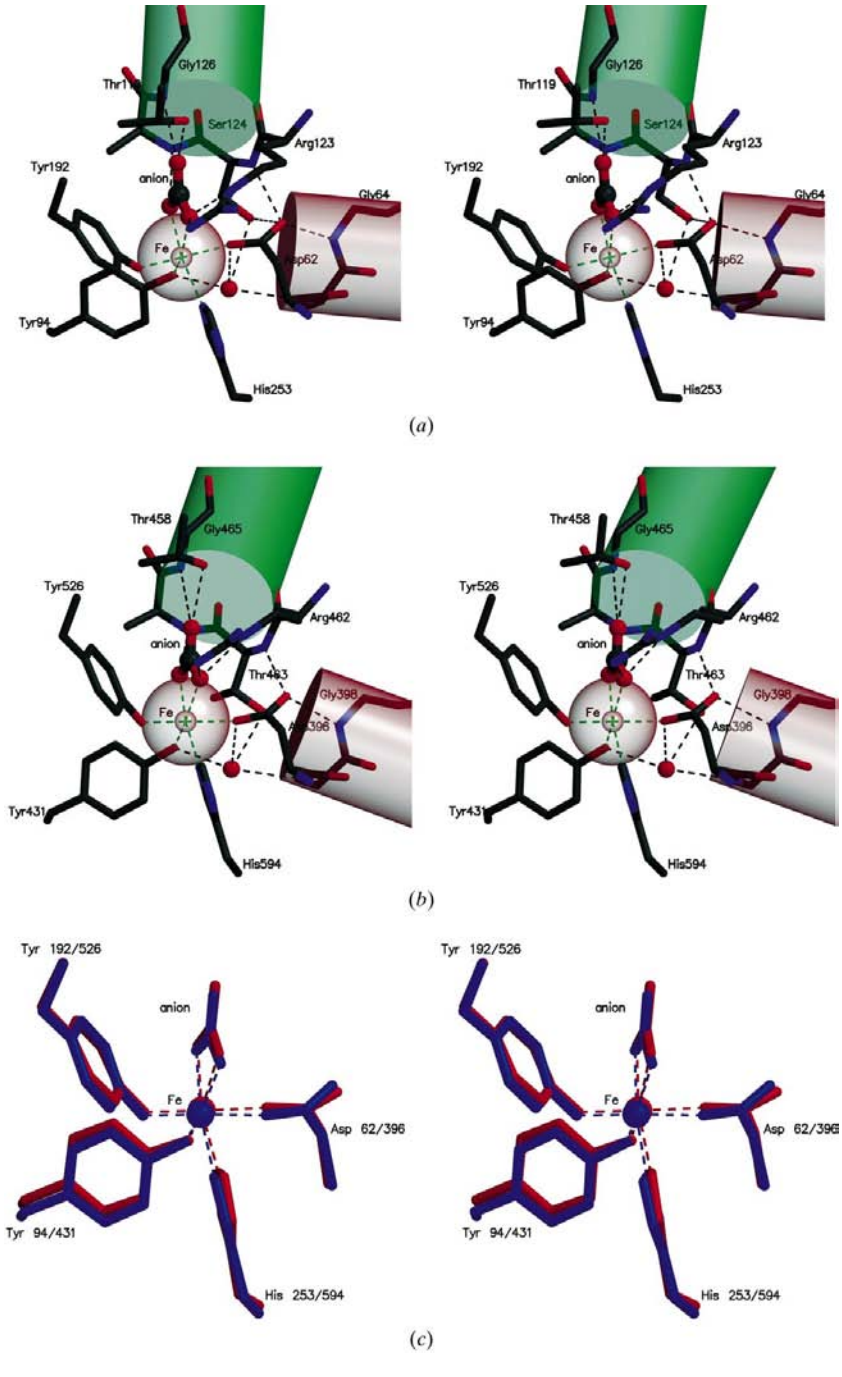

Figure 4
The iron-binding sites of pST: (*a*) the N-lobe, (*b*) the C-lobe. Note the Asp ligand binding to iron and hydrogen bonding between the two N-termini of helices 3 (red) and 5 (green) and (*c*) the overlay of the iron ligands of each lobe, where the components of the N-lobe are in red and those of the C-lobe in blue.

