Molecular Structure of Serum Transferrin at 3.3-Å Resolution†

Susan Bailey,‡ Robert W. Evans,§ Richard C. Garratt,‡ Beatrice Gorinsky,‡ Samar Hasnain,∥ Christopher Horsburgh,‡ Harren Jhoti,‡ Peter F. Lindley,*,‡ Assanah Mydin,‡ Robert Sarra,‡ and John L. Watson‡

Department of Crystallography, Birkbeck College, University of London, Malet St., London WC1E 7HX, U.K., United Medical and Dental Schools Division of Biochemistry, Guy's Hospital, London SE1 9RT, U.K., and SERC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, U.K.

Received December 17, 1987; Revised Manuscript Received March 22, 1988

ABSTRACT: Serum transferrin is a metal-binding glycoprotein, molecular weight ca. 80 000, whose primary function is the transport of iron in the plasma of vertebrates. The X-ray crystallographic structure of diferric rabbit serum transferrin has been determined to a resolution of 3.3 Å. The molecule has a $\beta\alpha$ structure of similar topology to human lactoferrin and is composed of two homologous lobes that each bind a single ferric ion. Each lobe is further divided into two dissimilar domains, and the iron-binding site is located within the interdomain cleft. The iron is bound by two tyrosines, a histidine, and an aspartic acid residue. The location of the 19 disulfide bridges is described, and their possible structural roles are discussed in relation to the transferrin family of proteins. Mapping of the intron/exon splice junctions onto the molecule provides some topological evidence in support of the putative secondary role for transferrin in stimulating cell proliferation.

Various mammalian transferrins have been identified [for recent reviews see Brock (1985) and Huebers and Finch (1987)], which include serum transferrin in blood plasma, lactoferrin found in milk, other secretory fluids and leucocytes, ovotransferrin found in egg white, and, most recently, melanotransferrin in human melanoma cells (Rose et al., 1986). All the transferrins, with the possible exception of melanotransferrin, sequester and solubilize iron, thereby controlling "available" iron levels with the attendant bacteriostatic and conservation effects, but, in addition, serum transferrin is the principal iron-transporting protein of vertebrates. The protein is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilization, primarily by the hemopoietic system. Rapid iron transfer to cells is effected when the protein binds to specific cell surface receptors on which it is internalized, the iron removed, and the apoprotein recycled into the blood stream. Serum transferrin may also have a further role in stimulating the growth of cells, which is unrelated to its transport function (Trowbridge & Lopez, 1982; James & Bradshaw, 1984).

Spectroscopic studies have long suggested that the nature of iron coordination in the transferrin family of proteins is unusual. In particular, the synergistic binding of ferric iron and the anion is believed to be unique in biology. Recently, the crystal structure analysis of human lactoferrin at a resolution of 3.2 Å has been reported (Anderson et al., 1987). However, the lactoferrins have primarily a bacteriostatic role and do not appear to function as iron donors to general body tissues. They differ in chemical structure and immunological properties from the serum transferrins.

We report here the first X-ray structural analysis of a serum transferrin, at a resolution of 3.3 Å. The overall organization of the rabbit transferrin molecule in terms of its secondary and tertiary structure is described. The topological fold of the molecule is very similar to that reported for human lactoferrin (Anderson et al., 1987). At present, only an incomplete amino acid sequence is available for rabbit serum transferrin, and the assignment of the ligands at the iron binding sites is therefore based on arguments of homology between transferrins and the published structure for human lactoferrin. The distribution of the disulfide bridges and their possible structural roles are discussed. Mapping of the exon distribution to the structure has resulted in the identification of a supersecondary structural unit that correlates with exon 2. This region of transferrin has been reported to be homologous to the lymphocyte-derived transforming proteins ChBlym-I (Goubin et al., 1983) and HuBlym-I (Diamond et al., 1984), and we suggest it may control serum transferrin's possible secondary role as a growth stimulant.

Experimental Procedures

Diferric serum transferrin was isolated from rabbit plasma and crystallized (Al-Hilal et al., 1976) in space group $P4_32_12$ with cell dimensions $a = b = 127.5$ (1) Å, $c = 144.8$ (2) Å and eight molecules per unit cell. The crystals have a solvent content of 68% (v/v) and are unstable at room temperature; all X-ray studies were carried out in the temperature range 4–6 °C.

The structure was solved by using a combination of multiple isomorphous replacement (MIR) and solvent-flattening techniques. Three heavy atom derivatives were prepared by soaking the native crystals in (i) 8 mM $HgCl_2$ for 24 h, (ii) 10 mM UO_2Ac_2 for 9 days, and (iii) 12 mM $NaAuCl_4$ for 3 days. Three-dimensional intensity data for the native protein and the mercuric chloride derivative were collected by using Ni-filtered copper radiation, $\lambda = 1.5418$ Å, produced by an Elliott GX1 rotating anode generator operated at 35 kV and 40 mA. Data for the uranyl acetate and chloroaurate derivatives were collected on the SERC synchrotron radiation

† Support provided by the Medical Research Council, United Kingdom, for S.B., B.G., and H.J. and by the Science and Engineering Research Council (SERC), United Kingdom, for R.C.G. Facilities at the SRS, Daresbury, where some of the X-ray data measurement and computations were undertaken, were made available by the SERC. This research is currently supported by the SERC and The Wellcome Trust, United Kingdom.

‡ Birkbeck College.
§ Guy's Hospital.
∥ SERC Daresbury Laboratory.

0006-2960/88/0427-5804$01.50/0

Table I: Final Multiple Isomorphous Replacement Statistics

(A) Refinement of Heavy Atom Sites[a]

atom	occupancy (relative)	x	y	z	atom	occupancy (relative)	x	y	z
$HgCl_2$					UO_2Ac_2				
Hg 1	0.925	0.125	0.485	0.176	U 1	0.674	0.245	0.230	0.093
Hg 2	0.268	0.278	0.030	0.186	U 2	0.237	0.392	0.012	0.451
Hg 3	0.176	0.231	0.243	0.283	U 3	0.177	0.081	0.351	0.443
$NaAuCl_4$					U 4	0.178	0.251	0.018	0.388
Au 1	0.450	0.126	0.486	0.180	U 5	0.110	0.222	0.301	0.111
Au 2	0.301	0.273	0.024	0.171					
Au 3	0.223	0.140	0.305	0.374					
Au 4	0.418	0.431	0.084	0.386					

(B) Phase Refinement Statistics as a Function of Resolution[b]

	resolution (Å)									
	14.74	8.51	6.59	5.57	4.91	4.45	4.09	3.81	3.58	3.38
$HgCl_2$										
N_{hkl}	510	1084	1389	1643	1840	2011	2188	2348	2480	2505
F_H	210	203	192	181	172	162	153	146	137	129
E	128	115	90	99	110	109	102	94	84	77
UO_2Ac_2										
N_{hkl}	526	1095	1393	1650	1854	2041	2219	2382	2514	2530
F_H	201	182	167	157	148	142	135	129	123	117
E	148	141	115	123	142	156	143	125	108	82
$NaAuCl_4$										
N_{hkl}	529	1079	1349	1619	1815	2008	2174	2337	2467	2474
F_H	144	146	139	132	125	118	112	105	101	94
E	121	126	98	113	131	135	126	113	101	90
FOM	0.73	0.72	0.66	0.64	0.59	0.59	0.55	0.54	0.53	0.52

[a] Only centric zone reflections were used; all atoms were assigned an isotropic temperature factor, $U = 20$ Å^2, but these were not refined. [b] F_H is the root mean squared heavy atom structure factor, E is the root mean squared lack of closure, and FOM is the average figure of merit.

source at Daresbury (2 Gev and average ring currents of 100 mA) on beam-line 7.2 and with $\lambda = 1.488$ Å. The effective resolutions for the native protein data and the heavy atom derivatives were 3.0 and 3.3 Å, respectively.

