Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of buffalo lactoferrin at 3.3 Å resolution at 277 K

S. Karthikeyan, S. Yadav, M. Paramasivam, A. Srinivasan and T. P. Singh*

Department of Biophysics, All India Institute of Medical Sciences, New Delhi 110 029, India

Correspondence e-mail: tps@medinst.ernet.in

Received 19 January 2000
Accepted 4 April 2000

PDB Reference: buffalo lactoferrin, 1biy.

The three-dimensional structure of diferric buffalo lactoferrin has been determined at 3.3 Å resolution. The structure was solved by molecular replacement using the coordinates of diferric human lactoferrin as a search model and was refined by simulated annealing (*X-PLOR*). The final model comprises 5316 protein atoms for all 689 residues, two Fe^{3+} and two CO_3^{2-} ions. The final *R* factor was 21.8% for 11 711 reflections in the resolution range 17.0–3.3 Å. The folding of buffalo lactoferrin is essentially similar to that of the other members of the transferrin family. The significant differences are found in the dimensions of the binding cleft and the interlobe orientation. The interlobe interactions are predominantly hydrophobic in nature, thus facilitating the sliding of two lobes owing to external forces. The interdomain interactions are comparable in the N and C lobes.

1. Introduction

Lactoferrin is a glycoprotein with a molecular weight of approximately 80 kDa. It is a prominent member of the transferrin family. It consists of a single polypeptide chain folded into two homologous N and C lobes. The N and C lobes both contain about 345 amino acids and are made up of two domains: N1, N2 and C1, C2, respectively. These proteins serve a general role in controlling the levels of free iron and possibly other elements in the body fluids of animals by sequestration of the bound metal ion. Iron binding is a common function of these proteins and is achieved by similar interactions involving residues from both domains, leading to an overall similar folding of the iron-bound states. These proteins are also involved in several other specific functions, presumably using other specifically designed structural sites. Such functional requirements may be attributable to specific elements of the three-dimensional structures, highlighting the importance of structural comparisons. The detailed three-dimensional structures of iron-saturated forms of rabbit serum transferrin (RST; Bailey *et al.*, 1988), hen ovotransferrin (HOT; Kurokawa *et al.*, 1995), duck ovotransferrin (DOT; Rawas *et al.*, 1996), human lactoferrin (HLF; Anderson *et al.*, 1989), bovine lactoferrin (CLF; Moore *et al.*, 1997) and mare lactoferrin (MLF; Sharma, Paramasivam *et al.*, 1999) have revealed that the overall folding of the polypeptide chains in these proteins are similar in that both lobes adopt closed conformations where domains N1, N2 and C1, C2 interact with each other through several interdomain hydrogen bonds. However, the structures of their native iron-free (apo) forms were found to fold differently. In duck apo ovotransferrin (Rawas *et al.*, 1997) both lobes adopt open conformations and in human apo lactoferrin the N-lobe is in the open conformation and the C lobe is in the closed conformation

Table 1
Data-collection statistics.

Crystallization conditions	0.025 *M* Tris–HCl with 19% ethanol at pH 8.0	
Space group	$P2_1$	
Unit-cell dimensions (Å, °)	a = 56.8, b = 101.4, c = 76.3, β = 104.9	
Number of molecules in the cell	2	
Solvent content (%)	55	
Diffraction limit (Å)	3.3	
Number of observed reflections	42104	
Number of unique reflections	11711	
	Overall	Outer shell
Resolution range (Å)	17.0–3.3	3.5–3.3
Completeness (%)	98.5	60.0
R_{sym}†	5.1	9.3
Average $I/\sigma(I)$	15.1	5.4

† $R_{sym}(I) = \sum|I - \langle I\rangle|/\sum|I|$.

(Anderson *et al.*, 1990), whereas in mare apo lactoferrin (Sharma, Rajashankar *et al.*, 1999) both lobes are found in closed conformations. Therefore, studies of lactoferrins from further species and in different states will help in the understanding of their functional differences.

In this context, the BLF provides a good test because its iron binding and release occur at a lower pH than CLF but at a higher pH than HLF and MLF (Sharma, Bhatia *et al.*, 1999). It shows a sequence identity of the order of 70% with other lactoferrins, but the sequence identity in the interlobe region is quite low (40%). It has four possible glycosylation sites compared with the three in human, three in mare, five in bovine, five in goat, three in porcine and four in camel lactoferrins.

Here, we present the three-dimensional structure of buffalo lactoferrin determined by the X-ray diffraction method at 3.3 Å resolution. In addition to providing new insights into the roles of interdomain and interlobe interactions in lactoferrin, several new features of various loop regions have been revealed.

