Indian Journal of Biochemistry & Biophysics
Vol. 40, February 2003, pp. 14-21

Crystal structure of human seminal diferric lactoferrin at 3.4 Å resolution

Janesh Kumar[1], Wolfgang Weber[2], Sabine Münchau[2], Savita Yadav[1], S Bhaskar Singh[1], K Saravanan[1], M Paramasivam[1], Sujata Sharma[1], Punit Kaur[1], A Bhushan[1], A Srinivasan[1], Christian Betzel[3] and T P Singh[1]*

[1]Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110029, India
[2]Universitätsklinikum Hamburg-Eppendorf, IMBM, Martinistrasse 52, D-20246 Hamburg, Germany
[3]Institute of Medical Biochemistry and Molecular Biology, C/o DESY, Geb 22a Notkestrasse 85, 22603, Hamburg, Germany

Received 29 October 2002; revised 16 January 2003

Lactoferrin was purified from human seminal fluid obtained from the semen bank. The purified samples were saturated with Fe^{3+} and crystallized by microdialysis method. The crystals belong to orthorhombic space group $P2_12_12_1$ with a = 55.9 Å, b = 97.2 Å, c = 156.1 Å and Z = 4. The structure was determined with molecular replacement method and refined to an R factor of 18.7% for all the data to 3.4 Å resolution. The overall structure of seminal lactoferrin is similar to human colostrum lactoferrin. The amino acid sequence of seminal lactoferrin shows that it has one amino acid less than human colostrum lactoferrin and the structure of its N-terminal region is far more ordered than other lactoferrins. The structure of the iron-binding site and its immediate surroundings indicate well defined features.

Lactoferrin an 80 kDa, single-chain, iron-binding glycoprotein consists of two similar sized homologous N and C lobes, which are further divided into two similar sized domains: N1 and N2 in the N-lobe and C1 and C2 in the C-lobe. The iron-binding sites are located within the interdomain cleft of each lobe. It has the ability to bind, very tightly but reversibly, two Fe^{3+} ions, together with two CO_3^{2-} ions. The synergistic relationship between metal binding and anion binding is one of its unique features not observed in other metal binding proteins[1,2,3]

Crystal structures of diferric forms of human[4,5], bovine[6], equine[7] and buffalo[8,9] lactoferrins reveal that the two domains are closed over an Fe^{3+} ion. For the iron-free apo forms, X-ray crystallographic studies on human[10], equine[11,12] and camel[13] lactoferrins and solution scattering analysis[14] have shown that the apolactoferrins assume conformations with variable domain arrangements. The intermediates in the iron-binding pathway using camel apolactoferrin as a starting point[3] and iron-release pathway using goat diferric lactoferrin as a starting point[15] have revealed the existence of iron-bound tetrahedral open and closed forms separately.

Although, the lactoferrins from various species share many structural and functional features, they differ in several others[16]. The behaviour of iron-binding and iron-release against *p*H shows remarkable variations. The numbers of carbohydrate attachments and their locations differ greatly[17]. The processes of domain-closing and domain-opening show remarkable differences. There are observed variations in functions other than the iron-binding and iron-release. So far, the lactoferrin structures from six species[4,6,7,8,13,15] have been determined, but all of them have used colostrum as the source. The primary consideration for choosing colostrum has been due to high concentrations of lactoferrin in it. However, there are a number of other secretary sources of lactoferrin such as tears, saliva, sweat, genital secretions, synovial fluid etc, but no structural report is available to highlight the similarities and differences between lactoferrins from different secretions. In order to understand the role of lactoferrin in human reproductive physiology and to evaluate the tissue related structural and functional variations, we have isolated lactoferrin from human seminal fluid and have determined its detailed three-dimensional structure. The results of structure analysis of seminal human lactoferrin (sLf) are presented here.

Materials and Methods

Isolation and purification

Samples of human semen obtained from a laboratory specialized in fertility analyses

*For correspondence
E-mail: tps@aiims.aiims.ac.in
Tel: +91-11-653 3931
Fax: +91-11-686 2663