In each case crystals were mounted about the *a* axis to minimize crystal slippage and 1° through 90° oscillations recorded on an Arndt-Wonnacott camera with crystal to film distances in the range 90–100 mm. Crystal stability typically allowed 6–8° of data to be collected on the rotating anode source (exposure time 20 000 s/deg), but 12–20° was possible with the synchrotron beam (exposure times of 100 s/deg). For the native protein data 14 crystals (96 film packs) enabled some 195 700 measurements to be made to 3.3 Å, and subsequent merging of these data yielded 18 257 unique *hkl* values with a conventional merging *R* factor of 0.076 on intensities; 23, 9, and 5 crystals were used respectively for the $HgCl_2$, UO_2Ac_2, and $NaAuCl_4$ derivatives, giving values of R_{merge} of 0.106, 0.151, and 0.124.

The heavy atom sites were located from difference Patterson and Fourier syntheses and refined by using only the centric reflections; in each case isotropic thermal parameters were held constant at 20 Å^2. Protein phases were calculated from the final parameters by standard methods to give an overall figure of merit of 0.62; the final multiple isomorphous replacement statistics are listed in Table I. The "best" phases and figures of merit were used to calculate an electron density synthesis at 3.3-Å resolution.

Although the resulting map clearly enabled the identification of the molecular boundary and allowed some of the elements of secondary structure to be putatively assigned, it remained impossible to unambiguously trace the polypeptide chain. In the N-lobe a high electron density peak was identified as an iron atom, but this was less easy to identify in the C-lobe. However, the iron atom sites were readily confirmed from a Fourier synthesis computed by using the native protein data, collected with copper Kα radiation, phase angles $\alpha_p - \pi/2$, and coefficients $||F_{p+}| - |F_{p-}||$, where the + and − signs refer to Freidel related reflections recorded on the same oscillation film in order to minimize errors due to absorption and scaling. Figure 1 shows sections of this anomalous difference density map through the two highest peaks (more than twice the height of other peaks), which correspond to the N- and C-lobe iron atoms, respectively.

In order to improve the quality of the MIR electron density map, solvent-flattening techniques were employed. Initially, a solvent flattening technique where the molecular boundary was determined manually was used. All solvent density was set to zero, all negative regions within the molecular envelope were multiplied by 0.1, and the resulting map was back-transformed into reciprocal space. Combined phases (Bricogne, 1976) were used to calculate a new map, and the cycle was repeated to convergence. After five cycles, the mean differences in the MIR and combined phases and the MIR and back-transformed phases were 36° and 63°, respectively. This procedure led to a much improved map in which the secondary structure was better defined, particularly the helical regions, and many connections were clarified.

In addition, Leslie's reciprocal space modification (1987) to the method of Wang (1985) was employed to calculate further solvent-flattened maps. Averaging spheres of radii of 8 and 10 Å with estimated solvent contents of 55% and 65% (v/v) were used, leading to a series of maps of very similar quality. Typically, four cycles were required for convergence with mean overall phase differences from the MIR phases of some 34°. However, none of these maps showed as well-defined secondary structure as obtained previously, although many of the connecting surface loops were more readily traced.

At this stage the structure of human lactoferrin was made available to us prior to publication (Baker, private communication), thus appreciably accelerating the process of interpreting the electron density map in terms of the polypeptide backbone. Examination of our original MIR map together with the two types of solvent-flattened maps enabled an almost complete interpretation of the molecular configuration. A model structure was built, by using first the automatic chain tracing program BONES (Jones & Thirup, 1986) and then a

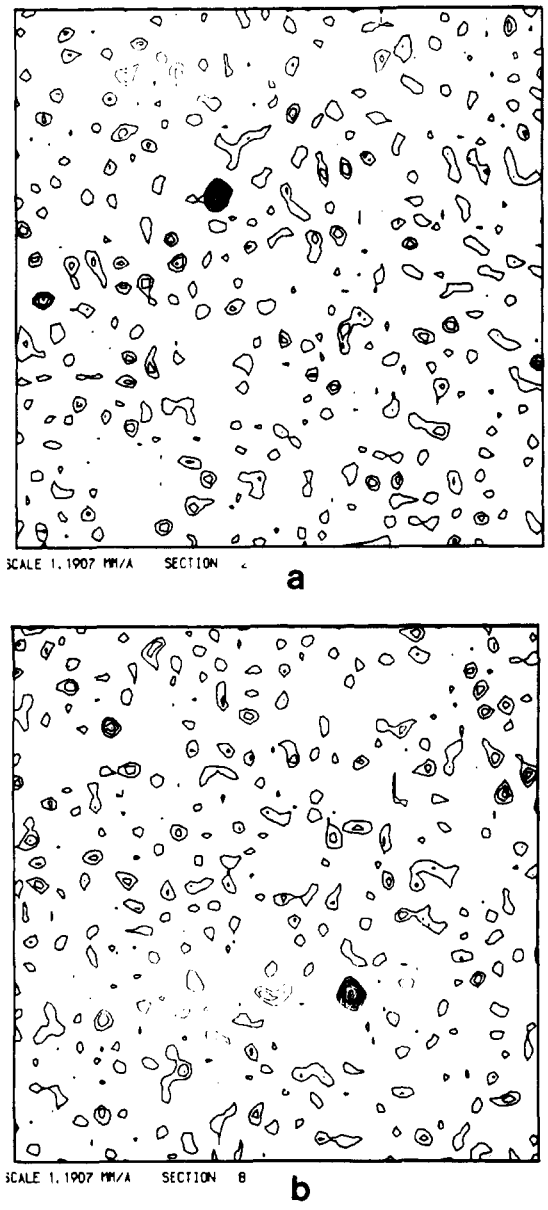

FIGURE 1: Location of the iron atoms in diferric rabbit serum transferrin. (a) and (b) are sections through the N- and C-terminal iron atom positions, respectively, from an electron density map computed by using Δ_{ano} as coefficients and phases $[\alpha_p - \pi/2]$. The Δ_{ano} were derived from intensity data collected photographically on an oscillation camera with Cu Kα radiation.

fragment-fitting facility, both of which were available in FRODO (Jones, 1978), version 6.0, implemented on an Evans and Sutherland PS300 interactive computer graphics system. To date, only a little of the primary structure of rabbit serum transferrin has been reported. However, complete sequences are available for four other members of the transferrin family, namely, human lactoferrin (Metz-Boutigue et al., 1984), human serum transferrin (MacGillivray et al., 1982, 1983; Yang et al., 1984), hen ovotransferrin (Jeltsch & Chambon, 1982; Williams et al., 1982), and human melanotransferrin (Rose et al., 1986). Of these, human serum transferrin is expected to most closely resemble the rabbit serum protein (vide infra regarding, for example, the disulfide bridging pattern). In building the molecular structure into the electron density maps we therefore decided to adhere to the human serum sequence wherever possible. Small insertions and deletions were included only in regions that were poorly conserved areas in all four sequences and where it was necessary in order to adequately account for the electron density. Side chains based on the human serum sequence were included for residues that showed well-defined electron density.