2. Methods

2.1. Purification of lactoferrin

A modified procedure of Law & Reiter (1977) was used for the purification of lactoferrin (Raman *et al.*, 1992). Fresh colostrum from Murrah buffalo was diluted three times with warm distilled water and was defatted. The casein was precipitated using 10% HCl at pH 4.6. The clear whey was separated from the precipitated casein by filtration. The whey was diluted twice with 0.05 *M* Tris–HCl buffer at pH 8.0. The contaminating whey proteins bound on cation-exchange CM Sephadex C-50 were eluted at 0.2 *M* NaCl in 0.05 *M* Tris–HCl buffer. On further elution with 0.5 *M* NaCl in 0.05 *M* Tris–HCl buffer at pH 8.0, the dark red–brown coloured protein lactoferrin was obtained. This showed a single band at 80 kDa on SDS–PAGE.

2.2. Iron-saturated lactoferrin

1 m*M* native lactoferrin solution was prepared in 0.1 *M* sodium bicarbonate/0.1 *M* sodium citrate pH 8.0. 2 m*M* ferrric chloride hexahydrate ($FeCl_3.6H_2O$) reagent was prepared in 0.1 *M* sodium bicarbonate/0.1 *M* sodium citrate (pH 8.0). 1 m*M* protein solution was equilibrated in 1.2 ml ferric chloride hexahydrate reagent for 24 h at 298 K. Excess ferric chloride reagent was removed by dialyzing the protein against distilled water.

2.3. Crystallization

Crystals suitable for X-ray intensity data collection were obtained using the microdialysis method by equilibrating 50 mg ml^{-1} of iron-saturated lactoferrin in 0.025 *M* Tris–HCl against the same buffer containing 19%(*v*/*v*) ethanol at pH 8.0. The crystallization experiments were carried out at 277 K. The dark red–brown crystals grew to dimensions of 0.5 × 0.4 × 0.3 mm in three weeks.

2.4. Diffraction data

The crystals of iron-saturated lactoferrin were not stable in the X-ray beam and dissolved within 2 min on irradiation. They were transferred to various different buffering conditions, but were found to be stable only when 30%(*v*/*v*) MPD was added to the crystallization buffer. The intensity data collection was carried out at 277 K using a MAR Research imaging-plate scanner mounted on an RU-200 rotating-anode generator with a graphite monochromator. Data-collection and processing statistics are summarized in Table 1. The data were integrated using *MARXDS* and scaled using *MARSCALE* (Kabsch, 1988). Crystals belong to space group $P2_1$, with unit-cell parameters a = 56.8, b = 101.4, c = 76.3 Å, β = 104.9° and one molecule in the crystallographic asymmetric unit. The data had an R_{sym} of 5.1% and a completeness of 98.5% to 3.3 Å resolution.

2.5. Software

Molecular replacement was carried out using *AMoRe* (Navaza, 1994) and the maps were calculated in *X-PLOR* (Brünger, 1990). Model building was performed with *O* (Jones *et al.*, 1991). Refinement was carried out using *X-PLOR*. Structural quality checks of the model were performed with *PROCHECK* (Laskowski *et al.*, 1993). Additional graphical work was carried out with *O* and the figures were prepared using *MOLSCRIPT* (Kraulis, 1991).

2.6. Sequence determination

At the time when structure-analysis work was initiated, the sequence of buffalo lactoferrin was not known. Therefore, it was decided to determine the sequence before the structure analysis was undertaken. The complete cDNA sequence was determined. A lactating buffalo mammary gland was obtained from the slaughterhouse. The isolation of polyA$^+$ mRNA and cDNA syntheses were performed following the manufacturer's protocols (Stratagene, Germany). The conserved