(Hamburg University hospital) were frozen after collection. For the purification of sLf, a pool of samples (typically 50-100 ml) was thawed, mixed with an equal volume of phosphate buffered saline, and centrifuged for 30 min at 25,000 *g*. Benzamidine hydrochloride (10 m*M* final concentration) and solid $(NH_4)_2SO_4$ (25% saturation) were added to the supernatant. The mixture was stirred for 1 hr on ice, followed by another centrifugation step (30 min, 25,000 *g*) to remove precipitated material. The supernatant was extensively dialyzed against 50 m*M* Tris-HCl *p*H 8.0 using dialysis tubing with a 6 kDa molecular weight cut-off and applied to a CM Sephadex C-50 column (40 ml bed volume) equilibrated with the same buffer. A gradient of up to 0.7 *M* NaCl served to elute the bound proteins. The sLf containing fractions were eluted in the conductivity range between 40 and 55 μSiemens, pooled and concentrated to about 10-fold by $(NH_4)_2SO_4$ precipitation (75% saturation). After dialysis against 20 m*M* Hepes buffer (*p*H 8.0), these were mixed with 4 ml DNA-cellulose gel (Sigma) and equilibrated with the same buffer. The incubation was carried out for 2 hr by rotating the suspension in a chromatographic column. A step-wise elution of the column with 0-200-700 m*M* NaCl in 20 m*M* Hepes (*p*H 8.0) yielded two fractions of sLf. As indicated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis, the low salt fraction contained a few low molecular weight protein contaminants, while the high salt fraction was a single band corresponding to 80 kDa. The latter fraction (comprising about one third of the lactoferrin bound to the column) was concentrated by lypholization and dialyzed against 0.05 *M* Tris-HCl buffer at *p*H 8.0.

Sequence determination

The tissues from human seminal vesicles were obtained from the mortuary of the All India Institute of Medical Sciences, hospital. The isolation of poly A^+ mRNA and cDNA synthesis were performed following the manufacturer's protocol (Stratagene, Germany). The conserved nucleotide sequences in human (X53961)[18], bovine (X57084)[19], goat (X78902)[20], porcine (M92089)[21], (M81327)[22], (L77887)[23], buffalo (AJ005203)[9], equine (AJ010930)[7] and camel (AF165879)[13] were used for the synthesis of primers. The PCR was performed with *Taq* polymerase (Promega, USA) using M J Research thermal cycler, model PTC-100. The nucleotide sequencing was performed either directly on the amplified DNA fragments or on the cloned double stranded DNA (pGEM-T) using automatic sequencer of Shimadzu model DSQ2000L. Both the strands were used for sequencing. The human seminal lactoferrin cDNA reported here is 2136 base pairs (bp) in length. It comprises 13 bp in the 5' UTR, the open reading frame encoding for 711 amino acid residues and part of the 3' UTR (170 bp). The nucleotide and derived amino acid sequences of human seminal lactoferrin are given in Table 1. The nucleotide sequence has been deposited in the gene bank with accession number AY 178998.

Preparation of iron-saturated lactoferrin

The purified lactoferrin was dissolved in 1 ml of 0.1 *M* sodium bicarbonate, buffered with sodium citrate at *p*H 8.0 to obtain a concentration of 1 m*M*. To this solution, 1.2 ml of ferric chloride reagent (2 m*M* $FeCl_3.6H_2O$ in 0.1 *M* sodium bicarbonate buffered with sodium citrate at *p*H 8.0) was added and equilibrated for 24 hr at 298 K. The excess of ferric chloride reagent was removed by passing the preparation through a Sephadex G-50 column (50 × 2 cm) and eluting with distilled water. The iron saturation of human seminal lactoferrin was estimated to be 92%.

Crystallization

The purified iron-saturated samples of human seminal lactoferrin were used for crystallization. The crystals were obtained by microdialysis method with protein concentration of 70 mg/ml in 0.025 *M* Tris-HCl at *p*H 8.0. This solution was dialyzed against the same buffer containing 13% (v/v) ethanol at *p*H 8.0. All crystallization experiments were carried out at 277 K. Rectangular shaped dark brown crystals grew to the dimensions of 0.3 × 0.16 × 0.12 mm^3 in 3 weeks.