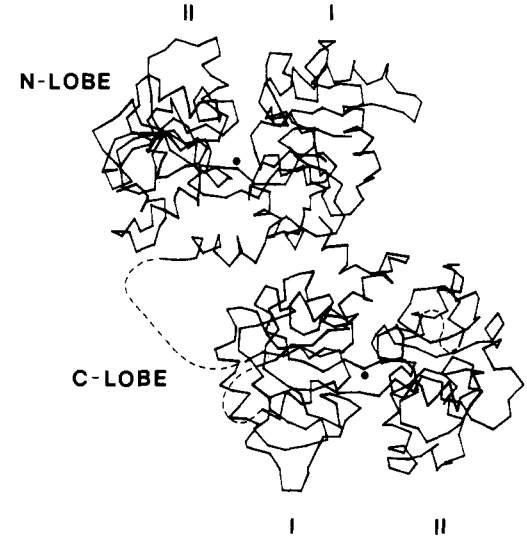

FIGURE 2: Overall organization of the rabbit serum transferrin molecule. The polypeptide chain is folded into two lobes, each containing some 330 amino acids and a single iron-binding site. The shape of each lobe can be described by a prolate ellipsoid of approximate semiaxial dimensions 21 × 25 × 35 Å with the major axes of the N- and C-lobes running antiparallel to one another at an angle of 155°. In turn each lobe is comprised of two dissimilar domains, I and II. Broken lines denote poorly defined regions in the current model. Picture produced by using a computer program written by A. M. Lesk and K. D. Hardman (1982, 1985).

Refinement of this model structure, initially using rigid-body least-squares refinement (RESTRAIN; Moss & Morphew, 1982) followed by computer graphics model building, is now in progress.

RESULTS

Overall Organization of the Molecule. The overall organization of the serum transferrin molecule is shown in Figure 2. The polypeptide chain is folded into two lobes, each containing some 330 amino acids and a single iron-binding site. The shape of each lobe can be described by a prolate ellipsoid (Taylor et al., 1983) of approximate semiaxial dimensions 21 × 25 × 35 Å with the major axes of the N- and C-lobes running almost antiparallel to one another. Each lobe is comprised of two dissimilar domains with the iron-binding site situated at the domain interface. The polypeptide chain within the two lobes can be superposed (Sutcliffe et al., 1987) by a rotation of approximately 167° with a translation of 23 Å to give a rms deviation of 1.6 Å for 147 pairs of equivalent α-carbon atoms (Figure 3). The helices and strands superpose rather well, whereas the major differences lie in the loops connecting these elements of secondary structure. A similar folding pattern of secondary structure is thus observed within the N- and C-lobes, consistent with a gene duplication event (Williams, 1982a).

Each domain consists of a mixed right-handed β-pleated sheet in which the strands are connected through loops and helices. Figure 4 is a schematic drawing of the secondary structure elements in the N-lobe. In domain I the sheet has six strands, *a*, *b*, *c*, and *d* pointing toward the second domain and *j* and *k* running antiparallel. Strands *a* and *b* together with helix 1 form a $\beta\alpha\beta$ motif. In domain II the sheet has five strands, *f*, *g*, *h*, and *i* orientated toward domain I with *e* antiparallel. Domain II is thus formed from a contiguous region of the polypeptide chain, residues 97–245 inclusive, in contrast to domain I, which is composed of two regions sep-

FIGURE 3: Superposition of the N- and C-terminal lobes. A similar folding pattern of secondary structure is observed within the N- and C-terminal lobes, consistent with a gene duplication event. The polypeptide chain within the two lobes can be superposed (Sutcliffe et al., 1987) by a rotation of approximately 167° coupled with a translation of 23 Å. Initially, 88 topologically equivalent α-carbon atoms were chosen from helices 2, 3, 5, 7, and 8 (see Figure 4), but iteration gave a best fit of 1.6 Å for 147 pairs of atoms by including all equivalent α-carbon atoms within 3.5 Å. Most of the helices superpose rather well, but it is the β-sheets, which compose the core of the domains, that give the best fit.

arate in the primary structure.

Nature of the Iron-Binding Sites. Spectroscopic measurements on the iron-containing protein and on the metal-substituted protein have implicated tyrosine and histidine residues in metal binding together with the synergistic anion (Schlabach & Bates, 1975). It has also been proposed that arginine and/or histidine is responsible for anchoring the anion to the protein [see Brock (1985) and references cited therein]. These results are broadly consistent with chemical modification experiments (Rogers et al., 1977, 1978; Williams, 1982b) and with studies based on model compounds (Pecoraro et al., 1981). However, different techniques have often led to disagreement over the numbers of each type of ligand involved. This is probably a consequence of the use of substitute metals that may occupy slightly different coordination sites to iron and the failure of chemical modification agents to access buried residues not directly involved in metal binding.

Figure 5 is a stereo view of the C_α chain in the N-terminal lobe of the rabbit serum protein, but with the addition of the five residues implicated in iron binding in the human lactoferrin molecule (Anderson et al., 1987). Domain I provides only one of these five protein residues, Asp-63(392 in the C-lobe),[1] on the loop between strand *c* and helix 3. Domain II contributes two residues, Arg-124(456)[1] at the N-terminus of helix 5 and Tyr-188(517)[1] at the N-terminus of helix 7. The arginine residue, which is conserved throughout the transferrins, may serve to anchor the (bi)carbonate anion to the cation, but there is no clear evidence for this at this stage in the analysis of rabbit serum transferrin. The remaining two residues, Tyr-95(426)[1] and His-249(585),[1] arise from β-strands *e* and *i*, respectively, which form a bridge between the two domains. These five protein residues are oriented toward the iron atom so that a water molecule could readily occupy a sixth coordination site within the cleft between the two domains. Evidence for this water molecule has been provided by NMR (Koenig & Schillinger, 1969), EXAFS, and EPR spectroscopy (Hasnain et al., 1987).