nucleotide sequences in bovine (Pierce *et al.*, 1991), goat (Provost *et al.*, 1994) and porcine (Lydon *et al.*, 1992) lactoferrins were used for the synthesis of primers. The PCR was performed with *Taq* polymerase (Promega, USA) using MJ 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 model ABI-377. Both strands were used for sequencing. The buffalo lactoferrin cDNA reported here is 2307 base pairs (bp) in length. It

```
                                                                                            ▼
1          CAGGACCTCAGAC ATG AAG CTC TTC GTC CCC GCC CTG CTG TCC CTT GGA GCC CTT GGA CTG TGT CTG GCT GCC CCG AGG AAA AAC GTT    88
1                        Met Lys Leu Phe Val Pro Ala Leu Leu Ser Leu Gly Ala Leu Gly Leu Cys Leu Ala Ala Pro Arg Lys Asn Val    25

89    CGA TGG TGT ACC ATC TCC CAA CCC GAG TGG CTC AAA TGC CAC CGA TGG CAG TGG AGG ATG AAG AAG CTG GGT GCT CCC TCT ATC ACC TGT    178
26    Arg Trp Cys Thr Ile Ser Gln Pro Glu Trp Leu Lys Cys His Arg Trp Gln Trp Arg Met Lys Lys Leu Gly Ala Pro Ser Ile Thr Cys    55

179   GTG AGG AGG GCC TTT GTC TTG GAA TGT ATC CGG GCC ATC ACG GAA AAA AAG GCA GAT GCT GTG ACC CTG GAT GGT GGC ATG GTG TTT GAG    268
56    Val Arg Arg Ala Phe Val Leu Glu Cys Ile Arg Ala Ile Thr Glu Lys Lys Ala Asp Ala Val Thr Leu Asp Gly Gly Met Val Phe Glu    85

269   GCA GGC CTG GAC CCC TAC AAA CTG CGG CCA GTA GCA GCA GAG ATC TAT GGG ACC AAA GAG TCT CCC CAA ACC CAC TAT TAT GCT GTG GCT    358
86    Ala Gly Leu Asp Pro Tyr Lys Leu Arg Pro Val Ala Ala Glu Ile Tyr Gly Thr Lys Glu Ser Pro Gln Thr His Tyr Tyr Ala Val Ala    115

359   GTC GTC AAA AAG GGC AGC AAC TTT CAG CTG GAC CAG CTG CAA GGC CGG AAT TCC TGC CAT ACG GGC CTT GGC AGG TCT GCT GGG TGG AAC    448
116   Val Val Lys Lys Gly Ser Asn Phe Gln Leu Asp Gln Leu Gln Gly Arg Asn Ser Cys His Thr Gly Leu Gly Arg Ser Ala Gly Trp Asn    145

449   ATC CCT ATG GGA ATC CTT CGC CCG TAC TTG AGC TGG ACA GAG TCA CTC GAG CCC TTC CAG GGA GCT GTG GCT AAA TTC TTC TCT GCC AGC    538
146   Ile Pro Met Gly Ile Leu Arg Pro Tyr Leu Ser Trp Thr Glu Ser Leu Glu Pro Phe Gln Gly Ala Val Ala Lys Phe Phe Ser Ala Ser    175

539   TGT GTT CCC TGC GTT GAT AGA CAA GCG TAC CCC AAC CTG TGT CAA CTG TGC AAG GGG GAG GGG GAG AAC CAG TGT GCC TGC TCC CCC CGG    628
176   Cys Val Pro Cys Val Asp Arg Gln Ala Tyr Pro Asn Leu Cys Gln Leu Cys Lys Gly Glu Gly Glu Asn Gln Cys Ala Cys Ser Pro Arg    205

629   GAA CCA TAC TTC GGC TAT TCT GGT GCC TTC AAG TGT CTG CAG GAC GGG GCT GGA GAC GTG GCT TTT GTC AAG GAG ACG ACA GTG TTT GAG    718
206   Glu Pro Tyr Phe Gly Tyr Ser Gly Ala Phe Lys Cys Leu Gln Asp Gly Ala Gly Asp Val Ala Phe Val Lys Glu Thr Thr Val Phe Glu    235

719   AAC TTG CCA GAG AAG GCT GAC AGG GAC CAG TAT GAG CTT CTC TGC CTA AAC AAC ACT CGG GCA CCA GTG GAT GCA TTC AAG GAG TGC CAC    808
236   Asn Leu Pro Glu Lys Ala Asp Arg Asp Gln Tyr Glu Leu Leu Cys Leu Asn Asn Thr Arg Ala Pro Val Asp Ala Phe Lys Glu Cys His    265

809   CTG GCC CAG GTC CCT TCT CAT GCT GTC GTG GCC CGA AGT GTG GAT GGC AAG GAA GAC TTG ATC TGG AAG CTT CTC AGC AAG GCG CAG GAG    898
266   Leu Ala Gln Val Pro Ser His Ala Val Val Ala Arg Ser Val Asp Gly Lys Glu Asp Leu Ile Trp Lys Leu Leu Ser Lys Ala Gln Glu    295

899   AAG TTC GGA AAA AAC AAG TCT GGG AGC TTC CAG CTC TTT GGC TCT CCA CCC GGC CAG AGG GAC CTG CTA TTC AAA GAC TGT GCT CTT GGG    988
296   Lys Phe Gly Lys Asn Lys Ser Gly Ser Phe Gln Leu Phe Gly Ser Pro Pro Gly Gln Arg Asp Leu Leu Phe Lys Asp Cys Ala Leu Gly    325

989   TTT TTG AGG ATC CCC TCG AAG GTA GAT TCG GCA CTG TAC CTG GGC TCC CGC TAC TTG ACC GCC TTG AAA AAC CTC AGG GAA ACT GCG GAG    1078