3.4. The dilysine 'trigger'

One of the major differences between serum transferrins and lactoferrins lies in the N-terminal lobes, away from the iron-binding site, at the interdomain interface; this is the so-called dilysine 'trigger' of serum transferrins (Baker & Lindley, 1992; Dewan *et al.*, 1993). This trigger is thought to provide a pH-sensitive switch for domain opening, permitting and accounting for the more acid-abile release of iron from the N-lobes of serum transferrins (Aisen *et al.*, 1978; Nurizzo *et al.*, 2001). However, a primary role in the iron-release mechanism has been questioned (Jeffrey *et al.*, 1998), since it is unlikely that domain opening will occur while the metal-binding site is occupied and all four protein ligands and the carbonate anion are coordinated to the metal. It is more likely that protonation of the carbonate anion (El Hage Chahine & Pakdaman, 1995; MacGillivray *et al.*, 1998) and the coordinated histidine (Jeffrey *et al.*, 1998) will constitute the primary stages, but the dilysine trigger may then be important when the iron coordination has been weakened. Conservation of the residues involved in the dilysine trigger is observed in the N-lobes of hST, pST, rST and cOT. In the structure of bLF both N-lobe lysine residues are conserved, but their conformation follows that of hLF and a dilysine trigger is not formed (Moore *et al.*, 1997).

The dilysine trigger in pST is shown in Fig. 5. The N^{ζ} atoms of Lys210 (domain NII) and Lys300 (domain NI) are 2.72 Å apart; the corresponding distance in rST is 2.5 Å, whilst a very short separation of 2.3 Å has been reported in cOT (Dewan *et al.*, 1993). Lys210 N^{ζ} also forms a hydrogen bond of length 3.2 Å with Ser302 $O^{\gamma 1}$ in domain NI, another domain–domain interaction. For Lys300, the Nζ atom has several interactions other than with Lys210. These are hydrogen bonds to the OH groups of Tyr84 (3.2 Å) and Tyr192 (3.3 Å) in domains NI and NII, respectively, and an amino-aromatic hydrogen bond to the ring of Tyr94 (approximately 3.4 Å). Similar interactions are observed for rST. In both hLF and bLF, however, a simple flip of the peptide 302–303 allows Lys301 in domain NI to form a salt bridge with Glu216 in domain NII. In addition, in hLF Lys210 is an arginine, which has comparable hydrogen bonding to the lysine, with the exception that

there is an extra amino-aromatic hydrogen bond to Tyr84 (Tyr82 in pST). The dilysine interaction therefore appears to be absent in the lactoferrins, even though in bLF there are two lysine residues.

The side chain of lysine typically has a pK_a of 10.4 and at the pH outside the cell (pH 7.4) or within the endosome after proton pumping (pH 5.0) should be protonated. The proximity of the N$^{\zeta}$ atoms of Lys210 and Lys300 should not occur since they should repel each other owing to their protonation. In the iron-bound form, therefore, both of them cannot be protonated and one must have a lowered pK_a value. It could well be that Lys300 is protonated owing its close proximity to two tyrosine residues; in addition, Glu82 is only 4.55 Å distant from Lys300 N$^{\zeta}$. The lowering of pK_a is not unknown and is a consequence of the microenvironment; for example, the lysine in the active site of acetoacetate decarboxylase is in a non-polar region where protonation is inhibited and its pK_a lowered to 5.9 (Fersht, 1985 and references therein; Highbarger *et al.*, 1996). The pK_a at the protein surface can also influence the pK_a of the interior and as such it may be possible that upon binding to TfR and the lowering of pH in the endosome the surface of the protein may be altered sufficiently to cause protonation of the dilysine trigger. The protonation of both lysines should force these residues apart and consequently help to separate the domains, whilst the protonation of the carbonate anion and His253 should aid the release of iron from the site.

