Crystallographic and physicochemical studies on anion-binding and deglycosylation of human lactoferrin

A thesis presented in partial fulfilment of the requirements for the degree of Master of Philosophy in Chemistry at Massey University

Heather Mary Baker

1995

ABSTRACT

Lactoferrin, a monomeric 80 kDa glycoprotein, is a major component of human milk and is found in many other exocrine secretions as well as in the neutrophilic granules of white blood cells. A member of the transferrin family of iron-binding proteins, lactoferrin has the ability to bind tightly but reversibly 2 Fe^{3+} ions with the concomitant binding of $2CO_3^{2-}$ ions. Crystal structure studies clearly demonstrate that the polypeptide chain is folded into two similar lobes, representing the N- and the C-terminal halves of the protein, and that each lobe contains one of the two very similar iron-binding sites. Transferrins also show considerable versatility in their binding properties, being able to bind many metal ions in place of Fe^{3+} and anions in place of CO_3^{2-}. Differences between the two sites become more pronounced, however, with the substitution of non-native metals and anions, and the origins of this inequivalence have long been debated.

To investigate the means by which lactoferrin can accommodate anions larger than carbonate, the diferricdioxalatolactoferrin ($Fe_2(C_2O_4)_2$Lf) complex was prepared and crystallised. The crystals, which were isomorphous with those of $Fe_2(CO_3)_2$Lf, were used to collect a complete 2.4 Å data set at the Photon Factory (Japan) synchrotron source. The structure was refined by restrained least squares methods to a final R factor of 0.196 for all 31758 reflections in the resolution range 8.0 to 2.4 Å. The polypeptide folding and domain closure were identical to those of the native $Fe_2(CO_3)_2$Lf. In contrast to the carbonate complex, however, in which the two binding sites appear almost identical, with the carbonate coordinating in a symmetrical bidentate mode to each iron, when oxalate is the anion, the coordination around the metal differs between the N- and the C-lobe. In the C-lobe, the oxalate has a symmetrical 1,2-bidentate coordination to the iron, but in the N-lobe this coordination is quite asymmetric (O_{1ox}-Fe = 1.87 Å, O_{2ox}-Fe = 2.55 Å). Analysis of the structure indicates that the stereochemistry of the oxalate coordination to the iron is influenced by the position of the anion-binding arginine. The position this arginine can adopt in each lobe is, in turn, influenced by residues more remote from the iron site and which differ between the N- and C-lobes.

All lactoferrins so far characterised are glycoproteins, but the importance of the glycan chains for structure and/or function has yet to be established. Enzymatic methods were used to deglycosylate human and bovine lactoferrins, and the native deglycosylated forms of the human protein were compared with respect to CD spectra, iron binding and release, stability to proteolysis and heat stability.

Deglycosylation was carried out at pH 6.0 on the iron-free form of lactoferrin, using an endoglycosidase preparation from *Flavobacterium meningosepticum*, comprising PNGase F and Endo F. Deglycosylation was rapid for human lactoferrin, being essentially complete within 12-24 hr. Only partial deglycosylation of bovine lactoferrin could be achieved under the same conditions, however, and this is attributed to the relative inaccessibility of at least one of the glycosylation sites.

The CD spectra of native and deglycosylated human lactoferrins were found to be essentially identical in the range 250-350 nm, implying the same three dimensional structures. Both also bind iron in identical fashion; 2 Fe^{3+} ions are bound and binding is complete within 1 minute. The release of iron as the pH was lowered from 8.0 to 2.0 also showed no significant difference, the pH at which 50% release had taken place being 3.2 and 3.0 respectively for native and deglycosylated proteins. Susceptibility to proteolytic digestion by bovine trypsin over a period of 24 hr showed similar fragmentation patterns and a similar time course for the reaction for both species. Iron binding ability as a function of temperature was used as a measure of heat stability; melting temperatures derived from these experiments were 64°C for native and 63°C for deglycosylated lactoferrin. Comparison of the three dimensional structures of glycosylated iron-lactoferrin with deglycosylated apo-lactoferrin are consistent with these results, showing only a small increase in flexibility near the glycosylation site, when the carbohydrate is removed.

Conclusions are that the *in vitro* physicochemical properties of lactoferrin are unaffected by the presence or absence of its glycan chains. *In vivo* studies may be necessary to establish the importance, if any, of glycosylation.

Acknowledgements

This thesis was made possible by the contributions and help of a number of people. My thanks are due in three quarters.

To my supervisors, Professors Andrew Brodie and Sylvia Rumball for their advice and encouragement throughout the course of this work.

To all members of the protein crystallography group (past and present), who contributed to the friendly and helpful working environment which made the course of this work so enjoyable.

To my family for their continuous support and encouragement from the beginning to the end of this project.

I am very grateful to Andrew and Sylvia who always made themselves available for discussion in spite of their busy schedules. Their expertise, thoughtful help, and time are very much appreciated.

While the members of the Chemistry and Biochemistry Department are gratefully acknowledged, special thanks go to all the protein crystallography group for their friendship, support, kindness and practical help. In particular I thank Dr Bryan Anderson for his help with the computing, Dr Gillian Norris for help in many aspects of the protein work and deglycosylation, Drs Clyde Smith and Musa Shongwe who shared a laboratory, and interest in the oxalate work, Richard Kingston who also shared a laboratory, many helpful discussions, books, and his computing skills and Jean King for help in preparation of this thesis.

I am especially grateful to my family, both parents and children, but above all I would like to express a particular debt of gratitude to Ted for first engendering a love of research, for all his encouragement and special support. Thanks Ted.

Table of Contents

List of Figures

List of Tables

Abbreviations

apoLf	metal-free human lactoferrin
apoTf	metal-free human serum transferrin
β-me	β-mercaptoethanol
bLf	bovine lactoferrin
CD	circular dichroism
cOTf	chicken ovotransferrin
$Cu_2(C_2O_4)_2Lf$	dicupricdioxalatolactoferrin
$Cu_2(CO_3)(C_2O_4)Lf$	dicupric-(carbonato-oxalato)lactoferrin
Cu_2Lf	dicupriclactoferrin
EDTA	ethylenediaminetetraacetic acid
Endo F	endo β-N-acetylglucosamidase F_1
EPR	electron paramagnetic resonance
ESEEM	electron spin echo envelope modulation
EXAFS	extended-X-ray absorption-fine-structure
$Fe_2(C_2O_4)_2Lf$	diferricdioxalatolactoferrin
$Fe_2(CO_3)(C_2O_4)_2Lf$	diferric-(carbonato-oxalato)lactoferrin
Fe_2Lf	diferric human lactoferrin
Fe_2Tf	diferric transferrin
$FeLf_N$	recombinant human N lobe lactoferrin, with iron
hTf	human transferrin
Lf	lactoferrin
mLf	mouse lactoferrin
MPD	2-methyl-2,4 pentanediol
msTf	Tobacco Hornworm transferrin
MTf	melanotransferrin
nmr	nuclear magnetic resonance
NTA	nitrilotriacetic acid
OTf	ovotransferrin
ovoTf	ovotransferrin
PAGE	polyacrylamide gel electrophoresis
PNGase	peptide-N^4-(N-acetyl-β-D-glucosaminyl)asparagine amidase F
rTf	rabbit transferrin
SDS	sodium dodecylsulphate
Tf	serum transferrin
UV/vis	ultraviolet and visible electronic absorption spectroscopy
XTf	Xenopus transferrin

CHAPTER I

INTRODUCTION

Iron is widely used in living systems to provide an essential metal ion in many biological processes, usually as a functional component of metalloproteins. The potential dilemma, arising from the very low solubility of iron in physiological conditions, has been solved by a system of proteins exquisitely designed to solubilise, store, sequester and transport iron. Fe(III), the thermodynamically stable form of iron under physiological conditions, is precipitated as insoluble ferric hydroxides at concentrations greater than 10^{-18} M (Aisen & Listowsky, 1980), but in vertebrates it is safely compartmentalised into catalytic and carrier proteins such as haemoglobins and cytochromes, storage proteins such as ferritin, and proteins of transport and regulation, namely the transferrins.

I.1 TRANSFERRIN FAMILY: Historical overview§

Proteins of the transferrin family, which include serum transferrin, ovotransferrin (or conalbumin) present in egg white, and lactoferrin found in milk and exocrine secretions, have been studied since the beginning of this century.

Osborne and Campbell (1900) isolated conalbumin from egg whites, but it was not until 1946 (Atherton, 1946) that its iron-binding properties were recognised. Meanwhile, with the stimulus of war research into blood plasma fractionation, serum transferrin was isolated and characterised (Schade & Caroline, 1946).

The most striking feature of transferrin chemistry, the unique synergistic cation-anion relationship, was first outlined in a landmark paper by Schade *et al* (1949). Two Fe^{3+} ions were found to be bound to the molecule but it was recognised that specific binding only occurred if two HCO_3^- (or CO_3^{2-}) ions were also bound. By 1962, when it was realised that the only difference between hen serotransferrin and conalbumin was in the attached carbohydrate moiety (Williams, 1962), conalbumin was renamed ovotransferrin.

Lactoferrin, first noted as the "milk red protein" from bovine milk (Sorenson & Sorenson, 1939), was purified and characterised as an iron-binding protein by Groves (1960) and Johanson (1960). Montreuil gave the name lactotransferrin, but now it is generally known as lactoferrin, indicating its initial identification in milk and the fact that although it is similar to transferrin, it is a distinct protein encoded by a separate gene.

Thus by 1960, the three major members of the transferrin family, ovotransferrin, serum transferrin and lactoferrin had been purified, characterised and recognised as belonging to a larger family known collectively as transferrins. Members of this family

§ Using Bezkorovainy (1980), Brock (1985), Aisen & Harris (1989) as secondary references unless otherwise stated.

appear to be present in all vertebrates so far examined.

Transferrins have also recently been characterised in several invertebrates, eg the hornworm moth *Manduca sexta* (Bartfeld & Law, 1990) and the cockroach (Jamroz *et al*, 1993). Finally a further distinct member of the transferrin family was identified with the discovery that a membrane-bound antigen, p97, from human melanoma cells, had a substantial amino acid sequence identity with the other transferrins (Rose *et al*, 1986). This protein has now been renamed melanotransferrin and has been found to be present in at least one other species (chicken).

I.2.1 **Molecular Properties**

All the transferrins are monomeric glycoproteins of 680-700 amino acids with molecular weights of approximately 80,000 Da. The exception to this is melanotransferrin which, while homologous to the other transferrins, has a 25 amino acid C-terminal extension believed to be involved in anchoring the protein to the plasma membrane of melanocytes (Rose *et al*, 1986). The isoelectric points of ovotransferrins, transferrins and melanotransferrin range between 5.0-6.0, while the lactoferrins are distinguished by much higher isoelectric points around 8.5-9.0 (Brock, 1985).

Transferrins exhibit an internal amino acid sequence homology of about 40% between the N- and C-terminal halves, which is reflected in their bilobal tertiary structure (Gorinsky *et al*, 1979; Anderson *et al*, 1987). The most outstanding feature of transferrin chemistry, however, is the synergistic anion-cation binding relationship mentioned earlier. Two ferric ions are bound together with two (bi)carbonate ions. Neither ion is bound strongly without the other, but taken together binding is extremely strong with a stability constant ~10^{20} molar (Aasa *et al*, 1963). Despite the high stability constant, this binding is readily reversible, both in *vivo* and *vitro* and indeed many of the biological functions of the transferrins depend on this reversibility.

I.2.2 **Biological Roles of Lactoferrin**

The common iron binding ability of the transferrins means that they all share certain general roles. These include a role in regulating iron levels in the body, and the ability to depress bacterial growth by iron deprivation (see below). However, the differences in both the bodily location and the individual molecular properties of lactoferrin give rise to certain distinct biological roles.

(a) *Location and expression of lactoferrin*

Lactoferrin is found in a wide variety of mammalian milks, and other exocrine secretions, such as tears, saliva and mucus. It is also in the secondary granules of neutrophils (Table I.1).

Table I.1: The occurrence of lactoferrin in various human bodily secretions[a]

Secretion	Concentration (mg/L)
milk	1550 - 5000[b]
serum	0.8 - 3.5
saliva	8
tears	1400
bile	1600
seminal fluid	550
urine	0.3

[a]Derived from Bezkorovainy (1980).
[b]Human colostrum contains a very high level of lactoferrin which typically falls to 1500-2000 mg/L about 10 days postpartum.
(From Smith (1992) with permission.)

In most locations lactoferrin is largely devoid of iron. In fact most of the iron in human milk is not, as originally assumed, bound to lactoferrin, but is present in the lipid and low molecular weight fraction (Fransson *et al*, 1980).

Both the level and timing of lactoferrin expression in the mammary gland vary greatly between species. In some animals such as the rabbit, dog and rat, lactoferrin is virtually absent from milk and is replaced by a variant of serum transferrin with modified carbohydrate (Baker *et al*, 1968).

In humans, on the other hand, colostrum has up to 5 mg/ml of lactoferrin, although later in lactation the level falls to 1-2 mg/ml (Lonnerdal *et al*, 1976). Similarly there is a variation in the amount of lactoferrin expressed in normal and cancerous tissues (Charpin *et al*, 1985; Rado *et al*, 1987). These variations in the level of expression, taken together with the finding of Pentecost and Teng (1987) that the major oestrogen-stimulated protein synthesised in the uterine tissue of the mouse is lactoferrin, suggest that the lactoferrin genes are hormonally regulated.

While sequence homology is high between serum transferrin and lactoferrin (~60%) and the three-dimensional structures are also very similar (Baker & Lindley, 1992), there are important differences in chemical properties between the two proteins. Both have an extremely high affinity for Fe^{3+} ions ($K_d = 10^{-20}$) (Brock, 1985; Aisen & Listowsky, 1980) but the binding constant of lactoferrin is 50-90 times greater than that of transferrin (Harris, 1986). The difference between the two iron binding sites is more marked in transferrin than lactoferrin (Aisen *et al*, 1978; Ainscough *et al*, 1980; Mazurier *et al*, 1983)

with perhaps one of the most important differences being the pH dependence of the release of iron. Transferrin begins to release iron when the pH is lowered below 6.0, whereas lactoferrin retains iron down to pH 3.5 (Mazurier & Spik, 1980). Lactoferrin does not bind to transferrin receptors (Cox *et al*, 1979; van Bockxmeer & Morgan, 1982) and thus protein-receptor interactions may differ and factors other than pH could stimulate iron release.

(b) *Bacteriostatic and bactericidal properties of lactoferrin*

The most extensively studied and widely accepted role of lactoferrin is as a bacteriostatic/bactericidal agent. All microorganisms, with the possible exception of lactobacilli, require iron for growth. Lactoferrin, by virtue of its high affinity for iron, has the potential for depriving bacteria of this essential element and thereby retarding their growth.

One of the first studies to show the inhibitory effect of milk and lactoferrin on bacterial growth was by Bullen *et al* (1972). They showed that *in vitro* the growth of *E. coli* 0111 (associated with enteritis in infants) was depressed by milk or lactoferrin, and that this effect could be abolished by saturating the protein with iron. (It is of interest to note that milk is more effective than lactoferrin on its own, suggesting a synergistic action with other proteins.) A large number of independent investigators have confirmed that lactoferrin has a bacteriostatic effect on a diverse spectrum of gram negative bacteria (Bishop *et al*, 1976; Reiter, 1983), with the common hypothesis being that lactoferrin produces an iron deficient environment that limits bacterial growth.

Studies showed that apolactoferrin had a direct bactericidal effect on *S. mutans* and and *V. cholerae*, an effect that could be reversed by the addition of iron (Arnold *et al*, 1977). The range of organisms that could be killed was later extended (Arnold *et al*, 1980), but these results have been questioned by Finkelstein *et al* (1983) who could only show a bacteriostatic effect.

By the 1980s the bacteriostatic properties of lactoferrin were well established *in vitro*. However to show that this property could exist *in vivo*, it was necessary to establish that lactoferrin could remain biologically active in the intestine. Work by Mason (1962), and Britten and Koldovsky (1980) suggested it was reasonable to expect at least a fraction of lactoferrin to survive digestion, and Prentice *et al* (1987) showed that the amount of lactoferrin in the stool of breast fed babies was greater than in formula fed infants, thus indicating that lactoferrin did indeed survive intact through the intestinal tract. However, *in vivo*, the faecal flora of infants fed formula was not altered by the addition of bovine lactoferrin to the formula (Wharton *et al*, 1992), suggesting that bovine lactoferrin on its own does not alter bacterial growth in the infant.

While iron binding by lactoferrin is generally agreed to be the reason why it is bacteriostatic, there remains controversy over the reported bactericidal effect of lactoferrin.

The research by Morinaga Milk Industry Company is of particular interest. A microbicidal peptide has been isolated from both bovine and human lactoferrin, which is 18 times more potent than the intact lactoferrin (Tomita *et al*, 1992). The active peptide is a loop of 18 amino acids (residues 20-37 of human lactoferrin) (Bellamy *et al*, 1992), the bactericidal effect of which is independent of iron concentration but which requires contact with bacteria. It is proposed that because this peptide has a high number of basic residues, like many other antimicrobial peptides, it is capable of increasing cell permeability, leading to disruption and eventual destruction of the cell. Further evidence suggests that active peptides produced by enzymatic digestion of lactoferrin may contribute to host defence against microbial disease (Tomita *et al*, 1992). This area of research is of much interest because of the many potential commercial applications.

(c) *Lactoferrin and Iron Absorption*

A possible role of lactoferrin in milk is to bind iron so tightly that it is unavailable for bacterial growth, but still available to the infant. Exclusively breast-fed babies maintain adequate levels of iron longer than infants fed on bovine milk formula (Saarinen, 1977), and furthermore, lactoferrin has been shown to enhance uptake of iron in suckling pigs and weanling mice (Fransson *et al*, 1983a, 1983b).

The mechanisms of uptake of lactoferrin or lactoferrin-bound iron are poorly characterised. Lactoferrin is able to donate iron to intestinal tissues but not to reticulocytes (Mazurier *et al*, 1985). A 100 kDa protein, identified as a lactoferrin receptor has been found on rabbit brush-border cells (Mazurier *et al*, 1989), rhesus monkey brush border cells (Davidson & Lonnerdal, 1988, 1989) and brush border sections of the mouse small intestine (Hu *et al*, 1988).

If lactoferrin does act as a carrier of iron then not only must it bind to cells via a specific receptor but there must also be a specific mechanism for the release of iron from the protein. Unlike transferrin, which will lose iron below pH 6.0, lactoferrin binds iron until pH 3.5 (Mazurier & Spik, 1980), so will not lose iron spontaneously at the intracellular pH of 5-6. A possible model could resemble the receptor-mediated release of iron from transferrin, which involves not only pH, but also the action of the receptor itself (Bali & Aisen, 1991; Bali & Aisen, 1992).

(d) *Lactoferrin and inflammation*

Iron has been shown to stimulate the formation of potentially damaging free radicals. Gutteridge *et al* (1979), and Monteira and Winterbourn (1988) have shown that both transferrin and lactoferrin could play a major role in preventing this occurring by sequestering the iron released by degrative processes, but as lactoferrin retains iron at a lower pH, it could be more effective than transferrin. Further work (Molloy & Winterbourn, 1990) suggested neutrophil lactoferrin may bind iron released by

microorganisms killed at the site of inflammation, thereby not only preventing catalysis of the formation of free radicals, but also suppressing the growth of surviving bacteria.

(e) *Lactoferrin as a regulator of myelopoeisis*

Neutrophil components are involved in *in vitro* regulation of granulocyte-macrophage progenitor cells. Lactoferrin has been identified as being responsible for this action (Broxmeyer *et al*, 1978) and consequently has been implicated as a regulator of white blood cell maturation (myelopoeisis) *in vivo*. Broxmeyer *et al* (1978) reported that apolactoferrin was inhibitory at concentrations down to 10^{-7} M, but iron-saturated lactoferrin had the same effect at concentrations as low as 10^{-17} M. Lactoferrin was inhibitory because it actually bound to the monocytes and macrophages involved; the difference between apo and iron saturated lactoferrin was attributed to the different three-dimensional structures of the two proteins (Broxmeyer *et al*, 1978, 1980).

The involvement of lactoferrin in this process remains to be conclusively demonstrated however. Winton *et al* (1981) have been unable to repeat the above experiments. Furthermore copper-lactoferrin has been shown not to have the same effect as iron-lactoferrin and yet the two proteins are isostructural (Smith *et al*, 1992), casting doubt on the attribution of the difference in the activity of the apo- and iron-lactoferrin to the difference in three-dimensional structure. In spite of this, the general consensus of opinion in the literature supports the role of lactoferrin as a regulator of myelopoeisis.

I.2.3 **Physicochemical Properties**

A wealth of spectroscopic studies, extensively reviewed by Aisen and Harris (1989), has been brought to bear on the characterisation of the metal binding sites. Early work by Warner and Weber (1953) indicated that two or three tyrosines acted as ligands for the metal ion; the colour change seen on iron binding was ascribed to iron-phenolate bonding (phenolate (π)$\rightarrow$metal (dπ^*) transition). The position of the absorption maximum, generally between 400-500 nm, depended on both the particular metal ion and the nature of the synergistic anion.

NMR (Alsaadi *et al*, 1981) and EPR spectroscopy (Zweier & Aisen, 1977; Zweier, 1980; Ainscough *et al*, 1983) both indicated that a histidine also acted as a ligand for the metal. EPR spectra have proved particularly useful not only because of the variety of paramagnetic metal ions that can be used, but also because of their sensitivity to the ligand environment of the metal.

Aspartate, being more difficult to characterise spectroscopically as a ligand, was revealed to be a metal ligand only when the crystal structure of diferric human lactoferrin was determined (Anderson *et al*, 1987).

The nature of the anion binding site proved more difficult to establish. One landmark paper was that of Schlabach and Bates (1975) which led to the proposal that the

anion acted as a bridge between the metal and the protein. The direct coordination of the

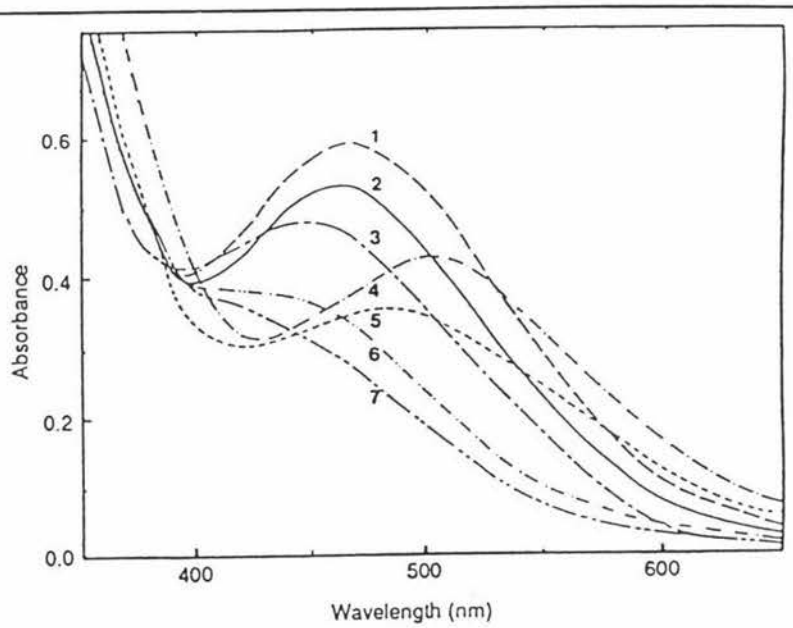


Fig. I.1 The visible absorption spectra of diferric transferrin complexes with various synergistic anions: 1,nitrilotriacetate; 2, carbonate; 3, salicylate; 4, thioglycolate; 5, glycine; 6, glyoxylate, and 7, glycolate. From Schlabach and Bates, 1975.

metal to the anion, as suggested by the anion dependence of the visible spectra of transferrins (Fig.I.1), has been supported by many studies. These include EPR studies of VO^{2+}-transferrins with a variety of anions (Campbell & Chasteen, 1977), electron-spin-echo studies of Fe^{3+}-, Cu^{2+}- and VO^{2+}-transferrins with oxalate and carbonate as the synergistic anion (Zweier *et al*, 1979, 1982; Eaton *et al*, 1989, 1990) and EXAFS studies of anion binding to Fe^{3+}-transferrin (Schneider *et al*, 1984). In addition it was established that arginine was probably associated with the binding of the anion (Rogers *et al*, 1978; Zweier *et al*, 1981).

PROTEIN STRUCTURE

I.3.1 Primary Structure

The complete amino acid sequences of human serum transferrin, (MacGillivray *et al*, 1983), chicken ovotransferrin (Williams *et al*, 1982; Jeltsch & Chambon, 1982), and human lactoferrin (Metz-Boutigue *et al*, 1984; Mazurier *et al*, 1984) established the homologous relationship between these three proteins, and the existence of the family now known collectively as the transferrins. cDNA sequencing has expanded the number of proteins known to belong to this family to include porcine (Baldwin & Weinstock, 1988), hornworm (Bartfield & Law, 1990), *Xenopus* (Moskaitis *et al*, 1991) and cockroach

(Jamroz *et al*, 1993) transferrins, murine (Pentecost & Teng, 1987), bovine (Mead & Tweedie, 1991), and porcine (Lydon *et al*, 1990) lactoferrins, and melanotransferrin (Rose *et al*, 1986). All show a remarkable degree of similarity, with approximately 70% identity between the lactoferrins and 50-60% between lactoferrins and transferrins (Baker, 1994). Melanotransferrin and cockroach transferrin are slightly less similar (~40% identity with human lactoferrin and transferrin).

Within each transferrin there is a significant sequence identity (~40%) between the N-terminal and C-terminal halves of the molecule, a feature strongly suggestive of ancestral gene duplication. Knowledge that stable half molecules can be made, each binding one iron, and the existence of one primitive 40,000 Da transferrin, from an ascidian, *Pyura stolonifera* (Martin *et al*, 1984), is in further agreement with the concept of two-sited transferrins having developed by gene duplication from a 40,000 Da, single-iron, precursor molecule.

All the transferrins are glycoproteins, with the carbohydrate attachment being N-linked through the sidechains of asparagine residues. The number of carbohydrate moieties varies from protein to protein, from one in ovotransferrin (Metz-Boutigue *et al*, 1984) to four in bovine lactoferrin (Pierce *et al*, 1991), and distribution over the molecular surface of the protein also shows a wide variation. The primary structure of the glycan chains differs between different transferrins and between different species of the same transferrins. Furthermore there is heterogeneity of carbohydrate chains even in the same species. Thus for human lactoferrin, at least 5 different glycan chains have been identified, but only 2 glycosylation sites. Although deglycosylation experiments have suggested that carbohydrate helps stabilise the iron binding site (Legrand *et al*, 1990), these results are in doubt, given the severe methods used to deglycosylate the lactoferrin, and the wide variations of the glycans mentioned above.

I.3.2 **Fragment Studies**

Further evidence of the bilobal structure of transferrins, suggested in the amino acid sequence, came from the isolation of half molecule (and smaller) iron-binding fragments resulting from limited proteolysis. Thus hen ovotransferrin can be cleaved into N- and C-terminal half molecules, each being stable and able to bind iron reversibly (Williams, 1974; 1975)). These also associate strongly (Ikeda *et al*, 1985; Williams & Moreton, 1988) to give a non-covalent complex capable of delivering iron to cells via the transferrin receptor. Half molecule fragments of bovine transferrin have also been prepared, with similar but not identical spectroscopic properties (Brock *et al*, 1976). This highlights one of the problems in examining two-sited proteins such as the transferrins, where the sites are very similar; namely the difficulty of separating the individual sites and establishing the interactions which may exist between them. Lactoferrin may also be cleaved by trypsin but this results in an N-terminal fragment of 30 kDa and a C-terminal fragment of 50 kDa

(Legrand *et al*, 1990), and hence the properties of these fragments are not necessarily those of the true half molecules.

Among the most interesting of the many transferrin fragments produced by proteolysis, are the quarter-molecule fragments of lactoferrin (Legrand *et al*, 1984) and ovotransferrin (Jhoti *et al*, 1987) which represent a domain of each protein (see Section I.3.3a), and which bind iron, but with different spectroscopic properties; the iron-saturated fragment showed a shift in the visible absorption maximum from 465 nm to 425 nm.

Recombinant DNA studies have opened the way to the production of half molecules of transferrins, and mutagenesis experiments should help expand our knowledge of the properties of each individual site and the factors which control metal- and anion-binding selectivity.

Recombinant N-terminal fragments of transferrin (Funk *et al*, 1990) and human lactoferrin (Day *et al*, 1992), each of which is terminated just prior to the connecting peptide, have been produced, and the structures determined (Day *et al*, 1993). Mutagenesis studies are also in progress. Difficulties in the expression of the recombinant C-terminal half-molecule have so far prevented any direct comparison of the two half molecules.

I.3.3 Crystallographic Studies

Crystallographic studies began over 20 years ago on a number of transferrins, including rabbit serum transferrin (Al-Hilal *et al*, 1976), human lactoferrin (Baker & Rumball, 1977), human transferrin (De Lucas *et al*, 1978) and hen ovotransferrin (Abola *et al*, 1982). A low resolution (6 Å) map of rabbit serum transferrin (Gorinsky *et al*, 1979) clearly showed the bilobal nature of the molecule, but it was not until 1987 that a high resolution three-dimensional structure of a transferrin, human diferric lactoferrin at 3.2 Å resolution, was solved (Anderson *et al*, 1987). This structure was then used to help solve the structure of diferric rabbit serum transferrin (Bailey *et al*, 1988). Since then the structures of a number of half molecules, the N-terminal halves of rabbit serum transferrin (Sarra *et al*, 1990), chicken ovotransferrin (Dewan *et al*, 1993) and human lactoferrin (Day *et al*, 1993) as well as two forms of human apolactoferrin (Anderson *et al*, 1990; R. Faber personal communication) have been reported.

The structure of human diferric lactoferrin, initially solved at 3.2 Å resolution by multiple isomorphous replacement (Anderson *et al*, 1987) and now extended to 2.2 Å resolution with a crystallographic R-value of 0.179 (Haridas *et al*, in press), remains the most accurately determined three-dimensional structure and can be taken as representative of the whole transferrin family.

(a) *Diferric human lactoferrin*

A striking feature of the lactoferrin molecule, seen even at 6 Å resolution, is its bilobal nature. The polypeptide is folded into two globular lobes. The N-terminal half of the polypeptide forms one lobe, the N-lobe (residues 1-333), and the C-terminal half, the C-lobe (residues 345-691), with about 40% sequence identity between the two lobes. A short 3-turn α helix (residues 334-344) links the two lobes.

Each lobe is further subdivided into two dissimilar domains; N1 and N2 in the N-lobe and C1 and C2 in the C-lobe. The domains are each based on an α/β framework, having either a five-stranded (N2, C2) or six-stranded (N1, C1) mixed β sheet with helices packed on either side. The combined metal and anion binding sites, one in each lobe, lie at the inner end of the cleft between the two domains of the lobe. Although deeply buried, ~10 Å from the surface, these are accessible to the solvent via water molecules bound in the cleft (Anderson *et al*, 1989).

Fig. I.2 Ribbon diagram of diferric lactoferrin, showing the organisation of the molecule, with the N-lobe on the left, and the C-lobe on the right. The positions of the two Fe^{3+} ions are indicated by stippled circles. The interdomain 'backbone' strands in each lobe can be seen behind the Fe^{3+} ions. The diagram was prepared using the program Ribbons (Richardson, 1985; Priestle, 1988).

