Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of recombinant human lactoferrin expressed in *Aspergillus awamori*

Xiao-Lin Sun,† Heather M. Baker,† Steven C. Shewry,† Geoffrey B. Jameson and Edward N. Baker*†

Departments of Biochemistry and Chemistry, Massey University, Palmerston North, New Zealand

† Present address: School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand.

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 24 March 1998
Accepted 25 August 1998

PDB Reference: recombinant human lactoferrin, 1bol.

Human lactoferrin (hLf) has considerable potential as a therapeutic agent. Overexpression of hLf in the fungus *Aspergillus awamori* has resulted in the availability of very large quantities of this protein. Here, the three-dimensional structure of the recombinant hLf has been determined by X-ray crystallography at a resolution of 2.2 Å. The final model, comprising 5339 protein atoms (residues 1–691, 294 solvent molecules, two $Fe^{3+}$and two CO_3^{2-} ions), gives an R factor of 0.181 (free R = 0.274) after refinement against 32231 reflections in the resolution range 10–2.2 Å. Superposition of the recombinant hLf structure onto the native milk hLf structure shows a very high level of correspondence; the main-chain atoms for the entire polypeptide can be superimposed with an r.m.s. deviation of only 0.3 Å and there are no significant differences in side-chain conformations or in the iron-binding sites. Dynamic properties, as measured by B-value distributions or iron-release kinetics, also agree closely. This shows that the structure of the protein is not affected by the mode of expression, the use of strain-improvement procedures or the changes in glycosylation due to the fungal system.

1. Introduction

Lactoferrin is an iron-binding protein which is the second most abundant protein in human milk (at levels in excess of 1 g l^{-1}) and is also expressed in other external secretions such as tears, saliva, nasal secretions, intestinal secretions and genital secretions (Masson *et al.*, 1966; Yu & Chen, 1993) as well as in white blood cells (Masson *et al.*, 1969). Lactoferrin is a member of the transferrin family of proteins (Brock, 1985; Baker, 1994), with which it shares a large number of structural and functional features. Like other transferrins, it is a glycoprotein of about 78 kDa (691 residues), it binds two Fe^{3+} ions with high affinity ($K_d \simeq 10^{-20}$) together with two CO_3^{2-} ions and it contains a twofold internal amino-acid sequence repeat that results in a bilobal protein with similar N- and C- terminal halves. X-ray crystallographic studies have previously established the three-dimensional structure of native human lactoferrin purified from human milk; these studies have defined the structures of not only the diferric form of hLf (Anderson *et al.*, 1989; Haridas *et al.*, 1995) but also an iron-free (apo) form (Anderson *et al.*, 1990) and numerous metal and anion-substituted forms (*e.g* Smith *et al.*, 1992; Baker *et al.*, 1996), the N-terminal half-molecule (Day *et al.*, 1993) and numerous site-specific mutants (*e.g.* Faber *et al.*, 1996).

Lactoferrin appears to be a multifunctional protein, with many demonstrated or proposed biological activities (Sanchez *et al.*, 1992). It has well recognized antibacterial properties that may operate through several different mechanisms; firstly, a

© 1999 International Union of Crystallography
Printed in Denmark – all rights reserved

bacteriostatic effect in which its high iron-binding affinity deprives bacteria of the iron that is essential for growth (Bullen *et al.*, 1972) and, secondly, a direct bactericidal activity that may derive from a highly cationic bactericidal domain on the N-terminal lobe of the molecule (Bellamy *et al.*, 1992). Lactoferrin also appears to regulate systemic immune responses, perhaps through its ability to bind to a wide variety of cells (Birgens *et al.*, 1983) and to regulate cytokine release (Crouch *et al.*, 1992). Among other activities, lactoferrin binds heparin and glycosaminoglycans (Mann *et al.*, 1994), acts as a promoter of cell growth (Hashizume *et al.*, 1983) and acts as an antioxidant by suppression of damaging free radicals (Baldwin *et al.*, 1984).

