Characterization and Structural Analysis of a Functional Human Serum Transferrin Variant and Implications for Receptor Recognition†

Robert W. Evans,*,‡ James B. Crawley,‡ Richard C. Garratt,‡,§ J. Günter Grossmann,∥ Margarete Neu,∥,⊥ Alastair Aitken,▽ Kokila J. Patel,‡ Andrew Meilak,‡ Collin Wong,‡ Juswinder Singh,# Adrian Bomford,◇ and S. Samar Hasnain*,∥

Division of Biochemistry & Molecular Biology, UMDS, Guy's Hospital, London SE1 9RT, U.K., Molecular Biophysics Group, Daresbury Laboratory, Warrington, Cheshire WA4 4AD, U.K., Laboratory of Protein Structure, National Institute for Medical Research, Mill Hill, London NW7 1AA, U.K., Institute of Liver Studies, King's College School of Medicine & Dentistry, London SE5 9 PJ, U.K., Instituto de Fîsica de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560 São Carlos, Brazil, and Parke-Davis Pharmaceutical Research, 2800 Plymouth Road, Ann Arbor, Michigan

Received June 2, 1994; Revised Manuscript Received July 20, 1994®

ABSTRACT: The nucleotide and amino acid substitutions leading to the only known functional variant of human serum transferrin have been characterized by sequencing of a peptide produced by cyanogen bromide digestion and genomic PCR coupled with cycle sequencing, respectively. There is an amino acid substitution at position 394 (Gly → Arg) resulting from a mutational transition, G → A, in the first nucleotide of the codon GGG. The Zn^{2+}-, Al^{3+}-, and Cu^{2+}-binding properties of the variant, ascertained by UV difference spectra and, in the case of copper, protein fluorescence quenching, confirm that these metals bind to only one of the two sites. Solution X-ray scattering measurements indicate that the lobe (the C-lobe) containing the mutation remains "open" in the iron-bound state, and modeling studies suggest that this is a consequence of the formation of a salt bridge between Arg394 in the variant protein and Asp392, one of the iron-binding ligands in the C-lobe. This rationalizes for the first time the observed reduction in receptor affinity of the diferric variant protein for PHA-stimulated lymphocytes [Young, S. P., *et al.* (1984) *Br. J. Haematol. 56,* 581–587] and here repeated with K562 cells. These data lend support to the hypothesis that the closed conformation for both lobes contributes to receptor recognition.

The transferrins are a group of homologous monomeric glycoproteins of molecular mass around 80 000 Da. The function of these proteins is related to their high affinity for iron ($K_d \sim 10^{-22}$) providing a means of depriving the extracellular environment of iron, hence negating the damaging effects of iron-catalyzed free radical cascades and exerting bacteriostatic effects through the denial of bacterial iron. The serum transferrins provide a translocative function, transporting iron from sites of absorption and storage to sites of utilization.

Screening of a number of human serum samples by 6 M urea–polyacrylamide gel electrophoresis led to the identification of a heterozygous individual possessing both a variant transferrin with a differential electrophoretic mobility, corresponding to loss of iron from the C-site, and a normal transferrin (Evans *et al.*, 1982). Although the variant protein was found to bind two ferric ions per molecule (Evans *et al.*, 1982), a substantial blue shift was observed (λ_{max} of 448 versus 470 nm) for the diferric protein, and iron dissociated from the C-site upon electrophoresis in the presence of 6 M urea. The iron-free C-lobe was also found to be less resistant to both urea and thermal denaturation. Furthermore, the protein has demonstrated a 10-fold reduction in affinity for phytohemagglutinin (PHA)-stimulated lymphocytes *in vitro* (Young *et al.*, 1984a), a cell type that expresses transferrin receptor in large quantity (Bomford *et al.*, 1983). Here we report the nucleotide and amino acid substitution and further aberrant physicochemical properties of the variant protein with respect to the normal (henceforth referred to as wild type) protein, also isolated from the subject's serum. Solution X-ray scattering data for the variant protein show that the molecule undergoes a less pronounced conformational change upon iron binding consistent with a "one lobe closed and one lobe opened" conformation. The implications of these results are discussed in terms of receptor recognition.

MATERIALS AND METHODS

Materials. All chromatography matrices were from Pharmacia. Cyanogen bromide and iodoacetamide were from Sigma. Neuraminidase was from BDH. Trypsin was purchased from Boehringer Mannheim. The sheep polyclonal anti-human serum transferrin antibodies were a gift from Dr. J. Williams, University of Bristol.

Protein Purification. The normal and variant transferrins were isolated from the subject by passing the serum through an immunoaffinity chromatography column containing sheep polyclonal anti-human serum transferrin antibodies linked to a Sepharose matrix. The protein was eluted using 1 M ammonia solution, the fractions were analyzed by UV absorption at 280 nm, and the protein-containing fractions

† This work was supported in part by grants from the Special Trustees of Guy's Hospital (R.W.E.), SERC (S.S.H.), and MRC (A.A.). We also thank the Wellcome Trust for a Vacation Scholarship to R.C.G. and MRC for providing studentship to J.B.C.

* To whom all correspondence should be addressed.
‡ Guy's Hospital.
∥ Daresbury Laboratory.
▽ Mill Hill.
◇ King's College.
§ Universidade de São Paulo.
# Parke-Davis.
⊥ Present address: Biological Structures and Biocomputing Group, EMBL Heidelberg, Postfach 10.2209, 69012 Heidelberg, Germany.

® Abstract published in *Advance ACS Abstracts,* September 15, 1994.

0006-2960/94/0433-12512$04.50/0

were pooled and desalted through a Sephadex G50 column equilibrated in 0.05 M ammonium bicarbonate. The protein fraction was freeze-dried and dissolved in 50 mM sodium acetate/10 mM calcium chloride/0.1 M sodium chloride, pH 5.6, to give a protein concentration of 10 mg/mL. The protein was treated with neuraminidase (50 units) for 24 h at 4 °C to remove the sialic acid. The protein was then fully saturated with iron, using Fe(III)$(NTA)_2$, after $NaHCO_3$ had been added to a final concentration of 0.4 M. The protein was further purified either on a Pharmacia HP Q-Sepharose column using a 0–0.3 M NaCl linear gradient in 20 mM Tris, pH 8, or using preparative polyacrylamide gel electrophoresis in 6 M urea (Evans *et al.*, 1982). Proteins were made iron-free by dialysis against 0.1 M citric acid, pH 4.7 (Garratt *et al.*, 1991).

The C-terminal monoferric of normal transferrin was isolated by trypsin digestion of protein that had been 30% iron-saturated with iron nitrilotriacetate (Evans & Williams, 1978). The N-terminal fragment of normal transferrin was isolated by subtilisin digestion of iron-saturated protein. Monoferric preparations of both normal transferrin and unresolved transferrin from the heterozygous individual were prepared by the method of Baldwin and de Sousa (1981).

Isolation of Variant Cyanogen Bromide Peptide. Protein samples were dissolved in 70% (v/v) formic acid at a concentration of 20 mg/mL, and an equal volume of a solution of cyanogen bromide (40 mg/mL) in 70% (v/v) formic acid was added. All solutions were purged with nitrogen prior to their use. The reaction was allowed to proceed in the dark for 24 h, after which CNBr and formic acid were removed by vacuum desiccation over NaOH pellets. Nonreduced CNBr fragments were resolved by gel filtration on a Pharmacia Superose 12 column equilibrated with 100 mM HCl. An abnormally migrating fragment on Triton X-100–urea–acetic acid gels was reduced in 1% (v/v) mercaptoethanol for 4 h (Sutton & Brew, 1973) and carboxyamidomethylated prior to chromatography on a Pharmacia Pro RPC HR 5/2 reverse-phase column using a linear gradient from 0.1% trifluoroacetic acid (TFA) to 0.1% TFA and 40% acetonitrile. SDS–polyacrylamide gel electrophoresis and Triton X-100–urea–acetic acid–polyacrylamide gel electrophoresis were used to identify an abnormal peptide.