326   Phe Leu Arg Ile Pro Ser Lys Val Asp Ser Ala Leu Tyr Leu Gly Ser Arg Tyr Leu Thr Ala Leu Lys Asn Leu Arg Glu Thr Ala Glu    355

1079  GAG GTG CAG GCA CGG CGC GCG AGG GTC GTG TGG TGC GCG GTG GGA CCC GAG GAG CAG AAA AAG TGC CAG CAG TGG AGC CAG CAG AGC GGC    1168
356   Glu Val Gln Ala Arg Arg Ala Arg Val Val Trp Cys Ala Val Gly Pro Glu Glu Gln Lys Lys Cys Gln Gln Trp Ser Gln Gln Ser Gly    385

1169  CAG ATC GTG ACC TGT GCC ACG GCC TCC ACC ACC GAT GAC TGC ATC GCC CTG GTG CTG AAA GGG GAA GCG GAT GCC CTG AGC TTG GAT GGA    1258
386   Gln Ile Val Thr Cys Ala Thr Ala Ser Thr Thr Asp Asp Cys Ile Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Leu Ser Leu Asp Gly    415

1259  GGA TAT ATC TAC ACT GCG GGC AAG TGT GGT CTG GTG CCT GTC CTG GCA GAG AAC CGG AAA TCC TCC AAA CAC AGT AGC CTA GAT TGT GTG    1348
416   Gly Tyr Ile Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Arg Lys Ser Ser Lys His Ser Ser Leu Asp Cys Val    445

1349  CTG AGA CCA ACG GAA GGG TAC CTT GCC GTG GCA GTT GTC AAG AAA GCA AAT GAG GGG CTC ACT TGG AAT TCT CTG AAA GGC AAG AAG TCG    1438
446   Leu Arg Pro Thr Glu Gly Tyr Leu Ala Val Ala Val Val Lys Lys Ala Asn Glu Gly Leu Thr Trp Asn Ser Leu Lys Gly Lys Lys Ser    475

1439  TGC CAC ACC GCC GTG GAC AGG ACT GCA GGC TGG AAC ATC CCC ATG GGC CTG ATC GCC AAC CAG ACA GGC TCC TGC GCA TTT GAT GAA TTC    1528
476   Cys His Thr Ala Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Ile Ala Asn Gln Thr Gly Ser Cys Ala Phe Asp Glu Phe    505

1529  TTT AGT CAG AGC TGT GCC CCT GGG GCT GAC CCG AAA TCC AGA CTC TGT GCA TTG TGT GCT GGC GAT GAC CAG GGC CTG GAC AAG TGT GTG    1618
506   Phe Ser Gln Ser Cys Ala Pro Gly Ala Asp Pro Lys Ser Arg Leu Cys Ala Leu Cys Ala Gly Asp Asp Gln Gly Leu Asp Lys Cys Val    535

1619  CCC AAC TCT AAG GAG AAG TAC TAT GGC TAC ACC GGG GCT TTC AGG TGC CTG GCT GAG GAT GTT GGG GAC GTT GCC TTT GTG AAA AAT GAC    1708
536   Pro Asn Ser Lys Glu Lys Tyr Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asp Val Gly Asp Val Ala Phe Val Lys Asn Asp    565

1709  ACA GTT TGG GAG AAC ACG AAT GGA GAG AGC ACT GCA GAC TGG GCT AAG AAC TTG AAT CGC GAG GAC TTC AGG TTG CTT TGC CTC GAT GGC    1798
566   Thr Val Trp Glu Asn Thr Asn Gly Glu Ser Thr Ala Asp Trp Ala Lys Asn Leu Asn Arg Glu Asp Phe Arg Leu Leu Cys Leu Asp Gly    595

1799  ACC AGG AAG CCT GTG ACG GAG GCT CAG AGC TGC CAC CTG GCG GTG GCC CCG AAT CAC GCT GTG GTG TCT TTG AGC GAA AGG GCA GCT CAC    1888
596   Thr Arg Lys Pro Val Thr Glu Ala Gln Ser Cys His Leu Ala Val Ala Pro Asn His Ala Val Val Ser Leu Ser Glu Arg Ala Ala His    625

1889  GTG GAA CAG GTG CTG CTC CAC CAG CAG GCT CTG TTT GGG GAA AAT GGA AAA AAC TGC CCG GAC AAA TTT TGT TTG TTC AAA TCT GAA ACC    1978
626   Val Glu Gln Val Leu Leu His Gln Gln Ala Leu Phe Gly Glu Asn Gly Lys Asn Cys Pro Asp Lys Phe Cys Leu Phe Lys Ser Glu Thr    655

1979  AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG GCC AAA CTT GGA GGC AGA CCA ACG TAT GAA GAA TAT TTG GGG ACA GAG TAT GTC    2068
656   Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Lys Leu Gly Gly Arg Pro Thr Tyr Glu Glu Tyr Leu Gly Thr Glu Tyr Val    685

2069  ACA GCC ATT GCC AAC CTG AAA AAA TGC TCA ACC TCC CCG CTT CTG GAA GCC TGC GCC TTC CTG ACG AGG TAA AGCCTGCAAAGAAGCTAGCCTGC    2163
686   Thr Ala Ile Ala Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala Cys Ala Phe Leu Thr Arg ***                          709

2164  CTCCCTGGGCCTCAGCTCCTCCCTGCTCTCATCCCCAATCTCCAGGCGCGAGGGACCCTCCTCTCCCTTCCTGAAGTTGGATTTTTGCCAAGCTCATCAGTATTCACAATTCCCTGCTG    2282

2283  TCATCTTAGCAAGAAATAAAATTAG                                                                                                   2307
```