The main interactions between the two lobes, apart from the linking helix, are a set of hydrophobic interactions between N1 and C1 domains, which could allow for movement of one lobe relative to the other, and could also explain the difficulty of separating the half molecules produced by proteolysis.

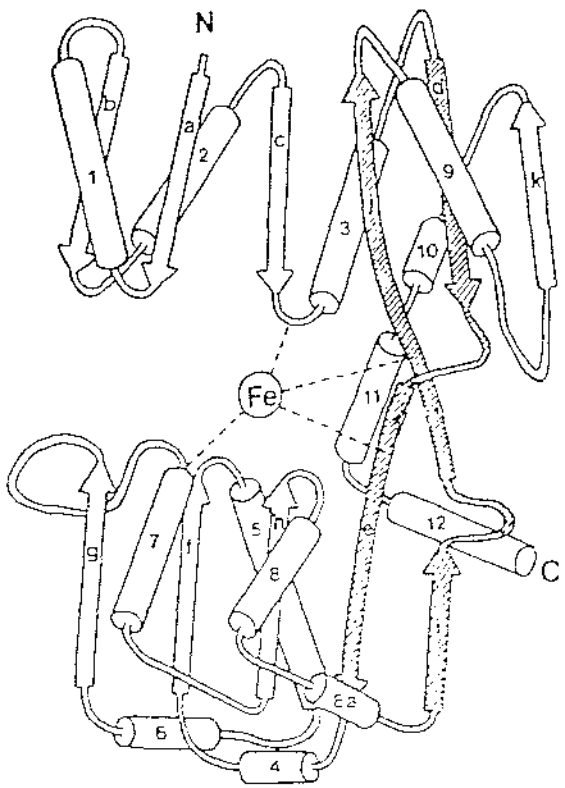


Fig. I.3 A schematic diagram of the folding pattern found in each lobe of human lactoferrin. Helices are shown as cylinders, while the β-strands are represented by arrows (from Baker *et al*, 1987).

The two domains in each lobe are not made up of a continuous piece of chain. The first ~90 residues form the first half of domain 1 (1-90; 345-433), then the chain crosses behind the iron site, via an extended β strand, to form the entire second domain (91-251; 434-595). It then crosses back, again as an extended β-strand to give a second "back-bone" strand behind the iron site, and the next 70 residues fold to complete domain 1 (252-320; 596-675). An α helix crosses back towards domain 2, and this together with the backbone strands forms the back of the iron site, and the principal links between the 2 domains. The C-terminus of the protein, a helix (676-691), corresponds to the linking helix between the two lobes.

The two lobes are almost identical in folding, as expected from a 40% sequence identity, so that when superimposed, 85% of C_α atoms match with an rms deviation of

~1.0 Å. Differences arise from the N-lobe being slightly more open than the C-lobe, and variations in amino acid sequence which alter some of the external loops change hydrogen-bonding patterns in some parts of the molecule, and give rise to three extra disulphide bridges in the C-lobe. The main difference between the two lobes arises from their external environment within the molecule. The molecule does not have two-fold symmetry; instead the two lobes are related by a screw axis, a rotation of ~180° and a translation of 25 Å. This means that equivalent parts of the two lobes are not in equivalent environments; for instance, the 'back' of the N1 domain butts on to the 'front' of the C1 domain (Fig. I.2)

(b) *Metal- and anion-binding sites*

The two binding sites, each having one Fe^{3+} and one CO_3^{2-} (not HCO_3^-) ion, are remarkably similar. In each case the carbonate ion is coordinated to the ferric ion in a bidentate manner, and is also hydrogen bonded to the protein via Arg 121 (465), Thr 117 (461) and the N-terminus of an α-helix, helix 5. The metal coordination is completed by four protein sidechains, through the imidazole nitrogen of a histidine (His 253; His 597 in the C-lobe), a carboxylate oxygen of an aspartate (Asp 60; Asp 395 C-lobe) and the phenolate oxygens from two tyrosines (Tyr 92 and Tyr 192; Tyr 435 and Tyr 528 in the C-lobe), to give a six-coordinate, approximately octahedral geometry around the metal.

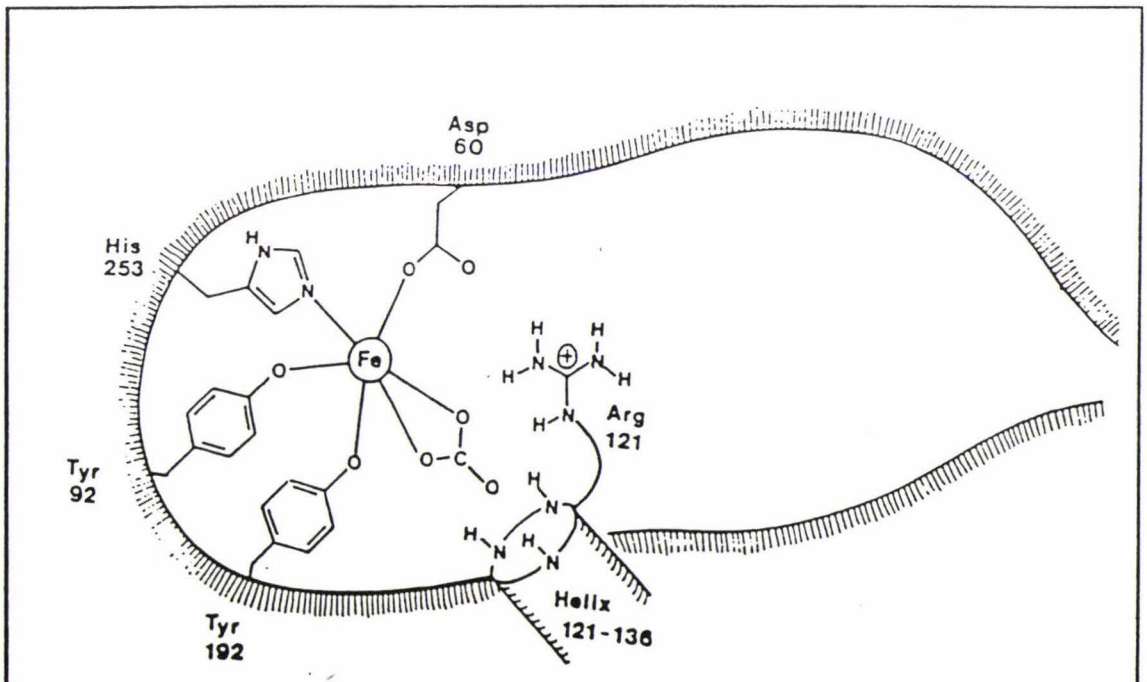


Fig. I.4 Schematic representation of the iron and anion binding sites of lactoferrin. (Taken from Anderson *et al*, 1989.)

These protein ligands are widely spaced along the polypeptide chain. One comes from each domain, and one from each of the backbone strands, an arrangement which should have advantages when large scale conformational changes take place on metal and anion binding and release. This arrangement is in marked contrast to many metalloproteins where the metal is an integral part of the protein. For example in the blue copper protein azurin, the ligands are contributed by a small loop of the polypeptide chain. The ligands used in lactoferrin, including those for the anion, are conserved in all transferrins except melanotransferrin and the serum transferrin from *Manduca sexta* (Table I.2), both of which have been observed to bind only one iron atom (Baker *et al*, 1992; Bartfeld & Law, 1990).

Table I.2: Conservation of metal- and anion-binding residues in transferrins

	Lf	hTf	cOTf	rTf	bLf	mLf	MTf	msTf
	Asp60	Asp	Asp	Asp	Asp	Asp	Asp	Asp
	Tyr92	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Phe
	Tyr192	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr
N-lobe	His253	His	His	His	His	His	His	Gln
	Arg121	Arg	Arg	Arg	Arg	Arg	Arg	Arg
	Thr117	Thr	Thr	Thr	Thr	Thr	Thr	Thr
	Asp395	Asp	Asp	Asp	Asp	Asp	Ser	Asp
	Tyr435	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Asn
	Tyr528	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Asp
C-lobe	His597	His	His	His	His	His	His	Arg
	Arg465	Arg	Arg	Arg	Arg	Arg	Ser	Thr
	Thr461	Thr	Thr	Thr	Thr	Thr	Ala	Ser

Lf, human lactoferrin; hTf, human transferrin; cOTf, chicken ovotransferrin; rTf, rabbit transferrin; bLf, bovine lactoferrin; mLf, mouse lactoferrin; MTf, melanotransferrin; msTf, *Manduca sexta* transferrin.
(From Baker, 1993 with permission.)

The anion, as well as providing two iron ligands and forming a bridge between the metal and the protein, occupies a positively charged pocket made by Arg 121 (465 in C-lobe) and the N-terminus of helix 5 (residues 122-125, N-lobe; 466-469, C-lobe), both of which have positive charges associated with them. The N-terminus of an α-helix has a +0.5 charge by virtue of the helix dipole, and is frequently the site of anion binding in

other proteins. Unlike the protein ligands to the metal, the ligands to the anion come from a short stretch of polypeptide chain located in just one domain (N2 for the N-lobe site and C2 for the C-lobe site) and thus the anion site should not be changed by domain movements.

The importance of anion binding by transferrins is underscored by a number of points. The first is that ferric ions cannot be bound tightly without the concomitant binding of a suitable anion. There is thus a synergistic relationship between cation and anion, in that each is required for the tight binding of the other. There is both kinetic (Cowart *et al*, 1982; Kojima & Bates, 1981) and crystallographic (Baker *et al*, 1987; Anderson *et al*, 1989) evidence to suggest that the anion binds first and prepares the site for metal binding by neutralizing the positive charge at the binding site. This would create, with the carboxylate and phenolate oxygens a highly anionic site, with high affinity for Fe^{3+}. The crucial role of the anion is further highlighted by both the invariance of the residues involved in anion binding and the almost perfect hydrogen bonding geometry for carbonate at the anion binding site. All residues involved in anion binding (117, 121-125 in the N-lobe, 461, 465-469 in the C-lobe) are conserved in all transferrins so far sequenced, except for those with reduced iron-binding ability; *Manduca sexta* transferrin and melanotransferrin. The hydrogen bonds from these conserved residues; contributed by two peptide NH groups (from residues 123 and 124 in the N-lobe, 467 and 468 in the C-lobe) and the sidechains of Arg 121 (465) and Thr 117 (461), are short, linear and exquisitely arranged to stabilise the carbonate anion binding to the protein (Fig. I.5).

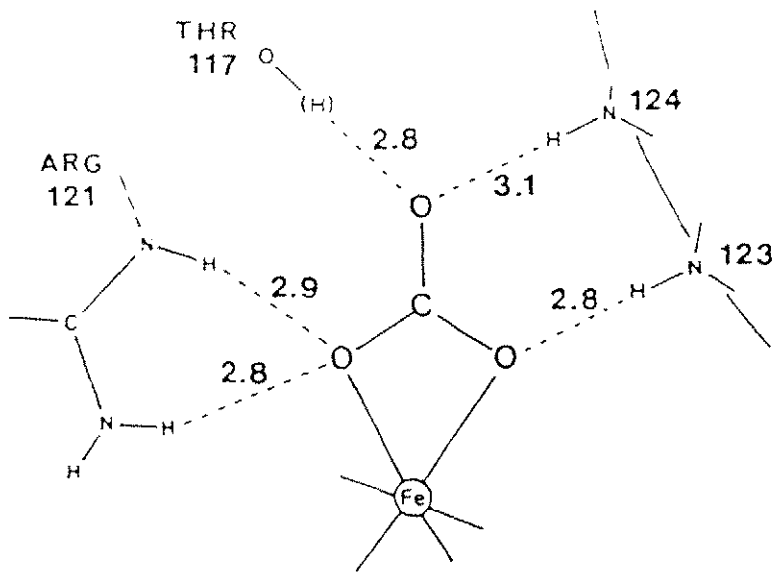


Fig. I.5 The hydrogen bonding environment of the carbonate anion in human lactoferrin (shown for the N-lobe site). Lengths of hydrogen bonds shown in Ångstroms.

An unexpected observation that further focuses on the importance of anion binding is the protein structural similarity between transferrins and certain bacterial periplasmic binding proteins (Anderson *et al*, 1987; Baker *et al*, 1987). These include those specific for L-arabinose (Gilliland & Quiocho, 1981), D-galactose (Vyas *et al*, 1983), sulphate (Pflugrath & Quiocho, 1985) and phosphate (Luecke & Quiocho, 1990). Each of these proteins is similar not only in size to a single lobe of lactoferrin, ~300 amino acids, but also in the folding of the polypeptide chain. The resemblance is most striking between lactoferrin and the sulphate binding protein. Both have exactly the same topology and the anion binding site is primarily located at the equivalent helix N-terminus in each case (Figs. I.6 and I.7). In the sulphate binding protein, however, in contrast to lactoferrin, other helices from both domains also contribute to the anion binding, and the anion site is slightly further from the backbone strands.

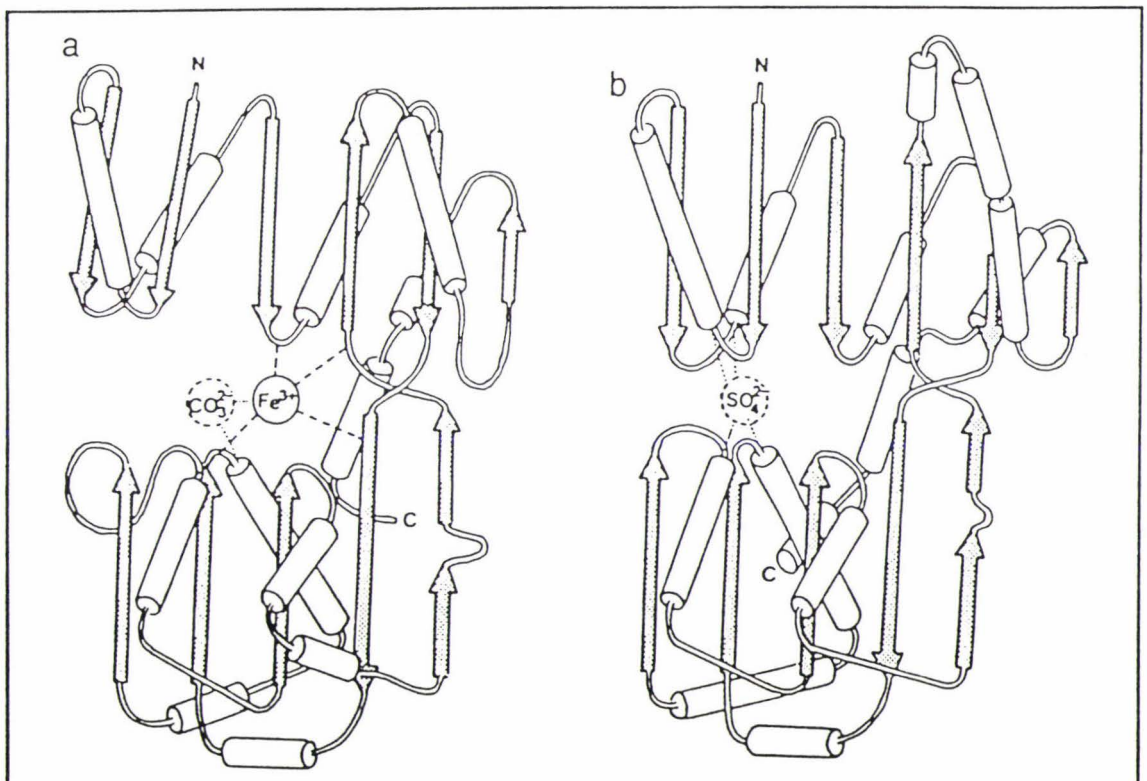


Fig. I.6 Comparison of the folding patterns of (a) the N-lobe of lactoferrin, and (b) the bacterial periplasmic sulphate-binding protein. (From Baker *et al*, 1987.)

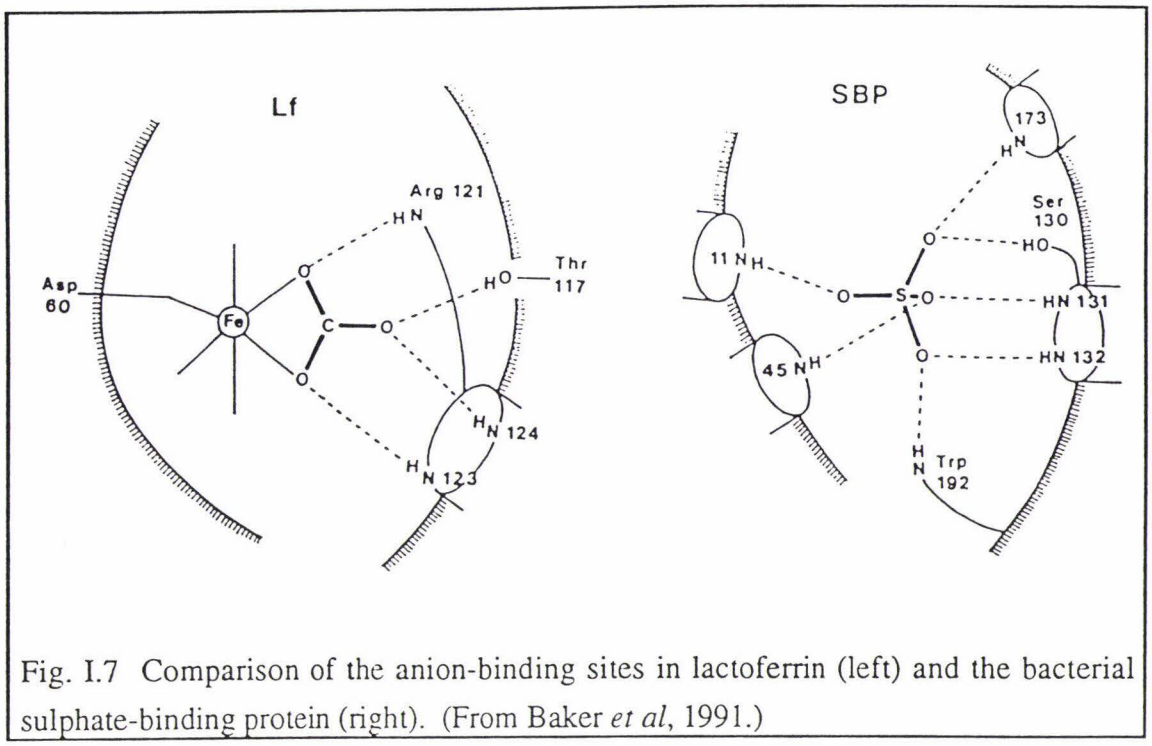


Fig. I.7 Comparison of the anion-binding sites in lactoferrin (left) and the bacterial sulphate-binding protein (right). (From Baker *et al*, 1991.)

The initial step in binding could be the binding of the anion to domain 2. In lactoferrin, this would be followed by the iron binding, and then the closure of domain 1 over the site, and in the periplasmic binding proteins, just the closure of domain 1.

Given the similarity between the periplasmic binding proteins and the transferrins, both in overall topology and in the anion site, it is conceivable that the two families may have diverged from a common anion-binding precursor molecule. Gene duplication would give the two-sited transferrin molecules. It seems likely therefore that the transferrins could equally well be thought of as anion binding proteins to which a metal binding functionality has been added.

(c) *Apo human lactoferrin*

While crystallographic studies of human diferric lactoferrin (Anderson *et al*, 1989) and rabbit transferrin (Bailey *et al*, 1988) characterised their binding sites and protein structure, physical studies (Rossenau-Montreff *et al*, 1971; Kilar & Simon, 1985) showed that a substantial conformational change accompanies iron binding and release. The nature and extent of this change were graphically illustrated by the X-ray analysis of deglycosylated human apolactoferrin at 2.8 Å (Anderson *et al*, 1990), now extended and further refined to 2.0 Å resolution, R = 0.176.

The most striking observation is the large-scale domain movement in which domain 2 of the N-lobe rotates 54° relative to domain 1. The domain moves as a rigid body, with all the deformations confined to the linking hinge region at the back of the iron site. These changes are basically restricted to the pivoting of helix 5 (in domain N2) on helix 11, and

the bending of the two β backbone strands. This causes the binding cleft to open wide and expose a number of basic residues previously buried in the interdomain cleft. It also separates the potential ligands so that the two tyrosine ligands, and the anion ligands remain with the N2 domain, whilst the histidine and aspartate ligands remain with N1. This structure suggests a model for the mechanism of iron binding, in which first the anion and then the metal binds to domain 2 (N2 and C2) with the cleft open, after which the domains close over the metal to complete its coordination to domain 1 (N1 and C1).

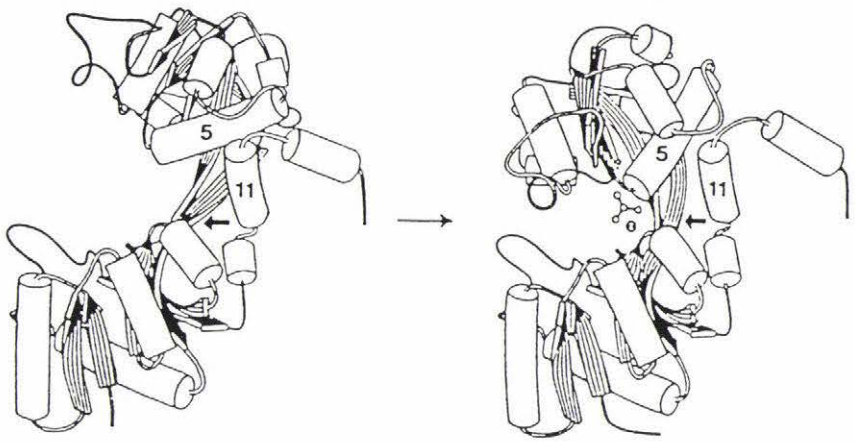


Fig. I.8 Schematic diagrams of the N-lobe of lactoferrin showing the change from the 'open' form of apolactoferrin to the 'closed' form of diferric lactoferrin.
(From Anderson *et al*, 1990.)

An equally interesting feature of the human apolactoferrin structure determined by Anderson *et al* (1990) is that whilst the N-lobe is open, the C-lobe is closed even though neither has metal bound. This was unexpected, for although disulphide 7, which links the C1 and C2 domains, has no counterpart in the N-lobe, and may give extra rigidity to the C-lobe, it was not expected to prevent opening of the lobe. A likely explanation lies in the flexibility of the molecule, with an equilibrium proposed to exist between open and closed forms in solution (Anderson *et al*, 1990). In fact a recent crystal structure of a second crystal form of apolactoferrin (R. Faber, personal communication) has shown that indeed both lobes can open simultaneously. The opening of the N-lobe is about 50°, but the C-lobe opening is less, ~15°, - supporting the hypothesis that the extra disulphide bridge increases the rigidity of the C-lobe. The apolactoferrin structures also show that even in the absence of any metal ion, the closed form must be a stable structure, and there must be very little difference in energy between the open and closed forms.

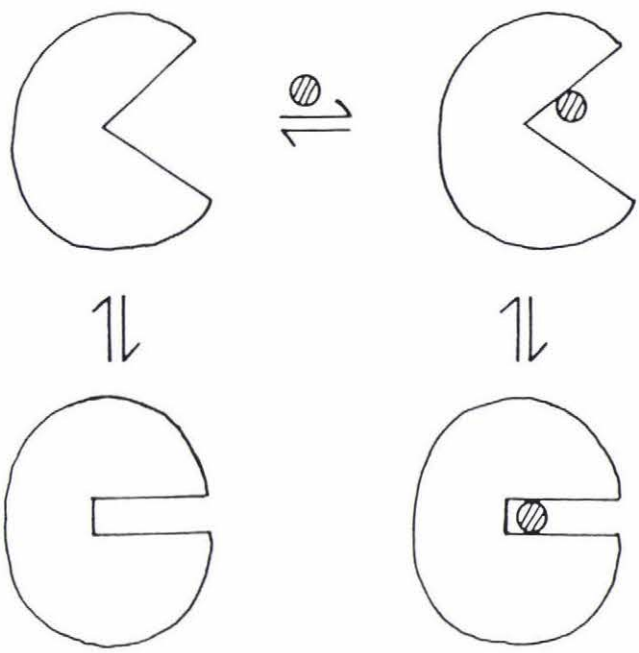


Fig. I.9 Schematic representation of the structural changes associated with binding and release by transferrins and by bacterial periplasmic binding proteins. For the unliganded form (left) an equilibrium is presumed to exist between open and closed states. Binding is to the open form, giving an intermediate in which the ligand is bound to one domain only; domain closure then gives the fully-closed liganded form (lower right). (From Baker *et al*, 1992 with permission.)

I.4 METAL AND ANION SUBSTITUTION

I.4.1 Metal substitution

The metal binding site seems ideally designed for the binding of Fe^{3+} both in terms of charge and size. The high-spin d^5 configuration of the Fe^{3+} ion is compatible with the quite distorted octahedral geometrical arrangement of the ligands provided by the protein, and the +3 charge is neutralised by the 3 anionic oxygen ligands (1 carboxylate and 2 phenolate oxygens). The sizes are such as to allow strong binding of the metal ion to the anion and very favourable hydrogen bonding of the anion to the protein. In spite of this there is clearly sufficient flexibility in the binding site to accommodate other metal ions.

Many solution studies, using a variety of spectroscopic techniques, have been carried out on a variety of metal- and anion-substituted transferrins. Other metal ions that can bind to the two specific sites include di- and tri-valent transition metal ions such as Cr^{3+}, Mn^{3+}, Co^{3+} (Aisen *et al*, 1969; Ainscough *et al*, 1979), Mn^{2+}, Co^{2+}, Cu^{2+}, Cd^{2+}, Zn^{2+} (Tan & Woodworth, 1969), trivalent lanthanides such as Tb^{3+}, Eu^{3+}, Pr^{3+}, Er^{3+}, Ho^{3+} (Luk, 1971) and other trivalent cations such as Ga^{3+}, and Al^{3+} (Gelb & Harris, 1980) and tetravalent actinides Th^{4+} and Pu^{4+} (Harris *et al*, 1981). Some of these are of

biomedical relevance. For example Ga^{3+} and In^{3+} transferrins are used in biomedical imaging, Al^{3+} has been implicated in Alzheimer's disease, and all manganese in human milk is reportedly bound to lactoferrin (Lönnerdal, 1984). It is probable that iron is not the only element whose level in biological fluids is controlled by binding to transferrins.

How can this variety of metal ions be accommodated? Only the aspartate ligand hydrogen bonds to other protein atoms, the remaining metal ligands are not so constrained, and the apolactoferrin structure shows that the domains could be less 'closed' to accommodate the bigger lanthanides. Lanthanide binding studies on transferrin have shown that the larger cations such as Nd^{3+} and Pr^{3+} only bind to one binding site (Gelb & Harris, 1980) and this is apparently the C-lobe site (Zak & Aisen, 1988). For lactoferrin, which binds metal ions more strongly than transferrin (Aisen & Leibman, 1972), all the lanthanide ions, from Yb^{3+} to La^{3+} bind with two metal ions per molecule (Smith, 1992). It is conceivable that protein complexes of larger metal ions may not exhibit the same closed structure as for Fe^{3+}. Spectroscopic monitoring of metal binding usually only reflects tyrosine coordination (Baker, 1994). In the 'open' apolactoferrin structure, both tyrosines are close together and are adjacent to the anion site on domain 2 (Anderson *et al*, 1990), but far away from the Asp and His ligands. Thus a lanthanide ion could bind to both tyrosines and the anion with the transferrin/lactoferrin in the open configuration, and water molecules complete the coordination of the lanthanide instead of the Asp and His ligands. Hence the degree to which lanthanide coordination stabilises the closed structure would not necessarily be detected spectroscopically. Ultimately crystallographic studies of lanthanide-substituted transferrins are needed to define the coordination of the larger metal ions (Baker, 1994).

The recent crystallographic studies of copper-substituted lactoferrin (Smith *et al*, 1991, 1992) complement the solution studies. These show that in this case, where the metal ion is about the same size as Fe^{3+}, the closed structure can indeed be formed. Moreover the extent of domain closure over the bound metal is exactly the same as for iron. The main difference is that there is a slight displacement of the copper atoms relative to the iron positions, ie. 1.0 Å in the N-terminal site, and 0.4 Å in the C-terminal site. This brings about changes in metal coordination geometry and increases the non-equivalence of the two sites. In the N-lobe, the anion, which remains in approximately the same position but rotated ~20°, becomes monodentate to the copper, and is very likely bicarbonate. Thus copper is bound in a 5-coordinate square pyramidal geometry in the N-lobe, but in a 6-coordinate distorted octahedral geometry in the C-lobe. The difference between the two sites was attributed to there being slightly more room for adjustment in the N-lobe, allowing the copper in this site to take on a more favourable geometry. EXAFS measurements with copper ovotransferrin are also consistent with 5-coordination in one lobe and 6-coordination in the other (Garratt *et al*, 1991). These results imply that while other metals may bind in the same sites as Fe^{3+}, their coordination need not be the

same, and furthermore, different metals may enhance the differences between the sites.

I.4.2 Anion Substitution

The binding of different anions to the transferrins has been explored by a variety of spectroscopic techniques, and this has led to a greater understanding of the mechanism of iron binding and release. Schlabach and Bates (1975), using visible absorption spectroscopy to explore the binding of over 30 different anions to transferrin, showed that many organic acids can act as the synergistic anion, but certain inorganic anions such as phosphate, nitrate and sulphate cannot do so. This confirmed and extended earlier EPR studies by Aisen *et al* (1967). They concluded that the common feature of all synergistic anions was the presence of a carboxylate group and a nearby electron-donor ligand (hydroxy, keto, amino or carboxylate), but no more than one bulky group present on the carbon α to carboxylate. A selection of synergistic anions is shown in Table I.3. An "interlocking sites model" was proposed in which the anion bound to the metal and simultaneously bound to a positively-charged protein group, suggested to be an arginine sidechain, through the carboxylate group (Fig. I.10).

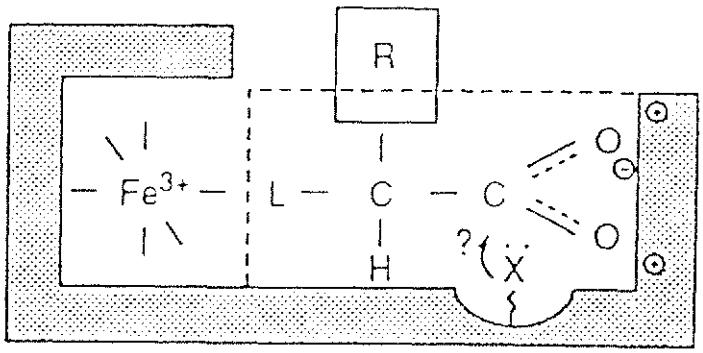


Fig. I.10 A schematic representation of the 'interlocking sites' model for the binding of anions to transferrins where L indicates the second (proximal) functional group. (From Schlabach & Bates, 1975.)