Until recently, the therapeutic use of lactoferrin has been limited by the lack of an efficient and cost-effective method for the production of the human protein in large quantities. This situation has changed with the demonstration that recombinant human lactoferrin can be produced in filamentous fungi (Ward *et al.*, 1992) with further development of the expression system, coupled with strain-improvement procedures, resulting in expression levels in excess of 2 g l^{-1} in the fungus *Aspergillus awamori* (Ward *et al.*, 1995). Expression levels of over 5 g l^{-1} are now reported (D. R. Headon, personal communication). The strain improvements do, however, involve rounds of mutagenesis and subsequent selection of higher-expressing strains, and the fungal expression system also results in changes to the composition of the N-linked carbohydrate chains (Ward *et al.*, 1995).

Here, we describe the three-dimensional structure of the recombinant human lactoferrin produced in *Aspergillus awamori*, in order to compare it with that of the native milk protein and so determine whether the structure of the recombinant protein is in any way affected by the method of production.

2. Experimental

2.1. Protein expression and purification

The protein used was provided by Agennix Inc. (Houston, Texas) from large-scale production of recombinant lactoferrin in *Aspergillus awamori* as described previously (Ward *et al.*, 1995). The particular transformant used was one giving an expression level of ~2 g l^{-1}. For crystallization the protein was first fully loaded with iron, using ferric nitrilotriacetate (FeNTA), dialysed into 20 m*M* Tris–HCl pH 7.0, 0.2 *M* NaCl, passed down a gel-filtration column (Pharmacia Hiload 16/60 Superdex 75) and concentrated to 80 mg ml^{-1} in 10 m*M* HEPES pH 8.0, 0.5 *M* NaCl.

2.2. Kinetics of iron release

Iron release was measured by a spectrofluorometric method adapted from that of Egan *et al.* (1993). The protein solution (80 µg of iron-saturated hLf in 10 µl of 0.05 *M* Tris–HCl pH 7.9) was mixed with 3 ml of buffer (0.05 *M* acetate buffer at the desired pH) containing 0.1 *M* NaCl and 0.01 *M* sodium citrate; the citrate acted as an iron 'sink' to prevent precipitation or re-binding of iron. Measurements were carried out at different pH values over the range (2.0–4.0) within which hLf releases iron (Ward *et al.*, 1995).

Table 1
Data-collection and refinement details.

Data collection
Figures in parentheses refer to the outermost (2.2–2.3 Å) shell.

Resolution range (Å)	40–2.2
Unique reflections	36329
Completeness (%)	82.3 (67.8)
Multiplicity	2.8 (3.1)
R_{merge}	0.046 (0.542)
I/σ	11.1 (1.5)

Refinement
Figures in parentheses refer to the free *R* calculation.

Resolution range (Å)	10–2.2
Number of reflections	32231 (3573)
R factor	0.181 (0.274)
Protein atoms	5339
Ions	2 Fe^{3+}, 2 CO_3^{2-}
Water molecules	294
R.m.s. deviations	
Bond distances (Å)	0.003
Angle distances (Å)	0.010

2.3. Crystallization

Crystals were grown by microdialysis of the above protein solution against 10 m*M* sodium phosphate pH 8.0 containing 10%(*v*/*v*) ethanol. Crystals of dimensions up to 1.5 × 1.0 × 0.5 mm grew in 3 weeks at 277 K. These crystals proved to be isomorphous with the native hLf crystals. The unit cell was orthorhombic, space-group $P2_12_12_1$, with dimensions a = 156.3, b = 97.5, c = 55.9 Å (*cf* native hLf crystals a = 156.2, b = 97.3, c = 55.85 Å).

2.4. Data collection

X-ray data were collected at room temperature with a Rigaku R-AXIS IIC image-plate detector on an RU200 rotating-anode generator using Cu $K\alpha$ radiation (λ = 1.5418 Å). Data were processed with *DENZO* (Otwinowski, 1990) and the programs *ROTAVATA* and *AGROVATA* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The data set of 36329 reflections was 82.3% complete (67.8% complete in the outermost, 2.3–2.2 Å, shell), with a redundancy of 2.8 and overall R_{merge} of 0.046. The mean fractional isomorphous difference between this data set and that of native milk hLf was 0.065. Further details are given in Table 1.