Subdigestion of Peptides with Trypsin. Trypsin digestion was carried out at a ratio of 1:30 proteinase substrate in 0.1 M *N*-ethylmorpholine acetate buffer (pH 8.5) for 2 h at 37 °C.

Peptide Sequencing. The normal and variant peptides together with their tryptic digests were sequenced using the Applied Biosystems Model 477A sequencer/Model 120A PTH analyzer system (Geisow & Aitken, 1989).

Genomic PCR and Cycle Sequencing. Epstein–Barr virus transformed lymphocytes were cultured in RPMI-1640 medium supplemented with 10% fetal calf serum (Gibco-BRL) and grown until sufficient for genomic DNA extraction. Genomic DNA was purified using phenol/chloroform extraction (Jeffries & Flavel, 1977). Primers were designed to amplify sequences adjacent to the mutated codon using the known exon sequences (Park *et al.*, 1985). The primer 5′ to the mutation had the sequence (5′ → 3′) AATGGAGAAGCT-GATGCCATGAGC. The primer 3′ to the mutation had the sequence (5′ → 3′) CAAAAGACGGTTCTGTCCCTGGTC. Exon 10 was amplified from the genomic DNA by PCR, and the amplified product was sequenced using the *fmol* DNA sequencing kit (Promega). A 6% urea–polyacrylamide gel was run in order to identify the mutation at the nucleotide level.

Ultraviolet Difference Spectra and Protein Fluorescence Quenching. Protein solutions of 1 mg/mL were obtained by dissolving the protein in 50 mM Hepes/0.1 M KCl/50 mM $NaHCO_3$ adjusted to the stated pH with NaOH. UV difference spectra were recorded on a Varian DMS 90 spectrophotometer. Protein fluorescence at 335 nm, after excitation at 295 nm, was measured using a Perkin-Elmer fluorescence spectrophotometer.

EPR Spectroscopy. EPR spectra were recorded on a Varian E-9 spectrometer. The EPR spectrum of the iron-loaded C-terminal site in the variant protein was computed by subtraction of the spectrum of the N-terminal iron-loaded fragment of the normal transferrin from that of the iron-loaded variant transferrin.

Transferrin Receptor Binding Studies. Iron-loaded proteins were labeled with ^{125}I to a specific activity of 300–400 cpm/ng of protein using solid-phase lactoperoxidase treatment (Young *et al.*, 1984b). The binding of labeled variant and wild-type proteins to K562 cells, a human leukemic cell line which expresses large numbers of transferrin receptors (Bomford *et al.*, 1986), was investigated in equilibrium binding experiments using methods described previously (Young *et al.*, 1984a; Bomford *et al.*, 1986). Cells were cultured as described previously (Bomford *et al.*, 1986).

Solution X-ray Scattering Data Collection and Analysis. X-ray scattering data were obtained using the small angle scattering station 8.2 (Bras *et al.*, 1993) at the Synchrotron Radiation Source, Daresbury, U.K. Measurements were performed with electron beam currents in the range of 180–260 mA with a ring energy of 2 GeV. The protein samples (5 mg/mL) were prepared in 50 mM MES and 50 mM HEPPS buffer (pH 7.5) containing 20 mM $NaHCO_3$. The iron-loaded transferrin solutions were prepared by addition of 10 mM Fe(III)$(NTA)_2$. The iron saturation of the samples was confirmed spectroscopically. In order to minimize background errors, buffer and protein samples were measured alternately for equal times. The total measuring time was 120 min each for both iron-loaded and iron-free samples and consisted of six independent data sets. Each data set comprised 20 min of buffer data collection followed by 20 min of protein sample data collection. The cell windows were changed after each set. Experiments were performed at room temperature (22–24 °C). The use of a 500-channel quadrant detector (Lewis *et al.*, 1988) with a sample-to-detector distance of 3 m resulted in a usable s range ($s = 2 \sin \Theta/\lambda$; 2Θ = scattering angle; λ = X-ray wavelength, here 1.5 Å) between 0.007 and 0.035 $Å^{-1}$. For data reduction we used the OTOKO software package (Boulin *et al.*, 1986). Guinier analysis (Guinier & Fournet, 1955) of the scattering curves at small s values ($s < 0.01$ $Å^{-1}$) yields the radius of gyration (R_g), which is a measure of the overall macromolecular size. Further details of experimental X-ray data collection and analysis are given elsewhere (Grossmann *et al.*, 1992, 1993a).

Computer Modeling. In order to interpret the solution scattering results, different overall conformations of serum transferrin have been modeled using the molecular graphics progam Insight II (version 2.2.0, Biosym Technologies, Inc., San Diego, CA) running on a Silicon Graphics Personal Iris Indigo XS24. In view of the high sequence similarity (79% identity) between rabbit and human serum transferrin, the use of the crystal structure of diferric rabbit serum transferrin (Bailey *et al.*, 1988) is made to model the iron-loaded wild-type protein. The transferrin molecule consists of N- and C-terminal lobes, each divided into two domains (I and II). In the holo form of the protein the domains close around the iron atom, while in the apo form they are open. According

to the previously reported procedure where a "fully opened" model of human apolactoferrin was obtained (Grossmann *et al.*, 1992), models of RST were achieved where both lobes adopt an open conformation or where one lobe is closed and the other is opened. These models rely on the identification of the large-scale conformational change in the N-lobe of lactoferrin induced by iron binding as revealed by the crystal structure determination of human apolactoferrin (Anderson *et al.*, 1990). The sequence homology among the transferrins as well as the internal homology between the N-lobe and the C-lobe is helpful in transfering structural and conformational features. Since the procedure of model building introduces strains at certain positions within the protein structure, e.g., due to the rupture of the polypeptide chain, all models were subjected to energy minimization using the program X-PLOR (version 3.0 by A. T. Brünger, Yale University, New Haven, CT). The positions of added water molecules forming a hydration shell around the protein were also optimized during the minimization procedure. This layer of bound water extends to about 3 Å from the molecular surface of the protein and is different from bulk water; thus it contributes to the X-ray scattering (Grossmann *et al.*, 1993a). There are two sugar side chains attached to serum transferrin which have not been taken into account, as their contribution is likely to be small due to their high mobility and disorder.

A detailed model of the iron-binding site in the C-terminal half of serum transferrin was constructed using the program TOM (Cambillau & Horjales, 1987). Protein structure information was extracted from the Brookhaven databank (Bernstein *et al.*, 1977; Abola *et al.*, 1987) and using SIRIUS, an automated method for the analysis of preferred packing arrangements between protein sidechains (Singh & Thornton, 1990).

Simulation of Scattering Profiles. The program DALAI (Pantos & Bordas, 1994) was used to compute the theoretical scattering profiles. It is based on the Debye formula (Debye, 1915) for the calculation of scattering intensities. The calculation was performed with the energy minimized and hydrated models of RST consisting of spheres at the atomic positions of all non-hydrogen atoms (5241 atoms in the case of iron-free RST, i.e., 675 amino acid residues, and 5251 atoms in the iron-loaded state, containing also two carbonate anions). The sphere radius for each atom used in our calculations was chosen to be 1.7 Å and correlates with the average van der Waals radius of non-hydrogen atoms in proteins (Grossmann *et al.*, 1993a). Hydration of the molecules in solution was considered by inclusion of a monolayer of water molecules (the hydration shell consisted of about 630, 670, and 650 water molecules for the fully closed model, the fully opened model, and the one lobe open and one lobe closed model, respectively).