Data collection and processing

Owing to the size, the crystals of sLf did not diffract beyond 3.4 Å. Despite low resolution, it was of enough interest to determine the structure of lactoferrin from a secretion other then colostrum/milk. X-ray intensity data were measured at 285 K using a MAR Research imaging-plate scanner having a diameter of 300 mm. Monochromatic CuKα radiation was produced by a graphite-crystal monochromator mounted on a Rigaku RU-200 rotating anode generator operating at 40 kV and 100 mA with a focal

Table 1—Nucleotide and amino acid sequences of human seminal lactoferrin.
[Amino acids are shown in three letter code. The triangle marks the N-terminal amino acid of the mature protein. Predicted glycosylation sites are underlined. The stop codon is indicated by ***]

▼

1	ATG AAA CTT GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC	87
1	Met Lys Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu Cys Leu Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys	29
88	ACC GTA TCC CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AGA GTG CGT GGC CCT CCT GTC AGC TGC ATA AAG	174
30	Thr Val Ser Gln Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Arg Val Arg Gly Pro Pro Val Ser Cys Ile Lys	58
175	AGA GAC TCC CCC ATC CAG TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT GGT GGT TTC ATA TAC GAG GCA	261
59	Arg Asp Ser Pro Ile Gln Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Ala Thr Asp Gly Gly Phe Ile Tyr Glu Ala	87
262	GGC CTG GCC CCC TAC AAA CTG CGA CCT GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT TAT GCC GTG GCT	348
88	Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr Ala Val Ala	116
349	GTG GTG AAG AAG GGC GGC AGC TTT GAG CTG AAC GAA CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA TGG	435
117	Val Val Lys Lys Gly Gly Ser Phe Glu Leu Asn Glu Leu Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly Trp	145
436	AAT GTC CCT ATA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT CCA CCT GAG CCC ATT GAG GCA GCT GTG GCC AGG TTC TTC TCA	522
146	Asn Val Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly Pro Pro Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser	174
523	GCC AGC TGT GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC	609
175	Ala Ser Cys Val Pro Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe	203
610	TCC TCC CAG GAA CCG TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA GAC GTG GCT TTT ATC AGA GAG AGC	696
204	Ser Ser Gln Glu Pro Tyr Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp Val Ala Phe Ile Arg Glu Ser	232
697	ACA GTG TTT GAG GAC CTG TCA GAC GAG GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG CCA GTG GAC AAG	783
233	Thr Val Phe Glu Asp Leu Ser Asp Glu Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys Pro Val Asp Lys	261
784	TTC AAA GAC TGC CAT CTG GCC CGG GTC CCT TCT CAT GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG AAT CTT	870
262	Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn Leu	290
871	CTC CGC GAG GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG AAA TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG	957
291	Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu	319
958	TTC AAG GAC TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC	1044
320	Phe Lys Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile	348
1045	CAG AAC TTG AGG AAA AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT GCG GTG GGC GAG CAG GAG CTG CGC AAG	1131
349	Gln Asn Leu Arg Lys Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala Val Gly Glu Gln Glu Leu Arg Lys	377
1132	TGT AAC CAG TGG AGT GGC TTG AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC GCC CTG GTG CTG AAA	1218
378	Cys Asn Gln Trp Ser Gly Leu Ser Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile Ala Leu Val Leu Lys	406
1219	GGA GAA GCT GAT GCC ATG AGT TTG GAT GGA GGA TAT GTG TAC ACT GCA GGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG AAC TAC	1305
407	Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Tyr Val Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr	435
1306	AAA TCC CAA CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG GAT AGA CCT GTG GAA GGA TAT GTT GCT GTG GCG GTG GTT AGG AGA TCA	1392
436	Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp Arg Pro Val Glu Gly Tyr Val Ala Val Ala Val Val Arg Arg Ser	464
1393	GAC ACT AGC CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC ACC GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG	1479
465	Asp Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr Ala Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met	493
1480	GGC CTG CTC TTC AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT GCC CCT GGG TCT GAC CCG AGA TCT AAT	1566
494	Gly Leu Leu Phe Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys Ala Pro Gly Ser Asp Pro Arg Ser Asn	522
1567	CTC TGT GCT CTG TGT ATT GGC GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAT GAG AGA TAC TAC GGC TAC ACT GGG GCT	1653
523	Leu Cys Ala Leu Cys Ile Gly Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr Tyr Gly Tyr Thr Gly Ala	551
1654	TTC CGG TGC CTG GCT GAG AAT GCT GGA GAC GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT AAC AAT GAC	1740
552	Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn Asn Asp	580
1741	GCA TGG GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC TGC	1827
581	Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys	609
1828	CAT CTT GCC ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG GTG GAA CGC CTG AAA CAG GTG TTG CTC CAC CAA CAG GCT	1914
610	His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala	638
1915	AAA TTT GGG AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT	2001
639	Lys Phe Gly Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr	667
2002	GAG TGT CTG GCC AGA CTC CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA GGC ATT ACT AAT CTG AAA AAG	2088
668	Glu Cys Leu Ala Arg Leu His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala Gly Ile Thr Asn Leu Lys Lys	696
2089	TGC TCA ACC TCC CCC CTC CTG GAA GCC TGT GAA TTC CTC AGG AAG TAA	2136
697	Cys Ser Thr Ser Pro Leu Leu Glu Ala Cys Glu Phe Leu Arg Lys ***	712

point of 3 × 0.3 mm^2. The crystals diffracted to 3.4 Å resolution and remained stable in the X-ray beam. Therefore, a complete intensity data set was obtained using one crystal.