2.5. Structure determination and refinement

The initial model was taken from the refined native milk hLf structure (Haridas *et al.*, 1995) with the omission of all solvent molecules. Refinement was initially with *TNT* (Tronrud *et al.*, 1987), with the final refinement cycles being performed with *SHELXL*97. Both refinements used 32231 reflections from the resolution range 10–2.2 Å, with another

3573 reflections (~10% of total) being set aside for the free R calculation. Initial cycles of rigid-body refinement (four rigid bodies, *i.e.* the four domains N1, N2, C1 and C2) were followed by cycles of restrained least-squares refinement interspersed with model checking by *PROCHECK* (Laskowski *et al.*, 1993) and manual adjustments on a computer-graphics system. Solvent molecules were added if peaks were present in both $2F_o - F_c$ and $F_o - F_c$ maps at levels $>1\sigma$ and $>3\sigma$, respectively, and the positions were geometrically and chemically reasonable. Bond lengths involving the Fe^{3+} and CO_3^{2-} ions were restrained early in the refinement, but these restraints were removed later; the Fe^{3+} and CO_3^{2-} positions were also checked several times by omitting them from the model and calculating omit maps.

The final model (5339 protein atoms, two Fe^{3+} ions, two CO_3^{2-} ions and 294 solvent molecules, all regarded as water) gave an R factor of 0.181 (free R = 0.274). Other details of the refinement parameters are given in Table 1.

3. Conclusions

3.1. Model quality

The final model is of good quality, as judged by the usual criteria. A Ramachandran plot (Fig. 1) shows that 85.0% of non-glycine residues are in the 'most favoured' regions defined by *PROCHECK* (Laskowski *et al.*, 1993) and a further 14.3% are in the 'allowed region'. Only two residues are in the normally disallowed region, but these are Leu299 and Leu642 which have excellent electron density and are the central residues in two γ-turns which appear to be a conserved feature in all transferrin structures determined to date (Haridas *et al.*, 1995); these γ-turns form part of one wall of the interdomain cleft. Bond lengths and angles in the protein structure have been restrained to the standard values of Engh & Huber (1991).

The overall mean B values are 43.4 Å^2 for main-chain atoms and 50.1 Å^2 for side-chain atoms. The final model fits the electron density well except in a few external loops and turns; these include the N-terminal residues 1–4, for which the electron density is very poorly defined, and the flexible loops 219–222, 281–285 and 416–425. No attempt has been made to model the carbohydrate (see below).

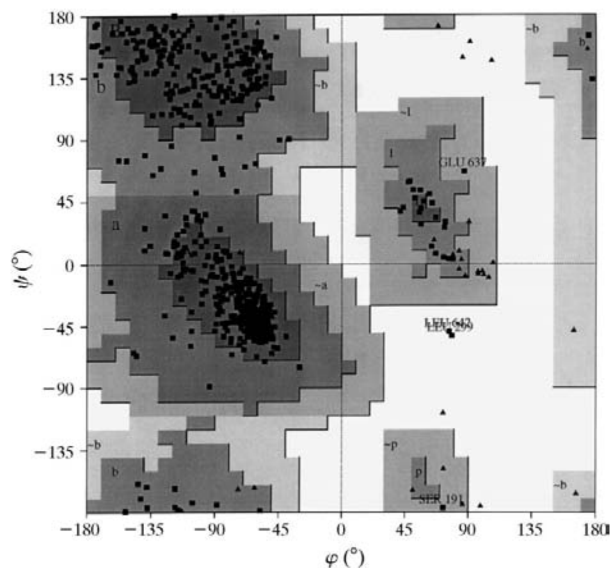

Figure 1
Ramachandran plot of main-chain torsion angles (φ, ψ) for recombinant hLf. Glycine residues are shown by triangles. Drawn with *PROCHECK* (Laskowski *et al.*, 1993).