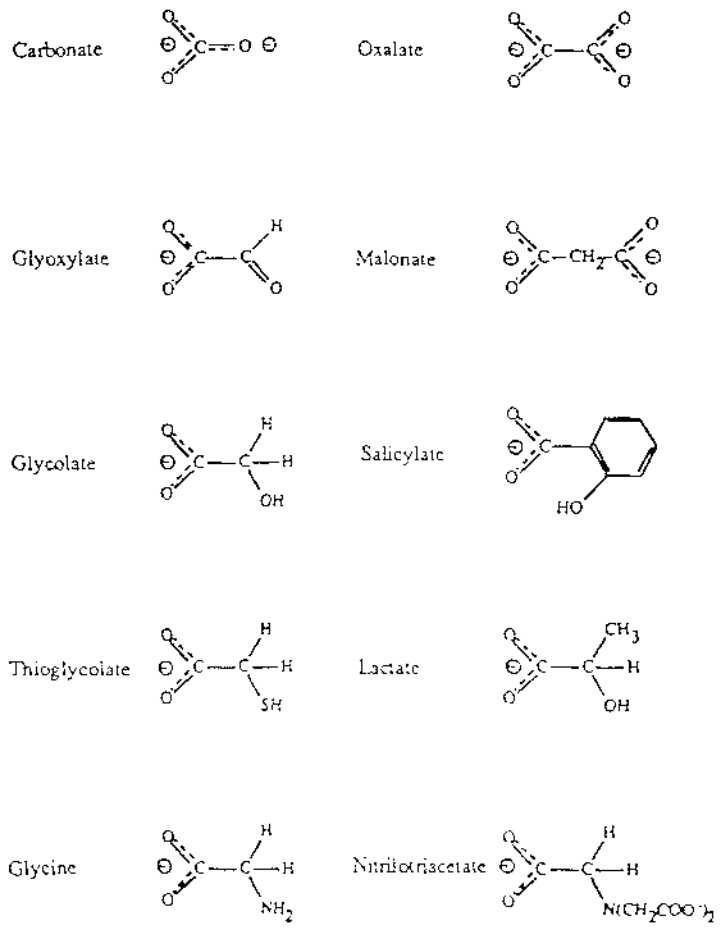


Table I.3 A selection of synergistic anions bound by transferrins. (From Schlabach & Bates, 1975.)

The EPR spectra of vanadyl-transferrin-anion complexes have been used to further probe the synergistic nature of a number of anions containing a carboxylate and a proximal polar group (Campbell & Chasteen, 1977). Anions with two carboxylate groups (oxalate, malonate, maleate) gave two classes of spectra, called A and B, but those with only one carboxylate group (salicylate, glycolate, thioglycolate) gave only class B. The transition from A to B in the first group was linked to the ionisation of a protein group with a pK_a ≈10.0. (This protein group was suggested to be a lysine, but with the knowledge of the three-dimensional structure, this group could be the anion-binding arginine.) It was concluded that these results were in agreement with the interlocking sites model, and that the anion bridged the metal and anion sites by coordinating to the metal through the proximal functional group.

EXAFS and Resonance Raman studies (Schneider *et al*, 1984), using thioglycolate and 2,3-hydroxybenzoate, also showed binding through the electron donor group L, as in the 'interlocking sites' model, but did not indicate if the carboxylate was simultaneously bound. More recent studies using NMR and ESEEM spectroscopy have given conflicting results. The inequivalence of NMR signals for ^{13}C oxalate bound to gallium transferrin was interpreted as indicating that one end of the oxalate was bound to the metal and the other end interacted with the protein, ie. a 1,1-bidentate coordination of oxalate (Bertini *et al*, 1986a). In contrast, ESEEM on Cu^{2+} and VO_2^+ complexes of human transferrin indicated that oxalate bound in a 1,2-bidentate fashion (Eaton *et al*, 1989). This work was further extended in a study involving the EPR and ESEEM spectra of 18 different iron-transferrin-anion complexes. These were all consistent with bidentate binding of the anion. and led to the proposal that all anions behave as bidentate ligands with coordination to the iron through both the carboxylate and proximal groups, and the carboxyl group acting as a bridge between the iron and a positively charged group on the protein (Dubach *et al*, 1991). (Note: these spectra were interpreted in the light of the diferric lactoferrin carbonate structure).

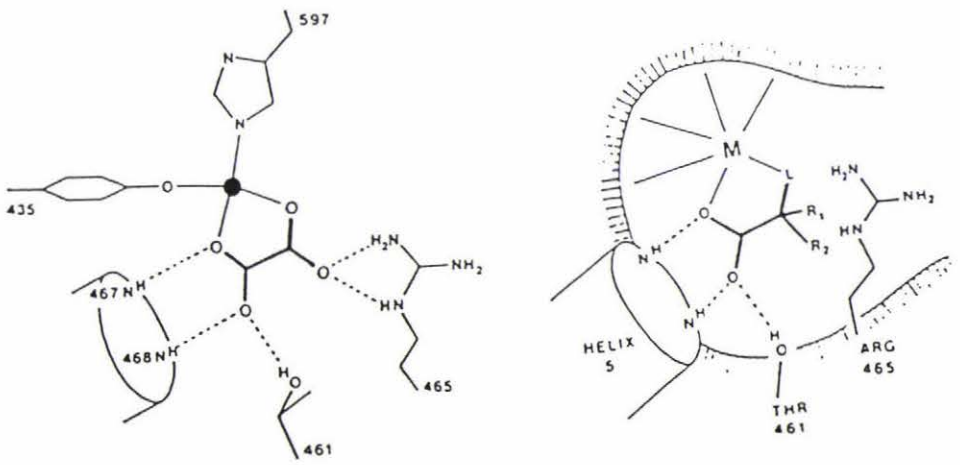


Fig. I.11 Binding modes for anions other than carbonate. (a) Oxalate binding as seen in the crystallographic analysis of a hybrid oxalate-carbonate complex of dicupric lactoferrin. (b) A generalised model for synergistic anion binding to transferrins which is essentially a revised version of the 'interlocking sites' model of Schlabach and Bates. (From Shongwe *et al*, 1992.)

The only crystallographic study of a transferrin with an anion other than carbonate has been a complex of lactoferrin in which Cu^{2+} is the metal ion and carbonate is bound in the N-lobe anion site and oxalate in the C-lobe site (Shongwe *et al*, 1992; Smith *et al*, 1994). This showed oxalate to be bound in a symmetrical 1,2-bidentate mode; one carboxylate binds to the metal, the N-terminus of helix 5, and to Thr 461, while the other carboxylate also coordinates to the metal as well as hydrogen bonding to Arg 465 (Fig. I.11).

The larger size of the oxalate ion, compared with carbonate, does however cause local rearrangement beyond the anion site, in that the sidechain of Arg 465 is displaced ~2 Å away from the anion and the nearby Tyr 398 sidechain ~1.5 Å. This movement is possible because of the solvent-filled cavity beyond the arginine, and appears to be the only change resulting from the introduction of the larger anion.

The implications of this structure led to the modification and extension of the original 'interlocking sites' model of Schlabach and Bates. In the more general model for the coordination of other synergistic anions to transferrins, the carboxylate group of the synergistic anion binds to the N-terminus of a helix (helix 5), and metal coordination is through both a carboxylate oxygen and the electron donor group L on the proximal carbon (Fig. 1.11).

This generalised anion binding model implies that a major factor in the strength of the anion binding is whether or not the arginine hydrogen bonds to the anion, and this depends on the substituents on the proximal carbon. The favourable interactions of oxalate with the arginine is probably the reason why this is the only anion that appears to have a binding affinity close to that of carbonate. The necessity for a larger anion to push aside the arginine may also contribute to the lower affinity for anions other than carbonate.

I.5 DIFFERENCES BETWEEN THE TWO SITES

A feature of transferrin chemistry has been extensive debate as to the degree of equivalence of the two binding sites and the origin of observed differences in their properties.

When Fe^{3+} is the metal ion and carbonate the associated anion, there is little evidence of inequivalence from either spectroscopic or structural data. The refined structure of human diferric lactoferrin shows, for example, that not only are the metal and anion binding groups the same in both sites (see also Section I.3.3(b)), but they also have a very similar geometrical arrangement. Superposition of 81 common atoms around the binding site, plus metal and carbonate, shows that the rms deviation between the N- and C- sites of diferric human lactoferrin is only 0.3 Å (Baker, 1994). This close similarity is reflected in their spectroscopic properties, and in general, spectroscopic properties of the individual N- and C-terminal sites where they have been compared with the

'physiological' Fe^{3+} and CO_3^{2-} ions bound, are so similar as to make it difficult to identify which site is which (Aisen & Harris, 1989).

On the other hand, it is known that in the case of transferrin at least there are distinct thermodynamic and kinetic differences between the sites. Thus the C-terminal site has been shown to bind iron more strongly and release it more slowly (Kretchmar & Raymond, 1986). The two sites also differ in their acid stability, with the N-terminal site releasing iron at a pH of ~6.5, but release from the C-terminal site occurring at pH ~5 (Princiotto & Zapolski, 1975). The biphasic acid-induced release of iron seen for transferrin does not appear to be shared by lactoferrin although it has been reported that in the presence of EDTA release can be biphasic (Mazurier & Spik, 1980).

Differences between the two sites become more pronounced for metal ions other than the native Fe^{3+}, and anions other than CO_3^{2-}. The differences are more apparent for larger metal ions such as lanthanides. For transferrin, some lanthanides appear to bind only at one of the binding sites, and for lactoferrin, U.V. difference titrations show that although binding occurs in both sites, the second metal ion binds much more weakly. A biphasic release of Ce^{4+} from lactoferrin contrasts with the monophasic release of Fe^{3+} (Smith *et al*, 1994b). Differences are also seen when smaller metal ions, such as those from the first transition series, with a similar size to Fe^{3+}, are bound to lactoferrin or transferrin. When Cr^{3+} is bound to lactoferrin, one Cr^{3+} ion is more readily displaced by Fe^{3+} than the other (Ainscough *et al*, 1980), and different EPR signals are seen for the two sites. Similarly the EPR spectra of VO^{2+}-substituted transferrin, indicate different metal configurations in each site (Chasteen *et al*, 1977) as do the ^{13}C NMR studies of Co^{2+}-substituted ovotransferrin (Bertini *et al*, 1986). The crystal structure of copper-lactoferrin (Smith *et al*, 1992) also showed that the coordination of the copper and carbonate could be different in each site. It thus appears that the sites are optimised for the binding of Fe^{3+}, and when different metals are bound, with different stereochemical requirements and different sizes, the sites can adjust with small movements, but these adjustments can be different in the two sites.

Anions other than the physiological CO_3^{2-} also reveal the differences between the two sites, especially when combined with metals other than iron. Thus the use of oxalate as the synergistic anion gives rise to spectroscopically distinct sites for ovotransferrin with Al^{3+}, Ga^{3+} and Zn^{2+} with the spectra indicating that the anion arrangement is very similar but not identical in the two sites (Bertini *et al*, 1986). When copper is the metal ion, transferrin will bind oxalate in just the N-lobe (Zweier & Aisen, 1977). On the other hand, copper-lactoferrin and copper-ovotransferrin each bind two oxalate ions, but binding occurs preferentially in the C-lobe (Zweier, 1980; Shongwe *et al*, 1992).

I.6 QUESTIONS TO BE ADDRESSED BY THIS THESIS

This thesis seeks to investigate two areas of lactoferrin structure and function. The first is the means by which lactoferrin can accommodate anions larger than carbonate, and the second is the effect of glycosylation on various physical and chemical properties of the protein. Given

(i) the biological importance of iron binding,

(ii) the prevalence of many different anionic metabolites (eg. oxalate, glycolate) as well as carbonate which can act as the synergistic anion, and

(iii) the synergistic relationship between the anion and the metal ion,

it was decided to investigate the effects of binding oxalate as the synergistic anion with Fe^{3+} as the associated cation. This first section aimed at determining how the anion binds and whether its binding is the same in both sites, as well as seeking to explain the apparent strong preference for carbonate over oxalate when Fe^{3+} is the metal ion.

To this end the oxalato complex of diferric lactoferrin, $Fe_2(C_2O_4)_2Lf$, was prepared. Solution studies were carried out in order to determine the conditions under which binding occurred, and to establish the spectroscopic signatures of the resulting lactoferrin complexes. Crystallographic analyses were then necessary to resolve the detailed mode(s) of interaction of oxalate with lactoferrin and to provide the basis for relating the solution studies to the three-dimensional structure.

The roles of the carbohydrate moieties of glycoproteins are a subject of continuing research and debate. Lactoferrin is no different in this respect. The primary structures of the lactoferrin glycans and their points of attachment have been determined, but evidence that the carbohydrate has roles in iron binding, stability receptor binding and antibacterial properties has been contradictory and inconclusive. Further doubts about the proposed roles of the glycans are raised by recent structural studies of native and deglycosylated lactoferrin.

Thus the purpose of the second section, was not to answer all these questions, but to compare a limited number of physicochemical properties of native and deglycosylated lactoferrin in an attempt to relate these to possible roles of the carbohydrate. A fuller discussion of the roles of the glycans on glycoproteins in general, and in lactoferrin in particular, is given at the beginning of Chapter IV.

CHAPTER II

PREPARATION OF DIFERRICDIOXALATOLACTOFERRIN

II.1 EXPERIMENTAL PROCEDURES

A particular feature with metal and anion binding studies on a protein such as lactoferrin, with its very high affinity for iron under physiological conditions (when atmospheric CO_2 provides adequate CO_3^{2-} for complexation), is the need for strict control over the presence of both metal ions and anions.

II.1.1 Glassware and Dialysis Tubing

All glassware was soaked in Pyroneg, then washed using DECON 50, and rinsed thoroughly with deionised water. It was subsequently treated with A.R. concentrated nitric acid, and thoroughly rinsed with Milli Q-water and dried at 100°C.

To minimise contamination from metal ions, the dialysis tubing was heated at 80°C for 10 minutes in a solution of 0.6M $NaHCO_3$, 1 mM EDTA (disodium salt). This was repeated using a fresh solution, and the tubing was then rinsed 4 times with Milli-Q water. Finally the tubing was heated to 90°C in two changes of Milli-Q water and stored in a 50:50 ethanol:water solution at 4°C.

II.1.2 Solutions

(a) *Tris-HCl buffer:*

The appropriate amount of tris(hydroxymethyl)aminoethane (SERVA Feinbiochemica) was dissolved in Milli-Q water, and the solution adjusted to the required pH with A.R. hydrochloric acid. Sufficient NaCl was added to the Tris-HCl solution to make it the desired molarity before adjusting the pH.

(b) *Phosphate buffer:*

This was prepared by titrating a 0.01M solution $Na_2HPO_4.2HO$ (Reidel-deHaen) with a 0.01M solution of NaH_2PO_4 (Reidel-deHaen) until the desired pH was reached.

(c) *HEPES buffer:*

Prepared by dissolving N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid (U.S. Biochem Corp.) in Milli-Q water and adjusting the pH with concentrated HCl.

(d) *FeNTA solution:*

This was prepared by dissolving the appropriate amount of NTA (disodium salt) (Sigma) in water to give a concentration of 0.01M. One equivalent of $Fe(NO_3)_3.9H_2O$,

A.R. grade (Reidel-deHaen) was added to give a 1:1 FeNTA complex. The solution was freshly prepared and used within 24 hours. The concentration of this solution was determined by atomic absorption spectrometry.

II.1.3 **Gel Electrophoresis**

SDS polyacrylamide gel electrophoresis was carried out following the procedure of Laemmli (1970). 0.75 mm slab gels were cast, with a stacking gel of 4% acrylamide, 1.7% crosslinking and a resolving gel of 9% acrylamide with 2.7% crosslinking. Protein samples were denatured and given a uniform charge by heating at 100°C for 2 minutes in a solution of 0.125M Tris-HCl pH 6.8, 4% SDS 20% glycerol, and a total of 4 µg of protein loaded per well. The reservoir of the electrophoresis apparatus was filled with buffer (0.025M Tris-HCl pH 8.3, 0.192M glycine, 0.1% SDS), and the electrophoresis carried out under constant current conditions, using 5mA until the dye front had penetrated the resolving gel, and 10 mA thereafter.

Gels were stained overnight using Coomassie Blue stain (0.125% Coomassie Blue R-250, 50% methanol, 10% acetic acid) then destained by soaking for 1 hour in 50% methanol 10% acetic acid, followed by several hours in 7% acetic acid, 5% methanol.

II.1.4 **Electronic Spectra**

(a) *Electronic Absorption:*

Absorption spectra in the range 240-700 nm were recorded with a Hewlett-Packard 8452A diode array spectrophotometer interfaced with an IBM-compatible computer. The calibration of the instrument was regularly checked using a holmium oxide filter, and no fluctuation of wavelength was observed. Quartz cuvettes (1.5 ml with a path length of 1 cm) were used to hold the protein samples.

(b) *Electron paramagnetic resonance spectroscopy:*

Electron paramagnetic resonance (EPR) spectra were recorded at 110K, with protein concentrations of approximately 15 mg/ml, using a Varian E-104A spectrometer equipped with a Varian E-257 variable temperature accessory. The protein sample (usually about 0.5 ml) was loaded into an EPR tube of 2 mm diameter back-filled with argon and slowly cooled to 110°K by passing a stream of oxygen-free nitrogen, cooled by liquid nitrogen, into the cavity and over the sample tube. Spectral g-values were calibrated with diphenylpicrylhydrazyl (DPPH) as standard.

Spectral g-values for the resonances were calculated using the equation:

$$g_{sample} = \frac{g_{std} \times B_{std}}{B_{sample}}$$

where g_{std} = 2.0036 (for DPPH)

B_{std} = the observed magnetic field for the DPPH standard

B_{sample} = the observed magnetic field for the metal ion complex

II.1.5 **Purification of Apolactoferrin**

All buffers used had been previously passed down a BioRad Chelex-100 column in order to minimise contamination by adventitious metal ions.

Apolactoferrin was prepared at 4°C following the method of Norris *et al* (1989). Fresh human colostrum, collected 1-4 days post-partum, was normally used, although colostrum which had been stored at -20°C for less than 3 months could be used without any noticeable difference to the properties of the lactoferrin prepared. The colostrum (typically 100-400 ml) was gently stirred while sufficient $(NH_4)_2SO_4$ was added over the period of an hour to bring the solution to a concentration of 2 M. Stirring was continued for a further hour to ensure the complete precipitation of the caseins and immuno-globulins. The solution was centrifuged at 12,000 r.p.m. for thirty minutes to remove the fat and casein. The supernatant was dialysed successively against 2 changes of Milli-Q water (5 L each) and 2 changes of 0.025M Tris-HCl pH 7.8 0.2M NaCl (5 L each) over a total period of 24 hours. The crude lactoferrin was loaded on to a CM Sephadex C-50 column (20 cm x 4 cm) previously equilibrated in 0.025M Tris-HCl pH 7.8, 0.2M NaCl, and the column was washed until the eluent showed no absorbance at 280 nm. Lactoferrin, which remained tightly bound to the top of the column, was eluted using a 0.2-1.1M NaCl gradient (200 ml:200 ml), 10 ml fractions being collected using a Pharmacia fraction collector. Lactoferrin generally eluted as a single, well-resolved peak, although occasionally a smaller peak would elute just prior to the major band (Plowman, 1979). The column could be restored for further use by washing with 0.025M Tris-HCl pH 7.8, 1.5M NaCl, and then equilibrating with 0.025M Tris-HCl pH 7.8, 0.2M NaCl.

The concentration and iron saturation of the lactoferrin was estimated using $\mathcal{E}^{1\%}_{1\text{ cm }280\text{ nm}} = 10.7$ and $\mathcal{E}^{1\%}_{1\text{ cm }466\text{ nm}} = 0.51$ respectively (Aisen and Leibman, 1972). The purity of each fraction across the peak was checked using SDS polyacrylamide gel electrophoresis, and pure fractions combined. The preparation procedure gave an average yield of 2 mg per ml of colostrum, with the lactoferrin having a maximum iron content of 8-10% saturation. A comparison of the spectral ratios of $A_{280\text{ nm}}/A_{466\text{ nm}}$ and $A_{412\text{ nm}}/A_{466\text{ nm}}$ with those of fully iron-saturated lactoferrin (ratios of 20-23 and 0.70 to 0.75 respectively) provided a further check on the purity and iron binding properties of the lactoferrin (Norris *et al*, 1989).

II.1.6 **Preparation of Diferricdioxalatolactoferrin** ($Fe_2(C_2O_4)_2Lf$)

Three separate methods were used in attempts to prepare oxalate-substituted diferric lactoferrin complexes, as outlined below. Two involved the use of carbonate-free solutions, but the third was based on attempts to displace carbonate directly with oxalate (Shongwe *et al*, 1992).

(a) *Metal-binding titration*

5 ml of apolactoferrin (10 mg/ml in 0.025M Tris-HCl, 0.2M NaCl, pH 8.0) were placed in a Schlenk tube, 6.3 mg of AR sodium oxalate (BDH) (50 fold excess) were dissolved in the solution, and the pH was gradually lowered to 4.3 with 0.10M metal-free HCl. At this stage a light precipitate appeared which was removed by filtration. Release of CO_2 was promoted by subjecting the solution to a small vacuum (10-12 mmHg) over a period of two hours during which time the Schlenk tube was intermittently flushed with argon. The pH of the solution was restored to 8.0 with CO_2-free NH_3, prepared by bubbling ammonia gas through Milli-Q water which had also been evacuated and flushed with argon.

Using a spectrophotometer to monitor the titration, 2 µl aliquots of 0.01M FeNTA solution were added to 1 ml of the carbonate-free apolactoferrin/oxalate solution in a sealed 1.5 ml quartz cuvette, back-filled with argon.

(b) *Anion-binding titration*

In a similar manner to the above, a titration of carbonate-free apolactoferrin/Fe^{3+} solution with 0.01M sodium oxalate solution was carried out. In this case the pH of the apolactoferrin solution was lowered to 4.3, the solution subjected to a gentle vacuum, and the pH raised to 8.0 as before. A 2 molar excess of ferric citrate was added (citrate is a non-synergistic anion) and the titration of the carbonate-free apolactoferrin/Fe^{3+} solution with 0.01M sodium oxalate was carried out.

(c) *Displacement method*

Three tubes with $Fe_2(CO_3)_2Lf$ (10 mg/ml in 0.025M Tris-HCl, 0.2M NaCl pH 8.0), one with a 50-fold, one with a 100-fold, and the third with a 200-fold excess of sodium oxalate, were left at 4°C for a period of 6 weeks. The UV-visible spectrum of each solution was measured daily for the first week and then weekly thereafter.

II.1.7 **Preparation of Diferricdioxalatotransferrin** ($Fe_2(C_2O_4)_2Tf$)

Diferricdioxalatotransferrin was prepared using human apotransferrin (Sigma) under carbonate-free conditions as in section II.1.6(a). This was done in order to be able to compare the spectra of the oxalate complexes of lactoferrin and transferrin, prepared and measured under identical conditions.

II.1.8 Preparation of Diferricdioxalatolactoferrin Solution for EPR Spectroscopy

To minimise contamination from excess FeNTA, the $Fe_2(C_2O_4)_2$Lf solution was quickly passed down a BioRad Chelex-100 column (1 cm x 10 cm) previously equilibrated in 0.025M Tris-HCl, 0.2M NaCl pH 8.0.

II.1.9 Crystallisation of Diferricdioxalatolactoferrin

Successful crystallisations of various lactoferrin species in this laboratory include the following species: native human Fe_2Lf (Baker and Rumball, 1977), native human apoLf (Norris *et al*, 1989), deglycosylated human apoLf (Norris *et al*, 1989), native bovine Lf (Norris *et al*, 1986), native human Cu_2Lf (Smith *et al*, 1991), native human $Cu(CO_3)(C_2O_4)$Lf (Smith *et al*, 1991), recombinant N-terminal half molecule, Lf_N (Day *et al*, 1992), deglycosylated Lf_N (Day *et al*, 1992), and the mutants Lf_N R121S, Lf_N R121E and Lf_N D60S. All these have involved very similar conditions, ie very pure protein, high protein concentration (typically 50-80 mg/ml) and dialysis from low ionic strength buffers, with a pH between 7.6-8.4. A very similar strategy was therefore followed in the present case, but as $Fe_2(C_2O_4)_2$Lf had been shown to slowly revert to the carbonato-complex (see results), buffers were made CO_2-free, and the crystallisations were set up under an atmosphere of argon.

(a) *Preparation of buffers for use in crystallisation*

In order to minimise contamination by CO_2 the solutions were subjected to the same repeated evacuation and flushing with argon as the protein solution. Thus 100 ml of 0.01M Na_2HPO_4 (pH 4.5) was placed in a round-bottom flask and had a small vacuum (10-12 mmHg) applied to it over a period of 2 hours during which time the flask was intermittently flushed with argon. The pH of the solution was then raised to 8.0 using CO_2-free ammonia solution.

AR methanol used in the dialysis solution for crystallisation, was also evacuated under argon as above.

(b) *Crystallisation*

The protein used had a purple-red colour (with a visible absorbance maximum at 482 nm, and a ratio of $A_{280\ nm}/A_{482\ nm}$ of 19.0. Using a microconcentrator (Centricon 30, Amicon) the protein was concentrated to 70 mg/ml and loaded into 30 µl microdialysis buttons. These buttons were set to dialyse against 20 ml of 0.01M sodium phosphate pH 8.0, 11% methanol, in vials which had been back-filled with argon, and then sealed securely to prevent CO_2 contamination.

II.2 RESULTS

II.2.1 Purification of Apolactoferrin

Lactoferrin generally eluted as a single well-resolved peak between 0.85 M and 1.0 M NaCl (Fig. II.1).

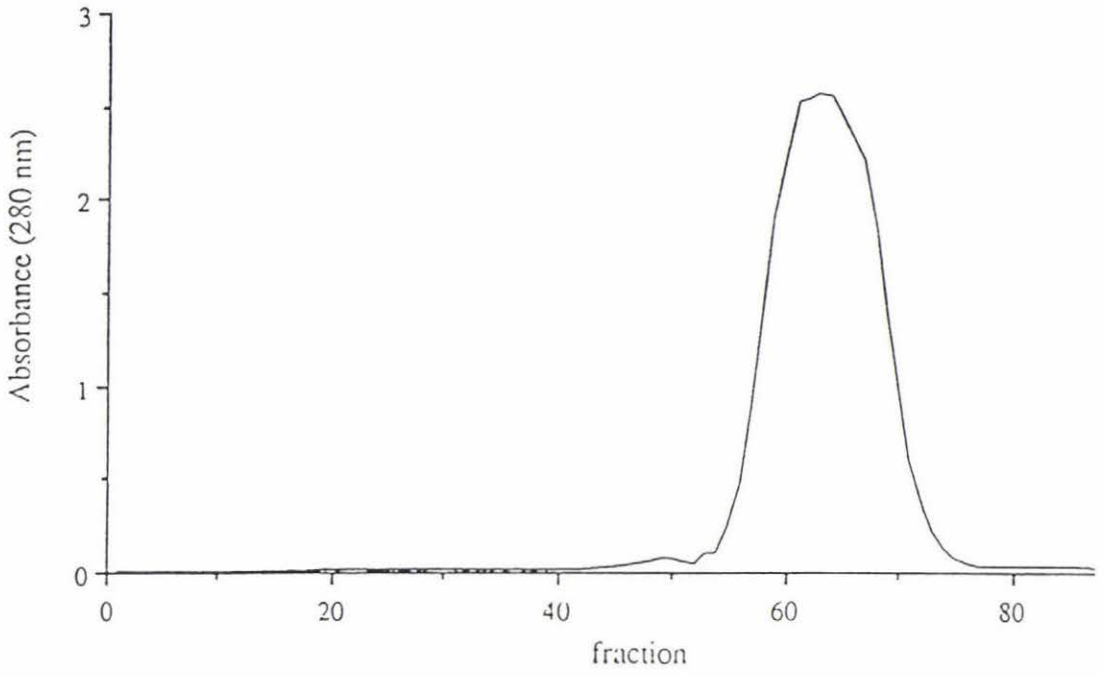


Fig. II.1. Elution profile for purification of lactoferrin by CM-Sephadex C-50 chromatography, using a 0.2-1.1 M NaCl gradient.

The preparation usually gave a yield of 2 mg of pure lactoferrin (as judged from SDS gels and spectral ratios) per ml of colostrum, with the lactoferrin having a maximum iron content of 8-10% saturation.

II.2.2 Preparation of Diferricdioxalatolactoferrin

(a) *From carbonate-free solution*

Spectrophotometric titrations carried out in the absence of carbonate clearly showed that 2 moles of Fe^{3+} and 2 moles of oxalate bound per mole of lactoferrin (Fig. II.2a, b).

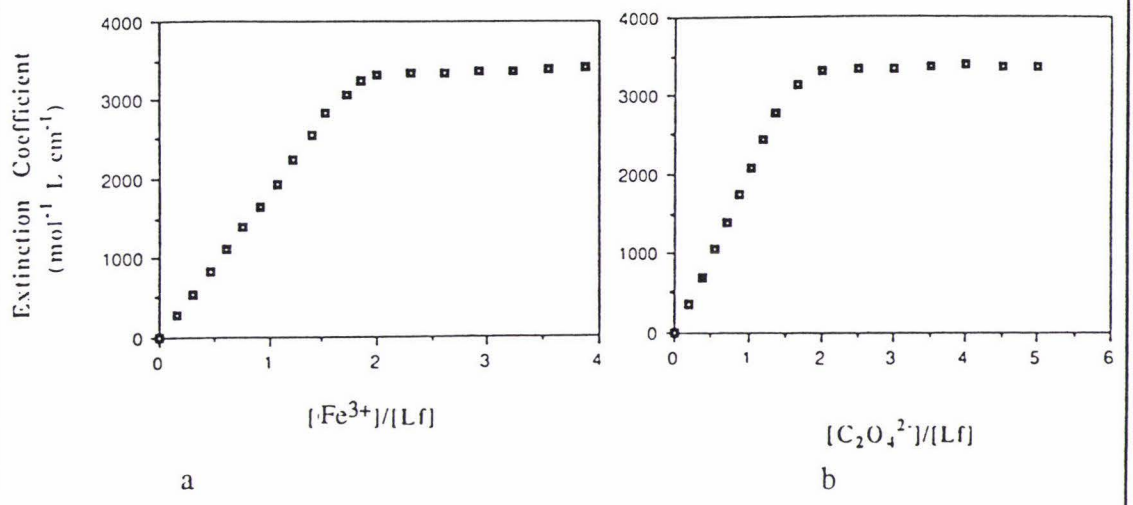


Fig. II.2. Fe^{3+}- and oxalate-binding titrations with carbonate-free human apolactoferrin. (a) Addition of Fe^{3+} to lactoferrin in the presence of excess oxalate. (b) Addition of $C_2O_4^{2-}$ to lactoferrin in the presence of excess Fe^{3+}.

(b) *By displacement*

All attempts to prepare oxalate complexes of lactoferrin by displacement of carbonate from $Fe_2(CO_3)_2Lf$ were unsuccessful, even when the protein was subjected to a 200-fold molar excess of oxalate for a period of six weeks. The wavelength of the visible absorption maximum increases from 466 nm for $Fe_2(CO_3)_2Lf$, to 474 nm for a mixed carbonate/oxalate complex, to 482 nm for $Fe_2(C_2O_4)_2Lf$ (see below). In these displacement experiments the visible absorption maximum remained unchanged at 466 nm, indicating that there had been no displacement of carbonate by oxalate.

II.2.3 **Electronic Spectra**

(a) *Electronic absorption spectra*

The substitution of carbonate by oxalate in diferric lactoferrin results in a distinct change in the wavelength of maximum absorption in the visible electronic spectrum. The dioxalate complex has a purple-red colour (λmax 482 nm) (Shongwe *et al*, 1992) compared with the brick-red carbonate complex (λmax 466 nm) (Fig. II.3). A mixed carbonate-oxalate lactoferrin complex, prepared by exposing $Fe_2(C_2O_4)_2Lf$ to atmospheric CO_2 for 2 weeks had a visible absorption maximum at an intermediate value of 476 nm. Furthermore the substitution of oxalate for carbonate gave a decrease by over 1000 mol^{-1} cm^{-1} in the molar absorption coefficient of the complex (Table II.1).