The HKL package and MARSCALE were used for the determination of unit cell parameters, data processing, scaling and merging[24]. The details of crystallographic data collection and processing statistics are given in Table 2.

Structure determination

The structure was determined by molecular replacement method employing the AmoRe program[25], incorporated in the CCP4[26] package, using the model of diferric human lactoferrin[5]. Since the sequences sLf and human milk lactoferrin (mLf) is identical, a full model was used for generating the rotation function. The model obtained from molecular replacement was refined using the program REFMAC[27].

Rounds of automated refinement were interspersed with manual rebuilding into 2Fo-Fc and Fo-Fc omit maps using the program O[28]. A bulk solvent correction was also applied in the final rounds of refinement and map calculations. The refined model of sLf has an R factor of 0.187 and the free R factor of 0.242 with a good geometry (Table 3).

Table 2—Crystallographic data on seminal lactoferrin

Space group	$P2_12_12_1$
a (Å)	55.9
b (Å)	97.2
c (Å)	156.1
Z	4
Matthew's coefficient *V*m (Å^3/Da)	2.72
Solvent content (%)	55
Resolution range (Å)	20.0-3.4
No. of observed reflections	85,368
No. of unique reflections	10,547
Overall completeness (20.0-3.4 Å)%	94.3
Completeness in the highest resolution shell (3.6-3.4Å) %	89.1
R_{sym} for all data (20.0-3.4 Å)%	14.7
R_{sym} in the highest resolution Shell (3.6-3.4 Å)%	39.6
Average I/σ(I)	6.4
I/σ(I) in the highest resolution Shell (3.6-3.4)	2.0

Table 3—Refinement statistics

PDB code	1N76
Space group	$P2_12_12_1$
Resolution range (Å)	20.0-3.4
R_{cryst} (for all data)%	18.7
R_{free} (for 5% randomly selected data)	24.2
R.M.S. deviations from ideal geometry	
bond distances(Å)	0.023
Bond Angles (°)	1.8
Torsion angles (°)	20.0
Overall average B factor (Å^2)	52.8
G-factor	–0.18
Residues in the most allowed regions (%)	82

Results and Discussion

Sequence analysis

The sequence of sLf (Table 1) is almost identical to that of human milk lactoferrin (mLf; X5396118). The only differences exist in the N-terminal region, where the first two residues are absent in sLf. Two more differences correspond to the residues at positions 10 and 28 (mLf structure numbering), where alanine and lysine in mLf have changed to threonine and arginine in sLf.

Overall structure

Since the resolution is limited to 3.4 Å, the structure was refined with stringent constraints. The final backbone electron density is continuous and is generally well defined for the side chains as well (Fig. 1). The electron density for the iron-binding ligands including Fe^{3+} ions is also clearly defined (Figs. 2a and b).

The molecule is folded into two globular lobes, the N-lobe, comprising the N-terminal half of the polypeptide chain (residue 2-333), C-lobe comprising C-terminal half of the polypeptide chain (residues 345-691) and a connecting helical peptide residues (334-344). Each lobe is further sub-divided into two domains of approximately 160 residues each (Fig. 3).

Metal and anion-binding sites

The two iron-binding sites situated one in each lobe at the end of each inter-domain cleft are 45 Å apart. The geometry of iron-binding coordination is a distorted octahedron in both lobes with four protein ligands occupying the four positions about the iron atoms (Figs. 2a and b). The composition of metal coordination distances involving protein ligands in