In the iron-binding cleft several other amino acid side chains lie in the neighborhood of the iron atom but do not appear to be directly involved in iron coordination. In particular, Lys-206 (534 in the C-lobe), situated on the loop between strand *h* and helix 8, approaches within 6 Å of the iron position on its opposite side from Arg-124. Figure 6 shows the electron density and its interpretation at the iron-binding site in the C-lobe for the rabbit serum molecule. Differential kinetic labeling experiments (Shewale & Brew, 1982) have indicated that this lysine undergoes a considerable change in reactivity

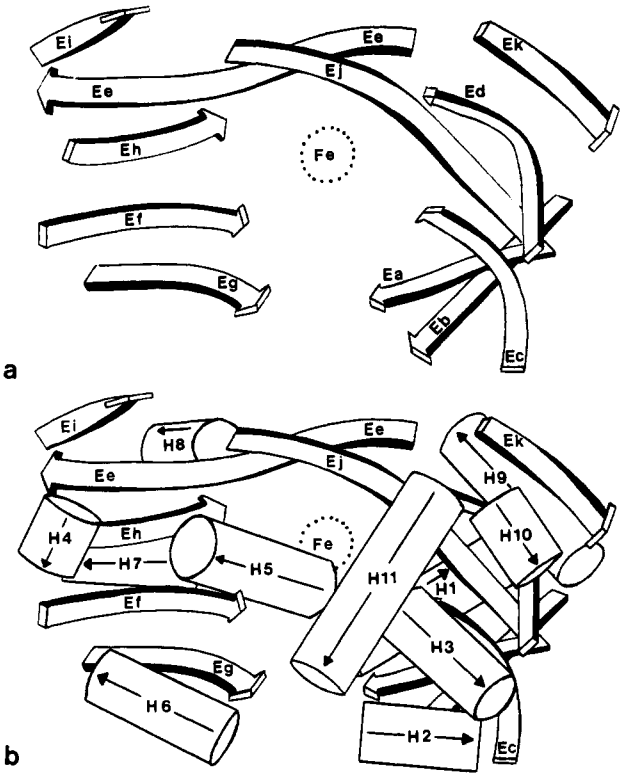

FIGURE 4: Secondary structure fold; schematic diagram of the overall topology in the N-lobe. A similar fold is observed in the C-lobe. In this figure the N-lobe is rotated about a horizontal axis in the plane of the figure by about 90° with respect to Figures 2 and 3. Each domain consists of a mixed β-sheet (a) with the strands connected by helices (b) and loops. Domains I and II contain 6- and 5-stranded β-sheets, respectively. The chain starts in domain I and initially folds to form the first four strands in a β-sheet: *a*, *b*, *c*, and *d*, which are connected through helices 1, 2, and 3, respectively. The chain then traverses into domain II where it folds into a β-sheet consisting of strands *e*, *f*, *g*, *h*, and *i* connected through helices 4, 5, 6, 7, and 8. After strand *i*, the chain returns to domain I to complete the β-sheet in this domain with strands *j* and *k* connected through helix 9. Helices 10 and 11 complete the fold in the N-lobe. The chain is anchored to domain II by disulfide bridge 10 prior to entering the C-lobe via the connecting peptide.

on iron binding. While this lysine is present in both lobes of human serum transferrin, it is not completely conserved across all transferrins. Its replacement by an arginine in the N-lobe of human lactoferrin and by a glutamine in the C-lobe of chicken ovotransferrin may be relevant to the observed chemical and spectroscopic differences between the iron-binding sites of the different molecules.

Disulfide Bridging Pattern. The transferrins share 15 disulfide bridges in common. The most ancient, from an evolutionary standpoint, are numbered 1–6 and occur in both lobes

[1] The numbering of the amino acid residues is according to the scheme for human serum transferrin. Equivalent residues for human lactoferrin are Asp-61, Arg-121, Tyr-93, Tyr-191, and His-252.

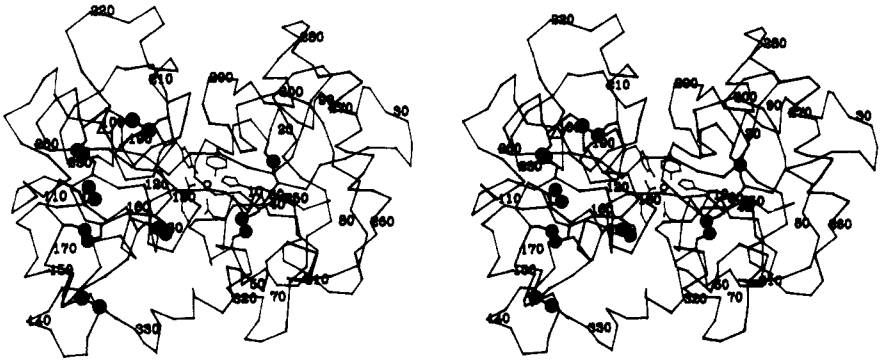

FIGURE 5: Stereo drawing of the α-carbon polypeptide chain in the N-terminal lobe of rabbit serum transferrin. The residue numbers correspond to the human serum transferrin sequence, whereas the black circles represent the sulfur atoms of the disulfide bridges in the N-lobe. The figure was produced by a computer program written by A. M. Lesk and K. D. Hardman (1982, 1985). The overall tertiary structure and iron-site geometry is similar to that found in human lactoferrin. At the iron-binding site, Asp-63, Tyr-95, Tyr-188, and His-249 bind directly to the iron. The disulfide bridge pattern for both the N- and C-lobes can be matched to that predicted for human serum transferrin.

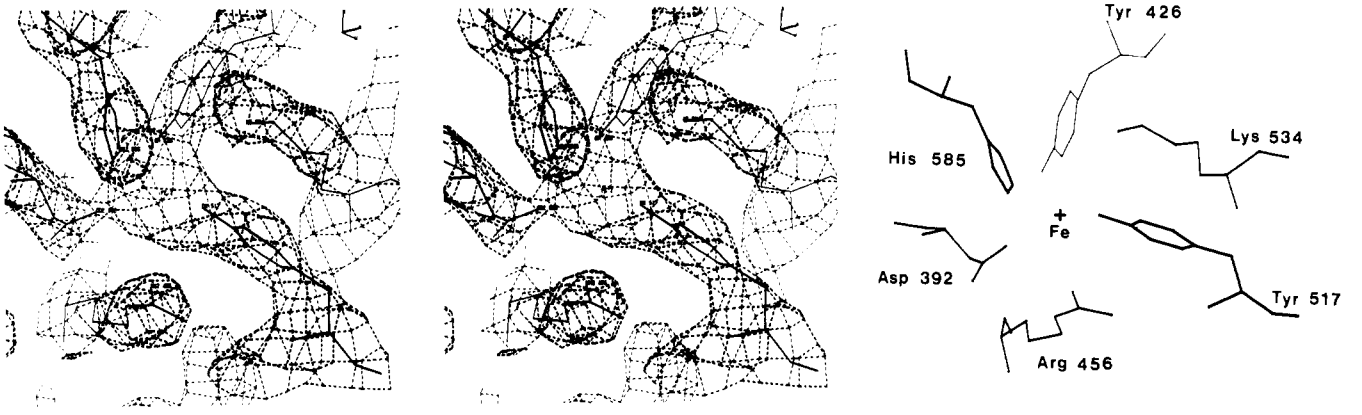

FIGURE 6: Electron density at the iron-binding site in the C-lobe of rabbit serum transferrin. The figure shows a stereo drawing of the electron density at the iron-binding site in the C-lobe. The residues directly involved are Asp-392 (63 in the N-lobe), Tyr-426 (95), Tyr-517 (188), and His-585 (249), all numbers referring to the human serum transferrin sequence. Arg-456 (124) and Lys-534 (206) are also within 6 Å of the ferric cation. The four protein ligands directly coordinated to the iron are conserved throughout both lobes of the serum transferrin, lactoferrin, and ovotransferrin family of proteins.