3.2. Protein structure

The folding of the recombinant hLf molecule is unchanged from that of the native milk hLf. When the two models are superimposed, the r.m.s. difference in atomic positions is only 0.31 Å for all main-chain atoms, omitting only residues 1–4. This high degree of correspondence can be clearly seen in Fig. 2. The N-terminal residues 1–4 have been modelled differently in the two structures, such that the N-termini are 8 Å apart, but we do not attach any significance to this difference since the B values are extremely high in both cases (greater than 100 Å^2). Clearly these residues are highly mobile and/or disordered. Only at one other location, the poorly defined external loop 417–423, are there differences of greater than 1.0 Å (maximum 1.2 Å). This, too, can be ascribed to disorder or flexibility. The agreement between the two structures is the same irrespective of whether the superposition is based on individual domains, individual lobes or the whole molecules. This shows that relative lobe and domain orientations are also the same between the native and recombinant proteins. All of the 16 disulfide bonds are correctly formed and most side-chain conformations are conserved between the two structures, the only apparent differences being in some of the flexible poorly ordered side chains on the protein surface. Likewise, most of the internal water molecules, including those in the two interdomain clefts, are found in the same positions in both structures.

The clone used to produce the recombinant lactoferrin contains one known sequence difference; at residue 10 (our numbering) Thr is substituted for Ala. This residue is partly

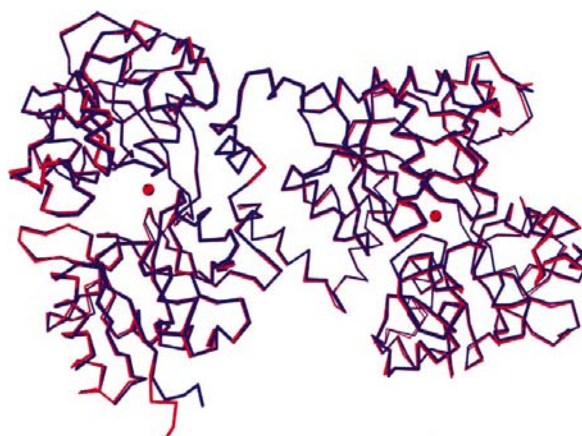

Figure 2
Cα plot showing superposition of recombinant hLf (blue) onto native milk hLf (yellow). The Fe atoms are shown by red spheres.

buried in one wall of the binding cleft, and the Thr side chain is accommodated easily in the recombinant protein without disturbance of the structure. The methyl group makes no contacts less than 3.5 Å and the hydroxyl group makes a hydrogen bond to the peptide carbonyl oxygen of Ile37, which is reoriented slightly (a displacement of 0.2 Å). No other sequence changes are visible in the electron density.

The binding sites for the two Fe^{3+} and two CO_3^{2-} ions are also identical within the experimental error of the two analyses. When the two structures are superimposed, the atoms of the two binding sites (two Fe^{3+} ions and two CO_3^{2-}, Asp60, Tyr92, Thr117, Arg121, Tyr192, His253, Asp395, Tyr435, Thr461, Arg465, Tyr528 and His597; total 130 atoms) have an r.m.s. difference of only 0.24 Å. Comparison of the bond lengths and angles in the binding sites shows an r.m.s. difference in metal–ligand bond lengths of only 0.09 Å and in bond angles of 6.0°. None of the differences appear significant and this is consistent with the near-identical UV–visible spectra of the two proteins; native milk diferric lactoferrin has a visible absorption maximum of 466 nm and the recombinant hLf from *A. awamori* has a maximum at 464 nm.

3.3. Glycosylation

Analysis of the glycosylation of the recombinant hLf from *A. awamori* has shown that although the extent of glycosylation is similar to that in native milk hLf, there are differences in the composition of carbohydrate chains; the oligosaccharides in the recombinant protein are high-mannose type compared with the complex carbohydrate structure of the milk hLf (Ward *et al.*, 1995). No interpretable density can be seen for the carbohydrate in the present structural analysis, however. At both glycosylation sites (Asn137 and Asn478), there is limited electron density extending a short distance from the Asn side chains. It is only sufficient for a single sugar in each case and only continuous at the 0.5σ level, and therefore we have made no attempt to model the carbohydrate. There are suggestions that the carbohydrate is slightly less well ordered than in native milk hLf, perhaps because of its simpler structure, but what is clear is that in neither case does the carbohydrate make any contribution to the protein three-dimensional structure.