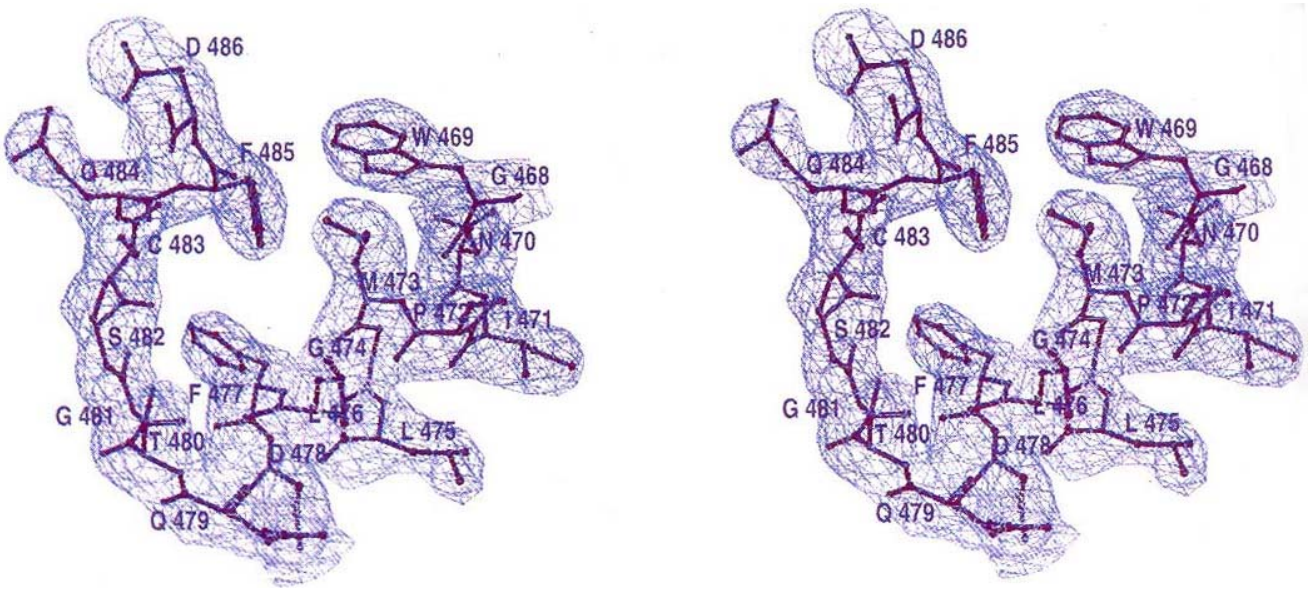

Fig. 1—Stereoview of a section of final (2Fo-Fc) electron density map contoured at 1σ level [The residue numbers are indicated]