RESULTS

Sequence Analysis of Normal and Variant Transferrin. Transferrin was isolated from the plasma of the subject by immunoaffinity chromatography (Evans *et al.*, 1982) and resolved into the normal and variant proteins by either preparative polyacrylamide gel electrophoresis in the presence of 6 M urea (Evans *et al.*, 1982) or anion-exchange chromatography on a Pharmacia HP Q-Sepharose column. It is worthy of note that the iron-loaded variant tranferrin elutes from the anion-exchange column after the normal protein, which is inconsistent with its electrophoretic behavior (Evans *et al.*, 1982) and the nature of the amino acid substitution (*vide infra*). This behavior suggests that the amino acid substitution could cause a conformational change within the variant protein which influences its interaction with the anion-exchange resin. The iron-free proteins were cleaved with cyanogen bromide, and the digest was analyzed, under reducing conditions, by polyacrylamide electrophoresis in the presence of Triton X-100/urea/acetic acid (Franklin *et al.*, 1980). Figure 1 shows the comparative electrophoretograms of the normal and variant CNBr peptides. These gels are identical except for one region where a fragment in the digest of the normal transferrin was replaced by a fragment with a higher mobility in the digest of the variant protein.

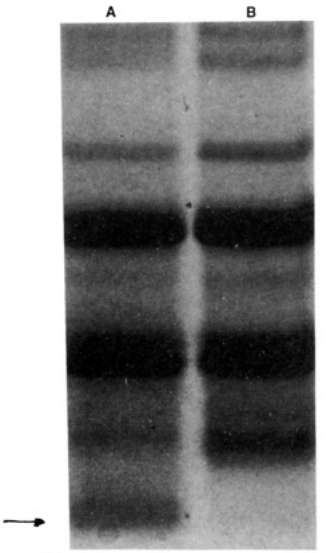

FIGURE 1: Comparative electrophoretograms of the CNBr digest of the variant (a) and normal (b) proteins. The high-mobility band migrates faster in the variant protein and is indicated by an arrow.

After fractionation of the cyanogen bromide digest by gel filtration, further analysis on Triton X-100/urea/acetic acid gels indicated that the abnormal fragment resided in cyanogen bromide fragment CN-A (MacGillivray *et al.*, 1983). CN-A from normal and variant transferrins was reduced, carboxyamidomethylated, and fractionated by reverse-phase chromatography. Analysis of fragments by both SDS and Triton X-100/urea/acetic acid polyacrylamide gel electrophoresis showed that the normal fragment corresponded to fragment CN-3 (MacGillivray *et al.*, 1983), comprising residues 390–464.

The amino acid sequence of the first 42 residues of CN-3 from the normal transferrin was determined by automatic sequencing and agreed with the published sequence (MacGillivray *et al.*, 1983). The sequence of CN-3 from the variant transferrin was found to be identical except at the fifth position, where a glycine residue at position 394 in the normal protein had been replaced by an arginine residue. The amino acid substitution was confirmed by sequence analysis of tryptic digests of CN-3 from the normal and variant transferrins. A number of residues were observed at each cycle corresponding to all of the theoretical tryptic peptides from the normal fragment. An additional sequence was detected in the variant fragment, Phe-Val-Tyr-Ile-Ala, due to cleavage at the Arg at position 394. Arginine 394 resides in exon 10 (Park *et al.*, 1985), and thus the subject's genomic DNA isolated from cultured EBV-transformed lymphocytes served as a template for the amplification of this exon. Cycle sequencing of this DNA revealed a G → A transition of the first nucleotide of codon 394 (Figure 2).

Metal-Binding Studies on Normal and Variant Transferrins. UV difference spectra at pH 7.5 of the Zn^{2+} complexes of the wild-type and variant transferrins (Figure 3) indicated

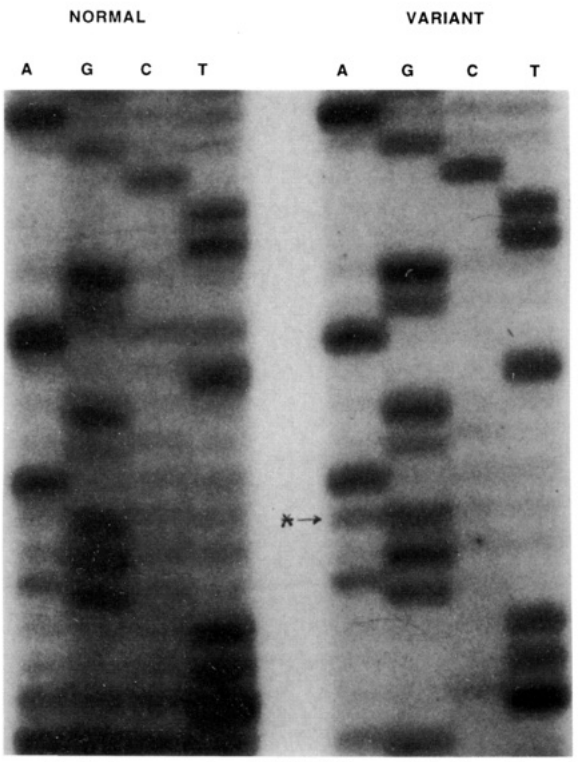

FIGURE 2: Cycle sequencing of this DNA revealed a G → A transition of the first nucleotide of the codon for residue 394. The heterozygous nature of this individual is reflected in the presence of A (marked by an asterisk) and G nucleotides in the first position of the codon encoding residue 394. A control from a normal individual is also shown, revealing only the GGG codon.

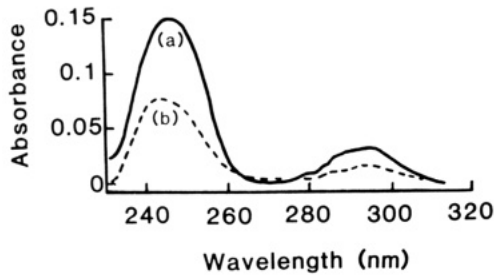

FIGURE 3: UV difference spectra at pH 7.5 of the Zn^{2+} complexes of the wild-type (a) and variant transferrins (b).

that the variant transferrin was only able to bind one zinc ion. $\Delta\epsilon_{247}$ values for the normal and variant proteins were found to be 15 790 and 7910, respectively. The value for the wild-type protein agrees well with that previously reported (Tomimatsu & Donovan, 1976). Titration of the variant with Al^{3+} also indicated that it was unable to bind this metal at both sites.

It has been shown that protein fluorescence can be used to monitor binding not only of Fe^{3+} but also of Cu^{2+} to ovotransferrin (Evans & Holbrook, 1975). This technique was therefore used to investigate the binding of Cu^{2+} to the wild-type and variant proteins. Experiments have shown that at pH 7.5 Cu^{2+} quenches the protein fluorescence of the normal and variant transferrins by 35% and 9%, respectively. At pH 6.3 the protein fluorescence of the normal protein was quenched by 9.5%, whereas no quenching was observed for the variant protein consistent with the presence of Cu^{2+} at the N-site only.