of the molecule (Figure 7). For this reason they are thought to have been present in the ancestral single lobe precursor prior to its duplication to give the parent protein from which all present-day transferrins evolved (Williams, 1982a). In addition to these six bridges, the C-lobe contains a further three disulfide bridges, numbers 7–9, which have no counterparts in the N-lobe and which presumably have arisen after gene duplication. This basic pattern of six plus nine bridges is that observed in chicken ovotransferrin. In the human serum transferrin, however, this pattern is augmented by two further bridges in each half of the molecule, numbers 10 and 11 in the N-lobe and numbers 12 and 13 in the C-lobe. None of the disulfide bridges cross from one lobe to the other, presumably explaining the relative ease with which single iron containing half-molecules can be isolated from many species of transferrin. The crystal structure of the rabbit serum protein suggests that the disulfide pattern is similar to that of its human counterpart.

(*a*) *Common or "Conserved" Bridges, Numbers 1–6.* Of the common and most ancient bridges (numbers 1–6 inclusive), five are local in nature (Thornton, 1981; Williams, 1982), with only number 3 spanning more than 40 residues. Number 3 is also the only one of the six to span an intron/exon boundary, suggesting that it is the most recently acquired of its class. None of them traverse from one domain to the other, nor do they join the two separate regions of the polypeptide chain that constitute domain I.

Disulfide bridges 1–3 attach the external helices, which pack around the domains (Figure 5), to the β-sheets, which form the domain cores. Specifically, in domain I, numbers 1 and 2 attach helices 2 and 1 to strands *a* and *b*, respectively; in domain II, disulfide 3 joins helix 7 to strand *f*. Bridges 4 and 5 occur together in a region of the lobe that has little secondary structure. In the N-lobe, this region includes a further bridge, number 11 (vide infra), forming a cluster of three bridges joined by short connecting loops. Several of the loops are as short as one or two residues, and not surprisingly, therefore, the local topology is dominated by this cluster rather than any secondary structure. In the C-lobe of the rabbit serum protein, the site of the carbohydrate attachment lies between the N-terminal half-cystines of bridges 4 and 5 (R. W. Evans, personal communication).

Disulfide bridge 6 connects the two ends of a surface loop that lies on the opposite side of the iron-binding site from the interdomain cleft. This loop folds back onto the surface of the molecule rather than extending into the solvent and is better conserved than most. In particular, it is of constant length in all known half-molecules, and the disulfide bridge is presumably important in maintaining its, as yet, unknown structural/functional role.

(*b*) *"Semiconserved" Bridges in the C-Lobe, Numbers 7–9.* Bridges 7–9 are only observed in the C-lobe and appear fundamentally different from the conserved bridges in that they are all long range in nature, each spanning approximately 200 residues.

Number 7 is unique in that it bridges the cleft between the two domains and serves to anchor them together. Evidence from dielectric dispersion and viscosity measurements (Ros-

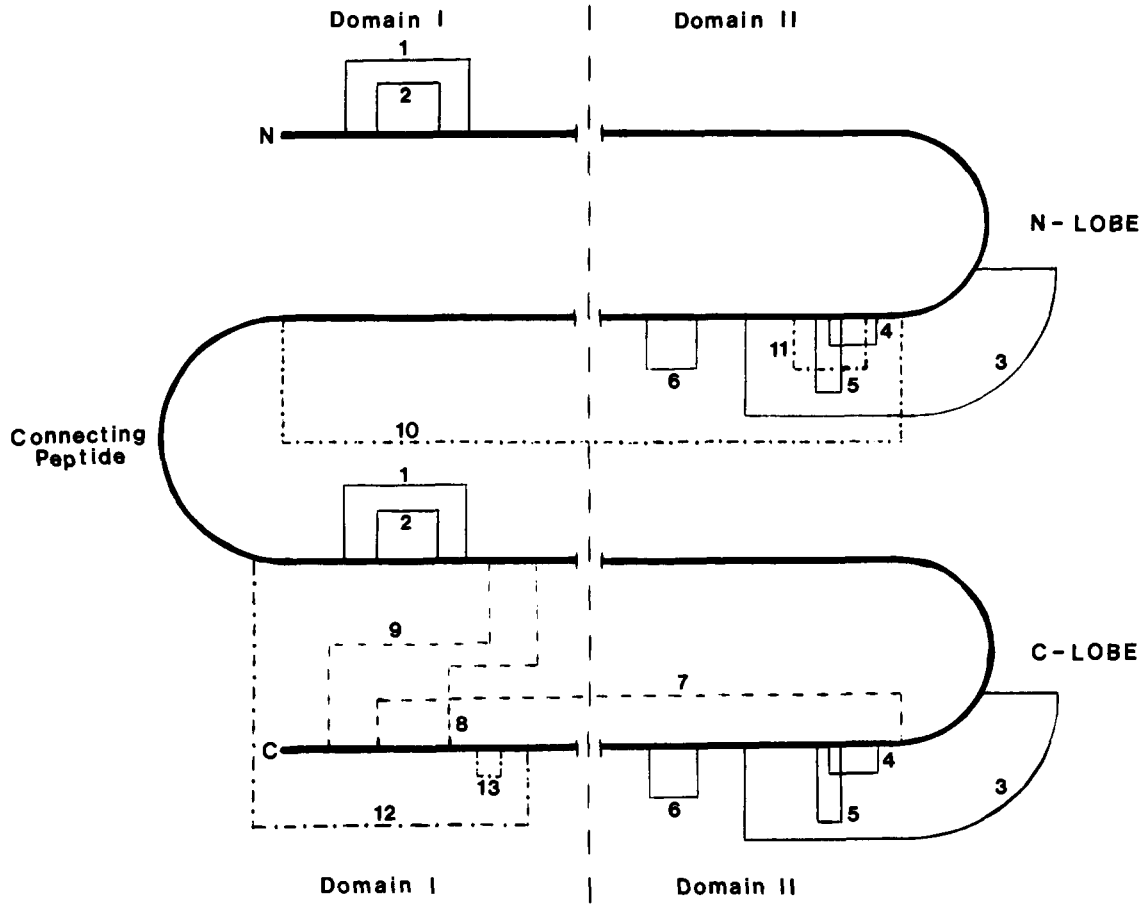

FIGURE 7: Disulfide bridge pattern in rabbit serum transferrin. Solid lines indicate the conserved disulfide bridges, numbers 1–6 inclusive. Dashed lines represent the semiconserved bridges 7–9 found in the C-lobe. Dashed-dotted lines show additional bridges 10 and 11 in the N-lobe and 12 and 13 in the C-terminal lobe.

seneu-Moutreff et al., 1971) suggests that the protein undergoes a considerable conformational change on the uptake and release of iron which may involve the opening and closing of the domain cleft (Anderson et al., 1987). By bridging the two domains, therefore, disulfide 7 may impose an additional constraint on such conformational changes and help to explain the differential stabilities and iron-binding properties of the two lobes (Brock, 1985; Huebers & Finch, 1987).