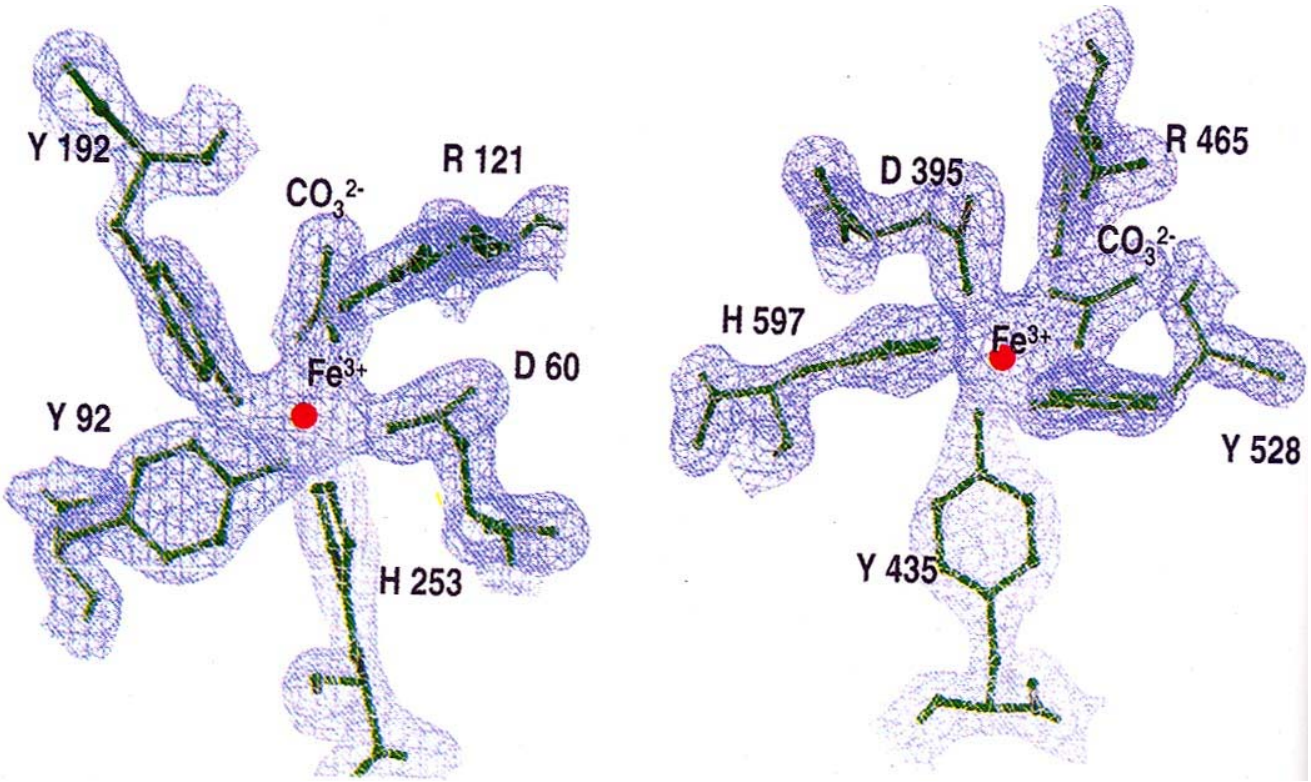

Fig. 2—View of the Fe^{3+} binding site showing difference electron density (Fo-Fc) map for iron and iron-binding protein ligands including CO_3^{2-} ion [(a): N-lobe; (b): C-lobe. The contours are drawn at 2σ level]

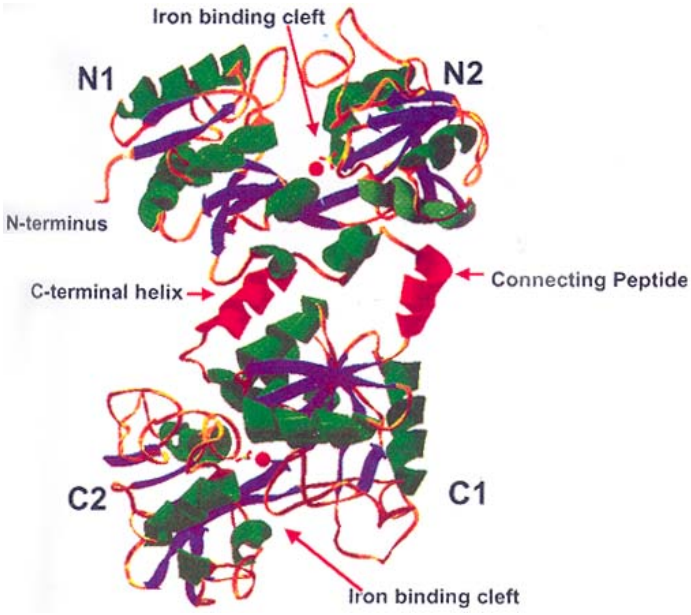

Fig. 3—Molscript[30] representation of the structure of human seminal lactoferrin [The N- and C-lobes and N1, N2, C1 and C2 domains are indicated. The interconnecting peptide and the C-terminal helix are indicated in red. The iron binding clefts are also shown]

Table 4—Distances in metal and anion binding

Metal–ligand	N-lobe	C-lobe
	Bond lengths (Å)	
Fe-OD1 60 (395)	1.9	2.0
Fe-OH 92 (435)	2.0	1.9
Fe-OH 192 (528)	1.9	2.0
Fe-NE2 253 (597)	2.2	2.4
Fe-O1 (carbonate)	2.4	2.0
Fe-O2 (carbonate)	2.0	2.2
Hydrogen bond lengths (Å)		
O1--------NE 121 (465)	3.0	3.1
O1--------NH2 121 (465)	2.8	3.4
O2--------N 123 (467)	3.2	3.2
O3--------Oγ1 117 (461)	2.8	3.3
O3--------N 124 (468)	3.5	3.5

two lobes shows an essentially similar binding environment (Table 4).

One of the main features of the construction of the iron sites is that the four protein ligating residues are widely spaced along the polypeptide chain and belong to distinctly different parts of the protein structure. As seen from Figs. 4a and b, Asp 60 belongs to the main body of domain I and Tyr 192 to domain II, while Tyr 92 and His 253 come from each of the two backbone strands, crossing between the two domains at the back of the iron site. This arrangement might be useful for a protein whose function is the binding and release of metal ions, with an associated large-scale conformational change. It is also clearly different from many other metalloproteins, where two, three or more of the ligands are contributed by a small loop of the polypeptide chain. It is indeed noteworthy that the carboxylate groups (of Asp 60 and Asp 395) appear to fill a particularly important structural role in each iron-binding site. While one oxygen, Oδ1, is bound to iron, the other is acceptor in two hydrogen bonds, from the main-chain NH groups of residues 62 and 122. Both NH groups are at the N-termini of helices 3 and 5. The carboxylate group thus has a dual role, binding to iron and forming a strong link between the two domains through interaction with the opposing helix N-termini. In contrast, the other iron ligands do not seem to be much constrained by the surrounding protein structure.