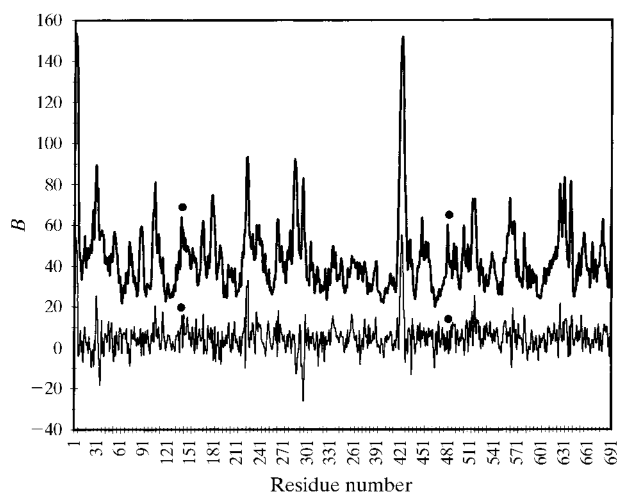

Figure 3
Distribution of main-chain B values (averaged values for N, Cα, C, O atoms in Å^2) for recombinant hLf, plotted as a function of residue number (upper trace). In the lower trace are plotted the differences in B values between recombinant hLf and native milk hLf, *i.e.* $B_{\rm rec.hLf} - B_{\rm hLf}$. The two glycosylation sites are marked with filled circles.

3.4. Dynamics

The crystallographic B values show a consistent pattern of agreement between the recombinant hLf and native milk hLf when plotted as a function of residue number (Fig. 3). Although B values can be influenced by model errors or other errors in refinement, they do give some measure of the flexibility of different regions of the molecule; the consistency of the pattern of B values between the two structures, despite different refinement protocols, therefore implies similar dynamic properties. Even at the two glycosylation sites (Asn137 and Asn478) there is no apparent difference in mobility, despite different glycosylation. (We note also that very few direct intermolecular contacts, which might restrict flexibility, are present in the crystals – only 22 out of 5339 atoms make direct intermolecular contacts for recombinant hLf and 32 out of 5330 atoms for milk hLf.)

The kinetics of iron binding or release also give some measure of dynamics. Iron-release curves for the two proteins are shown in Fig. 4. Although the rate appears slightly faster for the recombinant protein, the low pH (2.5) required makes these data sensitive to solution conditions and to the particular preparations used. The results show that the native and recombinant proteins have very similar iron-release kinetics, which are clearly differentiated from those of bovine lactoferrin.

3.5. Structural and functional implications

The results of this analysis show that neither the mutagenesis used in strain improvement, nor the method of expression, nor the altered glycosylation have any impact on the three-dimensional structure of lactoferrin. The polypeptide folding is the same, as are the hydrogen bonds and other

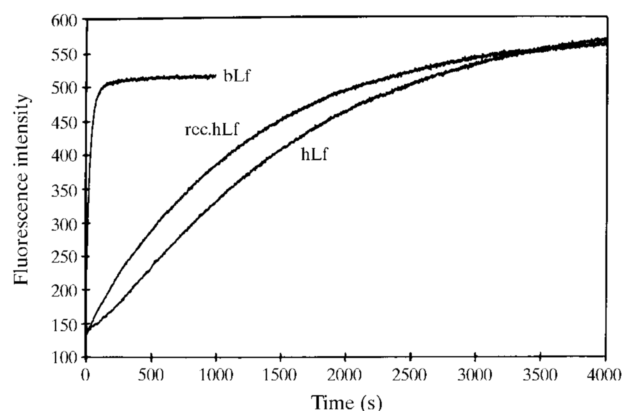

Figure 4
Kinetics of iron release for human milk Lf (hLf), recombinant human Lf from *A. awamori* (rec. hLf) and bovine lactoferrin (bLf), all at pH 2.5. The curves show simple first-order iron-release kinetics.