It has also been reported (Evans & Madden, 1984; Legrand et al., 1984) that it is possible to isolate a single-chain 18 000 iron-binding fragment from the N-terminal lobe of certain transferrins that corresponds to domain II. However, no report has yet appeared describing the isolation of such a fragment from a C-terminal lobe, probably because disulfide 7 covalently binds the two domains together. Yet, this disulfide is not deeply buried in the cleft and has been shown to be sufficiently exposed to be selectively reducible on treatment with dithiothreitol (Williams et al., 1985). Isolation of an iron-binding domain II fragment from a C-terminal lobe might be possible following reduction of disulfide 7.

Disulfides 8 and 9 are the only bridges to span the sequentially disparate regions of domain I and are the 2 of the 15 bridges common to all the serum, lacto-, and ovotransferrins that are not present in the membrane-bound melanotransferrin protein. Number 9 appears the more interesting of the two since it anchors the C-terminal helix of the molecule to the main body of the lobe. In melanotransferrin a C-terminal hydrophobic extension is thought to be responsible for anchoring the protein to the membrane (Rose et al., 1986). It has been suggested that the absence of disulfide 9 gives sufficient flexibility for the anchoring peptide to reach the membrane and adopt the appropriate conformation. Such a mechanism would most likely leave the interdomain clefts exposed and available for uptake of iron.

(*c*) *"Additional" Bridges, Numbers 10–13.* Disulfide bridges 10 and 12 are located at the ends of the interlobe connecting peptide and anchor its termini to the N- and C-lobes, respectively. The electron density is poorly defined in this region of the rabbit serum transferrin map, but there is no evidence for the prominent connecting helix observed in human lactoferrin (Anderson et al., 1987). Some form of extended structure appears more likely, and this is consistent with recent sequence data which show that the connecting peptide of rabbit serum transferrin consists of only seven amino acids, including a proline residue (H. MacKenzie, personal communication). A peptide of this size would be unable to span the distance between disulfide 10 and 12 by adopting an α-helical conformation. Furthermore, the equivalent residues in the lactoferrin sequence occur a few residues into the helix from each end, and it seems unlikely that cysteines at these positions could form disulfide bonds to the two lobes. Although, at present, this region of the rabbit serum molecule is not well determined, it appears that there are significant structural differences compared with the lactoferrin molecule. It should also be noted that the rotation and translation operations required to superpose the two lobes in the rabbit serum molecule are significantly different from the corresponding values in the human lactoferrin molecule, reported as 180° and 25 Å, respectively (Anderson et al., 1987).

Disulfide bridge 11 is a short-range bridge found only in the N-lobe, where it forms part of the cluster with bridges 4 and 5. It was not possible to interpret the well-defined electron

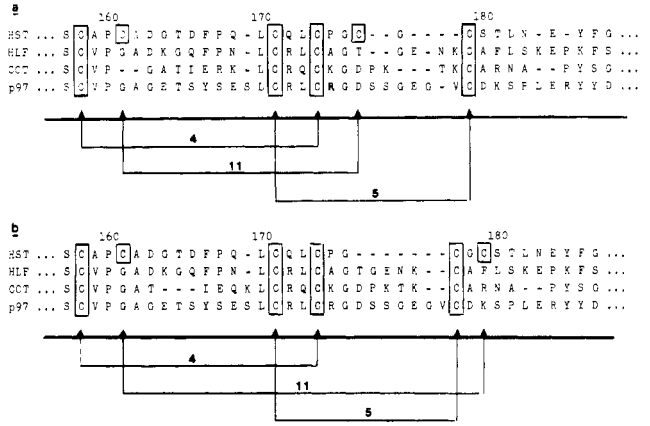

FIGURE 8: Connectivity of disulfide bridges 4, 5, and 11 in the N-lobe. The well-defined electron density at the cluster of disulfide bridges 4, 5 and 11 in N-lobe can best be interpreted by an interchange of the C-terminal half-cystines for bridges 5 and 11. (a) is the sequence alignment proposed by Metz-Boutigue et al. (1984) with the addition of the p97 melanotransferrin sequence. (b) is the sequence alignment predicted by Barton and Sternberg (1987) with a disulfide bridging pattern in agreement with the observed electron density. The abbreviations in the figure, HST, HLF, COT, and p97, refer to the amino acid sequences of human serum transferrin, human lactoferrin, chicken ovotransferrin, and human melanotransferrin, respectively.

density in this region according to the disulfide bridge pattern proposed by Metz-Boutigue et al. (1984). However, by exchanging the C-terminal half-cystines of bridges 5 and 11, as shown in Figure 8, all the features of the map could be successfully modeled. This interpretation is consistent with a new alignment of the known sequences in this region produced by using the algorithm of Barton and Sternberg (1987), also shown in Figure 8. The new alignment introduces just one insertion into the human sequence in place of four (Metz-Boutigue et al., 1984) and is consistent with the structure of lactoferrin in this region, which brings the equivalent residues to 161 and 179 (i.e., 159 and 180) into close proximity.

Disulfide 13 is local in nature within a large extended loop in the C-lobe. It is the only one of the additional bridges to be present also in lactoferrin and is therefore thought to have arisen prior to the divergence of mammalian iron-binding proteins into the transferrins and the lactoferrins.

The Carbohydrate Moiety. Rabbit serum transferrin has been reported to possess either one (Leger et al., 1978) or two (Strickland et al., 1979) biantennary carbohydrate chains terminating in *N*-acetylneuraminic acid residues. Incubation of the protein with neuraminidase followed by urea/polyacrylamide gel electrophoresis (R. W. Evans, personal communication) has shown a two-step decrease in the mobility of the protein consistent with a single, biantennary carbohydrate chain. The sequence of a carbohydrate-containing pentapeptide from rabbit transferrin (Strickland et al., 1979) has been shown to be Asn-Ser-Ser-Leu-Cys. In order to locate the precise position of the carbohydrate within the polypeptide chain a carbohydrate-containing CNBr fragment, molecular weight ca. 24 000, has been purified and its N-terminal sequence determined. The sequence shows that the carbohydrate-carrying pentapeptide corresponds to residues 491–495 (domain II in the C-lobe) in the human transferrin sequence. In the solvent-flattened maps calculated with the Wang algorithm, electron density in this region can be interpreted in terms of the glycan moiety, and one of the two antennae appears to bridge the solvent region betwen the two lobes and approach, via its terminal sialic acid, the N-terminal lobe in the vicinity of helix 9; this region is also the major heavy atom binding site for the uranyl acetate derivative. No definite functional significance has yet been assigned to the glycan moiety other than to signal transferrin clearance from the plasma (Regoeczi et al., 1974).