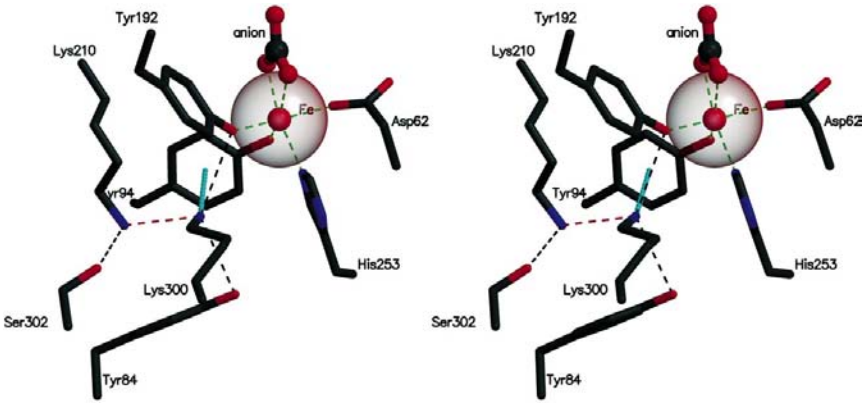

Figure 5
Stereoview showing the dilysine trigger (dashed red line) between lysines 210 and 300 in pST. Also shown are the iron ligands and the hydrogen bonds between Lys300 and the tyrosines 84 and 192 and the amino-aromatic hydrogen bond from Lys300 N$^{\zeta}$ to Tyr94 (cyan dashed lines).

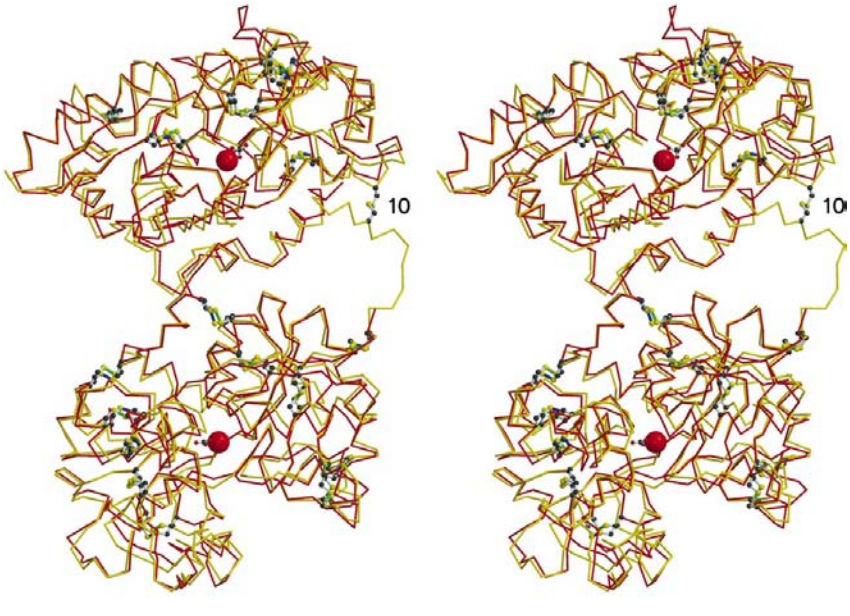

Figure 6
Stereoview showing the overlay of pST (red) and rST (yellow) showing the arrangement of disulfide bridges. Cysteines are illustrated in ball-and-stick mode, with the S atoms represented by yellow spheres. The covalent bonds between S atoms in pST are shown in green and those in rST are coloured blue. The only disulfide that is absent in pST is that found at the linker region of rST (labelled 10). The iron and carbonate illustrated are from pST.

The dilysine 'trigger' explains, at least in part, the *in vitro* release of iron more readily from the N-lobe compared with the C-lobe of serum transferrins and, moreover, the much more marked ease with which iron is released from transferrins with respect to lactoferrins. Not only is there no dilysine trigger, but lactoferrins also have more interdomain hydrogen bonds compared with those of the known serum transferrin structures. For iron release to occur more readily *in vivo* from the C-lobe than the N-lobe when it is receptor bound (Bali & Aisen, 1991), other factors must play a part, perhaps with Lys543 and Arg641 (3.81 Å between N$^{\zeta}$ and N$^{\eta 2}$) reorienting to form a comparable Lys-Arg trigger.

3.5. Interdomain interactions away from the binding sites

In a similar manner to other transferrin structures, interdomain hydrogen bonds in pST and rST are mainly found away from the iron-binding sites, at the tip and base of the binding clefts. Directly above the iron-binding site is the solvent-filled cavity through which protein–protein interactions are solvent mediated. For pST and rST (and the recombinant N-lobe of hST) there are no salt bridges present between the domains in the N-lobe, whilst two and four, respectively, are found between the domains in the C-lobe, as shown in Table 6. The presence of salt bridges only between the faces of the iron-binding cleft in the C-lobe for pST and rST agrees well with the known stronger affinity of iron binding at this lobe compared with the N-lobe (Aisen *et al.*, 1978). For hLF, salt bridges are found

at the interfaces of domains in both lobes, with four in the N-lobe and one in the C-lobe. In general, the lactoferrins have more extensive interdomain interactions away from the binding sites.