Interactions between the two lobes

The two lobes are covalently connected by a peptide segment 334-344, which forms a regular α-helix. It may be noted that the connecting peptide is on the outside and rather separate from the main area of contact between the two lobes. A major contribution to the interlobe regions comes from the C-terminal helix, residues 676-691, which runs across the surface of the C-lobe and finishes up against the surface of the N-lobe. This helix is in contact with residues 310-314 in the N-lobe and three other helices, numbers (315 to 321) and (321 to 332) in the N-lobe, and number (376 to 388) in the C-lobe.

Interactions between domains

In the N-lobe of lactoferrin, Glu 211 forms a salt bridge with Arg 89 in the other domain and is hydrogen bonded to the main chain of one crossover strand. In the C-lobe, Asp547 is hydrogen bonded to the crossover strand via a water molecule. Lys 301 forms an ion pair across the N-lobe domain interface with Glu216. An equivalent interaction does not occur in C-lobe.

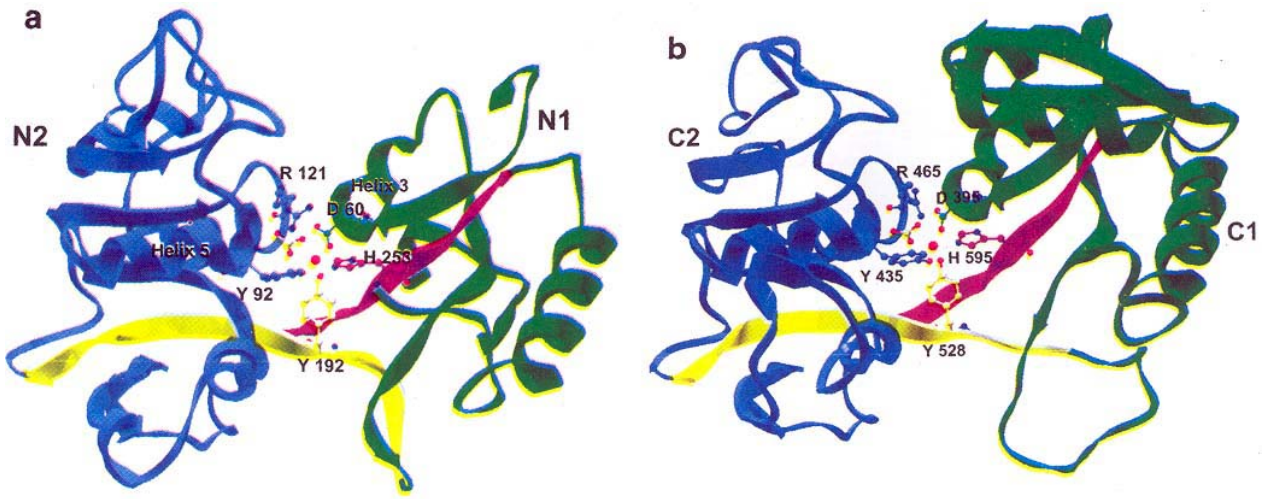

Fig. 4—(a): Showing the locations of iron-binding ligands in N-lobe; (b): Showing the locations of iron-binding ligands in C-lobe

Conclusions

The sequences of sLf and mLf differ by four residues. The structure of sLf is identical to that of mLf. The structure of sLf was superimposed with the structure of mLf in order to minimize the rms shift of the main chain atoms and was found to be 0.4 Å. The observed specific mutations in the sLf may pertain to some of its roles that are not required by the milk protein. The most striking differences between sLf and mLf are the absence of two residues at the N-terminus. The iron-binding geometry and its environment are also identical in the two proteins. There are two sites of carbohydrate attachment on sLf one in each lobe, but in neither case the electron density was observed. It may be due to several factors such as the low resolution, flexibility, disorder or the heterogeneity in the carbohydrate itself. The total carbohydrate content in sLf and mlf was reported to be 6.1% (ref. 29) and 6.4% (ref. 17), respectively. The regions of the lactoferrin structure that may be involved in cell receptor binding are difficult to identify.