The EPR spectrum of a monoferric preparation of the unresolved transferrin, obtained by selective removal of iron from the C-terminal site by the method of Baldwin and de Sousa (1981), was found to be identical to that of an N-terminal fragment of normal human transferrin. This result indicates

Table 1: Comparison of Transferrin Receptor Binding Affinities (nM) of the Variant Differic Protein with Those of the Diferric, Monoferric, and Apo Human Serum Transferrins

	proteins			
	wild type			variant
cell type	diferric	monoferric	apo	diferric
rabbit reticulocytes[a]	9	40	210	
PHA-stimulated lymphocytes[b]	1.6			16
K562 cells	3.8			15.4

[a] Data taken from Young et al. (1984b). In this case the protein is from rabbit serum. [b] Data taken from Young et al. (1984a).

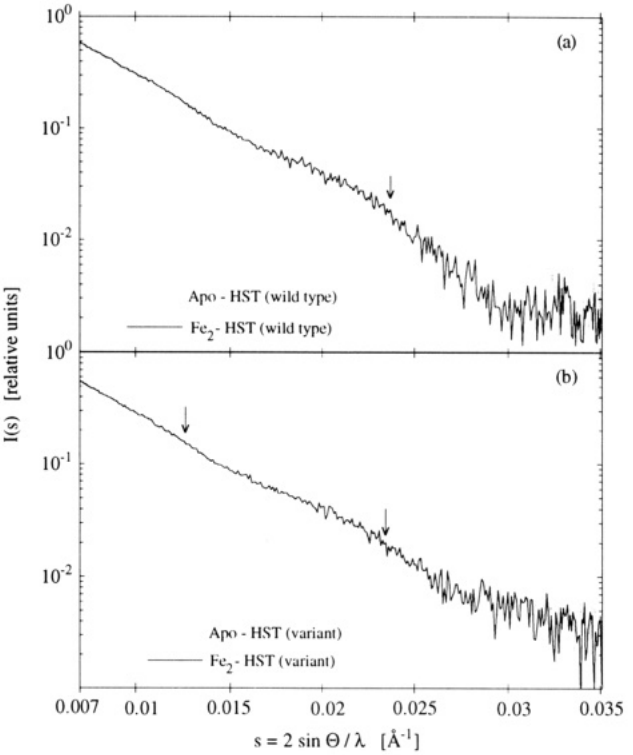

FIGURE 4: X-ray scattering curves of dilute solutions (5 mg/mL) of the wild-type (a) and variant human transferrin (b) recorded in the iron-free (light graph) and iron-loaded states (dark graph).

that the amino acid substitution in the C-terminal lobe has no effect on the N-terminal iron-binding site. The EPR spectrum of the iron-loaded C-terminal lobe in the variant transferrin, computed by subtraction of the EPR spectrum of the iron-loaded N-terminal fragment from that of the iron-loaded variant protein, confirmed the abnormal nature of the C-terminal site in the variant protein. The difference EPR spectrum closely resembles that of the 18-kDa fragment derived from duck ovotransferrin (Evans & Madden, 1984) which lacks the aspartate iron-binding ligand.

These results thus indicate that the metal-binding site in the C-lobe is abnormal even though it is able to bind an iron atom.

Receptor Binding of Wild-Type and Variant Proteins to K562 Cells. Table 1 presents the data from a number of experiments investigating the binding affinity of radiolabeled transferrins to different cell types. The data obtained here with K562 cells are very similar to those published previously using PHA-stimulated lymphocytes (Young *et al.*, 1984a). The variant transferrin show similar reductions in the binding affinity when compared to the wild-type diferric protein.

Solution X-ray Scattering and Modeling of Wild-Type and Variant Proteins. Figure 4 shows the X-ray scattering profiles of the iron-free and iron-loaded wild-type and variant proteins. The wild-type protein exhibits the characteristic

Table 2: Radii of Gyration for Wild-Type, Variant, and Modeled Proteins[a]

sample	R_g (Å)
Experiment (5 mg/mL Protein Solution)	
apo-HST (wild type)	32.6
apo-HST (variant)	32.4
Fe_2-HST (wild type)	31.8
Fe_2-HST (variant)	33.1
Simulation (Minimized and Hydrated RST Models)	
apo state (both lobes open)	30.3
holo state (both lobes closed)	29.9
variant model (C-lobe open and N-lobe closed)	31.4
model with N-lobe closed and C-lobe opened	29.0

[a] The Guinier approximation was applied to extract the experimental R_g values. Since the beam stop cuts down the innermost part of the scattering curves, the Guinier region was restricted to $0.007 \leq s \leq 0.01$ Å^{-1}. The error of the R_g values is about ±0.3 Å. The values for the models were calculated from the atomic coordinates using the formula $R_g^2 = \sum_i z_i R_i^2 / \sum_i z_i$, with z_i being the electronic charge of atom i at the distance R_i from the (electronic) center of mass.

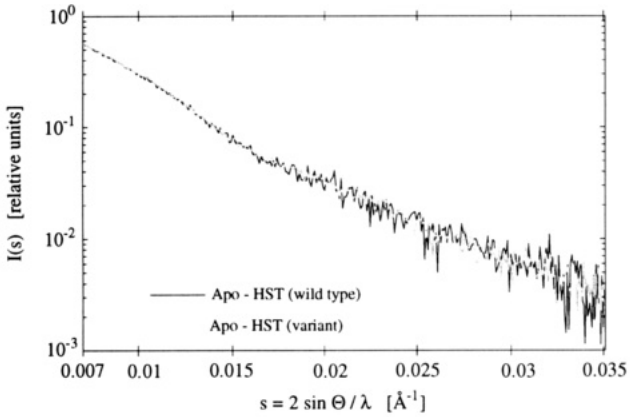

FIGURE 5: Comparison of the scattering patterns for the two iron-free human transferrin samples.

scattering features known from other transferrins in the apo and holo states (Grossmann *et al.*, 1992). The major change in the intensity curves when iron is bound is clearly evident in the s range between 0.015 and 0.035 Å^{-1} (note the use of a logarithmic intensity scale). A single intersection at $s \approx 0.023$ Å^{-1} underlines a conformational change, which is consistent with a structural transition where both N- and C-lobe turn from the opened into the closed state (Grossmann *et al.*, 1992; see also Figure 7a). In contrast, the variant protein clearly reveals a different scattering behavior. It not only shows a less pronounced difference for $s \geq 0.025$ Å^{-1} between the iron-free and the iron-loaded state but also exhibits a second intersection at low scattering angles ($s \approx 0.013$ Å^{-1}). These findings are further emphasized by a Guinier analysis of the low-angle scattering region (see Table 2). Unlike the iron-loaded wild-type protein, which represents the most compact structure (R_g = 31.8 Å), the overall conformation of the iron-bound variant protein appears to be even less compact (R_g = 33.1 Å) compared to the metal-free apoproteins ($R_g \approx 32.5$ Å). The comparison of the scattering curves for the two apoproteins (Figure 5) demonstrates that the mutation in the variant protein does not result in any significant difference in the molecular structure.