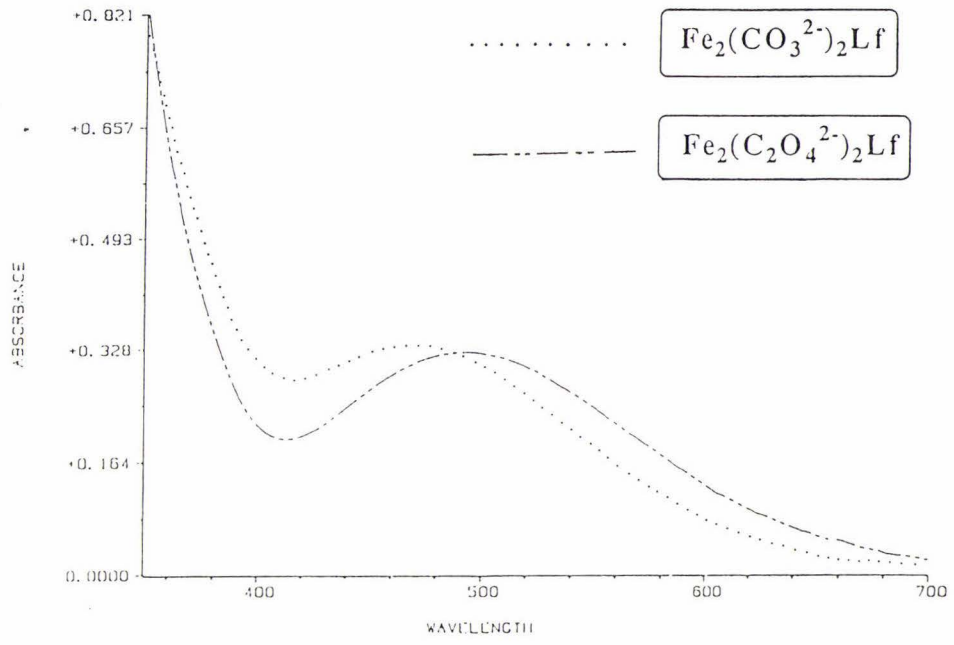


Fig. II.3(a). Electronic visible spectra of oxalate and carbonate complexes of diferric human lactoferrin.

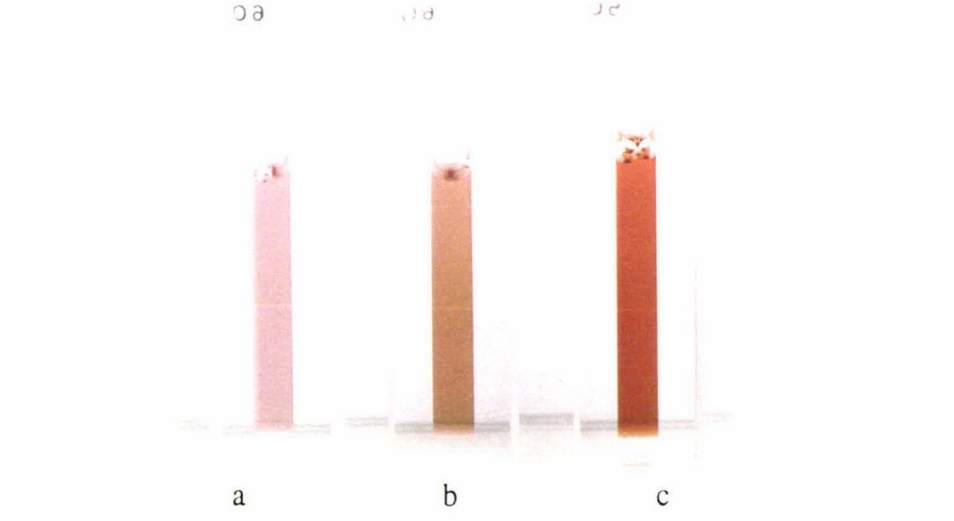


Fig. II.3(b). Solutions of oxalate and carbonate complexes of diferric lactoferrin. (a) $Fe_2(C_2O_4)_2Lf$, (b) mixed carbonate and oxalate diferric lactoferrin, (c) $Fe_2(CO_3)_2Lf$.

Table II.1. Charge transfer absorption maxima and extinction coefficients of diferric lactoferrin and transferrin complexes.

Complex	λmax (nm)	Σ ($mol^{-1}L\ cm^{-1}$)
$Fe_2(CO_3)_2Lf$	466	4600
$Fe_2(C_2O_4)_2Lf$	482	3600
$Fe_2(CO_3)_2Tf$	466	4400
$Fe_2(C_2O_4)Tf$	472	3300

(b) *Electron paramagnetic resonance spectra*

The EPR spectra (Fig. II.4) provided a more sensitive indication of the nature of the anion bound to lactoferrin than the visible absorption spectra. The EPR spectrum of $Fe_2(CO_3)_2Lf$ has a unique strong resonance near 1500 gauss ($g' \approx 4.3$), and a weaker resonance at 700 gauss ($g' \approx 9.7$), whereas the dioxalato complex has a broad peak centering on 1100 gauss. The mixed carbonate-oxalate complex mentioned above could be clearly identified from the presence of both types of spectrum.

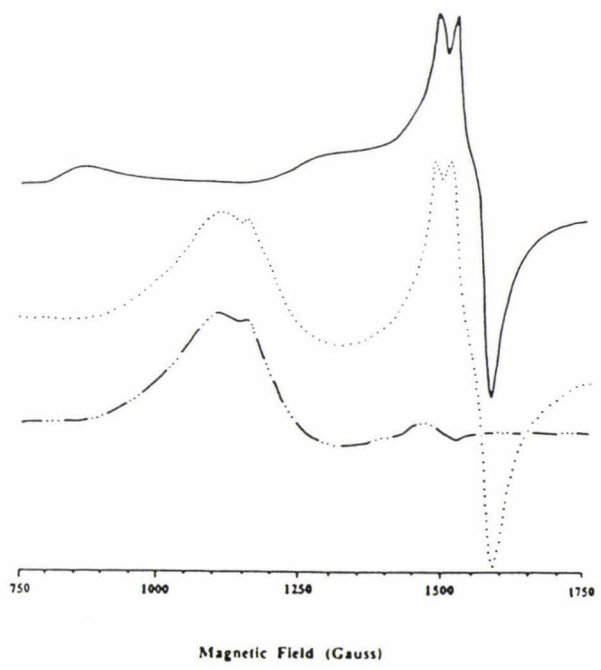


Fig. II.4. EPR spectra of carbonate, mixed carbonate/oxalate and oxalate complexes of diferric lactoferrin recorded at 110K and microwave frequency of 8,990 GHz. Spectra are $Fe_2(CO_3)_2Lf$ ——, mixed $Fe(CO_3)(C_2O_4)Lf$ ------, $Fe_2(C_2O_4)Lf$ — - - - —

II.2.4 **Crystallisation**

Large (0.5 x 0.5 x 1.5 mm) needle shaped, purple-red crystals (Fig. II.5) grew in 18-21 days if the vials were left undisturbed at 4°C. These crystals were characterised by precession photography and were found to be orthorhombic, with unit cell dimensions a = 155.46, b = 96.92, c = 55.85 Å $\alpha = \beta = \gamma = 90°$ and space-group $P2_12_12_1$ (deduced from systematic absences in the diffraction pattern). Assuming one molecule in the asymmetric unit of the crystal, the value of V_m is 2.63 $Å^3$ $Dalton^{-1}$ (Matthews, 1968) corresponding to a solvent content in the crystals of ~53%. The crystals are closely isomorphous with those of the native carbonate complex, $Fe_2(CO_3)_2Lf$ (a = 156.26, b = 97.40, c = 55.85 Å $\alpha = \beta = \gamma = 90°$).

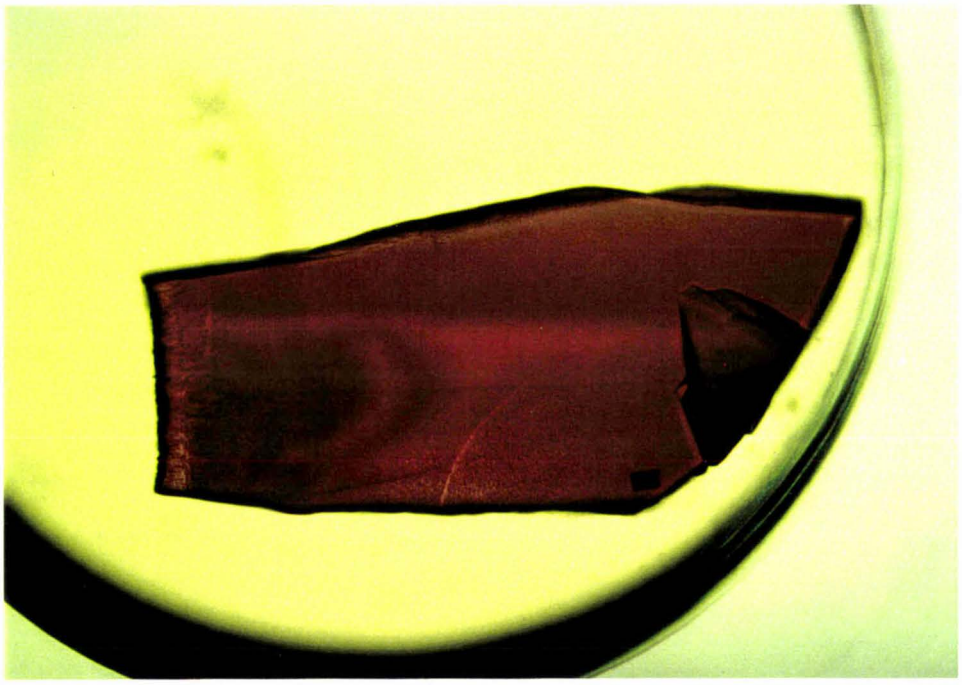


Fig. II.5. Crystal of $Fe_2(C_2O_4)_2Lf$. This was grown by microdialysis against 0.01M sodium phosphate, pH 8.0, containing 11% methanol.

II.3 **DISCUSSION**

The preparation and interconversion of carbonate, oxalate, and hybrid oxalate-carbonate complexes of diferric lactoferrin is summarised in Fig. II.6. Stringent carbonate-free conditions were necessary to prepare a pure dioxalato complex. This complex was stable only under CO_2-free conditions; carbonate begins to replace oxalate after one hour's exposure to atmospheric CO_2, with an approximately 1:1 carbonate:oxalate complex being formed after 2 weeks' exposure to atmospheric CO_2. $Fe_2(C_2O_4)_2Lf$ was completely converted to $Fe_2(CO_3)_2Lf$ after 1 day's incubation with 0.1M $NaHCO_3$. Both visible and EPR spectra were used to monitor the formation of the

oxalate and carbonate complexes, but the EPR spectra proved to be far more sensitive to small contamination of $Fe_2(C_2O_4)_2Lf$ by carbonate.

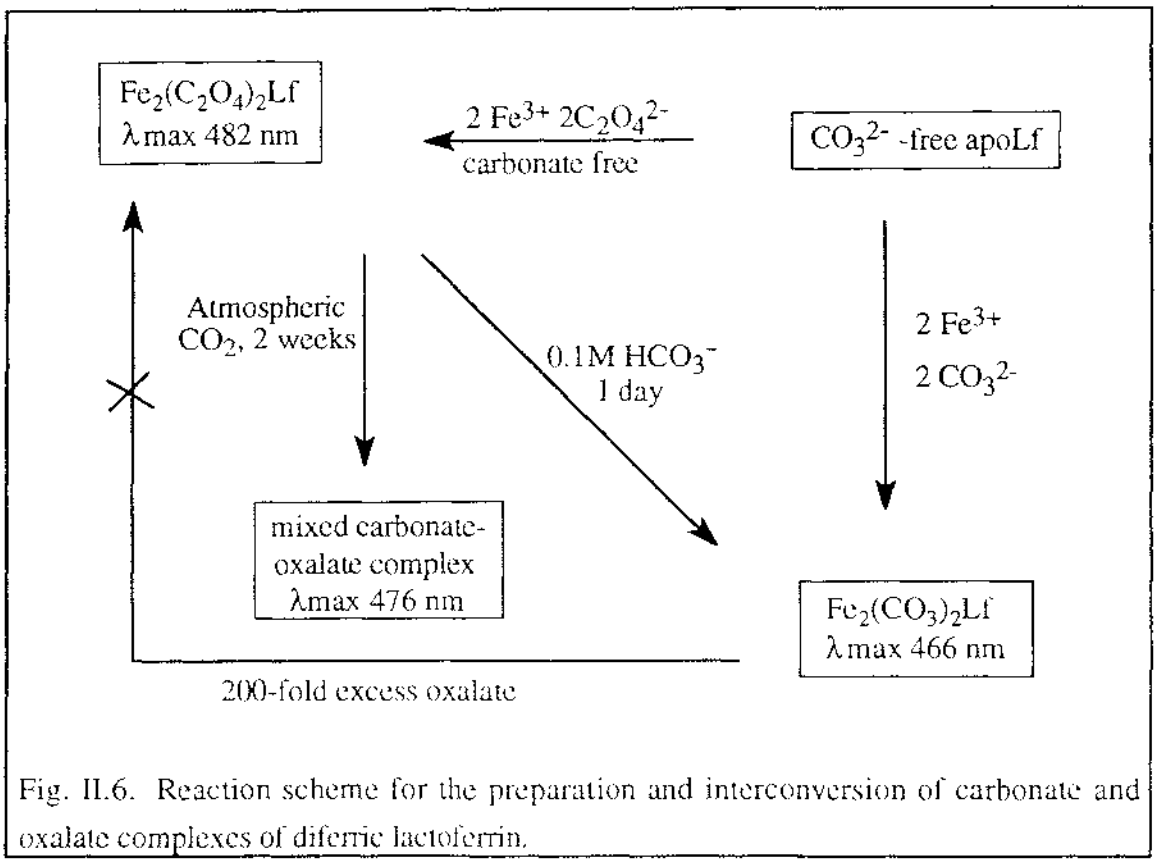


Fig. II.6. Reaction scheme for the preparation and interconversion of carbonate and oxalate complexes of diferric lactoferrin.

Further discussion and interpretation of the spectroscopic results in the light of the crystal structure of $Fe_2(C_2O_4)_2Lf$ will follow in the concluding chapter (Chapter V).

CHAPTER III

CRYSTAL STRUCTURE OF DIFERRICDIOXALATOLACTOFERRIN

III.1 METHODS

III.1.1 Stabilisation and Mounting of Crystals

Lactoferrin crystals are generally very difficult to handle. Having been grown at low ionic strength, in the cold, they are temperature-sensitive and have a tendency to dissolve in their mother liquor. Previous studies with $Fe_2(CO_3)_2$Lf (Anderson *et al*, 1987) showed that the crystals could be stabilised by replacing the volatile alcohol by a less volatile one such as 2-methyl-2,4 pentanediol (MPD). This also led to improved resolution in the diffraction pattern. Accordingly dialysis buttons with crystals of $Fe_2(C_2O_4)_2$Lf were successively transferred to vials containing phosphate buffer and the following methanol/MPD solutions: 10%:2%, 8%:4%, 6%:6%, 4%:8%, 2%:10%, 0%:12%, 0%:16%, allowing at least 24 hours to equilibrate in each solution. The solutions were subjected to a gentle vacuum and flushing with argon as in Section II.1.9, to prevent any carbonate contamination of the crystals. Careful checks were made on the crystals to ensure that they were not cracking or dissolving. The stabilised crystals were then mounted in thin-walled, siliconised capillary tubes, along with a little of the mother liquor, and the ends of the capillaries were sealed with wax.

III.1.2 Data Collection

X-ray diffraction data to 2.2 Å resolution were collected by screenless Weissenberg photography (moving crystal/moving film) using a Weissenberg camera equipped with imaging plates developed by Dr N. Sakabe, at the Photon Factory synchrotron radiation source, Tsukuba, Japan (Sakabe, 1991). This method, which is both rapid and efficient, allows many thousands of reflections to be recorded on each plate. Furthermore, the use of the high intensity X-ray beam of the synchrotron enables more data, often to higher resolution, to be collected from individual crystals; an important consideration when the crystals are few in number, small, or diffract only moderately to poorly.

Crystals were first aligned along a direct cell axis using small angle oscillation photographs taken with a polaroid cassette. Reflection data were then recorded on Fuji imaging plates and measured as laser stimulated luminescence with a BA100 image plate densitometer, as described in the Synchroton users guide (Rehse and Zubak, 1992). Two crystals were used in the data collection, one with the crystal aligned along its a-axis, and the second aligned along its b-axis. The camera settings and exposure times used to collect the data are summarised in Table III.1.

Table III.1 Image plate (IP) data collection statistics

Parameter	Value	
Source	Beam-line 6A2 Photon Factory	
Method	Screenless Weissenberg Photography	
Wavelength	1.0 Å	
Crystal to IP distance (mm)	430	
Coupling constant	2.5	
Oscillation range	7°	
Number of osc^{n}/IP	8	
ω scan speed (°/s)	8	
Exposure per image (s)	64	
Number of IPs	15 (a-axis)	14 (b-axis)
Total exposure(s)	960 (a-axis)	896 (b-axis)

III.1.3 **Data Processing**

The image plate data were processed using the program WEIS (Higashi, 1991). Each image was first oriented using the fiducial marks recorded on the image, and then a number of medium-to-strong reflections were used to refine the initial orientation matrix parameters. Initially data to 10.0 Å were used for refinement. Once the missetting parameters, DR and DS, (DR and DS are the *rms* differences between the observed and calculated spot positions in the x and y directions) had reached a minimum, the resolution of the data used for refinement of the orientation matrix was progressively extended to 3.5 Å. Although some images could be processed without manual intervention, others required a considerable amount of manipulation before a setting file could be obtained that could accurately predict the positions of all the spots. Some of the difficulties were due to both crystals giving split spots at higher resolution. This was especially marked in the data collected around the a-axis. All data within a sphere of 2.2 Å were processed. Details of the a-axis and b-axis data processing are given in Tables III.2a and III.2b respectively.

Table III.2a Data collection and processing statistics for the a-axis $Fe_2(C_2O_4)_2Lf$ images

ω range represents the oscillation range for which data were collected for each image. DR and DS give a measure of the missetting of the centre of each reflection. DR and DS are the *rms* differences between the observed and calculated spot positions in the horizontal and vertical directions respectively. RMS/BGD is the ratio of the standard error to the background level of density.

Image Plate	ω range	% expected reflections observed	DR	DS	RMS/BGD
1	-3.5 - 3.5	64.5	0.45	0.39	3.3
2	3.0 - 10.0	59.7	0.55	0.33	3.0
3	9.5 - 16.5	65.3	0.52	0.35	2.8
4	16.0 - 23.0	66.5	0.46	0.40	2.8
5	22.5 - 29.5	60.1	0.43	0.39	2.6
6	29.0 - 36.0	57.7	0.33	0.25	2.3
7	35.5 - 42.5	56.2	0.32	0.24	2.3
8	42.0 - 49.0	56.5	0.31	0.21	2.0
9	48.5 - 55.5	58.7	0.32	0.25	2.0
10	55.0 - 62.0	51.2	0.32	0.24	1.9
11	61.5 - 68.5	52.4	0.29	0.24	2.2
12	68.0 - 75.0	52.0	0.29	0.24	1.7
13	74.5 - 81.5	50.2	0.28	0.24	1.7
14	81.0 - 88.0	51.3	0.29	0.21	1.6
15	87.5 - 94.5	52.1	0.28	0.28	1.8

Total range of 98° (0.5° overlap per image).

Table III.2b Data collection and processing statistics for the b-axis $Fe_2(C_2O_4)_2Lf$ images

Image Plate	ω range	% expected reflections observed	DR	DS	RMS/BGD
2	4.0 - 11.0	50.1	0.31	0.33	2.1
3	10.5 - 17.5	55.4	0.22	0.34	2.6
4	17.0 - 24.0	55.3	0.34	0.33	2.6
5	23.5 - 20.5	55.1	0.32	0.32	2.6
6	30.0 - 37.0	52.8	0.33	0.35	2.5
7	36.5 - 43.5	59.6	0.29	0.33	1.8
8	43.0 - 50.0	63.0	0.31	0.33	1.5
9	49.5 - 56.5	67.0	0.38	0.35	1.0
10	56.0 - 63.0	70.1	0.39	0.36	1.0
11	62.5 - 69.5	70.1	0.39	0.36	1.0
12	69.0 - 76.0	73.5	0.38	0.36	0.9
13	75.5 - 82.5	67.4	0.41	0.39	1.1
14	82.0 - 89.0	66.2	0.38	0.38	1.1
15	-2.5 - 4.5	52.5	0.35	0.33	1.1

Total range of 91.5° (0.5° overlap per image).

The integrated intensities from each image were combined, and the data from the image plates scaled and merged using the CCP4 programs ROTAVATA and AGROVATA (the CCP4 Suite Programs, 1994). Table III.3 summarises the relevant statistics for data collected from crystal 1 and crystal 2 and the combined data from both crystals.

Table III.3 Merging of the intensity data

Source of data set	Numer of unique reflections for data with $I \geq 1\sigma_I$	R_{merge}	Resolution Range
Crystal 1 (a-axis)	27819	0.107	50 - 2.2 Å
Crystal 2 (b-axis)	27346	0.072	50 - 2.2 Å
Combined data (crystals 1 and 2)	38482	0.136	50 - 2.2 Å

$R_{merge} = \Sigma|I-\bar{I}|/\Sigma\bar{I}$

The final scaled and merged data set, using data from both crystals, was derived from 163844 measurements, with a multiplicity of 4.3. This gave 38482 unique reflections ($R_{merge} = 0.136$) for data $\geq 1\ \sigma_I$ in the 50-2.2 Å resolution range. Data were 88% complete overall, with a 62.5% completeness in the 2.28-2.20 Å shell.

Fig. III.1 shows the percentage of the total number of unique reflections measured with $I>1\sigma_I$, $2\sigma_I$ and $3\sigma_I$ as a function of resolution. There is a sharp decrease in the percentage of observed reflections (with $I>2\sigma_I$) at about 2.9 Å and the maximum effective resolution of the data (that used for restrained least squares refinement) can be taken as 2.4 Å, where about 50% of the reflections are observed with $I>2\sigma_I$. Statistics on the final data set for $Fe_2(C_2O_4)_2Lf$ are given in Table III.4.

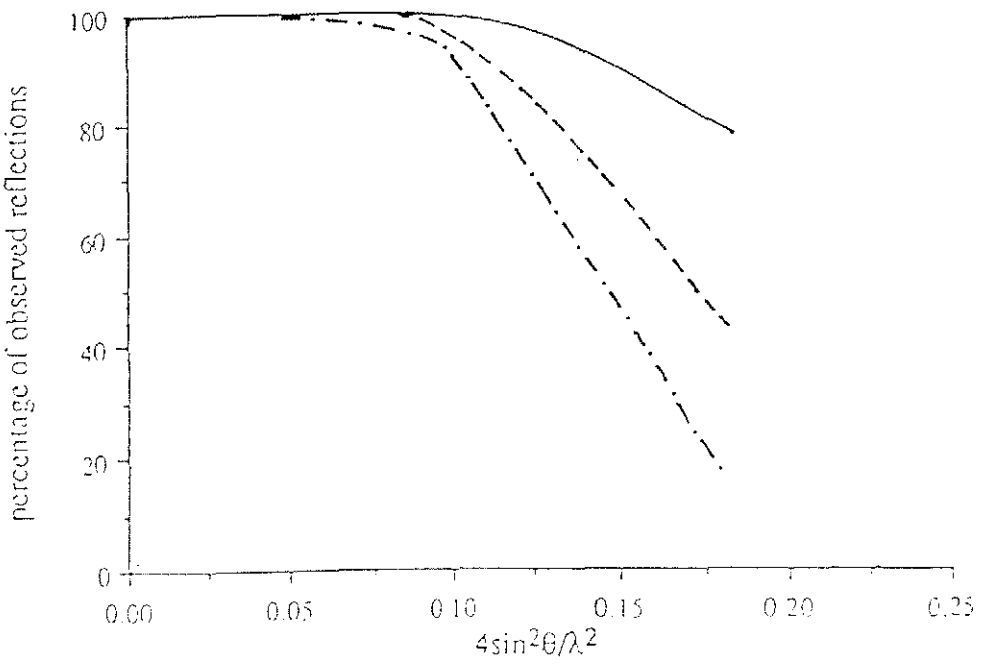


Fig. III.1 The percentage of the total number of unique reflections in each resolution shell in the final data set for $Fe_2(C_2O_4)_2Lf$, with $I>1\sigma_I$ (——), $I>2\sigma_I$ (----) and $I>3\sigma_I$ (— - — -) as a function of resolution

Table III.4 Statistics on the final data set for $Fe_2(C_2O_4)_2Lf$

Maximum measured resolution, d_{min} (Å)[a]	2.2
- number of expected reflections to d_{min}	43691
- overall completeness of data to d_{min}	73%
Maximum resolution used in refinement, d_{ref} (Å)[a]	2.4
- number of expected reflections to d_{ref}	33435
- observed reflections to d_{ref}	31758
- completeness of data to d_{ref}[a]	94.9%
Highest resolution shell (Å)	2.54-2.4
- unique reflections in this shell	4809
- completeness in this shell	83.3%

[a]The resolution of the reflection(s) measured at the highest value of θ

III.2 STRUCTURE DETERMINATION

Since the crystals of $Fe_2(C_2O_4)_2$Lf are isomorphous with those of $Fe_2(CO_3)_2$Lf (see Table III.5), the refined coordinates of $Fe_2(CO_3)_2$Lf determined at 2.2 Å (Haridas *et al*, in press) could be used as an initial model. As the major differences between the two structures were likely to be around the anion and metal binding sites, the eight sidechains that bind directly to the two metal ions, together with the anion-binding arginine (residues 121 and 465) and threonine (residues 117 and 461) sidechains were truncated to their C_β atoms, the two ferric ions, two anions and all solvent molecules were removed, and the individual temperature factors were all set to a uniform 25 $Å^2$. This model was used to initiate refinement.

Table III.5 Comparison of the unit cell dimensions of $Fe_2(C_2O_4)_2$Lf and $Fe_2(CO_3)_2$Lf crystals

Complex	Axial lengths (Å)	Angles	Space-group
$Fe_2(C_2O_4)_2$Lf	155.46 96.92 55.85	$\alpha = \beta = \gamma = 90°$	$P2_12_12_1$
$Fe_2(CO_3)_2$Lf	156.26 97.40 55.85	$\alpha = \beta = \gamma = 90°$	$P2_12_12_1$

III.2.1 The Refinement Problem

The crystallographic refinement of a protein structure poses particular problems because of the limited resolution of the data, and the small number of observations (X-ray data) compared with the number of parameters to be refined. For $Fe_2(C_2O_4)_2$Lf 32,000 reflections were recorded to 2.4 Å and approximately 22,000 parameters (x, y, z, B for ~5300 atoms) were to be refined, a ratio of only 1.4:1. In order to improve this ratio, known structural parameters such as ideal values of bond lengths, bond angles, torsion angles, planarities of groups, etc are included in the refinement as observations. This approach to refinement, in which the protein geometry is restrained, and atoms are not individually free, is referred to as restrained least squares refinement. The program used for refinement of $Fe_2(C_2O_4)_2$Lf was PROLSQ (Hendrickson and Konnert, 1980).

III.2.2 Refinement Strategy

The general philosophy of refinement was as follows. The structure was refined in a series of rounds, each comprising 10-30 cycles. During each round the restraints on the geometry were relaxed (to allow atoms to move) and then retightened to bring the structure back to proper geometry. (Note, however, that the *rms* deviation of bonds from standard values was not allowed to become too high, typically not allowed to exceed 0.05 Å.) At the end of each round, 2*Fo*-*Fc* and *Fo*-*Fc* electron density maps were calculated, and the fit of the model was examined on an Evans and Sutherland PS300 interactive graphics system. Where corrections could be made, the model was rebuilt manually using the

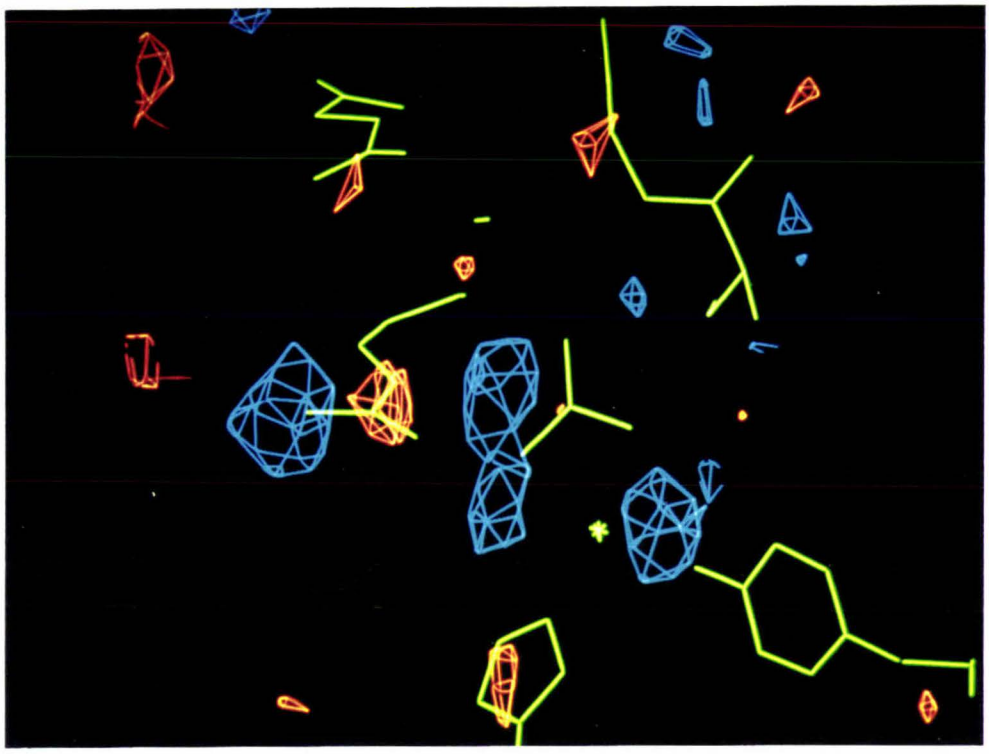
program FRODO (Jones, 1978). Regions that were in doubt (eg poorly defined sidechains) were omitted from the model, to be re-included later. Regions with high B values were several times systematically removed from the model and were then rebuilt into subsequent maps. Solvent molecules were not included until relatively late in refinement. When they were included, they were only added to the model if (1) the electron density was >2σ in the *Fo-Fc* map, and greater than 1σ in the 2*Fo-Fc* map, and (2) the position was structurally reasonable, ie within hydrogen bonding distance of other potential hydrogen bond partners (protein atoms or other solvent molecules). Several times during the course of refinement all water molecules with a B value greater than 70 $Å^2$ were removed, and the water model re-evaluated after another series of refinement cycles.

Towards the end of the refinement the program PROCHECK (Laskowski *et al*, 1993), which assesses the geometrical correctness of the model, both overall, and residue by residue, was used to help identify aspects of the structure that should be examined more closely.