Figure 1

Nucleotide and deduced amino-acid sequence of buffalo lactoferrin. The amino acids are shown as three-letter codes. The triangle marks the N-terminal amino acid of the mature protein. The stop codon is indicated by ***. The consensus sequence for polyA addition is boxed. Predicted glycosylation sites are underlined.

Table 2
Summary of refinement statistics.

R factor (%)	
After molecular replacement (10–5.0 Å)	44.9
After rigid-body refinement (10–4.0 Å)	28.5
Final (17–3.3 Å) for all the data	21.8
Free *R* factor (%)	29.6
Number of reflections	11711
Number of atoms (including Fe^{3+} and CO_3^{2-})	5326
R.m.s. deviations from ideal values	
Bond distances (Å)	0.010
Bond angles (°)	1.6
Improper angles (°)	0.7
Ramachandran plot (non-Gly and non-Pro)	
Residues in most favoured regions (%)	81
Residues in additionally allowed regions (%)	17.6
Residues in generously allowed regions (%)	1.0
Residues in disallowed regions (%)	0.3
Overall *G* factor	−0.01

comprises 13 bp in the 5′-UTR, the open reading frame encoding 708 amino-acid residues and part of the 3′-UTR (170 bp). The nucleotide and derived amino-acid sequences are given in Fig. 1 (Karthikeyan *et al.*, 1999).

2.7. Structure determination

The structure was solved by molecular replacement using *AMoRe* (Navaza, 1994) with a poly-Ala model of human lactoferrin (Haridas *et al.*, 1995). It gave a clear solution for both the rotation and translation functions. After rigid-body refinement, the *R* factor was 47.2% and the correlation coefficient was 50.8%.

2.8. Refinement

The initial model was built using *O* (Jones *et al.*, 1991). Refinement was carried out by the rigid-body, conjugate-gradient minimization, simulated-annealing and *B*-factor refinement protocol of *X-PLOR* (Brünger, 1990) using the Engh and Huber parameter libraries (Engh & Huber, 1991). Initially, rigid-body minimization was performed by considering four rigid groups corresponding to the four domains in the molecule. This reduced the *R* factor to 38.5%. The sequence of buffalo lactoferrin was compared with the map and the side chains were introduced into the electron density without much ambiguity. The electron density improved after each cycle of refinement and the *R* factor fell gradually. Repeated steps of model building using $2F_o - F_c$ and $F_o - F_c$ maps with *O* reduced the *R* factor to 24.6%. In view of the limited data, a conservative *B*-factor refinement was performed. The global *B* factor was initially calculated, followed by the refinement of one *B* factor per residue. The $2F_o - F_c$ and $F_o - F_c$ maps calculated at this stage showed good electron density at the iron-binding sites, to which Fe atoms and carbonate ions were fitted in both lobes (Fig. 2). The positions of the Fe atoms and carbonate ions were refined with distance constraints. The final *R* factor was 21.8% ($R_{\rm free}$ = 29.6%) (Table 2).