The interpretation of the experimental results required building of structural models based upon the crystallographic coordinates of rabbit serum transferrin (Bailey *et al.*, 1988) and human lactoferrin (Anderson *et al.*, 1989, 1990). We obtained a configuration for the fully-opened apoprotein as well as a model where the C-lobe remains open and only the N-lobe adopts the closed conformation. The models are shown

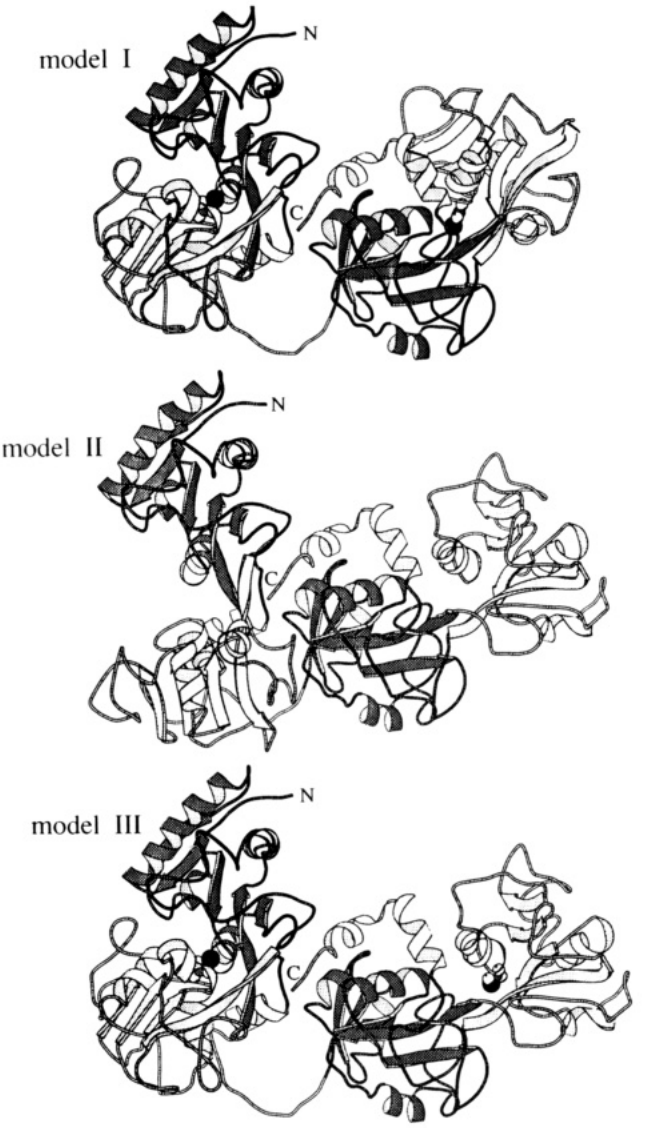

FIGURE 6: Ribbon drawings of models for serum transferrin used in the simulation of solution X-ray scattering data (see Figure 7). Model I: both interdomain clefts are closed, corresponding to the crystal structure of diferric RST (Bailey *et al.*, 1988). Model II: both lobes adopt the open conformation representing the unloaded, metal-free protein. Model III: the C-lobe remains open, and only the N-lobe is represented in the closed state. The drawings were prepared with the program MOLSCRIPT (Kraulis, 1991). Different gray levels have been used in order to emphasize the multidomain structure of serum transferrin. Iron atoms and carbonate anions are represented by atomic models.

in Figure 6. The scattering profiles arising from these models are depicted in Figure 7. Apart from the very low angle range, the experimental results for the wild-type (Figure 7a) and the variant protein (Figure 7b) are well reproduced quantitatively. In particular, the characteristic difference between apo and holo forms in the upper scattering region and the intersection points are consistent with the experimental findings. The discrepancy between simulation and experiment at the lower end of the scattering range is also represented in the R_g analysis (see Table 2). Despite showing the same trend, the experimental R_g values are about 6% higher than those calculated from the atomic coordinates of the model structures. The difference is likely to be partly due to the glycans which make up 5.8% of the total molecular weight of human serum transferrin (Spik *et al.*, 1988) and have not been included in the models. It may also be to some extent due to molecular aggregation. However, these factors will particularly modify the overall size of the molecule in solution (which is reflected

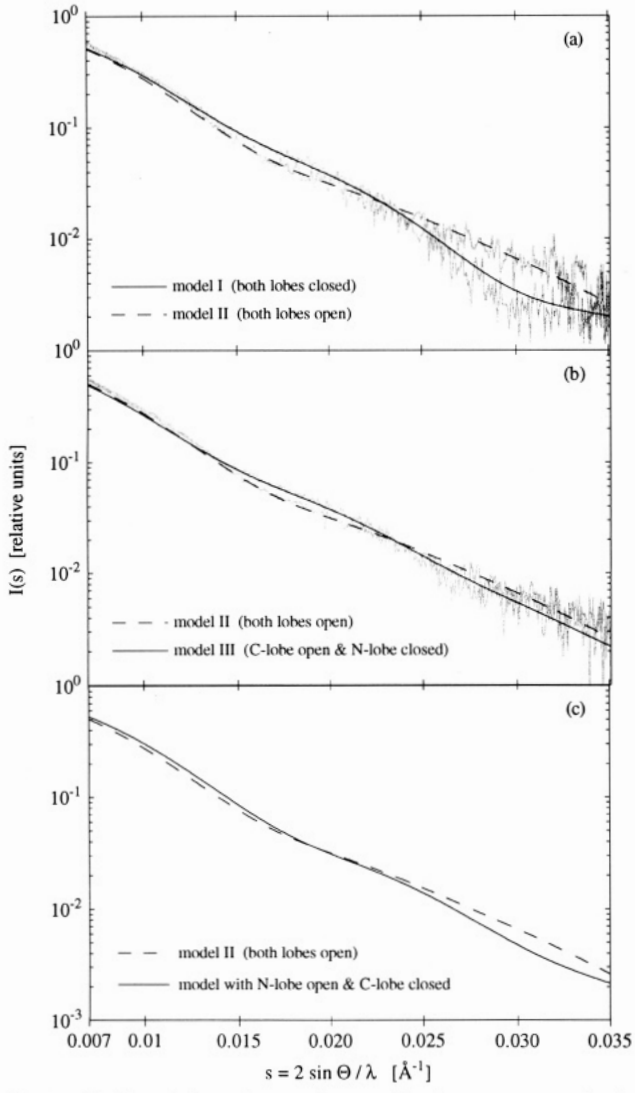

FIGURE 7: Simulation of scattering profiles for serum transferrin using energy-minimized and hydrated models of RST shown in Figure 6: (a) Comparison of curves resulting from model I and model II. (b) Curve for the fully opened conformation (model II) compared with the scattering pattern arising from model III. (c) Comparison of the profiles for the fully opened conformation (model II) and a model where the N-lobe remains open whereas the C-lobe is closed. Calculations were performed with all non-hydrogen atoms of rabbit serum transferrin including a hydration shell up to 3 Å from the molecular surface. Graphs a and b also contain the experimental results in light color corresponding to Figure 4.

in the low-angle scattering region) and will hardly affect the curvature of the scattering profiles in the higher angle region which is associated with the domain arrangement.

The simulations shown in Figure 7 clearly suggest that, in the C-lobe of the variant transferrin, the closure of the interdomain cleft upon iron binding does not take place. The simulation shown in Figure 7c and the changes in R_g (Table 2) exclude the possibility of a model where the N-lobe is open and the C-lobe is closed. The scattering data for the diferric variant protein is best simulated by an open C-lobe and a closed N-lobe molecule; thus the mutation Gly394 → Arg appears to have a decisive impact upon the domain closure in the C-lobe. This conformational behavior is rationalized at an atomic level in the following section.

Molecular Modeling of the Active Site in the Variant and Wild-Type Proteins. There are four protein ligands involved in iron coordination in the C-lobe, Asp392, Tyr426, Tyr517, and His585. In addition, the residues Thr452, Arg456, Ala458, and Gly459 are involved in anchoring the synergistic carbonate anion. Of these residues, Asp392 alone truly belongs to domain I. In view of the mutation observed for the variant protein, the feasibility of the formation of an ion pair between Asp392 and Arg394 has been examined. This problem was addressed in two ways, first by an examination of interacting Arg–Asp pairs observed in published crystal structures and second by molecular model building using computer graphics.