Conserved Regions. Sequence alignment of the homologous lobes of human serum transferrn (MacGillivray et al., 1982, 1983; Yang et al., 1984), hen ovotransferrin (Jelstch & Chambon, 1982; Williams et al., 1982), human lactoferrin (Metz-Boutigue et al., 1984), and the N-lobe of melanotransferrin (Rose et al., 1986) reveals the conserved regions. In addition to the iron-binding residues and 15 of the disulfide bridges, the most strongly conserved regions are the β-strands, while most variation occurs in the surface loops.

In the C-lobe of melanotransferrin both homologues of Asp-63 and Arg-124 have each been replaced by a serine residue. It seems unlikely that this site could bind iron in a similar manner to the other transferrin iron-binding sites.

The residues constituting the three central strands, *e*, *h*, and *f*, of the β-sheet in domain II in each lobe are very highly conserved; in human serum transferrin domain II shows a 56% homology between the N- and C-lobes, whereas the corresponding value for domain I is 46%. Domain II, in conjunction with the bridging strands, contributes three of the four protein residues directly involved in iron binding and can be isolated from the N-lobe of duck ovotransferrin and human lactoferrin as a single-chain iron-binding fragment (Evans & Madden, 1984; Legrand et al., 1984). In the intact lobe, the presence of domain I may serve not only to increase the stability of iron binding by the introduction of a further ligand (Asp-63[1]) but also to afford protection to the iron against hydrolytic attack on an otherwise exposed binding site.

Intron–Exon Distribution. The distribution of exons within the gene has been determined for both hen ovotransferrin (Cochet et al., 1979) and human serum transferrin (Park et al., 1985; Schaeffer et al., 1987). Each gene contains 17 exons separated by 16 intervening sequences. Exon 1 encodes most of the N-terminal leader sequence (or signal peptide), and the remaining 16 exons encode for the mature transferrin molecule. Fourteen of these constitute seven homologous pairs encoding corresponding regions in the two lobes, while the remaining two are unique to the C-terminal lobe.

Examination of the distribution of exons in relation to the molecular structure reveals that unlike hemoglobin, in which a single exon encodes for the heme-binding pocket (Fermi, 1975; Leder et al., 1978; Crick, 1979), there is no single exon on the transferrin gene that encodes the iron-binding region. Domain II, which is capable of reversible iron binding, is composed neither of a single exon nor of a discrete number of exons, and the essential metal-binding ligands are distributed throughout the polypeptide chain, which is a rather unusual feature for a metalloprotein. The vast majority of metal-binding sites in proteins include side- or main-chain ligands, at least two of which are separated by less than four residues in the amino acid sequence. It has been suggested that these may be nucleation sites facilitating metal binding and hence protein folding (Yazgan & Henkens, 1972; Waara et al., 1972; Liljas & Rossman, 1974). Iron is not an inherent component of the transferrin molecule and is thus not essential for correct polypeptide folding. During transport iron is both taken up and released, and the distribution of the ligands may be a reflection of this reversibility of iron binding.

Exon 2 (exon 9 in the C-lobe) is of particular interest, encoding residues −4 to 53 (Park et al., 1985; Schaeffer et al., 1987). This exon corresponds to the region of transferrin known to be homologous to the C-terminal 56 amino acids of ChBlym-I, the transforming protein from chicken B lym-

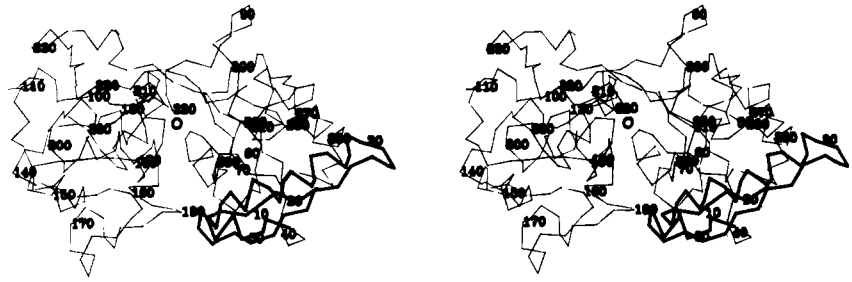

FIGURE 9: Region of the molecule in the N-lobe encoded by exon 2. A stereo drawing of the backbone trace of the N-lobe indicating the region encoded by exon 2 (bold type) is shown. In this figure the orientation of the N-lobe corresponds to that shown in Figure 4. This exon encodes the N-terminus of the molecule and is comprised of a $\beta\alpha\beta$ motif and an additional helix. It is a structurally discrete entity, distant from the iron site, which forms part of the domain I surface and is readily accessible to the solvent. This region of transferrin is homologous to the lymphocyte-derived transforming proteins ChBlym-I and HuBlym-I. The figure was produced by a computer program written by A. M. Lesk and K. D. Hardman (1982, 1985).

phocytes, which is similarly encoded by a single exon (Goubin et al., 1983). Furthermore, alignment of the transferrin sequences with those of ChBlym-I and HuBlym-I, a second transforming protein from Burkitt's lymphoma, has shown that the residues conserved among the transferrins tend to be those also conserved in the transforming proteins (Diamond et al., 1984). This part of the molecule forms an almost planar surface to domain I, as shown in Figure 9, on the opposite side of the lobe to the interdomain bridging strands. It is remotely situated from the iron-binding site and encodes none of the iron-binding ligands. The observation that in both human serum transferrin and ChBlym-I the homologous regions are encoded by a single exon suggests that this exon may have been incorporated into the transferrin gene by a process of exon shuffling and has retained a function independent from that of iron binding. The homology with the lymphocyte-derived transforming proteins and the knowledge that transferrin is an essential growth factor for most cells in serum-free medium (Brock, 1985) suggest further that exon 2 codes for a region of the molecule which may function to stimulate cell growth and proliferation. Some evidence already exists that this property of transferrin is not entirely dependent on its iron-transport capability (James & Bradshaw, 1984). The crystal structure lends some further credibility to this proposal in that this region is structurally discrete, distant from the iron site, forms part of the domain I surface, and is readily accessible to the solvent. It is therefore potentially suited to act as a recognition site for interaction with a cellular receptor. As has been suggested (Goubin et al., 1983), the ChBlym-I protein may exert its effect via cellular interactions more normally associated with transferrin.

CONCLUSIONS

The detailed analysis of the geometry of the iron-binding sites and comparison of the rabbit serum transferrin and human lactoferrin molecules must await further refinement of both X-ray structures. In this context, a proteolytic fragment, corresponding to the N-terminal lobe of the intact rabbit transferrin molecule, has been crystallized (Sarra & Lindley, 1986). This fragment diffracts to at least 1.8-Å resolution, and X-ray intensity data have already been collected to 2.3-Å resolution; molecular replacement studies are under way. Further, an 18000 fragment from duck ovotransferrin has been crystallized, and data have also been collected to 2.3-Å resolution (Jhoti et al., 1987). This fragment corresponds to domain II in the N-terminal lobe of the intact protein and retains the capacity to bind iron, although it cannot, by itself, provide all the ligands normally required for iron binding since it lacks the aspartic acid normally provided by domain I. The EPR and visible spectra are significantly different from those of the intact protein but resemble those obtained from the C-lobe of a human transferrin variant (Evans et al., 1982). Structural studies on this duck fragment may therefore help to explain the nature of the abnormalities in the human variant as well as provide higher resolution details of the iron-site geometry.