3. Results and discussion

3.1. Quality of the model

The final model (summarized in Table 2) consists of 5316 protein atoms from 689 residues, two Fe^{3+} and two CO_3^{2-} ions. The final crystallographic *R* factor was 21.8%. The structure is well defined and in the final $2F_o - F_c$ electron-density map there are no breaks in the main-chain density when contoured at the 1σ level. A section of the final $2F_o - F_c$ electron density is shown in Fig. 3. The average *B* factor for the structure is 35.0 Å^2, whereas in bovine lactoferrin the average *B* factor was found to be 71.4 Å^2. The root-mean-square (r.m.s.) coordinate error is estimated to be 0.38 Å from a Luzzati plot (Luzzati, 1952). A Ramachandran plot of the main-chain torsion angles (φ, ψ) (Ramachandran & Sasisekharan, 1968) showed that 81% of the residues were in the most favoured regions as defined in the program *PROCHECK* (Laskowski *et al.*, 1993). Only two residues were in the disallowed regions, but these (Leu299 and Leu640) are the central residues in two γ-turns (Matthews, 1972) and have (φ, ψ) values that are typical of such configurations, around (70, −50) (Baker & Hubbard, 1984). These two γ-turns are conserved in the N and C lobes of lactoferrins, transferrins and ovotransferrins (Bailey *et al.*, 1988; Haridas *et al.*, 1995; Kurokawa *et al.*, 1995; Moore *et al.*, 1997; Sharma, Paramasivam *et al.*, 1999), where they form part of one wall of each binding cleft.

3.2. Overall molecular structure

The amino-acid sequence of buffalo lactoferrin consists of a single polypeptide chain of 689 residues (Fig. 1). The overall folding of BLF is similar to HLF, MLF and CLF. The sequence identity of these proteins ranges from 69 to 80%. The secondary structures are conserved. The conformations of some of the loop regions differ substantially. The most prominent differences occurred in the N lobe at the mouth of the binding cleft, where the protein chains traversed significantly different paths. The protein chains from two domains (N1 and N2) approach each other to determine the size of the opening of the binding cleft. The opening to the cleft (the distance between the nearest C^α atoms from the opposite walls of the cleft) is smallest in mare lactoferrin (5.9 Å) and largest in buffalo lactoferrin (8.4 Å). The corresponding values in bovine and human lactoferrins are 6.3 and 7.7 Å, respectively. In this region, the BLF chain shows higher structural similarity with HLF, although the sequence homology is far higher with CLF. The individual domains of BLF compare well with those of HLF (r.m.s. differences of 0.9, 0.7, 0.7, 0.6 Å for the N1, N2, C1 and C2 domains, respectively), MLF (r.m.s. differences of 0.9, 0.9, 0.8, 0.6 Å for the N1, N2, C1 and C2 domains, respectively) and CLF (0.7, 0.5, 0.5, 0.6 Å for the N1, N2, C1 and C2 domains, respectively). The lobes as a whole superimpose less well, *i.e.* the N lobe in HLF (0.9 Å), MLF (1.0 Å) and CLF (0.7 Å) and the C lobe in HLF (0.7 Å), MLF (0.7 Å) and CLF (0.6 Å). This is because there are differences in the extent of domain closure for each lobe, *i.e.* when one pair of

Table 3
Distances in metal- and anion-binding sites.

	N lobe	C lobe
Metal–ligand bond lengths (Å)		
Fe–OD1 Asp60 (395)	2.0	1.8
Fe–OH Tyr92 (433)	1.7	2.0
Fe–OH Tyr192 (526)	2.2	2.0
Fe–NE2 His253 (597)	2.0	2.4
Fe–O1 (carbonate)	2.3	2.1
Fe–O2 (carbonate)	1.9	1.8
Hydrogen-bond lengths (Å)		
O1···NE Arg121 (463)	3.2	2.6
O1···NH2 Arg121 (463)	3.5	3.6
O2···N 123 (465)	3.0	3.1
O3···OG Ser117 (459)	2.9	3.0
O3···N 124 (466)	3.4	3.0

domains is superimposed (*e.g.* N1 domains) the other pair (N2 domain) do not match and an additional rotation is required to bring them into correspondence. By this analysis, the N lobe of BLF is oriented differently to that of HLF, MLF and CLF by 5.1, 4.8 and 3.0°, respectively, and the C-lobe of BLF is oriented differently to that of HLF, MLF and CLF by 2.4, 2.0 and 2.5°, respectively.

When the analysis was extended to other proteins of the transferrin family, *i.e.* HOT and DOT, the corresponding differences were found to be 1.1 and 3.0° for the N-lobe and 1.1 and 2.4° for the C-lobe for HOT and DOT, respectively.