Table 3: Distribution of Interacting Ion Pairs in Function of Residue Order and Separation[a]

no. of intervening residues	interacting			non-interacting	
	Asp–Arg	Arg–Asp	total	Asp–Arg	Arg–Asp
0	6	11	17		
1	23	5	28	12	33
2	21	9	30		
3	10	10	20		
4	4	3	7		
5	1	1	2		
>5	82	80	162		

[a] Also shown are the number of non-interacting pairs for the case of a single intervening residue.

Table 4: Occurrence of Asp-X-Arg and Arg-X-Asp Tripeptides in Different Types of Secondary Structure

	helix	sheet	type I turn or N-cap	other	total
Asp-X-Arg (interacting)	0	3	12	8	23
Asp-X-Arg (non-interacting)	5	0	3	4	12
Arg-X-Asp (interacting)	0	1	2	2	5
Arg-X-Asp (non-interacting)	11	2	0	20	33

Table 3 shows the sequence preference (which of the two charged residues appears first in the primary structure) and separation for 266 interacting pairs of arginine and aspartic acid residues observed in 94 proteins (resolution better than 2.5 Å). The criterion for interaction was that any two atoms of the interacting residues should lie within a distance given by the sum of their van der Waals radii plus 1 Å. A relatively large percentage of the interacting pairs (11%) are of the form $(i, i + 2)$, involving only a single intervening residue. Of these, Asp-X-Arg type interactions (where X is any amino acid) outnumber Arg-X-Asp interactions by approximately 5 to 1. Table 3 also shows for this sequence separation the number of non-interacting pairs observed in the data set. Over 60% of the Asp-X-Arg tripeptides involve self-interactions between the aspartate and the arginine, but this is true of little more than 10% of Arg-X-Asp tripeptides. These data suggest that Asp-X-Arg interactions of the type assumed for the variant transferrin are both common and favorable but that this is not so for the reverse sequence Arg-X-Asp.

The backbone conformation of the interacting Asp-X-Arg tripeptides shows a preference to be in a type I turn or at the N-terminus of an α-helix (Richardson & Richardson, 1988; Schultz & Schirmer, 1979; see Table 4). Several workers have shown that there is a preference for aspartic acid at position i of a type I turn (Baker & Hubbard, 1984; Wilmot & Thornton, 1988) or at the N-cap position (Richardson & Richardson, 1988) and attribute this to the ability of one of the carboxylate oxygens of the side chain to hydrogen bond to the amide nitrogen of the backbone at $i + 2$. This brings the carboxylate group into a position to form further hydrogen bonds with the guanidinium moiety of an arginine at position $i + 2$. Such a stable arrangement would not be possible for the sequence Arg-X-Asp.

The residues Asp392 to Phe395 in transferrin have a conformation similar to that of a type I turn (Figure 8) which includes a hydrogen bond between the carbonyl of Asp392 and the amide of Phe395. This conformation is more

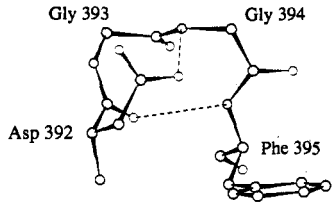

FIGURE 8: Residues Asp392 to Phe395 in transferrin form the first turn at the N-terminus of helix 3. Asp392 is in the N-cap position.

accurately described as having Asp392 at the N-cap position of an α-helix (helix 3), which is initiated by a turn which partially resembles a 3_{10} helix in that hydrogen bonds are formed between the carbonyl of 392 and the amides of both 395 and 396. A similar conformation is formed by the homologous residues (63–66) in the N-terminal lobe. One of the carboxylate oxygens is bound to the iron, while the other forms a hydrogen bond to the amide of Gly394. Given this conformation and the observed preference for the Asp-X-Arg interaction as described above, a model was constructed for the variant protein in this region which verifies the stereochemical feasibility of the proposed ion pair interaction between Asp392 and Arg394. Shown in Figure 9a, the model involves the aspartic acid being drawn away from the iron atom (compared to the wild type) and involved in strong bonds, one hydrogen bond to the backbone amide of Arg394, and the ion pair formation to the guanidinium group of the arginine residue, which itself involves two more hydrogen bonds in the preferred arrangement for Asp–Arg interactions (Singh, 1987).

Recently a G65R mutant of the half-molecule (N-lobe) has been reported (Woodworth *et al.*, 1991). This mutation is thus the N-lobe equivalent of that observed in the variant protein. Unlike the N-lobe mutants of the ligands Asp63 (Woodworth *et al.*, 1991) and Tyr95 (J. B. Crawley et al., in preparation), this does not show a significant blue shift in the optical spectrum, suggesting only a minor disruption in the coordination sphere. In particular, the similarity of the mutant's spectral properties to those of the wild type suggests that there is no reason to believe that the aspartic acid would not be bound to the iron and that it would therefore be expected to show a closed conformation in the iron-bound form. An examination of a 12-Å sphere around G65R and its equivalent residue in the C-lobe (G394R) reveals an interesting difference (Figure 9a,b,c). In the case of the N-lobe, Glu69 is in the near vicinity of G65R and it is feasible for the two residues to form a salt bridge in the apo conformation, thus leaving Asp63 free to coordinate to the iron atom in the normal fashion. This region of the transferrin molecule forms part of helix 3. Such salt bridges of the type $(i, i + 4)$ are possible in α-helices due to the fact that the two side chains appear on the same side of the helix (Sundaralingam *et al.*, 1987; Margusee & Baldwin, 1987; Lyu *et al.*, 1992).

We note that despite the lack of a significant blue shift, the G65R mutant protein behaves like the apoprotein and mutants of Asp63 on electrophoresis in the presence of 6 M urea (Woodworth *et al.*, 1991). Despite the proposed similar coordination of the aspartic acid in the N-lobe G65R mutant, modeling studies clearly demonstrate that it is not possible for the N-lobe to close around the iron in the "normal" way. This is due to the introduction of a large side chain in place of the normal glycine at position 65, which produces steric hindrance as a result of close contacts with residue Gly123 of domain II (Figure 9c and d). It therefore seems reasonable that the instability of the G65R mutant on 6 M urea electrophoresis may be due to incomplete closure of the lobe on iron binding while retaining normal iron coordination. A similar explanation is not applicable in the case of the variant protein because the equivalent residue to Glu69 in the C-lobe is Ile398, which cannot form the salt bridge with Arg394 (Figure 9a). In this case, as described above, we propose that the arginine would interact with the aspartic acid ligand thus making it unavailable for iron binding and resulting in an open configuration for the lobe.

DISCUSSION

We have shown that the variant transferrin contains an amino acid substitution at position 394 at which the glycine residue is replaced by an arginine. The crystal structures of both rabbit serum transferrin (Bailey *et al.*, 1988) and human lactoferrin (Anderson *et al.*, 1989) suggest that Gly394 is not directly involved in iron binding, nor does it seem likely to be involved in receptor binding given that it is buried deep within the cleft. The reduced affinity for iron binding appears to be that the arginine introduced at position 394 interacts with the neighboring Asp392, itself a ligand. The inability of the aspartate ligand to bind to iron in the variant protein is consistent with the observed blue shift in the visible spectrum of the C-site (Evans *et al.*, 1982) as judged by qualitative comparison with the mutants of Asp63 of the N-lobe protein (Woodworth *et al.*, 1991) and with an 18-kDa yellow fragment of duck ovotransferrin (Evans & Madden, 1984), which is known to lack the aspartate ligand, and from studies of model compounds which suggest a correlation between blue shift and basicity of the non-phenolate ligands (Ainscough *et al.*, 1980).