III.2.3 **Structure Solution and Refinement**

A difference Fourier synthesis was calculated using the observed structure amplitudes for $Fe_2(C_2O_4)_2$Lf and $Fe_2(CO_3)_2$Lf; ie the coefficients used were $||F_{Fe_2(C_2O_4)_2Lf}|obs-|F_{Fe_2(CO_3)_2Lf}|(obs)|$, with phases α calc ($Fe_2(CO_3)_2$Lf. The map was displayed on an Evans and Sutherland PS300 graphics station using FRODO (Jones, 1978). The largest positive and negative peaks were in the vicinity of the N-lobe and C-lobe metal- and anion-binding sites, with a positive peak adjacent to the anion site in each lobe, implying that a larger anion was now bound; the extra density arising from the 2 extra atoms of the oxalate ion (Fig. III.2). A pattern of positive and negative peaks similar to that seen in a $Cu_2(C_2O_4)(CO_3)$Lf-$Fe_2(CO_3)_2$Lf difference Fourier synthesis (Smith *et al*, 1994) indicated a movement of Arg 465 and an accompanying movement of Tyr 398 away from the C-lobe site. A similar pattern of positive and negative peaks was seen in the N-lobe, but here the pattern was more complex and difficult to interpret (Fig. III.2).

(a)



(b)

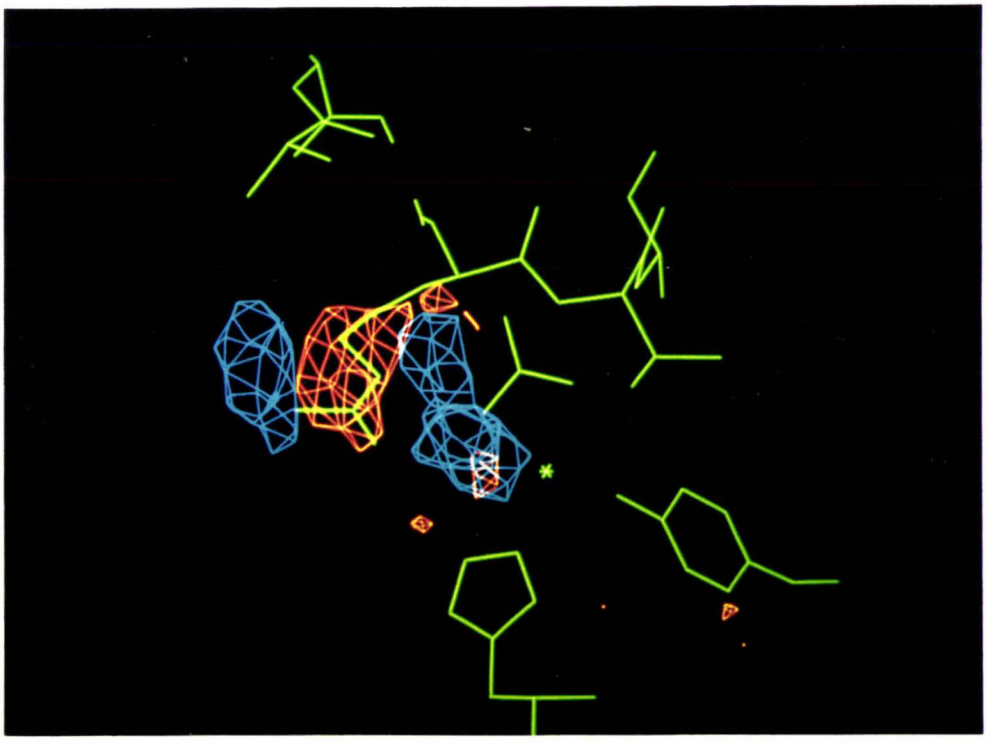


Fig. III.2 $Fe_2(C_2O4)_2Lf$ - $Fe_2(CO_3)_2Lf$ difference electron density map (calculated with coefficients $\Delta F = F_{(Fe_2(C_2O_4)_2Lf)} - F_{(Fe_2(CO_3)_2Lf)}$ in (a) the N-lobe and (b) the C-lobe sites. Negative density is shown in red and positive in blue. The Fe_2Lf model is superimposed in yellow.

(a) *The course of refinement*

The refinement of $Fe_2(C_2O_4)_2$Lf was carried out using PROLSQ (Hendrickson and Konnert, 1980). The starting model was as described in Section III.2. After the first phase of refinement (22 cycles), the metal ions and protein ligands were added to the appropriate 2*Fo*-*Fc* density, but the anion sites were left vacant, with the anion-binding arginine (121, 465) and threonine (117, 461) residues remaining truncated at C_β. Following a further 22 cycles of least squares refinement, the two oxalate ions, arginines 121 and 465, and threonines 117 and 461, were added to the appropriate 2*Fo*-*Fc* density.

At this stage it was noted that the electron density showed the oxalate ion fitted as a symmetrical 1,2-bidentate ligand to the iron in the C-lobe. This was not the case in the N-lobe, however. Here the electron density indicated an asymmetric 1,2-bidentate binding of oxalate to the metal. A cautious approach was adopted. The oxalate ion in the N-lobe was fitted in a very conservative manner, making it as symmetric to the iron as possible within the confines of the electron density. The following refinement cycles caused the oxalate ion to move back to being quite asymmetric ($O_{1(ox)}$-Fe = 1.9 Å, $O_{2(ox)}$-Fe = 2.4 Å). Several times during the course of the refinement the positions of both oxalate ions were checked by excluding them from the final 5-8 cycles of least squares refinement in a particular round, and calculating 2*Fo*-*Fc* and *Fo*-*Fc* maps to establish their true positions. Further checks on the position of the N-lobe oxalate ion were made either by manually placing the oxalate to have more symmetrical binding to the iron, or moving the iron atom to have symmetrical binding to the oxalate, but in all cases subsequent refinement cycles moved it back to being asymmetric (Fig. III.14).

Solvent molecules, all treated as water, were fitted into *Fo*-*Fc* density (contoured at 2 times the rms deviation of the map) following the criteria described above. To hasten the process of water addition, the refined $Fe_2(CO_3)_2$Lf structure was superimposed on the $Fe_2(C_2O_4)_2$Lf structure and the positions of the waters in the $Fe_2(CO_3)_2$Lf structure checked against the criteria listed above. The positions of about 95 water molecules were found in this way.

Major rebuilds of the structure were carried out during rounds 2, 3, 4, 5, 6 and 8 of refinement (see Table III.6). 'Omit maps' were especially useful in helping to establish the correct conformation or position of poorly defined sidechains and loops.

Individual B factors were refined after 91 cycles of xyz refinement (R = 22.7%). Restraints were placed on the B factors to prevent excessive shifts during refinement and to keep the B values of adjacent atoms within 5 $Å^2$ of each other. New water molecules were given a temperature factor of 25 $Å^2$ and treated in a similar manner to the protein atoms. Monitoring after refinement showed those with excessive increases in B values, and these were checked on the graphics system and removed if at all doubtful.

The progress of the refinement is summarised in Table III.6. Although the R factor did not reduce much in the last few stages, the quality of the structure, as judged by the

geometric and structural criteria (hydrogen bonding, bond lengths and angles, Ramachandran plot (Ramakrishnan and Ramachandran, 1965), distribution of Chi-1 vs Chi-2 angles, etc) was significantly improved. The program PROCHECK (Laskowski *et al*, 1993) was especially useful in helping to evaluate the model in terms of these criteria.

In the final stages, although it was possible to place some of the more poorly defined parts of the structure, it was also decided to remove the first three residues of the N-terminus because of their very high B factors and their poorly defined electron density.

Table III.6 Course of Refinement of $Fe_2(C_2O_4)_2Lf$

Round	Number of cycles	R Factor %	rmsΔ bonds	Comments
1	22	34.0 → 30.0	0.035	5-2.5 Å data, overall B = 25 $Å^2$ 5323 protein atoms
2	22	30.9 → 28.8	0.027	Added $2Fe^{3+}$, all ligand sidechains. Rebuilt 32 sidechains 5323 protein atoms, $2Fe^{3+}$ Refined with 5-2.5 Å data, xyz refinement
3	22	28.5 → 25.7	0.025	Added 2 oxalate ions, Arg 121, Arg 465, Thr 117, Thr 461 Rebuilt 23 sidechains. Fixed bad contacts. 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ Refined with 5-2.5 Å data, xyz refinement only.
4	25	26.0 → 22.1	0.021	Added 120 water molecules. Rebuilt loop 418-424 and 15 sidechains 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ Refined with 5-2.4 Å data, xyz refinement.
5	40	22.7 → 22.1	0.024	Added 40 water molecules, removed 1 water. Rebuilt N-terminus, 10 sidechains, checked oxalate positions with omit maps. 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ xyz B refinement with 5-2.4 Å data.
6	25	22.7 → 22.4	0.018	Added 33 water molecules, removed 6 water molecules. Omit map used to help rebuild loop 418-424. Rebuilt 9 sidechains. 5323 protein atoms, $2Fe^{3+}$, $2C^2O_4^{2-}$ xyz B refinement with 8-2.4 Å data.

cont. ...

Table III.6 Course of Refinement of $Fe_2(C_2O_4)_2$Lf (cont.)

Round	Number of cycles	R Factor %	rmsΔ bonds	Comments
7	40	22.6 → 19.23	0.019	Added 3 water molecules removed 16 water molecules. Checked oxalate position with omit maps adjusted sidechains. 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ xyz B refinement with 8-2.4 Å data.
8	30	20.14 → 19.78	0.019	Omit map used to check residue 512. Deleted 46 waters (B>70 $Å^2$ or too close contacts). Adjusted sidechains of loop 418-424. 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ xyz B refinement with 8-2.4 Å data.
9	25	19.92 → 19.24	0.014	Replaced 2 waters, deleted 8 waters. Omit map used to help rebuild Glu 221. Adjusted sidechains. 5323 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$. xyz B refinement with 8-2.4 Å data.
10	25	20.9 → 20.05	0.020	Ran PROCHECK on model from round 9. Adjusted sidechains, emphasis on Chi 1, Chi 2 angles and chirality. Checked oxalate in C lobe. Removed residues 1-3 of N-terminus 5275 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ xyz B refinement with 8-2.4 Å data.
11	30	19.91 → 19.75	0.018	Ran PROCHECK on model from round 10. Adjusted Arg 465 slightly. 5275 protein atoms, $2Fe^{3+}$, $2C_2O_4^{2-}$ xyz B refinement with 8-2.4 Å data.
12	12	19.74 → 19.68		Ran PROCHECK.

III.2.4 **The Quality of the Final Model**

The final $Fe_2(C_2O_4)_2$Lf model comprises the complete molecule, excluding the N-terminal residues 1-3, ie residues 4-691, 5275 protein atoms, 2 ferric ions, 2 oxalate ions and 121 solvent molecules. The stereochemistry of the model, summarised in Table III.7, is very close to ideal; the *rms* deviations from standard values of bond lengths and (1-3) angle distances are 0.018 Å and 0.059 Å respectively. The R factor is 19.6% for data with $I>1.0\sigma_I$ between 8 and 2.4 Å resolution.

Table III.7 Refinement statistics for $Fe_2(C_2O_4)_2Lf$	
Resolution limits (Å)	8.0 - 2.4
Initial R factor (%)	34.0
Final R factor (%)	19.6
Number of reflections used	31758
Number of protein atoms	5275
Number of solvent molecules	121
Other ions	$2Fe^{3+}$, $2C_2O_4^{2-}$
Number of carbohydrate residues	-
Number of refinement cycles	318
rms deviations from ideality	
Bond lengths (Å)	0.018
Angle distances (Å)	0.059
Planarity (Å)	0.011
Chiral volume ($Å^3$)	0.248
Non-bonded contacts Å	0.208
Average B value (main chain atoms)	45.3
Average B value (sidechain atoms and waters)	49.4

(a) *Agreement with X-ray data*

The average maximum coordinate error in the $Fe_2(C_2O_4)_2Lf$ model was estimated from the variation in the R factor with resolution as described by Luzzati (1952) (Figure III.3). Excluding points at the lowest resolution, where the higher values can be attributed to a failure to fully account for the solvent structure, and at the highest resolution, where a large proportion of the data were unobserved (see Fig. III.1) the maximum average error in the structure is ~0.3 Å. The theoretical variation of R factor with resolution was calculated for non-centric data.

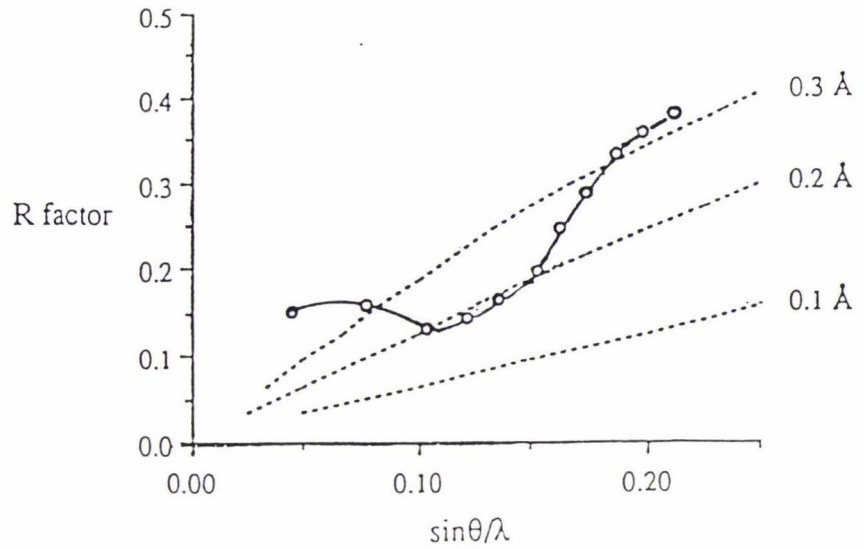


Figure III.3 A Luzzati plot of the crystallographic R factor (Luzzati, 1952) as a function of resolution for $Fe_2(C_2O_4)_2$Lf. The theoretical variation of R with resolution for *rms* errors of 0.1, 0.2, 0.3 Å in atomic positions, is shown for non-centric data.

The value obtained from the graph represents the maximum average error; the error would be somewhat less than this in the well-ordered internal regions (the area around the binding sites and secondary structure elements for example) but substantially more in the poorly-defined parts of the molecule such as some of the external loops connecting the β strands and helices. In spite of much effort to determine the correct conformation of these regions with the use of omit maps, some remained ill defined in comparison to the rest of the structure (see Table III.8).

(b) *Agreement with expected geometry*

The correctness of the model was also assessed by analysing the distribution of dihedral angles. A Ramachandran plot (Ramakrishnan and Ramachandran, 1965) of the main chain torsion angles (φ and ψ) (Fig. III.4) shows that virtually all lie within the allowed and partially allowed regions of conformational space, as defined by Laskowski *et al* (1993) in the program PROCHECK. The 'most favoured' regions account for 81.9% of non-glycine residues, and a further 16.6% of residues occupy the 'additionally allowed' regions. Only 2 residues, Leu 299 and Leu 642 are in normally disallowed regions; these are both central residues in γ-turns which are in equivalent positions in each lobe. These types of turns are characterised by mainchain conformational angles around φ = 70°, ψ = -50° (Baker and Hubbard, 1984). In the case of Leu 299 and Leu 642 these torsion angles are (68°, -50°) and (58°, -40°) respectively, similar to the characteristic value for classic γ-turns. The residues on either side, 298 and 300, 641 and 643, interact

via 1-3 hydrogen bonds (Baker and Hubbard, 1984). These two γ-turns with their highly conserved Leu-Leu-Phe sequence within the transferrin family are found in part of the wall of the binding cleft in each lobe, and seem to be a common feature of transferrin structure (Smith *et al*, 1994).

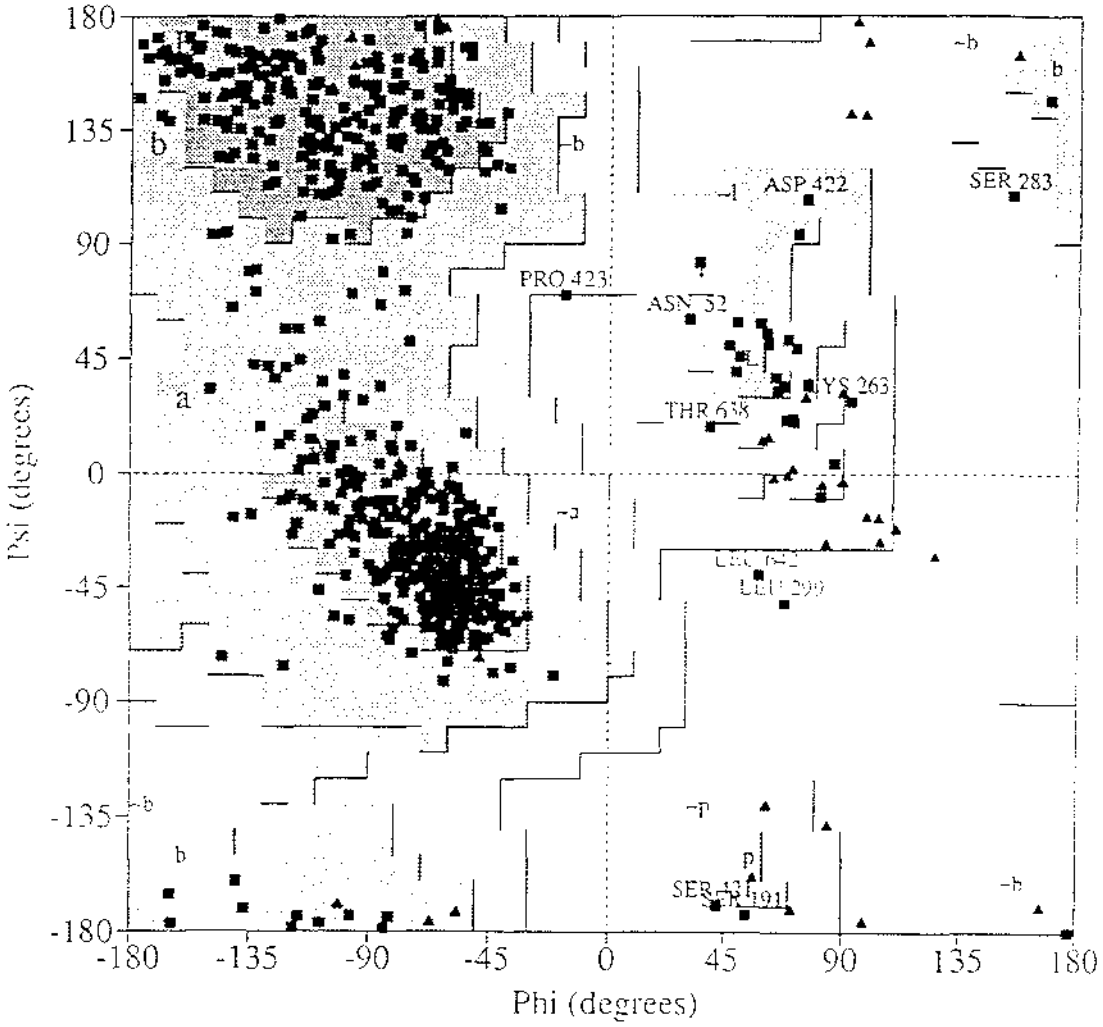
Ser 191 occupies a 'generously allowed' region, with mainchain (ϕ, ψ) torsion angles of 53°, -174°, designated as an ε conformation (Sibanda *et al*, 1989). This residue is in a well-ordered part of the structure, at the N-terminus of helix 7, and makes an important contribution to the N-lobe binding site, hydrogen bonding with the oxalate-binding Arg 121. The same ε configuration of Ser 191 is also seen in the lactoferrin structures with carbonate as the anion; $Fe_2(CO_3)_2$Lf (Anderson *et al*, 1989) and $Cu_2(CO_3)_2$Lf (Smith *et al*, 1994). The equivalent residue in the C-lobe is Gly 527 which has the same configuration. In all other transferrins whose sequences are known, there are glycine residues in this position in both lobes.

Further assessments of the correctness of the model are given by PROCHECK (Laskowski *et al*, 1993). These include comparisons of mainchain parameters, such as the quality of the Ramachandran plot, α-carbon tetrahedral distortion and hydrogen bond energies, with the average values of structures at the same resolution. These comparisons show that the structure is better than average in all respects except the standard deviations of the α-carbon tetrahedral distortions and hydrogen bond energies, although even here the deviation is still within acceptable limits (Fig. III.5). Sidechain parameters (standard deviations of Chi-1 gauche minus, Chi-1 trans, Chi-1 gauche plus, Chi-1 pooled standard deviations, and Chi-2 trans angle standard deviations) are all better than average (Fig. III.6). A plot of Chi-1 vs Chi-2 dihedral angles (Fig. III.7) should have most residues clustering around the crosses which represent one standard deviation for the gauche minus, trans, and gauche plus regions for Chi-1 and Chi-2 dihedral angles. It can be seen in Fig. III.7 that only 36 residues are more than 2.5 standard deviations from ideal, also indicating the overall good quality of the structure. A plot of residue properties (abs mean deviation of Chi-1, omega, and zeta), residue by residue is given in Appendix 1.

Fig. III.4

Ramachandran Plot

Diferricdioxalatolactoferrin



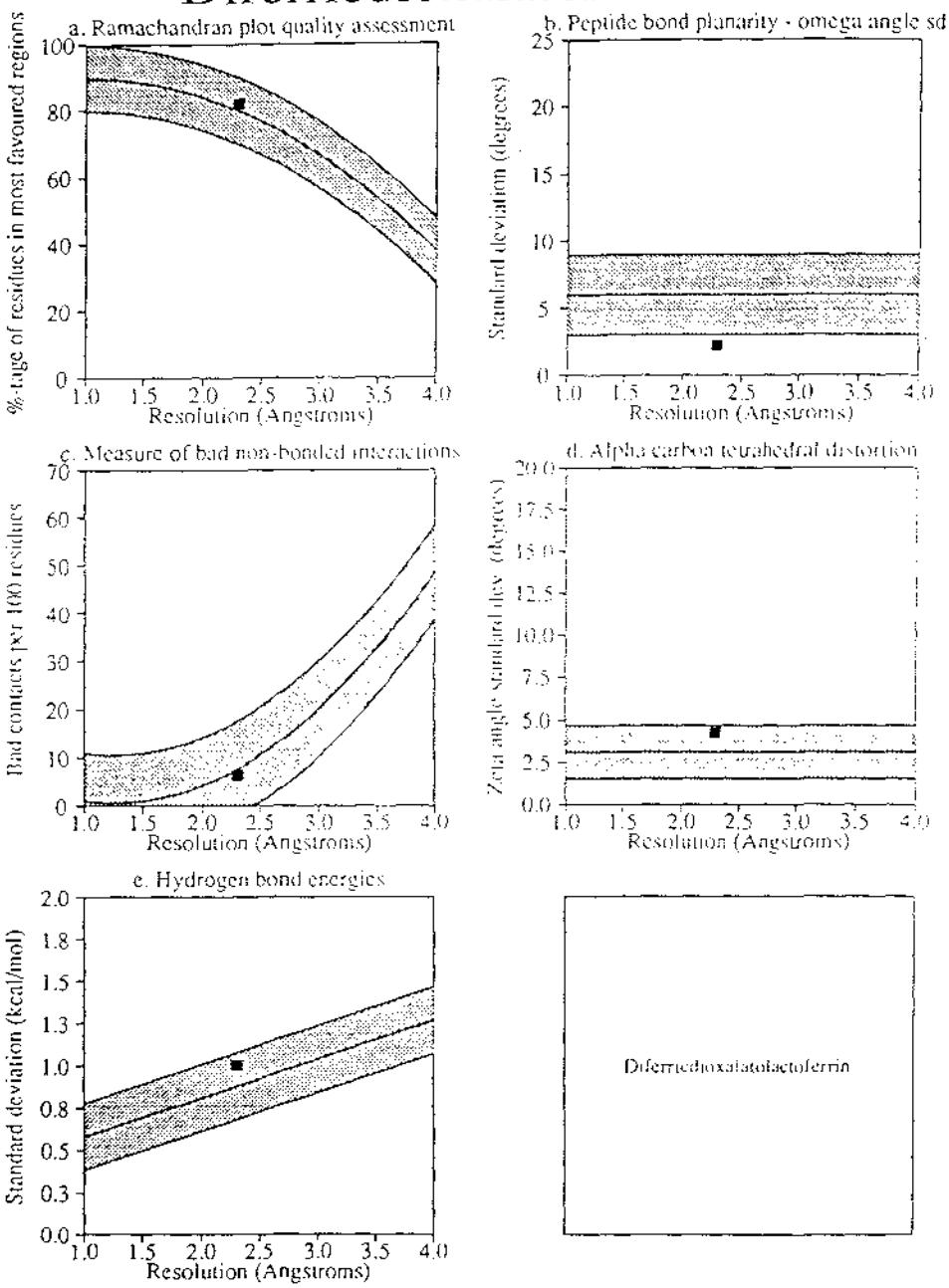
Plot statistics

Residues in most favoured regions [A,B,L]	490	81.9%
Residues in additional allowed regions [a,b,l,p]	99	16.6%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.2%
Residues in disallowed regions	2	0.3%
Number of non-glycine and non-proline residues	598	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	53	
Number of proline residues	35	
Total number of residues	688	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

Fig. III.5 Main-chain parameters

Diferricdioxalatolactoferrin

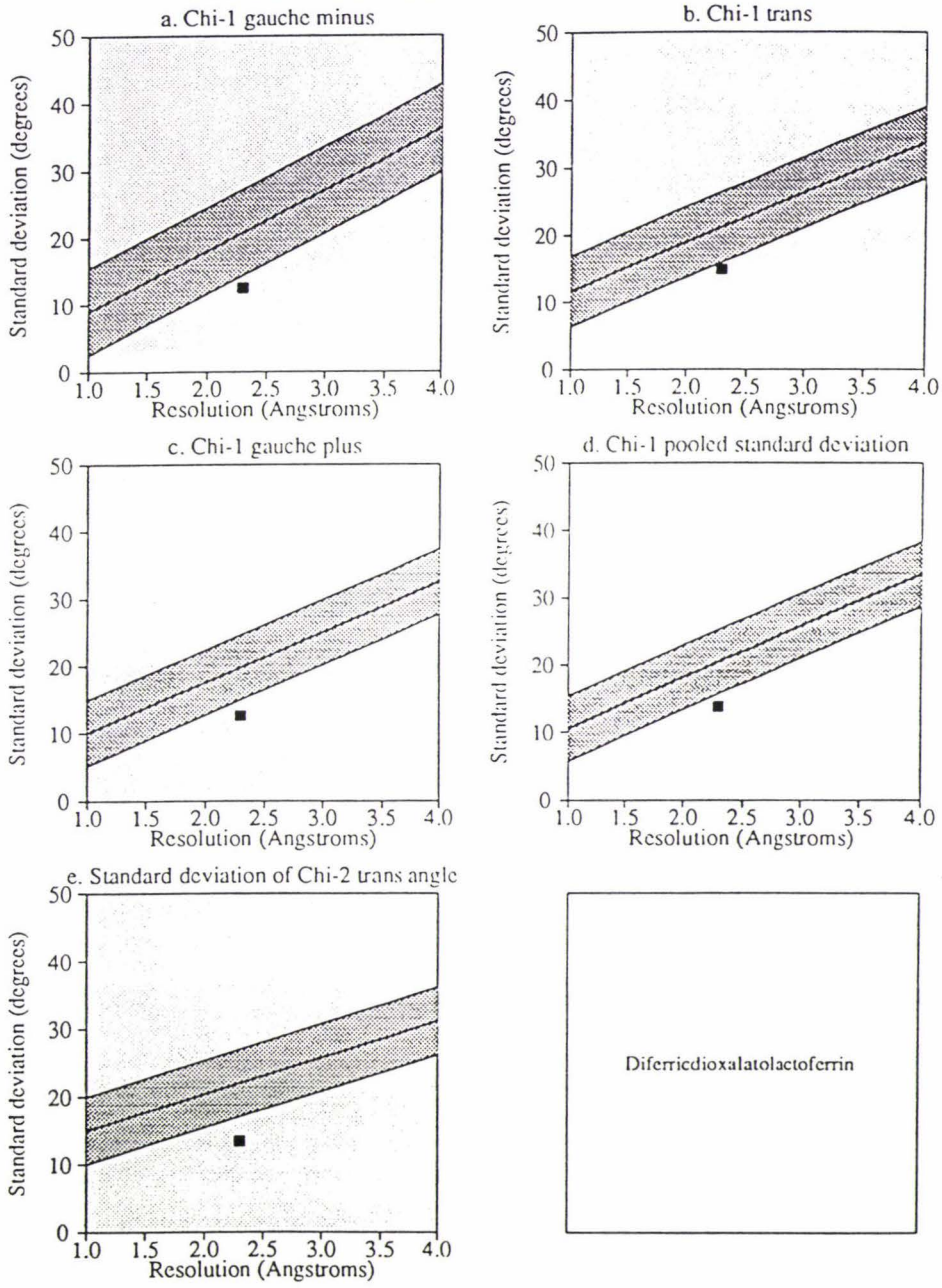


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	598	81.9	79.8	10.0	0.2	Better
b. Omega angle st dev	684	2.2	6.0	3.0	-1.3	BETTER
c. Bad contacts / 100 residues	44	6.4	7.6	10.0	-0.1	Better
d. Zeta angle st dev	635	4.3	3.1	1.6	0.7	Worse
e. H-bond energy st dev	412	1.0	0.9	0.2	0.6	Worse