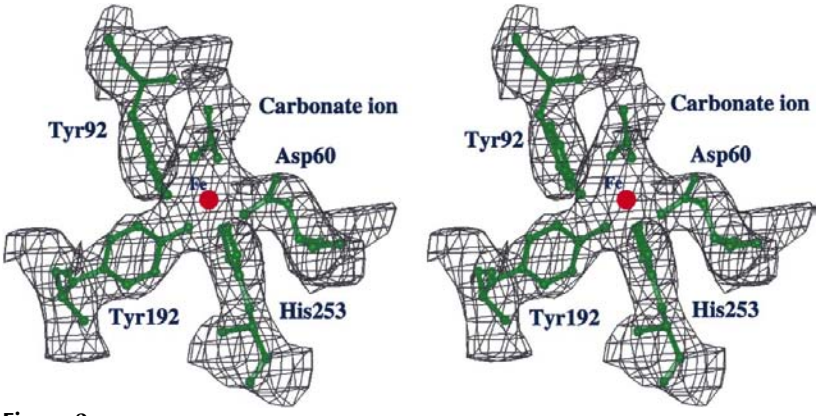

Figure 2
Stereoview of $2F_o - F_c$ electron-density map at the iron-binding site for the N lobe. The map is contoured at 1.0σ. Similar electron density is present at the C lobe binding site.

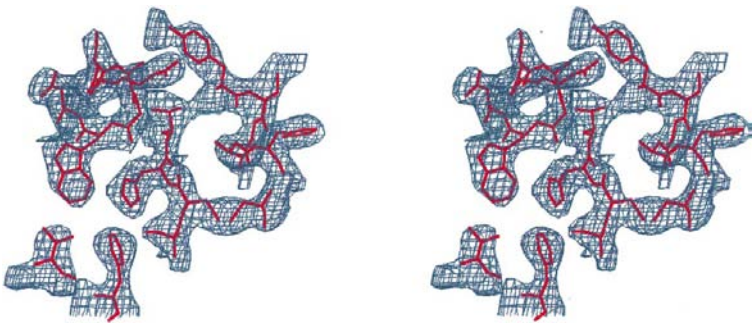

Figure 3
Stereoview of a section of a final $2F_o - F_c$ electron-density map contoured at 1.0σ.

3.3. Metal- and anion-binding sites

The two metal- and anion-binding sites are very similar to each other and to the corresponding sites in other lactoferrins. The iron ligands are Asp60, Tyr92, Tyr192, His253 and the bidentate CO_3^{2-} ion in the N-lobe and Asp395, Tyr433, Tyr526, His597 and the CO_3^{2-} ion in the C-lobe (Fig. 2). The carbonate ion occupies a positively charged pocket in the wall of domain 2 (N2 or C2). This pocket is formed by the N-terminus of an α-helix and the side chain of Arg121 (in the N lobe) and Arg463 (in the C lobe). The metal–ligand bonds and the hydrogen bonds involving the anion are listed in Table 3.

3.4. Interdomain interactions

The interdomain interactions in lactoferrins and transferrins were considered to be critical in enhancing or retarding the iron binding and release, which were believed to be associated with the large-scale conformational changes. This idea was consistent with the observations in diferric human lactoferrin (Anderson *et al.*, 1989) and human apo lactoferrin (Anderson *et al.*, 1990), where the conformations of the N lobe in the two crystal structures differed greatly. Similarly, in diferric duck ovotransferrin (Rawas *et al.*, 1996) and duck apo ovotransferrin (Rawas *et al.*, 1997), both lobes adopted closed and open conformations, respectively. Contrary to the above observations, the structures of both mare diferric (Sharma, Paramasivam *et al.*, 1999) and mare apo (Sharma, Rajashankar *et al.*, 1999) lactoferrins were found to be identical. This suggested that the large-scale conformational changes were not necessary for iron binding and release. Despite the different conformations of their apo forms, all the lactoferrins tend to have similar interdomain interactions in their iron-bound states.

3.5. Interactions between the lobes

The relative orientations of two lobes vary significantly in different lactoferrins. The connecting peptide in all of them adopts a α-helical conformation which may differ only slightly and may not be the only reason for the variation. A detailed examination of the interlobe interactions revealed that the two lobes interact mainly through hydrophobic forces. There is only one electrostatic interaction in BLF between the two lobes, involving Asp315 and Lys86, and more than 33 van der Waals interactions with distances less than 4.0 Å. The existence of the hydrophobic interactions is consistent with the observation that the cleaved molecular halves of lactoferrin tend to reassociate (Singh *et al.*, 1998; Sharma, Bhatia *et al.*, 1999), but this association can be abolished using detergents and organic solvents. The residues that are the main contributors are are Pro311, Lys313, Val314, Asp315, Leu318 and Ser322 on the N lobe and Leu385, Lys386, Pro679, Leu680,

Table 4
Relative orientations of two lobes compared with buffalo lactoferrin (°).