What remains to be explained is why a mutation in the interior of the molecule should result in a reduction in receptor binding. The solution X-ray scattering results demonstrate that the C-lobe is open in the diferric variant. Mutants of the N-terminal half-molecule of human serum transferrin revealed that the closure of the interdomain cleft does not take place when iron is bound in D63S, D63C, (Woodwoth *et al.*, 1991; Grossmann *et al.*, 1993b), Y95H, and H249Y (Crawley, 1993; J. B. Crawley *et al.*, in preparation), which also show a blue shift consistent with the loss of aspartate coordination. The binding of aspartate to iron has been shown to be linked with the trigger for closing the interdomain cleft (Grossmann *et al.*, 1993b); its loss as an iron ligand may, as in this case, lead to the failure of domain closure and thus give a diferric species with structural topology essentially identical to that of an N-terminal monoferric species, itself having a 10-fold lower affinity for the receptor (Young *et al.*, 1984b). This difference in topology may also explain why the variant protein in its iron-loaded form elutes from an anion-exchange chromatography column later than the normal protein.

We note that the variant protein and the N-terminal monoferric species have a receptor binding affinity similar to that of the C-terminal monoferric protein (Young *et al.*, 1984b). However, some recent data indicate that the primary receptor recognition site of human serum transferrin may only reside in the C-lobe (Zak *et al.*, 1993). In this case, the N-terminal monoferric species, with its C-lobe open, would be expected to have an affinity comparable to that of the apoprotein, which is some 30-fold lower than that of the diferric species (Young & Aisen, 1981; see also Table 1). This raises the question as to whether site–site cooperativity leads to a small conformational change perhaps through the interlobe connecting peptide allowing the C-terminal half to bind more tightly to the receptor. Alternatively, the receptor recognition elements may not be confined to the C-terminal lobe of the protein and may also involve motifs present in the N-lobe. The data provided here supplies new evidence suggesting that

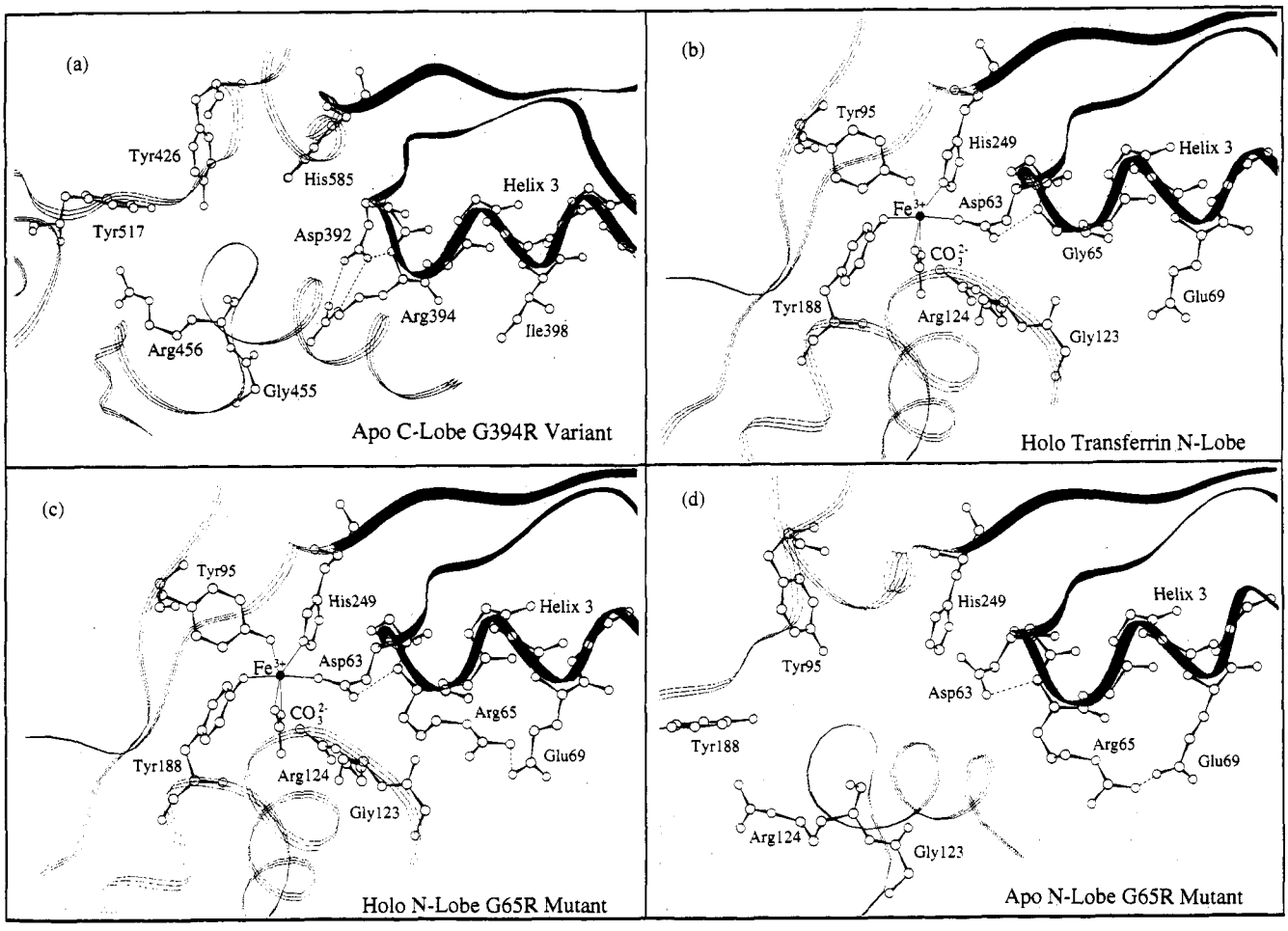

FIGURE 9: Structural models for (a) the C-lobe binding site of the variant protein G394R in the apo open conformation [modeled on the N-lobe of apolactoferrin (Anderson *et al.*, 1990)], (b) the normal human N-lobe in the iron-bound conformation [coordinates taken directly from the crystal structure of the N-lobe of rabbit serum transferrin (Sarra *et al.*, 1990)], (c) the N-lobe binding site of the G65R mutant in the iron-bound conformation (based on the N-lobe of rabbit serum transferrin), and (d) the N-lobe binding site of the G65R mutant in the apo conformation (based on the N-lobe of apolactoferrin). In (a) Arg394 forms a salt bridge with Asp392 covering the N-terminus of helix 3. This interaction disfavors the binding of iron to the CI domain since Asp392 is normally one of the iron-binding ligands. As a consequence the domains are expected to remain open on iron binding with the metal bound to Tyr426, Tyr517, the carbonate anion (not shown), and possibly His585. The consequent alteration in the iron ligand field is expected to be the source of the considerably altered spectroscopic properties. Incorrect closure of the lobes is also expected due to steric hindrance of the Arg394 side chain (from domain CI) with the carbonyl of Gly455 (from domain CII). This is normally avoided because of the conserved glycine at 394 and its homologous position in the N-termial lobe (b). In (c) the effect of a similar mutation (G65R) made in the N-lobe of human transferrin is shown. In this case, however, the arginine can form a salt bridge with Glu69, which is separated by approximately one turn of helix 3. As such, the aspartic acid ligand in the N-lobe (Asp63) remains free to coordinate iron, leading to a similar ligand field and as a consequence to the observed spectroscopic properties, which do not differ greatly from those of the native protein. A similar bridge is not possible in the variant protein, as the equivalent position in the C-lobe is occupied by Ile398; see (a). However, it is not possible for the two domains of the G65R variant to close around the iron in the normal fashion [as has been depicted in (c)] due to the shock between the arginine side chain and the carbonyl of Gly123. Thus, while it is imagined that the full complement of normal ligands participates in iron binding, it is assumed that domain closure is somewhat different from normal, leading to destabilization of the iron-bound conformation, which explains the abnormal migrational properties of the mutant in urea gel electrophoresis. Panel d shows that in the apo conformation the problem of steric hindrance no longer exists. Pictures were produced with the molecular graphics software Insight II (Biosym Technologies, Inc., San Diego, CA). Domain separation has been stressed by the filled and lined ribbon for domains I and II, respectively.

the closed conformation of both lobes plays a functional role in receptor recognition.