Acknowledgement

Financial support from the Department of Biotechnology, New Delhi is gratefully acknowledged. Janesh Kumar thanks the Council of Scientific and Industrial Research (CSIR), New Delhi for the award of a fellowship. Skillful technical assistance in protein purification by Sigrid Himple is gratefully acknowledged.

References

1 Aisen P & Harris D C (1989) in *Iron carriers and iron proteins,* Vol. 5., pp. 241-351, VCH Publishers, New York
2 Baker E N (1994) *Adv Inorg Chem* 41, 389-463
3 Khan J A, Kumar P, Srinivasan A & Singh T P (2001) *J Biol Chem* 276, 36817-36823
4 Anderson B F, Baker H M, Dodson E J, Norris G E, Rumball S V, Waters J M & Baker E N (1987) *Proc Natl Acad Sci USA* 84, 1769-1773
5 Haridas M, Anderson B F & Baker E N (1995) *Acta Crystallogr* D51, 629-646
6 Moore S A, Anderson B F, Groom G R, Haridas M & Baker E N (1997) *J Mol Biol* 274, 222-236
7 Sharma A K, Paramasivam M, Srinivasan A, Yadav M P & Singh T P (1999) *J Mol Biol* 289, 303-317
8 Karthikeyan S, Paramasivam M, Yadav S, Srinivasan A & Singh T P (1999) *Acta Crystallogr* D55, 1805-1813
9 Karthikeyan S, Yadav S, Paramasivam M, Srinivasan A & Singh T P (2000) *Acta Crystallogr* D56, 684-689
10 Jameson G B, Anderson B F, Norris G E, Thomas D H & Baker E N (1998) *Acta Crystallogr* D 54, 1319-1335
11 Sharma A K, Rajashankar K R, Yadav M P & Singh T P (1999) *Acta Crystallogr* D 55,1152-1157
12 Kumar P, Khan J A, Yadav S & Singh T P (2002) *Acta Crystallogr* D58, 225-232
13 Khan J A, Kumar P, Paramasivam M, Yadav R S, Sahani M S, Sharma S, Srinivasan A & Singh T P (2001) J *Mol Biol* 309, 751-761
14 Grossmann J G, Neu M, Pantos E, Schwab F J, Evans R W, Towns-Andrews E, Lindley P F, Appel H, Thies W G & Hasnain S S (1992) *J Mol Biol* 225, 811-819
15 Kumar P, Yadav S & Singh T P (2002) *Ind J Biochem Biophys* 39, 16-21
16 Karthikeyan S, Sharma S, Sharma A K, Paramasivam M, Yadav S, Srinivasan A & Singh T P (1999) *Curr Sci* 77, 241-255
17 Spik G, Coddeville B, Mazurier J, Bourne Y, Cambilaut C, Montreuil J (1994) *Adv Exp Med Biol* 357, 21-32

18 Rey M W, Woloshuk S L, de Boer H A & Pieper F R (1990) *Nucleic Acids Res* 18, 288-5288
19 Pierce A, Colavizza D, Benaissa M, Maes P, Tartar A, Montreuil J & Spik J (1991) *Eur J Biochem* 196, 177-184
20 Provost F L, Nocart M, Guerin G & Martin P (1994) *Biochem Biophys Res Commun* 203, 1324-1332
21 Lydon J P, O'malley B R, Saucedo O, Lee T, Headon D A & Conneely O M (1992) *Biochim Biophys Acta* 1132, 97-99
22 Alexander L J, Levine W B, Teng C T & Beattie C W (1992) *Anim Genet* 23, 251-256
23 Wang S R, Lin T Y & Weng C N (1997) *Am J Vet Res* 58, 1152-1158
24 Otwinowski Z & Minor W (1997) *Methods Enzymol* 176, 307-326
25 Navaza J (1994) *Acta Crystallogr* D50, 157-163
26 Collaborative Computational Project, No. 4 (1994) *Acta Crystallogr* D50, 760-763
27 Murshudov G N, Vagin A A, & Dodson E J (1997) *Acta Crystallogr* D53, 240-255
28 Jones T A, Zou J, Cowan S W & Kjeldgaard M (1991) *Acta Crystallogr* A47, 110-119
29 D'Andrea G, D' Alessandro A M, Salucci M L & Oratore A (1994) *J Protein Chem* 13, 31-36
30 Kraulis P J (1991) *J Applied Crystallogr* 24, 946-950