3.6. Disulfide bridging

The transferrins are found in extracellular fluids and as such are stabilized by many disulfide bridges. The positions of possible disulfide bridges predicted from the sequence of pST (Baldwin & Weinstock, 1988) and knowledge of the structure of rST were confirmed. All the bridges observed for rST with the exception of bridge 10 (Bailey *et al.*, 1988) are found in pST. Their position and functions have been described before (Bailey *et al.*, 1988), but the effects of disulfide 10 being absent from pST were unknown. This bridge in rST is the only disulfide that crosses between the domains of the N-lobe, doing so at the base of the cleft. In the context of domain closure of the N-lobe in pST, the absence of this disulfide does not result in any major altering of the degree of closure in the diferric form of this transferrin compared with that of rST (Fig. 6). Thus, 655 C^{α} atoms of rST can be superimposed on pST with an r.m.s. deviation of only 1.56 Å; the individual fits for the N- and C-lobes are 0.68 Å (317 C^{α} atoms) and 0.71 Å (327 C^{α} atoms), respectively. The most marked effect of the absence of this disulfide in pST occurs at the linking peptide between the N- and C-lobes. In the model of pST the linking peptide was not observed and is assumed to be highly mobile and disordered, whereas in rST a loop is observed, albeit with poor electron density. hLF also lacks disulfide 10, but the linking peptide is a short well defined helix. This highly flexible region in conjunction with the fact that there are few interlobe interactions in pST may permit reorientation of the lobes with respect to each other more readily than either rST or hST. The orientation of lobes with respect to each other may be important for receptor recognition and binding.

3.7. Carbohydrate structure

Transferrins are glycoproteins, although there appears to be no significant structural or functional role for the carbohydrate moieties with respect to iron metabolism since they are widely distributed on the protein surface in differing numbers in different transferrins (Baker & Lindley, 1992). The carbohydrates observed in transferrin structures have no significant interactions with the protein and are ill defined beyond the first two or three sugars. Based on the sequence of pST two possible carbohydrate-binding sites are present. These are at Asn25 and Asn497. Carbohydrate and sialic acid analyses on purified pST revealed that it contains one *N*-linked bi-antennary glycan chain (van Gelder *et al.*, 1995). Electron density representative of an *N*-acetyl glucosamine molecule was only observed at Asn497, with none found at the solvent-accessible Asn25 $N^{\delta2}$. It was not possible to model the glycan chain beyond the *N*AG molecule, suggesting that it is mobile and has no further interactions with the protein.

Table 6
Interdomain hydrogen bonds and salt bridges in pST and rST.

All distances $\leq$3.3 Å are given. Salt bridges are shown in bold type and the dilysine interaction in the N-lobe is shown in italics.

Domains NI and NII.

pST			rST		
Domain NI	Domain NII	Distance (Å)	Domain NI	Domain NII	Distance (Å)
Ser12 O^{γ}	Asn185 $N^{\delta2}$	2.7			
			His14 $N^{\delta1}$	Ser181 O^{γ}	3.3
Asn13 N	His186 $N^{\varepsilon2}$	2.9			
Asp62 $O^{\delta2}$	Ser124 N	2.8	Asp63 $O^{\delta2}$	Ser125 N	2.7
Asp62 $O^{\delta2}$	Ser124 O^{γ}	3.2	Asp63 $O^{\delta2}$	Ser125 O^{γ}	3.1
			Gly65 N	Ser125 O^{γ}	3.3
			Lys291 N^{ζ}	Glu212 O	3.1
			Asn292 $N^{\delta2}$	Glu212 $O^{\varepsilon2}$	3.2
Lys300 N^{ζ}	***Lys210 N^{ζ}***	***2.7***	***Lys296 N^{ζ}***	***Lys206 N^{ζ}***	***2.5***
Ser302 O^{γ}	Lys210 N^{ζ}	3.2	Ser298 O^{γ}	Lys206 N^{ζ}	3.0
Tyr323 OH	Ser124 O^{γ}	2.7	Tyr319 OH	Ser125 O^{γ}	2.5
			Tyr319 OH	Asn129 $N^{\delta2}$	3.1

Domains CI and CII.