Bovine lactoferrin	3.3
Mare	6.9
Human	8.0
Hen ovotransferrin	21.2
Duck ovotransferrin	22.2

Ala683, Pro686, Leu687 and Arg689 on the C lobe. The differences in the relative orientations of two lobes in various lactoferrins (Table 4) seem to arise from slight differences in the packings of the lobes at the interface. There are several changes in the sequences at the interface in various lactoferrins, thus causing easy variation in the lobe orientations.

3.6. Crystal packing

The molecules of BLF are packed loosely, as there are only limited direct protein–protein contacts with distances less than 3.5 Å. The amino acids which are involved in the intermolecular hydrogen bonds are Arg25, Arg47, Lys53, Lys100, Asn103, Glu140, Glu336 in the N-lobe and Glu535, Asp536, Asn565, Lys652, Glu658, Glu659, Lys674 in the C-lobe. The solvent content in the crystals of BLF is 55%, which is on the high side and might be responsible for the large number of solvent-mediated intermolecular interactions. Unfortunately, owing to the limited resolution, the positions of solvent molecules were not determined. Since the packing of the molecules is not very tight, intramolecular events such as orientations of the domains and lobes are unlikely to be affected by solvent-dependent intermolecular interactions in these crystals. This also explains the difficulties in obtaining good-quality crystals of BLF. It may be worth mentioning here that the BLF is the first protein of the transferrin family which has been crystallized in a monoclinic space group. Despite the high solvent content of BLF, the average *B* factor was found to be 35.0 Å^2, whereas in the case of bovine lactoferrin with a similar amount of solvent in the unit cell, the average *B* factor was 71.4 Å^2.

The sequence of the protein has been deposited in the EMBL database with accession No. AJ005203.

The authors thank the Department of Science and Technology and Department of Biotechnology (New Delhi) for financial support.

References

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., 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. Mol. Biol.* **44**, 97–179.

Brünger, A. T. (1990). *X-PLOR Manual, Version* 3.1. The Howard Hughes Medical Institute, New Haven, Connecticut, USA.

Engh, R. H. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–646.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Karthikeyan, S., Sharma, S., Sharma, A. K., Paramasivam, M., Yadav, S., Srinivasan, A. & Singh, T. P. (1999). *Curr. Sci.* **77**, 1926–1941.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kurokawa, H., Mikami, B. & Hirose, M. (1995). *J. Mol. Biol.* **254**, 196–207.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Law, B. A. & Reiter, B. (1977). *J. Dairy Res.* **44**, 595–599.

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

Lydon, J. P., O'Malley, B. R., Saucedo, O., Lee, T., Headon, D. R. & Conneely, O. M. (1992). *Biochim. Biophys. Acta*, **1132**, 97–99.

Matthews, B. W. (1972). *Macromolecules*, **5**, 818–819.

Moore, S. A., Anderson, B. F., Groom, C. R., Haridas, M. & Baker, E. N. (1997). *J. Mol. Biol.* **274**, 222–236.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Pierce, A., Colavizza, D., Benaissa, M., Maes, P., Tartar, A., Montreuil, J. & Spik, G. (1991). *Eur. J. Biochem.* **196**, 177–184.

Provost, F. L., Nocart, M., Guerin, G. & Martin, P. (1994). *Biochem. Biophys. Res. Commun.* **203**, 1324–1332.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

Raman, A., Bhatia, K. L., Singh, T. P., Betzel, Ch. & Mallik, R. C. (1992). *Arch. Biochem. Biophys.* **294**, 318–321.

Rawas, A., Muirhead, H. & Williams, J. (1996). *Acta Cryst.* D**52**, 631–640.

Rawas, A., Muirhead, H. & Williams, J. (1997). *Acta Cryst.* D**53**, 464–468.

Sharma, S., Bhatia, K. L. & Singh, T. P. (1999). *J. Dairy Res.* **66**, 81–90.

Sharma, A. K., Paramasivam, M., Srinivasan, A., Yadav, M. P. & Singh, T. P. (1999). *J. Mol. Biol.* **289**, 303–317.

Sharma, A. K., Rajashankar, K. R., Yadav, M. P. & Singh, T. P. (1999). *Acta Cryst.* D**55**, 1152–1157.

Singh, T. P., Sharma, S., Karthikeyan, S., Betzel, Ch. & Bhatia, K. L. (1998). *Proteins Struct. Funct. Genet.* **33**, 30–38.