Fig. III.6 Side-chain parameters

Diferricdioxalatolactoferrin

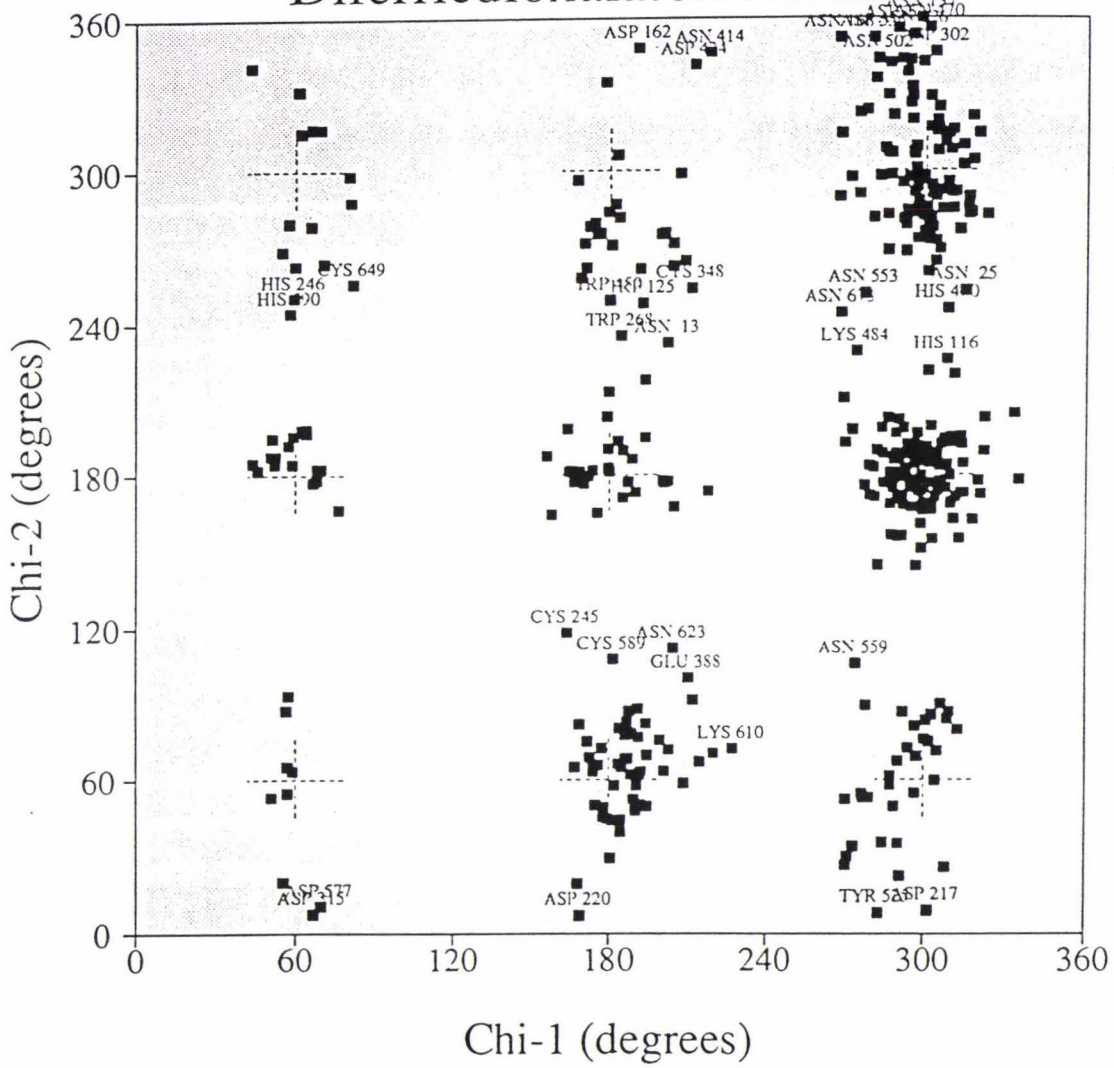


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	87	12.5	20.9	6.5	-1.3	BETTER
b. Chi-1 trans st dev	155	14.9	21.2	5.3	-1.2	BETTER
c. Chi-1 gauche plus st dev	291	12.6	19.8	4.9	-1.5	BETTER
d. Chi-1 pooled st dev	533	13.8	20.5	4.8	-1.4	BETTER
e. Chi-2 trans st dev	155	13.4	22.0	5.0	-1.7	BETTER

Fig. III.7 Plot of Chi-1 vs Chi-2

Diferricdioxalatolactoferrin



Total number of residues = 688

Number of residues plotted = 404

Number of labelled residues = 36

The dashed crosses represent the gauche minus, trans, and gauche plus regions for the Chi-1 and Chi-2 dihedral angles. The width of each cross is approximately one standard deviation. The points should therefore cluster around these crosses. Points more than 2.5 standard deviations from the ideal are labelled

(c) *Temperature factors*

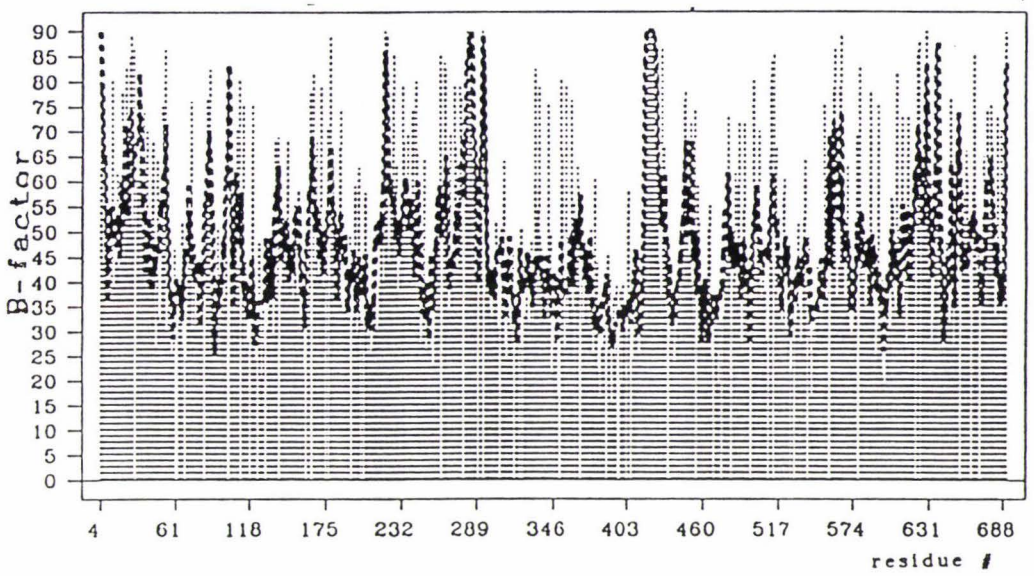
Fig. III.8(a) shows average mainchain and sidechain temperature factors as a function of residue number. The average mainchain B value of 45 $Å^2$ reflects the weakness of the high resolution data and the fact that all data were used in the least squares refinement. (Only about 52% of reflections have $I>2\sigma_I$ at 2.4 Å.) The high B values are also consistent with the overall B estimated from a Wilson plot (Wilson, 1942) which gives an average B of about 50 $Å^2$. It is clear that the residues with low B values correspond to α helices, β strands and residues involved in the metal- and anion-binding sites. These regions are, in general, internal and stabilised by a large number of hydrogen bonding and other interactions with other protein atoms and internal water molecules.

(d) *Fit to electron density*

The fit of the model to the electron density can be quantified by calculating correlation coefficients which give a residue by residue measure of the agreement between the model electron density and the final 2*Fo*-*Fc* electron density map. Real-space correlation coefficients for the individual residues, calculated using the real-space fit routine in the model building and refinement program O (Jones *et al*, 1991), show that in general the correlation of the model with the data is high (Fig. III.8(b)) with most of the residues having correlation coefficients greater than 0.8. The regions which have the poorest correlation(<0.7) include the external loops 220-222, 240-243, 280-285 and 293-294 in the N-lobe, the loops 417-424, 623-628 in the C-lobe, and the C-terminal residues 687-691. Well-defined regions of the model, including the metal- and anion-binding ligands, and those residues near the binding site all show good correlation between observed and calculated density (correlation coefficients between 0.85 and 0.90). The pattern is similar to that seen when temperature factors are plotted as a function of residue number (Fig. III.8(a)).

Most of the protein structure is represented by well-defined electron density. Examples of typical electron density in the well-defined regions are shown in Fig. III.9, while Fig. III.10 shows the density for several poorly-defined regions. The least satisfactory parts of the structure are also listed in Table III.8.

(a)



(b)

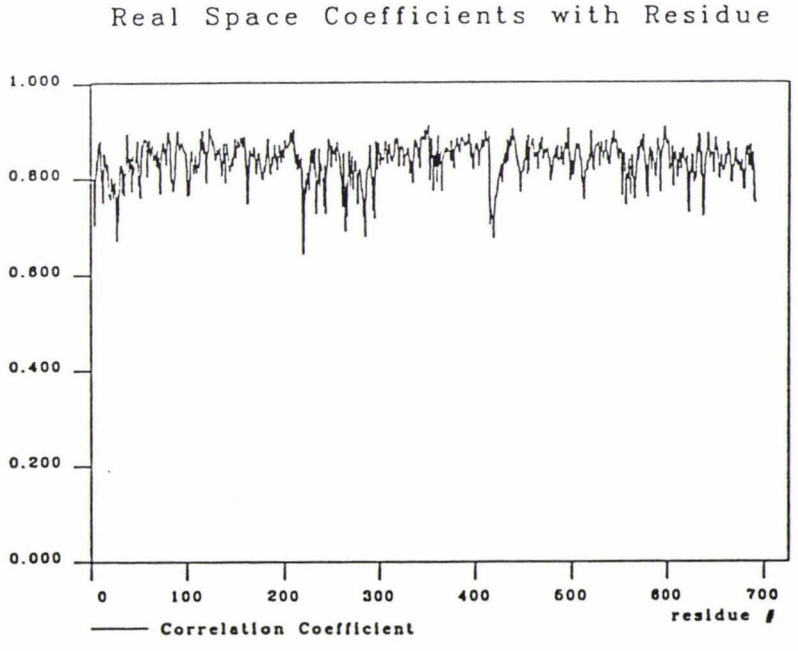
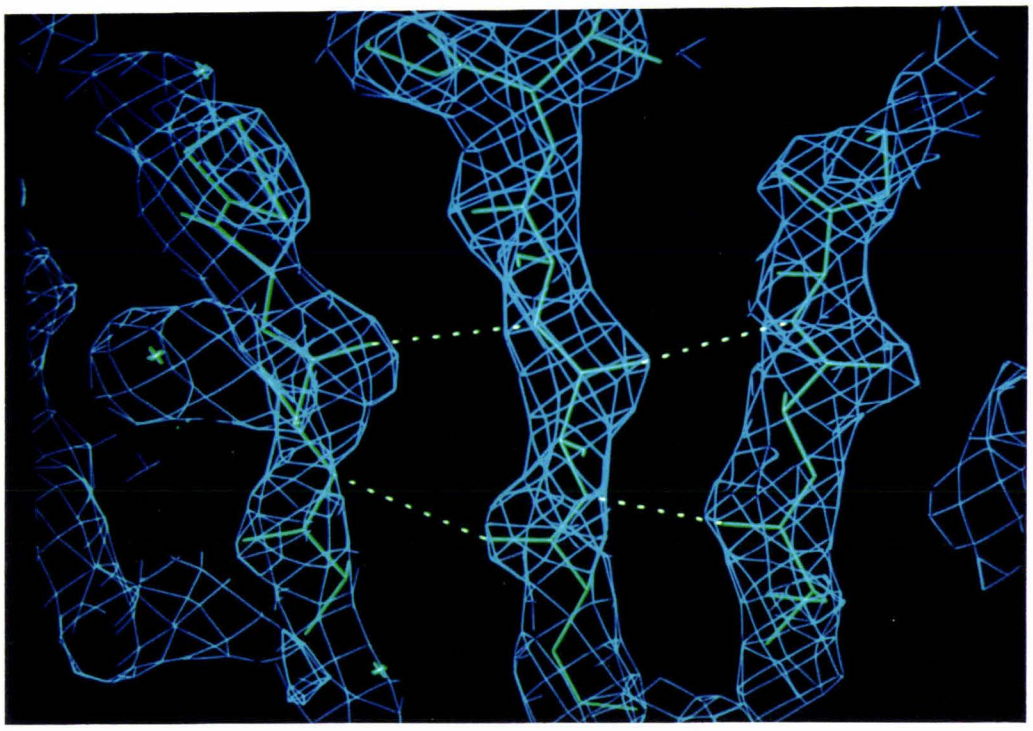


Fig. III.8 (a) Plot of mainchain temperature factors (B values) (bold) and sidechain temperature factors for each residue as a function of residue number,.(b) Plot of the real space correlation coefficient for the fit of the refined model of $Fe_2(C_2O_4)_2$Lf to the final 2*Fo*-*Fc* electron density, as a function of residue number.

(a)



(b)

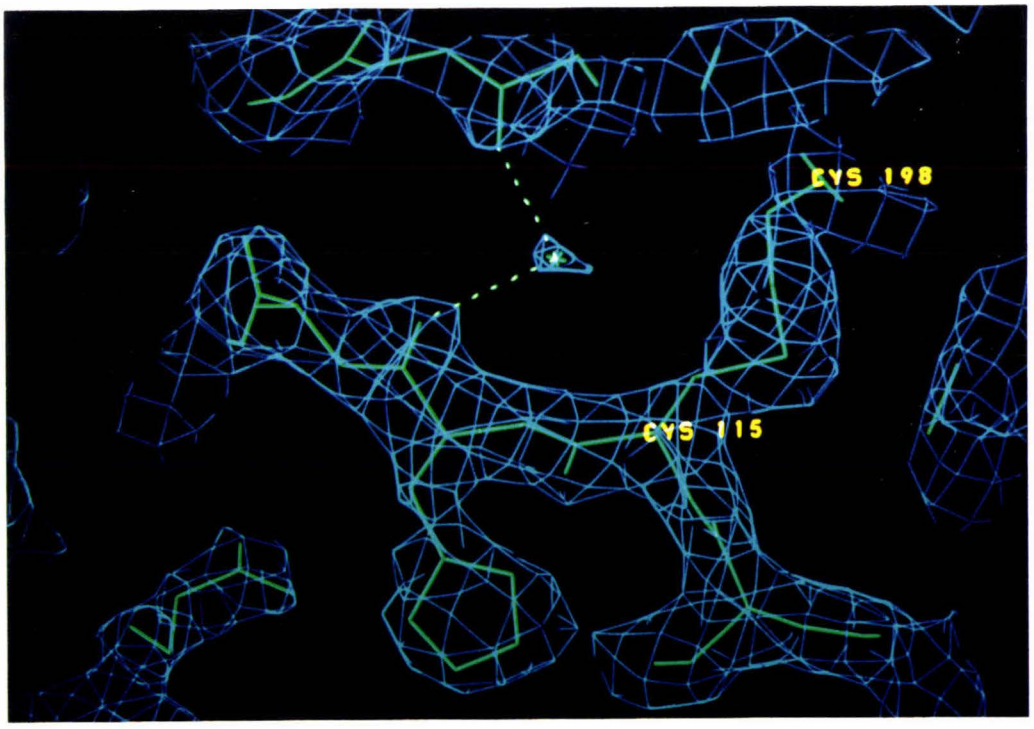


Fig. III.9 Examples of well-defined electron density in the final $2Fo$-Fc map for $Fe_2(C_2O_4)_2Lf$, contoured at 1.5σ. (a) Part of the N2 domain β-sheet, (b) Ser 114, His 116 and the disulphide bridge 115-198.

(a)

(b)

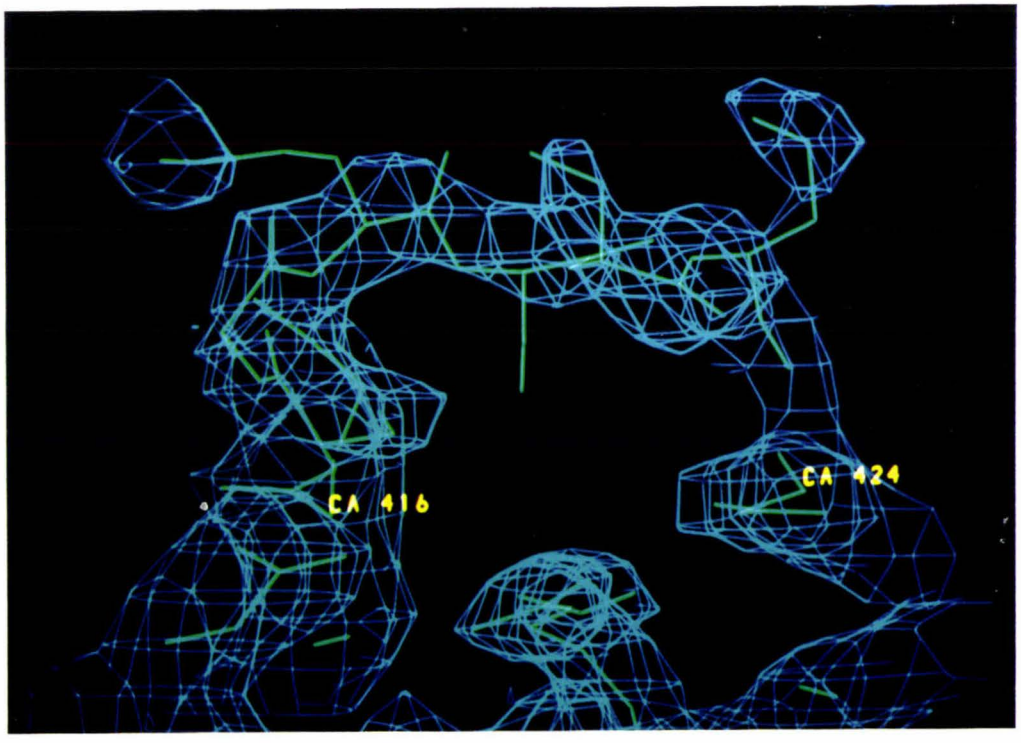


Fig. III.10 Examples of the most poorly-defined electron density in the final 2*F*o-*F*c map for $Fe_2(C_2O_4)_2$Lf, contoured at 0.7σ. (a) Residues 292-295, (b) Part of the loop 417-423. Although it was difficult to fit the model precisely to the electron density in these regions, the course of the polypeptide chain was unambiguous.

Most of the protein structure is represented by well-defined electron density; the least satisfactory regions of the structure are listed in Table III.8.

Table III.8 Poorly defined regions of the structure

Residues	Average B	Comment
1-3	>98 Å^2	Not modelled. N-terminus projects into solvent, electron density poorly defined. Could be both disordered and/or heterogeneous
220-222	80 Å^2	External turn, probably disordered.
293-294	75 Å^2	Part of a projecting β turn, probably disordered.
417-423	85 Å^2	Large flexible surface loop. No interactions with the rest of the molecule.
689-691	61 Å^2	C-terminus of molecule. Probably disordered.
Carbohydrate attached to Asn 137 and Asn 478		Not modelled. Electron density poorly defined and indicated little interaction with protein structure. Probably is both flexible and chemically heterogeneous.

In these poorly defined regions there is generally sufficient density to follow the polypeptide chain, but this density is rather weak and fragmented, and the atoms have high B values, indicative of disorder. Towards the end of refinement the first 3 residues of the N-terminus were removed because of very high B values and weak electron density. Sequencing has shown that there can be a chemical heterogeneity at the N-terminus (Stowell *et al*, 1991) and this, combined with the positional disorder, made accurate modelling of the N-terminus very difficult. Residues Ser 421, Asp 422, Pro 423, all of which occupy regions that are either 'generously allowed' or 'additionally allowed' in the Ramachandran plot, form part of a large extended surface loop (residues 417-423) with high B values. This makes no interactions with the rest of the protein and has little internal structure. Although there was electron density for the whole course of the loop, it was poorly defined and the difficulty encountered when trying to build into this density probably reflects the flexibility or disorder of the loop.

Two large clouds of electron density, corresponding to the position of the two glycan chains of lactoferrin, could be seen extending from the sites of their attachment; Asn 137 and Asn 478. Unfortunately although the boundaries were relatively well defined, the resolution of the electron density within each cloud was insufficient to be able

to place any carbohydrate moieties with any confidence. The two carbohydrate chains attached to human lactoferrin are known to be heterogeneous, with at least five chemical structures having been identified (Spik *et al*, 1982), and clearly there is spatial disorder or flexibility. The electron density indicates few interactions with the protein which might stabilise the carbohydrate chains.

III.3 THREE-DIMENSIONAL STRUCTURE

The general organisation of $Fe_2(C_2O_4)_2Lf$ is the same as that described previously for $Fe_2(CO_3)_2Lf$ (Anderson *et al*, 1989; Baker *et al*, 1987). In both structures the polypeptide chain is folded into two lobes (N-lobe and C-lobe) and each lobe is further subdivided into 2 similarly sized domains (N1, N2 and C1, C2). There is a 42% sequence identity between the two lobes of the molecule which is reflected in the similar fold of the polypeptide chain. A rotation of ~180° of the C-lobe, followed by a translation of ~25 Å allows the C-lobe to be superimposed on the N-lobe with 85% of the Cα atoms matching within an rms deviation of ~1.0 Å. (Superpositions were performed using the program LSQKAB (CCP4 program suite, 1994).) The structural agreement is even closer if the N1 and C1 domains and N2 and C2 domains are superimposed separately. The N-lobe is slightly more open than the C-lobe, so that if the N2 and C2 domains are superimposed the N1 and C1 domains are slightly out of register. For the individual domains, the *rms* deviations after superposition are ~0.8 Å (~133 Cα atoms).

Each lobe contains a single metal- and anion-binding site located in the interdomain cleft with the metal ion taking ligands from both domains and from two of the interdomain crossover strands. The anion, on the other hand, forms a bridge between the metal and just one domain of the protein, being liganded to a short stretch of the polypeptide chain of domain 2 (N2 or C2). A schematic diagram of the folding pattern and a stereo view of the Cα trace of $Fe_2(C_2O_4)_2Lf$ are shown in Figs. III.11 and III.12, while a ribbon diagram (Fig. III.13) shows the division into domains.

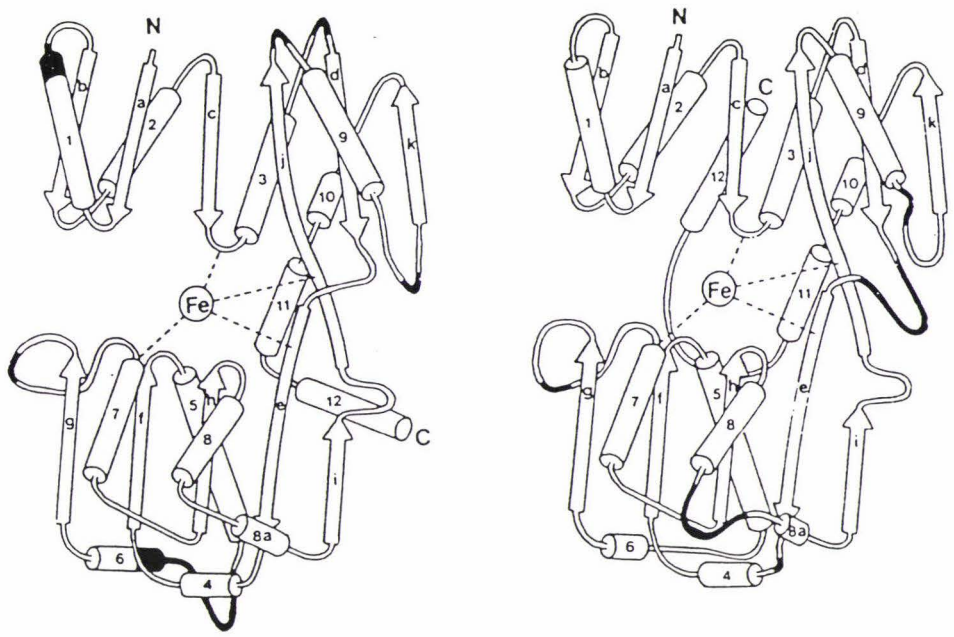


Fig. III.11 Schematic representation of the secondary structure elements for the N-lobe (left) and C-lobe (right) of $Fe_2(C_2O_4)_2$Lf. Helices are shown as cylinders and β strands as arrows. Sites of insertion in one lobe or the other are shown in black.

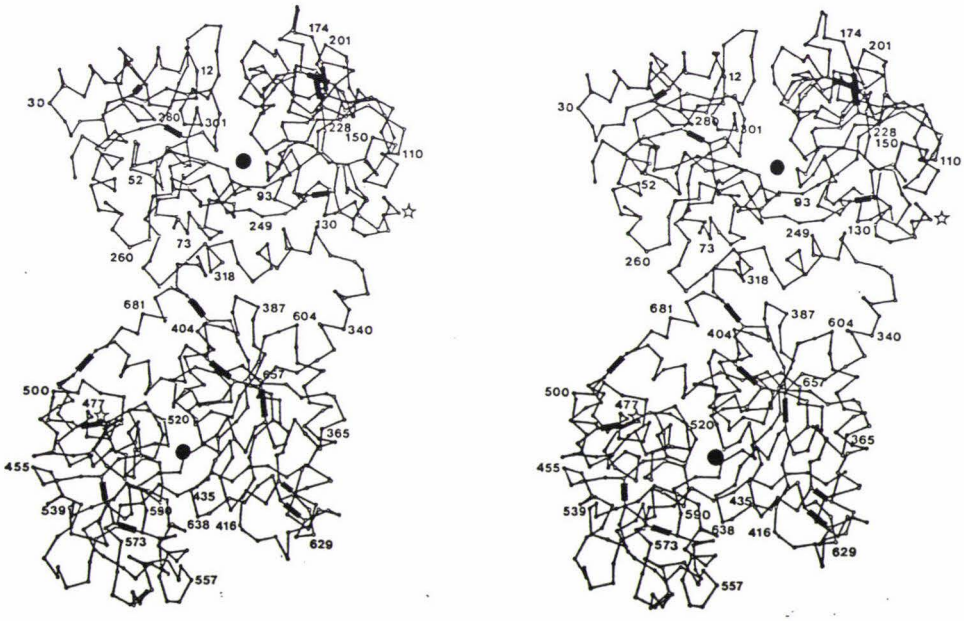


Fig. III.12 Stereo Cα plot of $Fe_2(C_2O_4)_2$Lf. The N-lobe is above, the C-lobe below and the domains are labelled N1, N2, C1 and C2. The iron atoms are shown with filled circles, and the 2 glycosylation sites are marked by stars. Disulphide bridges are shown as thickened lines.

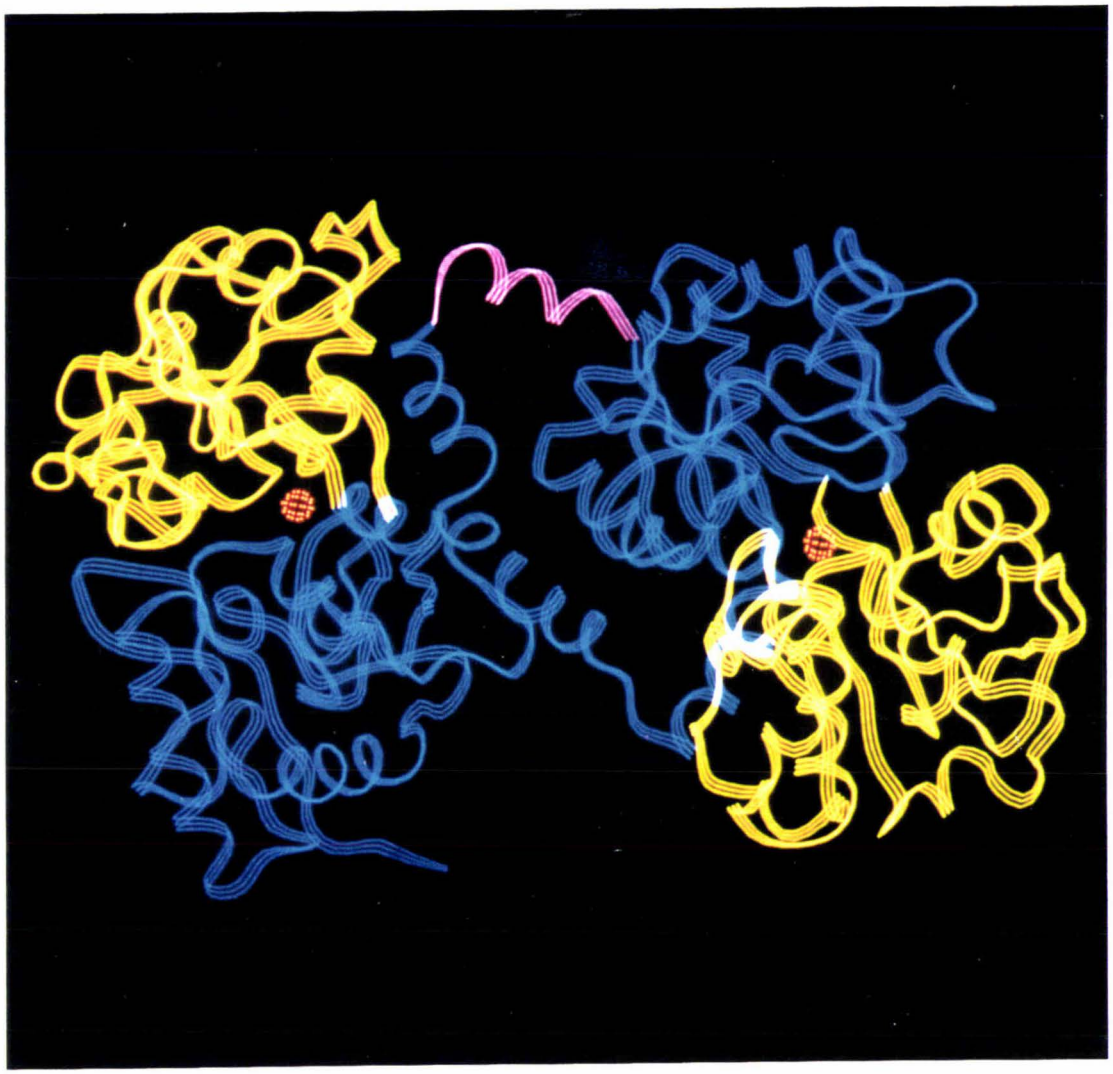


Fig. III.13 A ribbons diagram of $Fe_2(C_2O_4)_2$Lf showing the division into domains. The N-lobe is on the left and C-lobe on the right. N1 and C1 domains are shown in blue, N2 and C2 domains in yellow, the connecting helix purple, and the iron atoms as stippled red spheres.

III.3.1 **Secondary Structure**

All four domains have similar secondary structure, based on a mixed β-sheet with helices packed around it. Table III.9 lists the residues which make up the β-sheet and helical structure.

Table III.9 Residue assignments for the secondary structure of $Fe_2(C_2O_4)_2Lf$

β-strands[(a)]			Helices[(b)]		
N-lobe	label	C-lobe	N-lobe	label	C-lobe
5-10	a	345-350	12-31	(1)	351-365
33-39	b	368-374	41-53	(2)	376-388
54-59	c	389-394	60-69	(3)	395-407
75-82	d	408-415	105-108	(4)	449-453
90-100	e	435-443	121-136	(5)	465-481
112-116	f	456-460	144-153	(6)	484-488
153-158	g	489-494	166-170	(6a)	501-505
204-211	h	540-547	190-203	(7)	526-539
226-231	i	570-575	212-218	(8)	547-554
246-258	j	590-602	223-227	(8a)	567-571
304-310	k	647-653	241-245	(8b)	585-589
			263-279	(9)	605-621
			315-321	(10)	658-664
			321-332	(11)	664-678
			334-344	(12)	680-691

(a) Residues were included in β strands if they contributed to the hydrogen bonding of the β sheet or were part of a continuous sequence of residues with β-type (ϕ, ψ) angles.

(b) Helices were defined in terms of those residues involved in the continuous hydrogen bonding of the helix. This meant that the residues at the end of the helix may not have the characteristic torsion angles (ϕ, ψ) of a helix, although the intermediate ones do.

The topologies of the β-sheets and helices are the same as described previously for $Fe_2(CO_3)_2Lf$ (Haridas *et al*, in press), and the nomenclature for the helices and β-strands follows that used for both $Fe_2(CO_3)_2Lf$ and rabbit serum transferrin (Bailey *et al*, 1988). All are mixed β-sheets with both parallel and antiparallel strands. The domain 1 β-sheet in each lobe has four parallel strands (a, b, c, d strands) directed towards the interdomain

cleft, and two antiparallel, pointing to the exterior of the molecule, with the strand order being b-a-c-j-d-k. In the N2 and C2 domains there are four parallel strands pointed to the interdomain cleft (f, g, h and i/j) but only one antiparallel (e). The strand order is g-f-h-e-(i+j); thus the β sheet topology is the same in each domain but with an extra antiparallel strand in the domain 1 sheets. The two strands (e and i/j) which pass behind the binding site each contribute to the β-sheets of both domains. The inserted loops within each of these strands may contribute to the flexibility of the molecule, which is necessary for the opening and closing of each lobe.