The comparison of the high-resolution structures of transferrins isolated from different physiological compartments should provide important insights into chemical and functional differences that exist within this family of proteins.

ACKNOWLEDGMENTS

We gratefully acknowledge Drs. B. F. Anderson and E. N. Baker and their colleagues for making available to us details of the human lactoferrin structure prior to its publication and for their interest and encouragement. We thank Professor D. Blow and Dr. A. Wonnacott for the use and assistance of the X-ray film scanning facilities at Imperial College, University of London, and the MRC Laboratory, Cambridge. We also acknowledge the help and support of Dr. D. S. Moss for some of the computations.

Registry No. Asp, 56-84-8; Tyr, 60-18-4; His, 71-00-1; Fe, 7439-89-6.

REFERENCES

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.

Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M., & Baker, E. N. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1768–1774.

Barton, G., & Sternberg, M. J. E. (1987) *J. Mol. Biol. 198*, 327–337.

Bricogne, G. (1976) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A32*, 832–847.

Brock, J. H. (1985) *Metalloproteins, Part II* (Harrison, P., Ed.) pp 183–262, Macmillan, London.

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., & Chambon, P. (1979) *Nature (London) 282*, 567–574.

Crick, F. (1979) *Science (Washington, D.C.) 204*, 264–271.

Diamond, A., Devine, J. M., & Cooper, G. M. (1984) *Science (Washington, D.C.) 225*, 516–519.

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

Evans, R. W., Williams, J., & Moreton, K. (1982) *Biochem. J. 210*, 19–26.

Fermi, G. (1975) *J. Mol. Biol. 97*, 237–256.

Goubin, G., Goldman, D. S., Luce, J., Neiman, P. E., & Cooper, G. M. (1983) *Nature (London) 302*, 114–119.

Hasnain, S. S., Evans, R. W., Garratt, R. C., & Lindley, P. F. (1987) *Biochem. J. 247*, 369–375.

Huebers, H. A., & Finch, C. A. (1987) *Physiol. Rev. 67*, 520–582.
James, R., & Bradshaw, R. A. (1984) *Annu. Rev. Biochem. 53*, 259.
Jeltsch, J. M., & Chambon, P. (1982) *Eur. J. Biochem. 122*, 291–295.
Jhoti, H., Gorinsky, B., Garratt, R. C., Lindley, P. F., Walton, A. R., & Evans, R. W. (1987) *J. Mol. Biol. 200*, 423–425.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Jones, T. A., & Thirup, S. (1986) *EMBO J. 5*, 819–822.
Koenig, S. H., & Schillinger, W E. (1969) *J. Biol. Chem. 244*, 6520–6526.
Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M., & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 6187–6191.
Leger, D., Tordera, V., Spik, G., Dorland, L., Haverkamp, J., & Vliegenthart, J. F. G. (1978) *Febs. Lett. 93*, 255–259.
Legrand, D., Mazurier, J., Metz-Boutigue, M.-H., Jolles, J., Jolles, P., Montreuil, J., & Spik, G. (1984) *Biochim. Biophys. Acta 787*, 90–96.
Lesk, A. M., & Hardmann, K. D. (1982) *Science (Washington, D.C.) 216*, 539–540.
Lesk, A. M., & Hardmann, K. D. (1985) *Methods Enzymol. 115*, 81–390.
Leslie, A. (1987) *Acta Crystallogr., Sect. A: Found. Crystallogr. A43*, 134–136.
Liljas, A., & Rossman, M. G. (1974) *Annu. Rev. Biochem. 43*, 475–507.
MacGillivray, R. T. A., Mendez, E., Sinha, S. K., Sutton, M. R., Lineback-Zins, J., & Brew, K. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2504–2508.
MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J., & Brew, K. (1983) *J. Biol. Chem. 258*, 3543–3553.
Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J., & Jolles, P. (1984) *Eur. J. Biochem. 145*, 659–676.
Moss, D. S., & Morphew, A. (1982) *Comput. Chem. 6*, 1–3.
Park, I., Schaeffer, E., Sidoli, A., Barelle, F. E., Cohen, G. N., & Zakin, M. M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3149–3153.
Pecoraro, V. L., Harris, W. R., Carrano, C. J., & Raymond, K. N. (1981) *Biochemistry 20*, 7033–7039.
Regoeczi, E., Hatton, M. W. C., & Wong, K.-L. (1974) *Can. J. Biochem. 52*, 155.
Rogers, T. B., Gold, R. A., & Feeney, R. E. (1977) *Biochemistry 16*, 2299–2305.
Rogers, T. B., Boerresen, T., & Feeney, R. E. (1978) *Biochem. J. 17*, 1105–1109.
Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellstrom, K. E., & Brown, J. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1261–1265.
Rosseneu-Moutreff, M. Y., Soetewey, F., Lamote, R., & Peeters, H. (1971) *Biopolymers 10*, 1039–1048.
Sarra, R., & Lindley, P. F. (1986) *J. Mol. Biol. 188*, 188–189.
Schaeffer, E., Lucero, M. A., Jeltsch, J.-M., Py, M.-C., Levin, M. J., Chambon, P., Cohen, G. N., & Zakin, M. M. (1987) *Gene 56*, 109–116.
Schlabach, M. R., & Bates, G. W. (1975) *J. Biol. Chem. 250*, 2182–2188.
Shewale, J. G., & Brew, K. (1982) *J. Biol. Chem. 257*, 9406–9415.
Strickland, D. K., Hamilton, J. W., & Hudson, B. G. (1979) *Biochemistry 18*, 2549–2554.
Sutcliffe, M., Haneef, I., Carney, D., & Blundell, T. L. (1987) *Protein Eng. 1*, 377–384.
Taylor, W. R., Thornton, J. H., & Turnell, W. G. (1983) *J. Mol. Graphics 1*, 30–38.
Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261–287.
Trowbridge, I. S., & Lopez, F. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1175–1179.
Waara, I., Lovgren, S., Liljas, A., Kannan, K. K., & Bergsten, P. C. (1972) *Adv. Exp. Med. Biol. 28*, 169–187.
Wang, B. C. (1985) *Methods Enzymol. 115*, 90–112.
Williams, J. (1982a) *Trends Biochem. Sci. (Pers. Ed.) 7*, 394–397.
Williams, J. (1982b) *Biochem. J. 201*, 647–651.
Williams, J., Elleman, T. C., Kingston, I. B., Wilkins, A. G., & Kuhn, K. A. (1982) *Eur. J. Biochem. 122*, 297–303.
Williams, J., Moreton, K., & Goddearl, D. J. (1985) *Biochem. J. 228*, 661–665.
Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., Van Bragt, P. H., Baldwin, W. D., & Bowman, B. H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2752–2756.
Yazgan, A., & Henkens, R. W. (1972) *Biochemistry 11*, 1314–1318.