ACKNOWLEDGMENT

We thank Miss P.C. for her cooperation in this work. We are grateful to Mr. Trevor Jones of the U.K. Transfusion Service (UKTS) for the lymphocyte transformations and Dr. Philip Judson (UKTS) for making plasma samples available to us. The assistance of Mr. Alan Harris (NIMR) in protein sequencing is acknowledged. We are thankful to Professors Ted Baker and Peter Lindley for providing coordinates of lactoferrin and transferrin, respectively, prior to their deposition in the Protein Data Bank. We would like to acknowledge the help and support of our colleagues at the respective institutes.

REFERENCES

Abola, A., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., & Sievers, R., Eds.) pp 107–132, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester.

Ainscough, E. W., Brodie, A. M., Plowman, J. E., Bloor, S. J., Loehr, J. S., & Loehr, T.M. (1980) *Biochemistry 19*, 4072–4079.

Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W., & Baker, E. N. (1989) *J. Mol. Biol. 209*, 711–734.

Anderson, B. F., Baker, H. M., Norris, G. E., Rumball, S. V., & Baker, E. N. (1990) *Nature (London) 344*, 784–787.

Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S. S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A.,

Sarra, R., & Watson, J. L. (1988) *Biochemistry 27*, 5804–5812.
Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Molec. Biol. 44*, 97–179.
Baldwin, D. A., & de Sousa, D. M. R. (1981) *Biochem. Biophys. Res. Commun. 99*, 1101–1107.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, F. M., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Bomford, A., Young, S. P., Nouriaria, K., & Williams, R. (1983) *Br. J. Haematol. 55*, 93–101.
Bomford, A., Isaac, J., Roberts, S., Edwards, A. Young, S. P., & Williams, R. (1986) *Biochem. J. 236*, 243–249.
Boulin, C., Kempf, R., Koch, M. H. J., & McLaughlin, S. (1986) *Nucl. Instrum. Methods Phys. Res. A249*, 399–407.
Bras, W., Derbyshire, G. E., Ryan, A. J., Mant, G. R., Felton, A., Lewis, R. A., Hall, C. J., & Greaves, G. N. (1993) *Nucl. Instrum. Methods Phys. Res. A326*, 587–591.
Cambillau, C. C., & Horjales, E. (1987) *J. Mol. Graphics 5*, 174–176.
Crawley, J. B. (1993) Ph.D. Thesis, University of London.
Debye, P. (1915) *Ann. Phys. (Leipzig) 46*, 809–823.
Evans, R. W., & Holbrook, J. J. (1975) *Biochem J. 145*, 201–207.
Evans, R. W., & Williams, J. (1978) *Biochem. J. 173*, 543–552.
Evans, R. W., & Madden, A. D. (1984) *Biochem. Soc. Trans. 12*, 661–662.
Evans, R. W., Williams, J., & Moreton, K. (1982) *Biochem J. 201*, 19–26.
Franklin, S. G., Wolf, S. I., Ozdemir, Y., Yuregir, G. T., Isbir, T., & Blumberg, B. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5480–5482.
Funk, W. D., MacGillivray, R. T. A., Mason, A. B., Brown, S. A., & Woodworth, R. C. (1990) *Biochemistry 29*, 1654–1660.
Garratt, R. C., Evans, R. W., Hasnain, S. S., Lindley, P. F., & Sarra, R. (1991) *Biochem J. 280*, 151–155.
Geisow, M. J., & Aitken, A. (1989) Gas or pulsed-phase sequence analysis, in *Protein Sequencing: A practical approach* (Geisow, M. J., & Findlay, J. B. C., Eds.) pp 85–98, IRL Press, Oxford.
Grossmann, J. G., Neu, M., Pantos, E., Schwab, F., Evans, R. W., Townes-Andrews, E., Lindley, P. F., Appel, H., Thies, W.-G., & Hasnain, S. S. (1992) *J. Mol. Biol. 225*, 811–819.
Grossmann, J. G., Abraham, Z. H. L., Adman, E. T., Neu, M., Eady, R. R., Smith, B. E., & Hasnain, S. S. (1993a) *Biochemistry 32*, 7360–7366.
Grossmann, J. G., Mason, A. B., Woodworth, R. C., Neu, M., Lindley, P. F., & Hasnain, S. S. (1993b) *J. Mol. Biol. 231*, 554–558.
Guinier, A., & Fournet, G. (1955) *Small-Angle Scattering of X-rays*, John Wiley & Sons, Inc., New York.
Jeffries, A. J., & Flavel, R. A. (1977) *Cell 12*, 429–439.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Lewis, R., Sumner, I., Berry, A., Bordas, J., Gabriel, A., Mant, G., Parker, B., Roberts, K., & Worgan, J. (1988) *Nucl. Instrum. Methods Phys. Res. A273*, 773–777.
Lyu, P. C., Gans, P. J., & Kallenbach, N. R. (1992) *J. Mol. Biol. 223*, 343–350.
MacGillivray, R. T., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J., & Brew, K. (1983) *J. Biol. Chem.* **258**, 3543–3553.
Margusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5286–5290.
Pantos, E., & Bordas, J. (1994) *Pure Appl. Chem. 66*, 77–82.
Park, I., Schaeffer, E., Sidoli, A., Baralle, E., Cohen, G. N., & Zakin, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3149–3153.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Sarra, R., Garralt, R., Gorinsky, B., Jhobi, H., & Lindley, P. (1990) *Acta Crystallogr. B46*, 763–771.
Schultz, G. E., & Schirmer, R. H. (1979) *Principle of Protein Structure*, Springer-Verlag, New York.
Singh, J., & Thornton, J. M. (1990) *J. Mol. Biol. 211*, 595–615.
Singh, J., Thornton, J. M., Snarey, M., & Campbell, S. F. (1987) *FEBS Lett. 224*, 161–171.
Spik, G., Coddeville, B., & Montreuil, J. (1988) *Biochimie 70*, 1459–1469.
Sundaralingam, M., Sekharudu, Y. C., Yathindra, N., & Ravichandran, V. (1987) *Int. J. Quantum Chem., Quantum Biol. Symp. 14*, 289–296.
Sutton, M. R., & Brew, K. (1973) *Biochem. J. 139*, 163–168.
Tomimatsu, Y., & Donovan, J. W. (1976) *FEBS Lett. 71*, 299–302.
Wilmot, C. M., & Thornton, J. M. (1988) *J. Mol. Biol. 203*, 221–232.
Woodworth, R. C., Mason, A. B., Funk, W. D., & MacGillivray, R. T. A. (1991) *Biochemistry 30*, 10824–10828.
Young, S. P., & Aisen, P. (1981) *Hepatology 1*, 114–119.
Young, S. P., Bomford, A., Madden, A. D., Garratt, R. C., Williams, R., & Evans, R. W. (1984a) *Br. J. Haematol. 56*, 581–587.
Young, S. P., Bomford, A., & Williams, R. (1984b) *Biochem. J. 219*, 505–510.
Zak, O., Trinder, D., & Aisen, P. (1993) *11th International Conference on Iron and Iron Proteins*, Jerusalem, Israel, Abstract, p 29.