Lactoferrin is an α/β protein (Levitt and Chothia, 1976), with the helices generally running antiparallel to the β-strands. In domain 1, the three helices connecting the first four parallel β-strands thus have their N-termini pointing towards the interdomain cleft. The polypeptide chain then passes behind the metal-binding site (strand e) and a similar α/β structure makes up the bulk of domain 2. This domain has three extra helices (4, 6 and 8) compared to domain 1 but nevertheless the helices 5, 7 and 9 are oriented so that their N-termini all point towards the interdomain cleft. Finally the chain crosses back to complete the folding of domain 1, and leads to the final helix (in the N-lobe this is the connecting helix, in the C-lobe it is the C-terminal helix). It can be seen that a consequence of the α/β motif adopted by each domain is that the N-termini of the helices contribute to the interdomain cleft. It is proposed that the partial positive charge associated with these α helix N-termini (Hol *et al*, 1978) may have a role in attracting the anion into the binding site in lactoferrin (Anderson *et al*, 1989).

The majority of helices in $Fe_2(C_2O_4)_2$Lf contain essentially all α-type hydrogen bonds, but there are short pieces of 3_{10}-helix such as helices 4 and 8a. A notable feature of each lobe is the unusually wide first turn of helix 5. This also occurs in serum transferrin and the sequence in this turn is highly conserved . Helix 5, which binds the anion (oxalate or carbonate) at its N-terminus, begins with two π-type (1-6) hydrogen bonds (121O...126N and 122O...127N in the N-lobe, 465O...470N and 466O...471N in the C-lobe) instead of the usual α-type (1-5). The helix appears to play a key role in the binding site, hydrogen bonding not only to the anion, but also to the metal ligand, Asp 60 or Asp 395. In fact residues 114 to 130 (458 to 474 C-lobe) (the helix 5 N-terminus and the preceding β-strand) are amongst the most highly conserved regions in all transferrins (Metz-Boutigue *et al*, 1984). Helix 5 is further disturbed by a kink in the middle associated with a conserved proline (Pro 128, Pro 472), and the glycosylation sites Asn 137, Asn 478 are found at or close to its C-terminus.

III.3.2 **Comparison of Secondary Structure in the Two Lobes**

On the whole the β sheets are the most highly conserved secondary structure elements between the two lobes, forming their structural core. Equivalent β-sheets in each

lobe have essentially the same conformation and hydrogen bonding patterns.

The helices, however, are more variable. While some match closely, even having the same termination hydrogen bonding pattern, others differ, with deletions or insertions near the outside of the molecule. Helices 1 and 6 in the N-lobe both have 3 or 4 extra residues than their counterparts in the C-lobe. The only major folding difference between the two lobes is found in the orientation of the last helix in each lobe (helix 12). These helices are really not comparable structurally, do not show sequence homology and fulfil different roles. In the N-lobe, helix 12 points away from the main body of the lobe, to be the connecting helix to the C-lobe, whereas the C-terminal helix 12 folds back behind the binding site to contribute to the hydrophobic region between the two lobes. This helix is fastened down by two additional disulphide bridges, not present in the N-lobe (677 to 483 and 686 to 405) giving an extra constraint on the C-lobe domains.

As expected, turns are the most variable elements between the two lobes, with many of the insertions and deletions occurring in the loops between the secondary structure elements. Nevertheless it is interesting to note that amongst the conserved turns are the two classic γ-turns; 298-300 and 641-643, mentioned before. Such turns are quite rare (Nemethy and Printz, 1972; Matthews, 1972) but these two turns are conserved in all the transferrin family.

III.3.3 **Metal and Anion Sites**

The direct coordination of the anion to the metal means that in each case the metal- and anion-binding sites are highly interdependent. In contrast to the situation where carbonate is the synergistic anion, in which case the two binding sites are extremely similar (Haridas *et al*, 1995), the final model for the oxalate structure shows a clear difference between the binding sites of the two lobes.

In each lobe the iron atom makes the same four bonds with equivalent protein sidechains; Asp 60, Tyr 92, Tyr 192 and His 253 in the N-lobe and Asp 395, Tyr 435, Tyr 528 and His 597 in the C-lobe. These protein ligands, as well as binding to the iron atoms, make a number of hydrogen bonds. The non-coordinating carboxylate oxygen of Asp 60 (395) receives two important hydrogen bonds from the peptide nitrogens of Gly 62 (397) on helix 3 and Thr 122 (466), on helix 5; providing a direct protein-protein interaction between the two domains. Tyr 92 (435) and His 253 (597) are hydrogen bonded to water molecules and Tyr 192 receives a hydrogen bond from the guanidinium $N_{\eta 2}$ of Arg 210. In these respects the two binding-sites are identical to other lactoferrin structures (Anderson *et al*, 1989; Smith *et al*, 1994; Day *et al*, 1993, see Fig. I.4).

The most striking difference between the binding-sites of the two lobes (and from $Fe_2(CO_3)_2Lf$), is clearly related to the substitution of the carbonate ion by oxalate. In the C-lobe the oxalate ion is bound in a symmetrical 1,2-bidentate mode to the iron, with bond

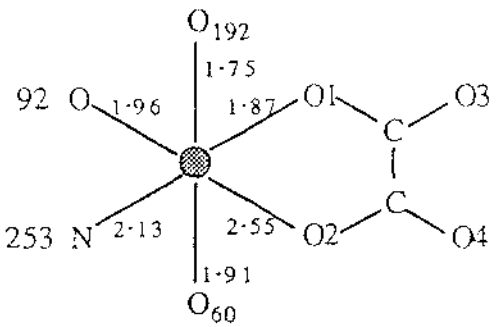
lengths O_{1ox}-Fe = 2.07 Å, O_{2ox}-Fe = 1.91 Å. In the N-lobe, however, the coordination is quite asymmetric, with bond lengths O_{1ox}-Fe = 1.87 Å, O_{2ox}-Fe = 2.55 Å; a lenthening of the O_{2ox}-Fe bond relative to O_{1ox}-Fe of ~0.65 Å. In each lobe the oxalate completes the octahedral coordination of the iron (Figs. III.14 and III.15).

The overall geometry of the two sites therefore differs also. In both lobes it can be best described as distorted octahedral (Table III.10). In the C-lobe the metal to ligand distances are all between 1.8-2.1 Å, close to the ideal value of 2.0 Å, with the sole exception being the His 597-Fe bond which is lengthened to 2.32 Å. The error in the Fe-ligand bond distances, estimated at ≈0.2 Å, is such that caution should be adopted in interpreting small differences in bond lengths. Nevertheless the lengthening of the His_{597}-Fe bond appears significant. The His_{597}-Fe-O_{1ox} axis thus becomes the long axis of the distorted octahedron, with a length of 4.4 Å, compared with the other axes, Asp_{60}-Fe-Tyr_{192} and O_{2ox}-Fe-Tyr_{92}, of ~3.8 Å each.

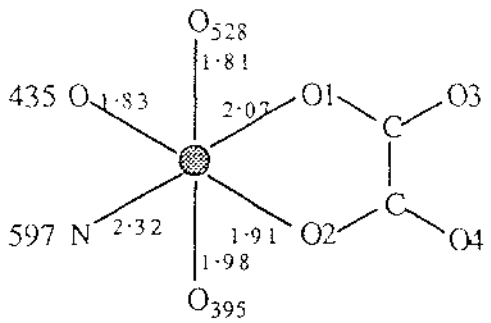
Table III.10 Bond lengths (Å) and angles (°) at the iron sites of $Fe_2(C_2O_4)_2Lf$ (Iron atoms shown as shaded spheres)

N-lobe

(a) Bond lengths (Å)



C-lobe



(b) Bond angles (°)

O_{60}-Fe-O_{92}	103.4	O_{395}-Fe-O_{435}	82.4
O_{60}-Fe-O_{192}	169.6	O_{395}-Fe-O_{528}	170.6
O_{60}-Fe-N_{253}	96.0	O_{395}-Fe-N_{597}	70.6
O_{60}-Fe-$O1_{ox}$	95.2	O_{395}-Fe-O_{1ox}	91.5
O_{60}-Fe-$O2_{ox}$	79.8	O_{395}-Fe-O_{2ox}	88.4
O_{92}-Fe-O_{192}	79.8	O_{435}-Fe-O_{528}	99.3
O_{92}-Fe-N_{253}	92.2	O_{435}-Fe-N_{597}	85.6
O_{192}-Fe-N_{253}	79.6	O_{528}-Fe-N_{597}	105.5
O_{1ox}-Fe-N_{253}	165.8	O_{1ox}-Fe-N_{597}	160.4
O_{1ox}-Fe-O_{2ox}	75.8	O_{1ox}-Fe-O_{2ox}	83.8

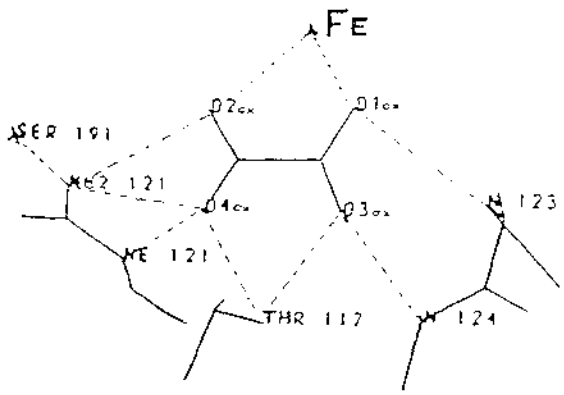
The significantly larger chelate angle (≈85° for oxalate compared with ≈64° for carbonate) allows a somewhat more regular octahedral geometry in the C-lobe site compared with that seen in the C-lobe of $Fe_2(CO_3)_2$Lf, but the angles still vary between 71° (Asp_{395}-Fe-His_{597}) and 105° (Tyr_{528}-Fe-His_{597}). (The ideal value is 90°, the error in angles is estimated to be ≈5°.)

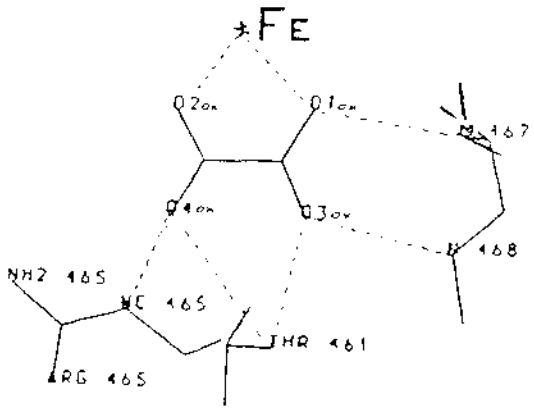
The N-lobe binding site geometry is profoundly affected by the asymmetry of the oxalate coordination to the metal. The long Fe-O_{2ox} bond of 2.55 Å is trans to the Fe-Tyr_{92} bond of 1.96 Å. This means that in the N-lobe, the O_{2ox}-Fe-Tyr_{92} axis is the long axis (4.5 Å) of the metal coordination sphere. The Fe-Asp_{60} bond of 1.8 Å is trans to the other short, Fe-Tyr_{192} bond of 1.8 Å, giving a short axis of 3.6 Å. The Fe-His_{257} bond, at 2.13 Å is, as expected, slightly longer than the Fe-O distances, but still considerably shorter than that seen in the C-lobe. The third axis of the metal coordination sphere His_{257}-Fe-O_{1ox} is an intermediate length of 4.0 Å.

The asymmetry of the oxalate bonding to the iron in the N-lobe is reflected in the reduction of the O_{1ox}-Fe-O_{2ox} chelate angle from ~84° in the C-lobe to ~76° in the N-lobe. Other angles, whilst close to the optimal 90°, vary from ~80° (Asp_{60}-Fe-O_{2ox} and Tyr_{92}-Fe-Tyr_{192}), to 103° (Asp_{60}-Fe-Tyr_{192}). A list of relevant bond lengths and angles in the two iron sites is given in Table III.10. Views of the two sites are shown in Figs III.14 and III.15.

What are the reasons for the different anion coordination in each lobe? The hydrogen bonding pattern of the oxalate in each site is quite different. In the C-lobe the oxalate is symmetrically bound to the metal and makes five hydrogen bonds with the surrounding protein structure, including one with the anion-binding arginine residue, Arg 465 (O_{4ox}...Nε_{465}). The other four hydrogen bonds comprise two with the main chain nitrogens of residues 467 and 468 (O_{1ox}...N_{467}, O_{3ox}...N_{468}) and two with Oγ1 of Thr 461 (O_{3ox}...O$\gamma1_{461}$, O_{4ox}...O$\gamma1_{461}$). In the N-lobe, on the other hand, the equivalent arginine, Arg 121, is in a different orientation. The oxalate cannot remain in a symmetrical position (relative to the iron) as this would bring it too close to the arginine. Instead the oxalate ion rotates 15° in its plane around O_{1ox} (see Fig. III.16 and Table III.11) and makes a set of three hydrogen bonds with the arginine; Nε_{121}...O_{4ox}, N$\eta2_{121}$...O_{2ox}, and N$\eta2$...O_{4ox}. The hydrogen bonds involving O_{1ox} and O_{2ox} remain the same as in the C-lobe, but the changed oxalate orientation means that the interactions with Oγ1 of Thr 117 are also changed, relative to the C-lobe. The primary difference in the two sites is that Arg 465 has been able to swing further away from the anion site, allowing the anion to remain symmetrically bound to the iron, and still remain hydrogen bonded to the arginine (Fig. III.17).

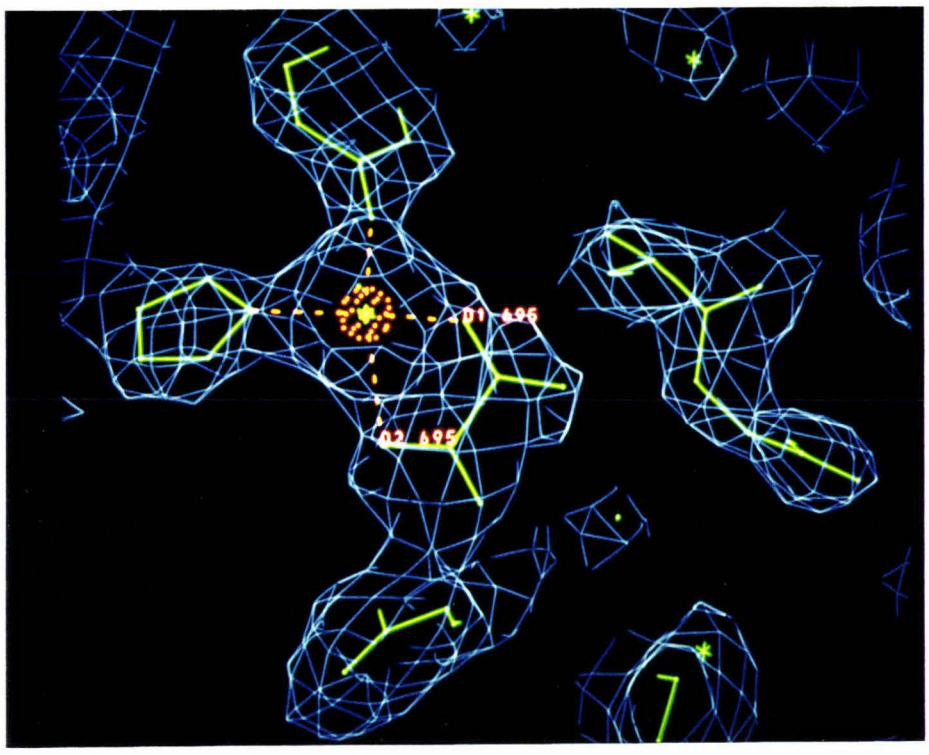
Table III.11 Hydrogen bonding interactions around the oxalate ions and the surrounding protein in $Fe_2(C_2O_4)_2Lf$





N-lobe		C-lobe	
O_{2ox}...$N\eta2_{121}$	2.51 Å		
O_{4ox}...$N\eta2_{121}$	2.65 Å		
O_{4ox}...$N\varepsilon_{121}$	2.51 Å	O_{4ox}...$N\varepsilon_{465}$	2.89 Å
O_{4ox}...$O\gamma1_{117}$	2.63 Å	O_{4ox}...$O\gamma1_{461}$	3.40 Å
O_{3ox}...$O\gamma1_{117}$	2.65 Å	O_{3ox}...$O\gamma1_{461}$	2.69 Å
O_{3ox}...N_{124}	2.62 Å	O_{3ox}...N_{468}	2.98 Å
O_{1ox}...N_{123}	3.10 Å	O_{1ox}...N_{467}	2.88 Å

(a)



(b)

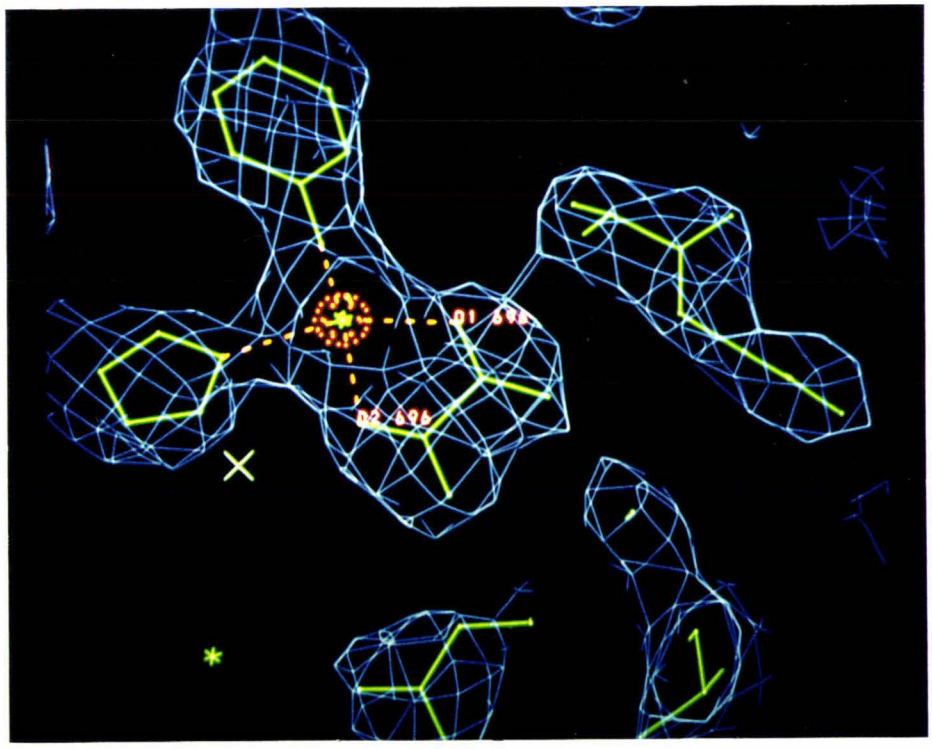


Fig. III.14 Views of the 2*Fo*-*Fc* electron density in the oxalate-binding site of the N-lobe (top) and C-lobe (bottom). The final model of the $Fe_2(C_2O_4)_2$Lf structure is superimposed in yellow. Oxalte coordinates in a symmetrical 1,2-bidentate mode in the C-lobe, but in the N-lobe the coordination is asymmetric.

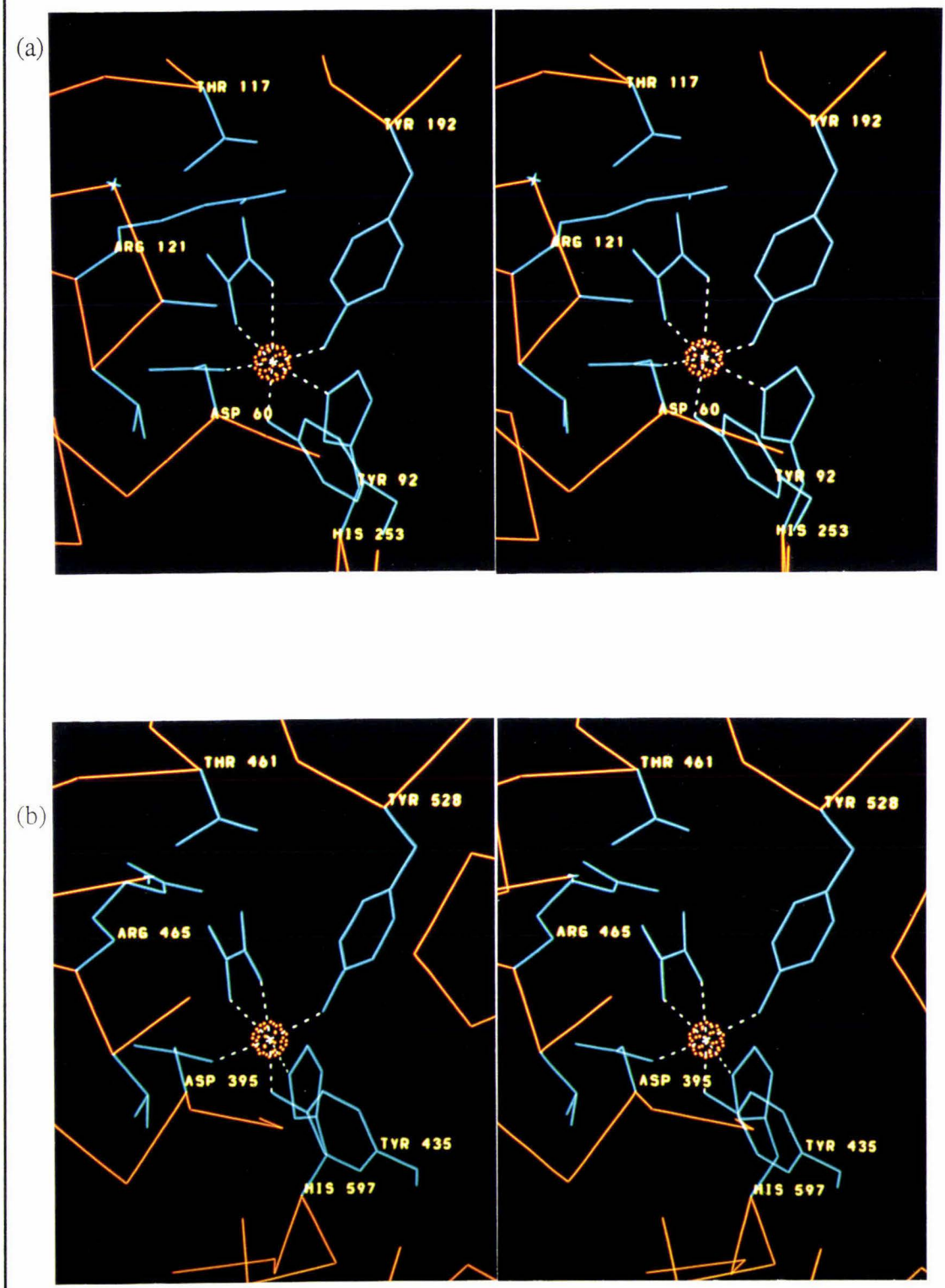


Fig. III.15 Stereo views of the iron-binding site of the N-lobe (top) and C-lobe (bottom) of $Fe_2(C_2O_4)_2$Lf.

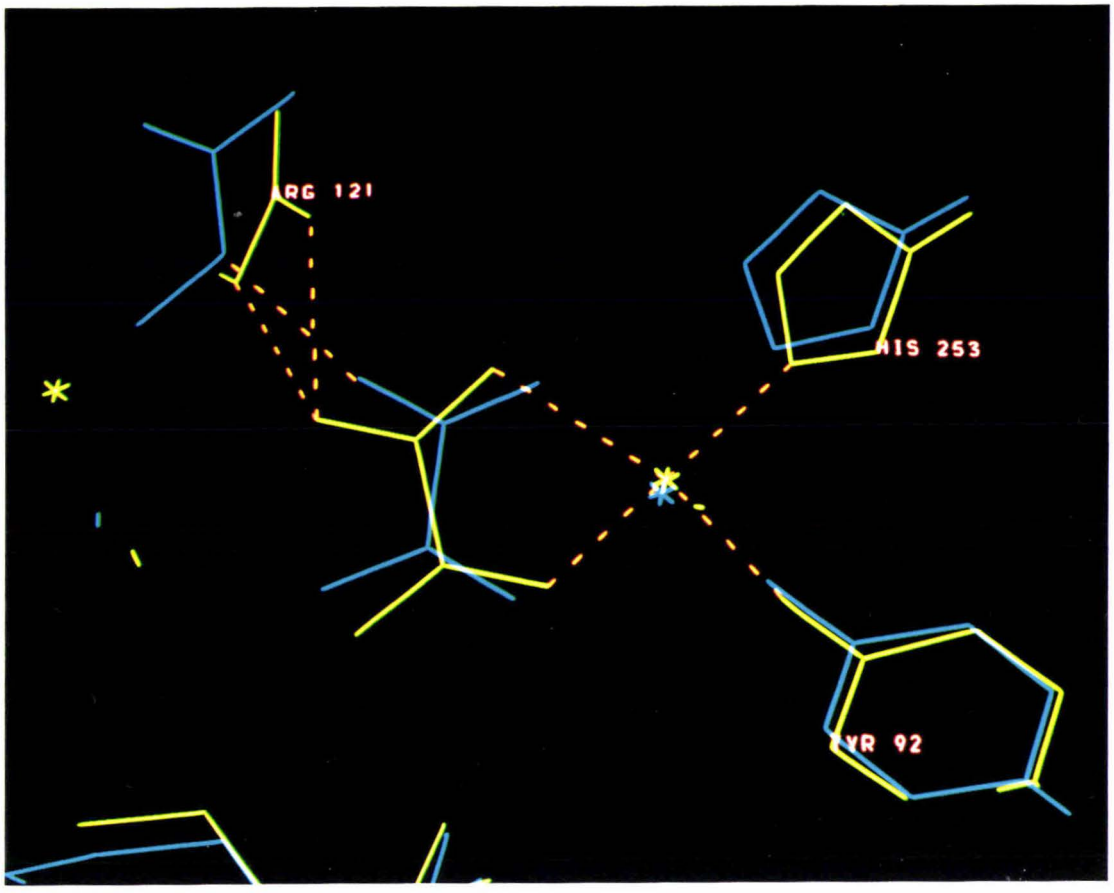


Fig. III.16 Superposition of the oxalate ions from the two binding sites (N-lobe yellow, C-lobe blue), showing the rotation of the oxalate ion in the N-lobe with respect to that in the C-lobe. Note also the different hydrogen bonding interaction with the arginine.

The reason for the difference in the orientation of these arginines (and therefore the oxalate ions) lies in residues more remote from the binding site. The N-lobe Arginine 121 is held in place by a hydrogen bond to O_γ of Ser 191. This hydrogen bond cannot be made in the C-lobe as the equivalent residue is a glycine. Furthermore if the Arginine 121 is moved as in the C-lobe a clash would occur with Phe 183. Such steric hindrance does not happen in the C-lobe as the residue equivalent to Phe 183 is a smaller proline (Pro 519) sited further from the iron site. This relative restriction of the orientation of the anion-binding arginine in the N-lobe is in contrast to the situation in the C-lobe. Here not only is Arginine 465 free to swing further away from the iron site (because of the lack of a restraining hydrogen bond or steric hindrance), but it can also adopt a position where it is strapped in place by two additional hydrogen bonds from Tyr 398 and Tyr 526. Such stabilising bonds cannot be made in the N-lobe where equivalent residues are both phenylalanines (Phe 63, Phe 190) (Fig. III.17).

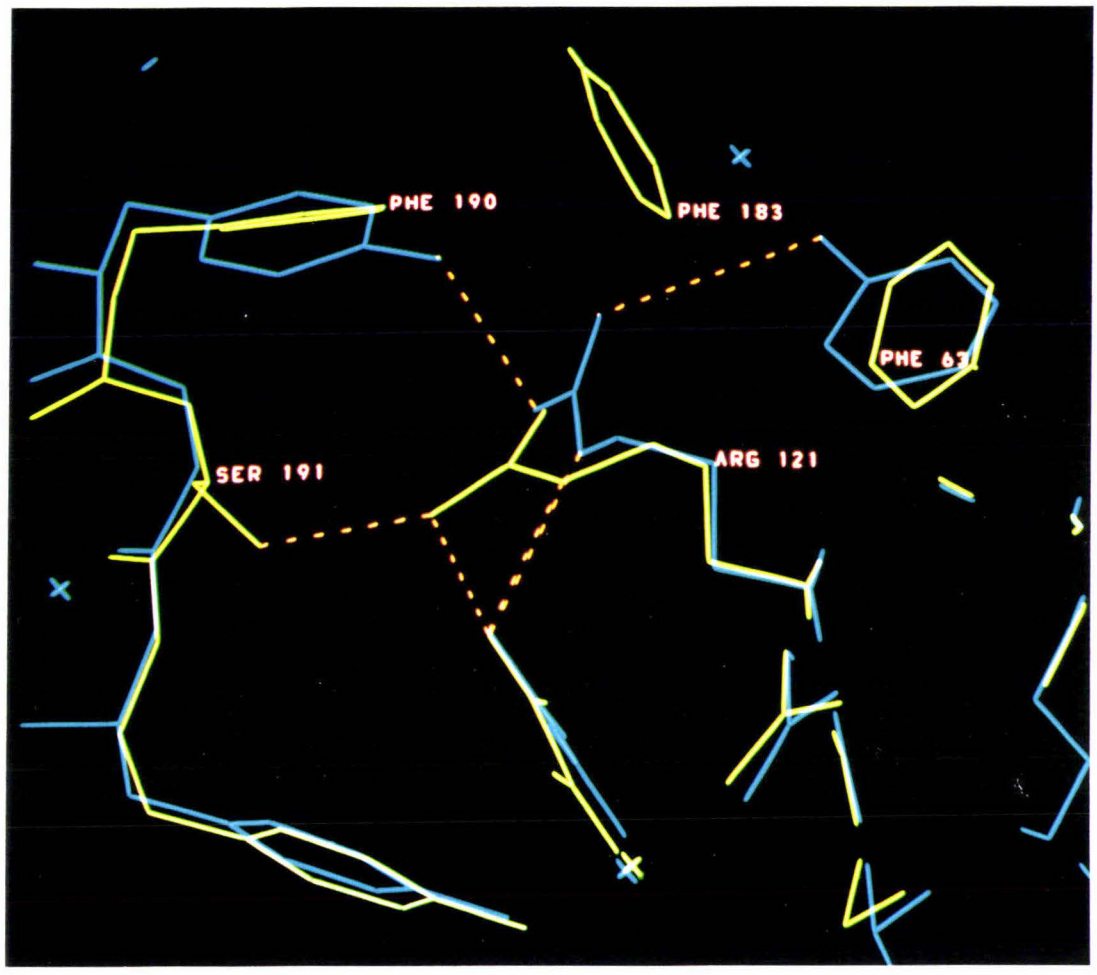


Fig. III.17 Superposition of the anion-binding arginine residues from the two anion-binding sites, Arg 121 (N-lobe, yellow) and Arg 465 (C-lobe, blue), showing their different orientations and interactions with oxalate (shown centre bottom), Ser 191 (Gly 527, C-lobe), Phe 190 (Tyr 526, C-lobe), Phe 183 (Pro 519, C-lobe) and Phe 63 (Tyr 398, C-lobe). (Labelling shown for N-lobe only.)

III.3.4 **Solvent Structure**

As described earlier (Section III.2.3(a)), all solvent molecules added were treated as water. Some peaks on the outside of the molecule are more diffuse and could represent phosphate ions or MPD molecules which were present in the crystallisation medium; however, at this resolution they could not be distinguished from disordered water. When the refined structure of $Fe_2(C_2O_4)_2$Lf was superimposed on the $Fe_2(CO_3)_2$Lf structure, approximately 95 waters were found to be in conserved positions. The remainder of the water molecules were located from analysis of $2F_o$-F_c and F_o-F_c density maps over several rounds of refinement and manual rebuilding. A total of 121 water molecules have

been included with an average B value of 57, of which 6 are 'second-shell' waters, interacting with only other solvent molecules. The remainder are first shell waters and have at least one protein hydrogen bond partner. The small number of water molecules included (121) for a protein of this size reflects the somewhat conservative approach adopted, and the exclusion of solvent molecules with large increases in B values on refinement. Clearly many less well-ordered water molecules will be associated with the $Fe_2(C_2O_4)_2$Lf structure but at the present resolution they cannot be reliably placed.