pST			rST		
Domain CI	Domain CII	Distance (Å)	Domain CI	Domain CII	Distance (Å)
Glu376 O	Asn518 $N^{\delta2}$	3.0	Glu372 O	Asn506 $N^{\delta2}$	2.8
			Glu372 $O^{\varepsilon2}$	**Arg508 $N^{\eta1}$**	**2.8**
Asp396 $O^{\delta1}$	Thr463 N	2.7	Asp392 $O^{\delta2}$	Thr456 N	2.6
Asp396 $O^{\delta1}$	Thr463 $O^{\gamma1}$	3.3	Asp392 $O^{\delta2}$	Thr456 $O^{\gamma1}$	3.0
Gly398 N	Thr463 $O^{\gamma1}$	3.2	Gly394 N	Thr456 $O^{\gamma1}$	3.1
			Leu426 O	Ser532 N	2.7
Lys429 O	Arg590 $N^{\eta1}$	2.9			
			Lys624 N^{ζ}	Asn538 $O^{\delta1}$	3.1
			Lys624 N^{ζ}	Asn543 $N^{\delta2}$	3.2
			Arg629 N^{ε}	**Lys531 N^{ζ}**	**3.3**
Asp643 $O^{\delta1}$	**Lys543 N^{ζ}**	**2.8**	**Asp631 $O^{\delta1}$**	**Lys531 N^{ζ}**	**2.6**
			Asp631 $O^{\delta2}$	**Lys531 N^{ζ}**	**3.3**
Asp643 $O^{\delta1}$	Gln549 $N^{\varepsilon2}$	3.2	Asp631 $O^{\delta1}$	Gln537 $N^{\varepsilon2}$	3.1
Leu659 O	Arg590 $N^{\eta2}$	2.9			
Asp662 $O^{\delta1}$	**Arg590 $N^{\eta2}$**	**2.6**			
Asp662 $O^{\delta2}$	**Arg590 $N^{\eta2}$**	**3.0**			
Asp662 $O^{\delta2}$	**Arg590 N^{ε}**	**2.9**			
Tyr663 OH	Thr463 $O^{\gamma1}$	2.7	Tyr652 OH	Thr456 $O^{\gamma1}$	2.9
Tyr663 OH	Asn467 $N^{\delta2}$	3.1	Tyr652 OH	Asn460 $N^{\delta2}$	3.1
Asn669 $N^{\delta2}$	Asn475 $O^{\delta1}$	2.9			

4. Conclusions

This work has resolved the structure of diferric porcine serum transferrin to 2.15 Å, providing a significant improvement on the quality of structure and accuracy for diferric serum transferrins. The pST structure has given a more exact representation of both metal-binding sites. The absence of disulfide bridge 10 appears to have no effect on the N-lobe structure compared with rST where it is present. For both proteins the poor definition of the linking peptide indicates significant flexibility and this may permit lobe–lobe orientations that are suited to transferrin receptor recognition and binding. The structural differences between pST and rST are not as marked as those between pST and hLF and explain some of the biochemical differences in iron binding and

release between species. The precision of the pST model may assist the understanding of iron-loaded transferrin–transferrin receptor (TfR) recognition and binding when a detailed structure of TfR becomes available; the current resolution is 3.2 Å (Lawrence *et al.*, 1999).

The authors would like to acknowledge the provision of synchrotron facilities at the SRS, Daresbury Laboratory, England by the Central Laboratory of the Research Councils in the UK. In particular, they would like to thank Miroslav Papiz, the beamline scientist for station 9.6. They also acknowledge the Medical Research Council, the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council for supporting the Structural Biology Group at Daresbury Laboratory. The figures were created with the help of *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merrit & Murphy, 1994).