interactions that stabilize the structures, including internal water molecules. The identity in the iron sites is also consistent with the similar iron-binding and iron-release characteristics reported for the recombinant hLf in comparison with native milk hLf (Ward *et al.*, 1995). A similar distribution of *B* values through the two molecules is also apparent. While for some residues (at the few crystal contact regions) these may be influenced by crystal packing, in most cases the *B* factors represent genuine molecular properties and are consistent with similar dynamic properties of the two proteins; this is consistent also with the close similarity of the iron-release kinetics.

We wish to acknowledge financial support from the US National Institutes of Health (grant HD-20859). The lactoferrin was a generous gift from Agennix Inc.

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.
Baker, E. N. (1994). *Adv. Inorg. Chem.* **41**, 389–463.
Baker, H. M., Anderson, B. F., Brodie, A. M., Shongwe, M. S, Smith, C. A. & Baker, E. N. (1996). *Biochemistry*, **35**, 9007–9013.
Baldwin, D. A., Jenny, R. R. & Aisen, P. (1984). *J. Biol. Chem.* **259**, 13391–13394.
Bellamy, W., Takase, M., Yamauchi, K., Wakabayashi, H., Kawase, K. & Tomita, M. (1992). *Biochim. Biophys. Acta*, **1121**, 130–136.
Birgens, H. S., Hansen, N. E., Karle, H. & Kristensen, L. O. (1983). *Br. J. Haematol.* **54**, 383–391.
Brock, J. H. (1985). *Metalloproteins*. Part 2, edited by P. Harrison, pp. 183–262. London: MacMillan Press.
Bullen, J. J., Rogers, H. J. & Leigh, L. (1972). *Br. Med. J.* **3**, 69–75.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Crouch, S. P. M., Slater, K. J. & Fletcher, J. (1992). *Blood*, **80**, 235–240.
Day, C. L., Anderson, B. F., Tweedie, J. W. & Baker, E. N. (1993). *J. Mol. Biol.* **232**, 1084–1100.
Egan, T. J., Zak, O. & Aisen, P. (1993). *Biochemistry*, **32**, 8162–8167.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Faber, H. R., Baker, C. J., Day, C. L., Tweedie, J. W. & Baker, E. N. (1996). *Biochemistry*, **35**, 14473–14479.
Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–646.
Hashizume, S., Kuroda, K. & Murakami, M. (1983). *Biochim. Biophys. Acta*, **763**, 377–382.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. (1993). *J. Appl. Cryst.* **26**, 283–291.
Mann, D. M., Romm, E. & Migliorini, M. (1994). *J. Biol. Chem.* **269**, 23661–23667.
Masson, P. L., Heremans, J. F. & Dive, C. (1966). *Clin. Chim. Acta*, **14**, 735–379.
Masson, P. L., Heremans, J. F. & Schonne, E. (1969). *J. Exp. Med.* **130**, 643–658.
Otwinowski, Z. (1990). *DENZO Processing Package*, Yale University, New Haven, Connecticut, USA.
Sanchez, L., Calvo, M. & Brock, J. H. (1992). *Arch. Dis. Child.* **67**, 657–661.
Smith, C. A., Anderson, B. F., Baker, H. M. & Baker, E. N. (1992). *Biochemistry*, **31**, 4527–4533.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Ward, P. P., Lo, J.-Y., Duke, M., May, G. S., Headon, D. R. & Conneely, O. M. (1992). *Biotechnology*, **10**, 784–789.
Ward, P. P., Piddington, C. S., Cunningham, G. A., Zhou, X., Wyatt, R. D. & Conneely, O. M. (1995). *Biotechnology*, **13**, 498–503.
Yu, L.-C. & Chen, Y.-H. (1993). *Biochem J.* **296**, 107–111.