References

Aisen, P., Leibman, A. & Zweiter, J. (1978). *J. Biol. Chem.* **253**, 1930–1937.
Al-Hilal, D., Baker, E., Carlisle, C. H., Gorinsky, B., Horsburgh, R. C., Lindley, P. F., Moss, D. S., Schneider, H. & Stimpson, R. (1976). *J. Mol. Biol.* **108**, 255–257.
Arnold, R. R., Cole, M. F. & McGhee, J. R. (1977). *Science*, **197**, 263–265.
Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A., Sarra, R. & Watson, J. L. (1988). *Biochemistry*, **27**, 5804–5812.
Baker, E. N. & Lindley, P. F. (1992). *J. Inorg. Biochem.* **47**, 147–160.
Baker, E., Shaw, D. C. & Morgan, E. H. (1968). *Biochemistry*, **7**, 1371–1378.
Baldwin, G. S. (1993). *Comput. Biochem. Physiol. B*, **106**, 203–218.
Baldwin, G. S. & Weinstock, J. (1988). *Nucleic. Acids Res.* **16**, 8720.
Bali, P. K. & Aisen, P. (1991). *Biochemistry*, **30**, 9947–9952.
Bewley, M. C., Tam, B. M., Grewal, J., He, S., Shewry, S., Murphy, M. E P., Mason, A. B., Woodworth, R. C., Baker, E. N. & MacGillivray, R. T. A. (1999). *Biochemistry*, **38**, 2535–2541.
Brünger, A. T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR.* New Haven, CT, USA/London: Yale University Press.
Castellano, E. E., Oliva, G. & Navaza, J. (1992). *J. Appl. Cryst.* **25**, 281–284.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 11–22. Warrington: Daresbury Laboratory.
Dewan, J. C., Mikami, B., Hirose, M. & Sacchettini, J. C. (1993). *Biochemistry*, **32**, 11963–11968.
Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
El Hage Chahine, J.-M. & Pakdaman, R. (1995). *Eur. J. Biochem.* **230**, 1102–1110.
Evans, R. W., Aitken, A. & Patel, K. J. (1988). *FEBS Lett.* **238**, 39–42.
Faber, H. R., Bland, T., Day, C. L., Norris, G. E., Tweedie, J. W. & Baker, E. N. (1996). *J. Mol. Biol.* **256**, 352–363.
Fersht, A. (1985). *Enzyme Structure and Mechanism*, p. 174. New York: W. H. Freeman & Co.
Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–46.
He, J. & Furmanski, P. (1995). *Nature (London)*, **373**, 721–724.
Highbarger, L. A., Gerlt, G. A. & Kenyon, G. L. (1996). *Biochemistry*, **35**, 41–46.
Jandl, J. H., Inman, J. K., Simmons, R. L. & Allen, D. W. (1959). *J. Clin. Invest.* **38**, 161–185.
Jeffrey, P. D., Bewley, M. C., MacGillivray, R. T. A., Mason, A. B., Woodworth, R. C. & Baker, E. N. (1998). *Biochemistry*, **37**, 13978–13986.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lawrence, C. M., Ray, S., Babyonyshev, M., Galluser, R., Borhani, D. W. & Harrison, S. C. (1999). *Science*, **286**, 779–782.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. (1991). *Crystallographic Computing V*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 27–38. Oxford University Press.
Lindley, P. F. (2001). *Handbook of Metalloproteins*, edited by A. Messerschmidt, R. Huber, T. L. Poulos & K. Wieghardt, pp. 793–811. Chichester: John Wiley & Sons.
MacGillivray, R. T. A., Moore, S. A., Chen, J., Anderson, B. F., Baker, H., Luo, Y., Bewley, M., Smith, C. A., Murphy, M. E. P., Wang, Y., Mason, A. B., Woodworth, R. C., Brayer, J. D. & Baker, E. N. (1998). *Biochemistry*, **37**, 7919–7928.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Moore, S. A., Anderson, B. F., Groom, C. R., Haridas, M. & Baker, E. N. (1997). *J. Mol. Biol.* **274**, 222–236.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1993). *Acta Cryst.* D**49**, 588–591.
Nurizzo, D., Baker, H. M., He, Q.-Y., MacGillivray, R. T. A., Mason, A. B., Woodworth, R. C. & Baker, E. N. (2001). *Biochemistry*, **40**, 1616–1623.
Otwinoski, Z. & Minor, W. (1997). *Methods Enzymol.* **234**, 129–132.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. & Lindley, P. (1990). *Acta Cryst.* B**46**, 763–771.
Singh, T. P., Sharma, S., Karthikeyan, S., Betzel, C. & Bhatia, K. L. (1998). *Proteins. Struct. Funct. Genet.* **33**, 30–38.
Thibodeau, S. N., Lee, D. C. & Palmiter, R. D. (1978). *J. Biol. Chem.* **253**, 3771–3774.
Tronrud, D. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 1–10. Warrington: Daresbury Laboratory.
Van Gelder, W., Huijskes, M. I. E., Huksorn, C. J., de Jeu-Jaspers, C. M. H., van Noort, W. L. & van Eijk, H. G. (1995). *Comp. Biochem. Physiol. B*, **111**, 171–179.
Zuccola, H. J. (1993). PhD thesis. Georgia Institute of Technology, Atlanta, Georgia, USA.