Whilst the more mobile regions of the structure are not associated with well-defined solvent sites, such well-defined water molecules are found in other regions where they are trapped within the protein structure. Of the 121 water molecules 41 are totally internal, removed from the bulk solvent. Such molecules make an important contribution to the overall stability of the molecule. Well-defined clusters of water molecules are also found within the interdomain cleft in each lobe. The anion-binding arginine projects into this cavity and these solvent molecules can be displaced by the arginine as it moves to permit the coordination of larger anions.

Several water molecules appear to play an important part in the hydrogen bonding networks of each binding site. One tightly bound water is found close to the iron atom in each site. In the N-lobe this water (OW 916) is hydrogen bonded to the phenolic oxygen of the ligand Tyr 92, as well as to Gly 61 N, Thr 122 $O\gamma_1$ and Pro 251 O, and is only 3.24 Å from the iron. In the C-lobe an equivalent water (OW 918) is hydrogen bonded to Tyr 435 OH, Thr 466 $O\gamma_1$ and Pro 595 O, and is 3.6 Å from the metal. These water molecules are conserved in $Fe_2(CO_3)_2$Lf (Anderson *et al*, 1989), Cu_2Lf (Smith *et al*, 1994), and FeLfN (Day *et al*, 1993), and in each case neither solvent molecule interacts with the interdomain network.

Another solvent molecule common to both sites, and all four structures is hydrogen bonded to $N_{\delta 1}$ of the histidine ligand (His 253 N-lobe, His 597 C-lobe). These water molecules make identical contacts in each lobe, namely to $O_{\varepsilon 1}$ and $O_{\varepsilon 2}$ of Glu 80 (N-lobe) and Glu 413 (C-lobe), and form part of the interdomain solvent network.

CHAPTER IV

DEGLYCOSYLATION OF LACTOFERRIN AND A COMPARISON OF SOME PROPERTIES OF THE NATIVE AND DEGLYCOSYLATED FORMS

IV.1 INTRODUCTION

Glycosylation is the most common post-translational modification of proteins. It is also the most diverse, with the particular glycans attached being amino acid specific, polypeptide specific and also tissue specific. The carbohydrate is most commonly N-linked, through the amide nitrogen of an asparagine sidechain, or O-linked through the hydroxyl oxygen of serine or theonine. Although the consensus sequence specifying N-glycosylation is well known; Asn-Xaa-Thr or Asn-Xaa-Ser where Xaa is any amino acid other than proline, an equivalent sequence for O-glycosylation has yet to be determined. Many more less common linkages such as the glycosyl-phosphatidylinositol anchor found in melanotransferrin have been more recently recognised.

IV.1.1 Functions of Glycoprotein Glycans

The intriguing question of the functional significance of the carbohydrate on glycoproteins has long been a subject of interest and still remains wide open. For a small number of proteins an alteration of physico-chemical properties or modulation of biological activity has been demonstrated, yet in others no marked effect of the presence of the carbohydrate could be detected. The role of the carbohydrate thus remains an enigma. It is apparent that the glycans do not have a single function but act in a variety of ways that vary from one protein to another. This means that each glycoprotein must be examined individually to determine the contributions of its attached carbohydrate to its properties and functions.

The carbohydrate moieties of glycoproteins form a large part of their total molecular mass (frequently 20-50%) and on occasions can hydrogen bond back to the protein surface. They thus have the potential to cover a large proportion of the polypeptide surface. For this reason the possible roles of the glycans fall into two main categories; the conferring of particular physico-chemical properties on the protein or signalling for cell-cell or cell-surface recognition processes (see Table IV.1).

Table IV.1. Functions of glycoprotein glycans

Type	Function
Physicochemical	Modify solubility, electrical charge, mass, size and viscosity in solution. Control protein folding. Stabilize protein conformation, confer rigidity to molecule. Confer thermal stability and protection against proteolysis.
Biological	Regulate intracellular traffic and location of glycoproteins. Determine lifetime of glycoproteins in circulation. Modify immunological properties. Modulate activity of enzymes and hormones. Act as cell surface receptors for lectins, antibodies, toxins, etc. Participate in cell-cell interactions.

From Lis, H. and Sharon, N. (1993).

Glycosylation may thus affect the solubility, effect of pH, conformation, rigidity, stability and susceptibility to proteolysis of proteins that are modified this way. Several recent studies have highlighted these points.

NMR studies have shown that the glycosylated forms of bovine ribonuclease (RNase B) have a greater overall dynamic stability and rigidity than the non-glycosylated form (RNase A) (Joao, *et al*, 1992). The increase in stability may not only affect those residues close to the attachment site, but others as far as 30 Å away, including the active site residues. It is therefore possible for the oligosaccharide to influence enzyme activity in two ways; by increasing the stability and rigidity of the molecule, or by steric hindrance of a functionally important site. For the same reasons, glycosylation can give increased resistance to proteolysis by limiting the access of the attacking proteases to vulnerable sites, or increasing the rigidity of the protein. Such resistance has been well documented (see Lis & Sharon, 1993) with perhaps the most extreme case being the glycocalyx with which some viruses such as HIV protect themselves from their host.

The relationship between heat stability and glycosylation has been shown for two different β-1,3/1,4 glucanases. The thermostability of the glycosylated forms of these enzymes (expressed in *Saccharomyces cerevisiae*) was considerably greater than their non-glycosylated counterparts (expressed in *E. coli*) (Olsen & Thomsen, 1991). Such

findings must be important to industry where thermostability is often a requirement.

Another well documented function of the glycans on many proteins is their involvement in the initiation of correct polypeptide folding in the endoplasmic reticulum and the subsequent maintenance of protein solubility and conformation. Many proteins that are incorrectly glycosylated are subsequently degraded.

On the other hand there are examples where prevention of glycosylation does not alter the expression of the protein, and others where removal of the carbohydrate from the mature protein is of little apparent consequence to its thermostability, sensitivity to proteolysis or function. All glycoproteins differ in their carbohydrate requirements. At present it is still impossible to predict the consequences of altering the glycosylation of a glycoprotein. The loss of a single N-linked glycosylation site in erythropoetin reduces activity by $\geq 70\%$ (Parekh *et al*, 1989), whereas tissue plasminogen activator actually has improved biological activity and a longer half life in the non-glycosylated form.

Equally (if not more) interesting than the effect of the glycans on the physico-chemical properties and biological activities of proteins, are their roles in biological recognition. Here their structural diversity provides precise protein targeting and cell-cell interactions. The carbohydrates can act as recognition determinants for a variety of biological processes, including the clearance of glycoproteins from the circulatory system (Ashwell & Harford, 1982), intracellular trafficking of enzymes (Kornfeld, 1992) and a wide range of cell-cell interactions, from the ABO blood group determinants (Koscielak, 1986), to the attachment of sperm to ova (Wassaram, 1990), and the role of selectins on the surface of leukocytes (Springer, 1990).

IV.1.2 Glycosylation and Recombinant Proteins

The advent of recombinant DNA technology has led to the rapid growth in industry centred around the production of recombinant proteins. The production of recombinant protein generally involves the expression of recombinant DNA in a cell line that is different from that which produced the native protein. Glycosylation may thus be altered or even absent. An understanding of the roles of the carbohydrate moieties is thus not only of theoretical interest and biomedical relevance, but also has a major impact on the biotechnology industry that produces glycoproteins for therapeutic applications. Erythropoietin, the first recombinant glycoprotein produced industrially for clinical use, is used to treat anaemia of patients on haemodialysis. This protein had sales exceeding US$645 million in 1991, while recombinant tissue plasminogen activator, a clotting agent, had sales of US$200 million (Lis & Sharon, 1993). In both these cases understanding the roles of the carbohydrates was essential in controlling their biological activity (see Section IV.1.1).

IV.1.3 **Glycosylation of Transferrins**

Almost all proteins of the transferrin family are glycosylated, the only known exceptions apparently being the transferrins of several of the cartilagenous fishes (Stratil *et al*, 1983), yet in spite of extensive work, the roles of the attached carbohydrates are not clear.

In the case of human lactoferrin, two biantenary complex oligosaccharides are attached, through Asn linkages, one in the N-terminal half of the molecule (Asn 137), and one in the C-terminal half (Asn 478) (Spik *et al*, 1982; Metz-Boutigue *et al*, 1984). In fact, at least five different carbohydrate moieties are associated with human lactoferrin (Spik *et al*, 1988) (see Fig. IV.1), and with only two different attachment sites, which means that there is a microheterogeniety such as is quite commonly found in glycoproteins.

Functions that have been attributed to these glycans on lactoferrin include the conferment of enhanced resistance to proteolysis (Spik *et al*, 1985) a role in receptor binding (Davidson & Lonnerdal, 1988) and a modification of iron binding (Legrand *et al*, 1990). With respect to iron binding it was shown that the full removal of the glycans from the 30 kDa N-tryptic fragment of lactoferrin led to a 50% loss in iron-binding capacity. This loss was ascribed to the reduction of the stability conferred on the protein by the glycan moiety. Deglycosylation of the whole molecule was not attempted.

The three dimensional structure of diferric lactoferrin showed the glycosylation sites to be remote from the iron-binding sites, some 20 Å away (Anderson *et al*, 1989). The carbohydrate was disordered, with only the first few residues able to be modelled, and as far as could be judged, the glycans extended from the protein surface with little direct hydrogen bonding back to the polypeptide. For these reasons it was difficult to see how such events as iron binding could be directly affected by the presence of the carbohydrate. A re-evaluation of the properties of deglycosylated lactoferrin was undertaken as part of this study.

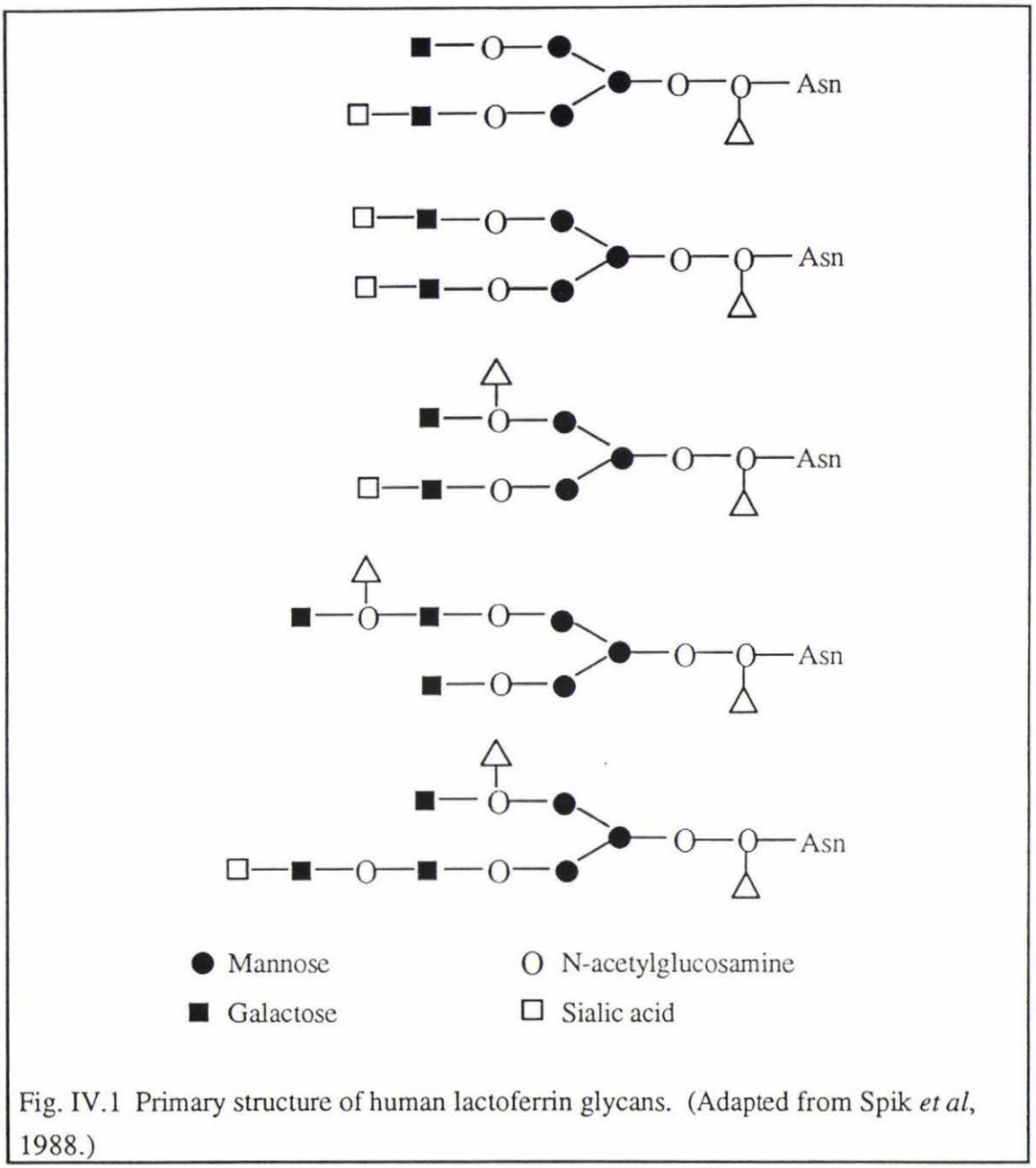


Fig. IV.1 Primary structure of human lactoferrin glycans. (Adapted from Spik *et al*, 1988.)

IV.1.4 **Methods of Deglycosylation**

A variety of approaches have been used to study the roles of the carbohydrates of glycoproteins. These include deglycosylation of the native protein, either chemically or enzymatically, or the prevention of glycosylation itself.

Chemically, hydrazinolysis, for example, can cleave N-glycosidic linkages, although with this method, care must be taken to prevent unwanted additional modification of the protein. Enzymatically, a set of exoglycosidases can be used for the stepwise hydrolysis of the major portion of the glycans. However, with the recent discovery of endoglycosidases it has now become possible to release susceptible carbohydrate chains

'en bloc' from the protein. Two classes of enzymes are available for the liberation of asparagine-linked chains; endo β-N-acetylglucosamidases (eg Endo F), which hydrolyse the glycosidic bond between the two N-acetylglucosamine residues adjacent to the asparagine (thus leaving a single N-acetylglucosamine on the protein) and the N-glycanases, eg PNGase, which hydrolyse the GlcNAc-Asn linkage. If the glycosylation site is not sterically hindered, the use of endoglycosidases can give rapid, complete deglycosylation with no chemical damage to the remaining polypeptide.

An alternative to deglycosylation of the native protein is the prevention of glycosylation, either by the use of specific inhibitors of transferases, eg tunicamycin, or the expression of recombinant proteins in cell lines with known defects in glycosylation, or in bacteria such as *E. coli* which produce completely non-glycosylated proteins. Site-directed mutagenesis is frequently used to abolish glycosylation sites (enabling the role of an individual sugar chain to be investigated). Problems associated with these approaches can include poor expression of the protein (the carbohydrate may be essential in the folding and transport of the nascent polypeptide - see Section IV.1.1). Furthermore the effects observed after removal of the glycosylation site might be due to the change in the amino acid sequence itself rather than the loss of carbohydrate.

The method chosen to deglycosylate lactoferrin, hydrolysis by a mixture of Endo F and PNGase, which had been prepared in our laboratory, was dictated partly by practical reasons. The presence of the two N-linked glycans on human lactoferrin suggested the use of either endoglycosidase as an ideal method of deglycosylation, but the cost of buying sufficient enzyme to deglycosylate the required amount of lactoferrin would have been quite prohibitive (~$100,000). These enzymes could, however, be prepared in our laboratory with small expense, apart from time and the unpleasant nature of the organism. There was also a second advantage, that of freshness of the enzyme preparation, which was at least 30 times more active than could be purchased. This ensured a rapid, safe deglycosylation of the protein; essential if one is to study the properties of the remaining polypeptide.

IV.2 EXPERIMENTAL

IV.2.1 Preparation of Endo F and PNGase

Endo F and PNGase are obtained from *Flavobacterium meningosepticum* which releases the enzymes into the culture medium during growth. A mixture of the two enzymes was prepared following the method of Elder and Alexander (1982). The culture of *Flavobacterium meningosepticum* was obtained from the National Health Institute, Porirua, Wellington. Because the strain was isolated locally it may not be identical to that used by Elder and Alexander.

Cells were grown to stationary phase in M9 medium (M9 medium consists of (per

litre) 5 g casamino acids, 3 g KH_2PO_4, 6 g Na_2HPO_4, 0.5 g NaCl, 1 g NH_4Cl, autoclaved and then 1 ml of sterile 1 M $MgSO_4$ and 10 ml of sterile 0.01 M $CaCl_2$ added). Usually 3.5 L batches were used, attempts to scale up the production to 30 L only giving a twofold increase in the amount of enzyme. The culture was maintained at 22°C, with a stirring rate of 450 rpm, and airflow of 800 cm^3/min. Stationary phase was usually reached after 30 hours. All subsequent work was done at 4°C.

The cells were removed by centrifugation at 8000 g for 40 minutes, and the medium brought to 95% saturation with ammonium sulphate. The suspension was then allowed to equilibrate for at least two hours. (If the solution had to be left over night at this stage it was also made 50 mM in EDTA.) The solution was centrifuged (13,800 g for 40 minutes) and the resulting very small pellet was suspended in 100 ml of 0.025 M Tris, 50 mM EDTA 50% ammonium sulphate, adjusted to pH 7.2 with HCl. This suspension was stirred for 1 hour before being centrifuged (17,300 g, 20 min) and the resulting pellet was resuspended in about 20 ml of 0.02 M Tris-HCl pH 7.2, containing 0.15 M NaCl and 5 mM EDTA. Not all the pellet could be dissolved and any insoluble material was removed by centrifugation (17,300 g, 20 minutes). The dark brown supernatant was then applied to an Ultrogel AcA_{44} (LKB, Uppsala, Sweden) column (2.5 x 100 cm). Three main protein peaks were identified and their constituent fractions were pooled. Each of these was concentrated by ultrafiltration to between 0.2 and 0.5 ml (Amicon YM-5 membrane). Initially glycerol was added to 50% (v/v) before storing the sample at -20°C, but it is now recognised that the protein stores better without the addition of glycerol.

IV.2.2 **Endoglycosidase Assay**

Each of the three peaks concentrated from the AcA_{44} column was assayed for endoglycosidase activity using a method developed by Dr G.E. Norris. In this method 5 µl of each peak were added to 100 µl of ovalbumin soln (2 mg/ml) in 0.1 M sodium phosphate buffer, 50 mM EDTA, 0.5% Nonidet P40, pH 6.1 (Nonidet buffer) and incubated at room temperature overnight. The samples were analysed using SDS polyacrylamide gel electrophoresis, using a 10% resolving gel, following the method of Laemmli (1970), and staining with Coomassie Brilliant Blue R_{250}. Endoglycosidase activity was shown by the splitting of the 45 kDa ovalbumin band into two discrete bands (Fig. IV.2).

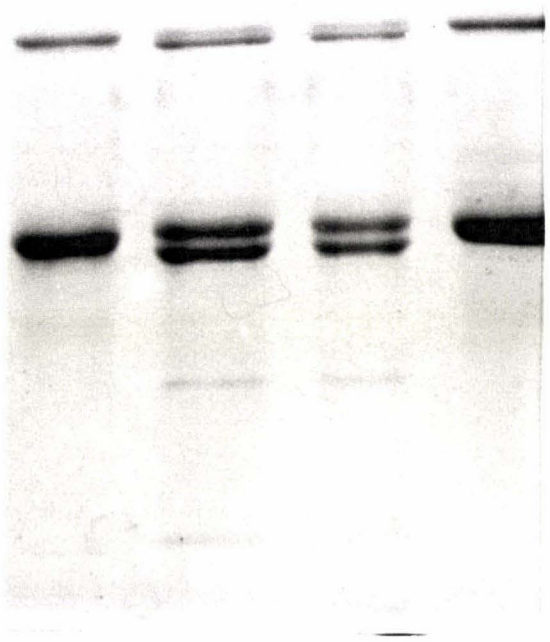


Fig. IV.2 Assay for endoglycosidase activity using ovalbumin as a substrate. Lanes 1 and 4 are untreated ovalbumin, lanes 2 and 3 are ovalbumin treated with Endo F and PNGase. The splitting of the 45 kDa band shows that both glycosylated and deglycosylated ovalbumin are present.

IV.2.3 **Preparation of Human Apolactoferrin**

Human apolactoferrin was prepared following the procedure outlined in Chapter II.

IV.2.4 **Deglycosylation of Lactoferrin**

Apolactoferrin, 200 mg in 40 ml of Nonidet buffer, was incubated at room temperature with 200 µl of the concentrated active endoglycosidase enzyme fraction until deglycosylation was judged to be complete (usually 12-36 hours depending on the activity of the enzyme preparation). For comparative purposes a 200 µl portion of the lactoferrin-Nonidet solution was removed before addition of the enzyme, and kept at room temperature.

The progress of deglycosylation was monitored using SDS polyacrylamide gel

electrophoresis (Chapter II) to observe the reduction in molecular weight (2.5 μg of protein loaded per well). The gels were stained with Coomassie Brilliant Blue R250, and then scanned with an Isco Gel Scanner (Model 1312) at 580 nm. To verify that the decrease in molecular weight was due to loss of carbohydrate and not unwanted protease activity, gels with a heavier loading of lactoferrin per well (20 μg) were also run as above, except that they were stained with thymol-H_2SO_4, and Fuschin Basic; stains which are specific for glycoproteins.

IV.2.5 **Purification of Deglycosylated Lactoferrin**

Gel filtration proved to be an ideal means of separating the deglycosylated lactoferrin (molecular weight 74 kDa) from the endoglycosidases (molecular weight ~35 kDa). Thus, after deglycosylation, the lactoferrin-incubation solution was dialysed against 0.025 M Tris-HCl, 5 mM EDTA, 0.15 M NaCl, pH 7.8 (AcA buffer), prior to precipitation by the slow addition of finely powdered ammonium sulphate. The fraction of protein precipitating between 55-77% ammonium sulphate saturation was centrifuged (17,300 g 15 min), the resulting pellet dissolved in the minimum volume (<10 ml) of AcA buffer, and subjected to gel filtration. The column used was an Ultrogel AcA_{44} column (2.5 x 100 cm) equilibrated in AcA buffer. The peak protein fractions were checked for purity by SDS polyacrylamide gel electrophoresis as before, and those fractions with just a single band migrating at 74 kDa (compared with 80 kDa for native lactoferrin) were pooled and concentrated by ultrafiltration (Amicon 8050 cell, XM50 membrane). Impure fractions were combined, dialysed against 0.025 M Tris-HCl, 0.2 M NaCl, pH 7.8 and subjected to ion exchange chromatography on CM Sephadex C-50 (Pharmacia) (2.5 x 20 cm column) equilibrated with the same buffer. The protein was eluted using a linear salt gradient as previously described in Chapter II. Homogeneous fractions of molecular weight 74 kDa were combined, and concentrated. All work was carried out at 4°C.

IV.2.6 **Determination of Protein Concentration**

Lactoferrin concentrations were determined spectrophotometrically using a Hewlett-Packard 8452 Diode Array Spectrophotometer as described in Chapter II.

IV.2.7 **Tests for Carbohydrate**

If accurate conclusions were to be drawn from this work it was crucial that the protein used in these studies was indeed fully deglycosylated. Therefore in addition to the thymol-H_2SO_4 and Fuchsin Basic stain used to follow the progress of deglycosylation on SDS polyacrylamide gel electrophoresis, two further quantitative tests were used to determine the completeness of deglycosylation.

(i) The phenol-sulphuric acid test (Ashwell, 1966) was used to determine the sugar content of native and deglycosylated lactoferrin. All samples were analysed in duplicate,

read against a blank containing water instead of the protein solution, and compared to a standard plot based on glucose standards. A control standard of lactoferrin, concentrated sulphuric acid and water in place of phenol was also used.

(ii) Amino acid analysis. Analysis of N-acetylglucosamine content was carried out by Drs G.E. Norris and J. Reid. To both native and deglycosylated lactoferrin (2 mg) were added 0.5 ml of 5.9 M HCl containing 1 mg/ml of phenol. Samples were sealed under vacuum and hydrolysed for 24 hours at 110°C. Amino acid analyses were carried out on samples of approximately 0.2 mg using a Beckmann 199 BL amino acid analyser.

IV.2.8 **Iron binding**

The iron content of the deglycosylated and native apolactoferrin solutions were determined as outlined in Chapter II, by comparing their spectral ratios $A_{280\ nm}/A_{466\ nm}$ and $A_{412\ nm}/A_{466\ nm}$ with those of the fully iron-saturated protein (ratios 20-22 and 0.70-0.74 for native lactoferrin and 19-20 and 0.70-0.74 for deglycosylated lactoferrin).

Iron-binding experiments were carried out at room temperature, and the iron binding followed spectrophotometrically by monitoring the change in absorbance at 466 nm using a Hewlett Packard 8452 Diode Array Spectrophotometer. Freshly prepared FeNTA solution (Chapter II) was added to a solution of apolactoferrin which was equilibrated in 0.05 M Tris-HCl, 0.025 M $NaHCO_3$, 0.2 M NaCl pH 8.0. Two mole equivalents of FeNTA were added to the lactoferrin (1 ml) in a quartz cuvette, using a microsyringe, and the rise in absorbance at 466 nm recorded.

IV.2.9 **Heat Treatment and Iron binding**

The thermostabilities of native and deglycosylated lactoferrin were compared by measuring the effect of heat treatment of each protein on iron-binding. Briefly, native and deglycosylated lactoferrin were heated to 50°C in a thermostatted water bath, held at this temperature for 1 hour, cooled to 20°C and then the iron-binding measured as in Section IV.2.8. Experiments were repeated at 55°, 60°, 65° and 70°.

IV.2.10 **pH-Mediated Iron Release**

The effect of pH on iron release from native and deglycosylated lactoferrin was determined by dialysis of protein samples against solutions of different pH. A series of buffers covering the pH range 8.0-2.3 were prepared (0.05 M Tris-HCl, pH 8-7.0; 0.05 M ammonium acetate, pH 6.5-3.5; 0.1 M glycine-HCl, pH 3-2.0 each containing 0.2 M NaCl). Samples were dialysed for 48 hours against each buffer (steps of 0.5 pH unit), and measurements made of the pH (Radiometer pHM82 Standard pH meter) and absorbance at 466 nm (Hewlett-Packard Diode Array spectrophotometer). The percentage of iron saturation at each pH was calculated as ($A_{466\ nm}$ at pH)/($A_{466\ nm}$ at pH 8.0).

IV.2.11 **Circular Dichroism Spectra**

Circular Dichroism (CD) spectra were recorded on a Jobin/Yvon V dichrograph linked to an Apple II computer and a Hewlett Packard plotter. Light from a Xenon arc source is passed through a double monochromater which produces a beam polarised in the horizontal plane. A birefringence modulator then splits the light into left and right polarised beams. These are focussed to a thin line in the sample cavity and a detector measures the absorbance of the two beams (A_L and A_R). Samples are held in cylindrical quartz cuvettes with a path length of 0.1 cm for the UV and far UV regions. The spectrum of the protein solution was first recorded, the zero being set at a wavelength at which there were no observed dichroic bands. The spectrum of the buffer was then recorded, using the same settings as for the protein, and subtracted from the protein spectrum.

The spectra were recorded in the 250-350 nm region using a step size of 0.5 nm, sensitivity 1 x 10^{-6}, response time 1 s, slit width 20 nm, and path length 0.1 cm. Apolactoferrin was dissolved in 0.02 M Tris-HCl/0.2 M NaCl, pH 7.8, with protein concentrations of 4.1 mg/ml for the native protein and 4.6 mg/ml for the deglycosylated protein. Molar ellipticities (θ) were calculated using the equation

$$[\theta] = \frac{33\ M\ (A_L - A_R)}{100\ l.c.} \text{ degrees cm}^2 \text{ d mole}^{-1}$$

where M is the molecular mass (74,000 Da for deglycosylated lactoferrin, 80,000 Da for native lactoferrin, calculated values) l = path length in dm, c = concentration in g/ml. The numbers 33 and 100 are included in the equation for historical reasons.

IV.2.12 **Proteolysis Experiments**

A comparison of the susceptibility of native and deglycosylated lactoferrin to proteolytic attack by trypsin was monitored using SDS polyacrylamide gel electrophoresis. Solutions of iron-saturated native and deglycosylated lactoferrin (1.6 mg/ml) in 0.05 M Tris-HCl, pH 8.0, were incubated at 37°C with bovine trypsin (Sigma T-1005). Two sets of experiments were carried out, one with an enzyme-to-substrate ratio of 1:50 w/w, the other with enzyme-to-substrate ratio of 1:25 w/w. Aliquots were removed after 10 min, 30 min, 1, 2, 3, 4 and 24 hours. The reaction was stopped by the addition of an equal volume of an SDS denaturing solution (0.125 M Tris-HCl, pH 6.8, 4% v/v SDS, 0.075 M dithiothreitol, 20% glycerol), followed by heating to 80°C for 2 minutes. Control samples of native and deglycosylated lactoferrin were incubated under the same conditions, but without the presence of trypsin. The incubated samples were analysed by SDS polyacrylamide gel electrophoresis, and stained with Coomassie Blue.

IV.2.13 **Structural Comparisons of Native and Deglycosylated Lactoferrins**

The glycans of some glycoproteins are involved in control and/or maintenance of protein folding and conformation, and may also confer dynamic stability and rigidity on the molecule. Structural effects caused by the removal of the carbohydrate moieties from lactoferrin could include:

(i) a change in relative domain orientations,

(ii) an alteration of the three dimensional structure within a domain,

(iii) an increase in flexibility of the molecule especially within the vicinity of the glycosylation sites.

Four crystal structures (two glycosylated, two deglycosylated) were used to compare the three dimensional structures and thermal parameters of glycosylated lactoferrins with deglycosylated lactoferrins. These were:

Diferric lactoferrin (Fe_2Lf)	glycosylated (Haridas *et al*, 1994)
Diferric dioxalatolactoferrin (Fe_2Lf)	glycosylated (This work)
Apolactoferrin (ApoLf)	deglycosylated (Anderson *et al* , 1987)
Recombinant N-lobe lactoferrin ($FeLf_N$)	deglycosylated (Day *et al* , 1992)

Relative domain orientations were compared for the N-lobe only, by superimposing $FeLf_N$ (deglycos) and Fe_2Lf (glycos). ApoLf could not be used in this comparison because iron binding itself causes a large rotation (54°) of one domain relative to the other (Anderson *et al*, 1990).

Superpositions of the Cα atoms of the N2 domains (residues 92-250, containing the glycosylation site at Asn 137) were used to compare three dimensional structures within that domain.

The effect of glycosylation on the flexibility of lactoferrin was analysed by comparing the thermal parameters (B factors) of the above proteins. The thermal parameter (B factor) is related to the amplitude of vibration of an atom about its mean position. In practical terms it includes contributions from the vibration of an atom and from any discrete disorder in its position, ie, it includes both dynamic movement and static disorder. An increase in the B factors of a section of polypeptide is therefore an indication of an increase in disorder or flexibility of that region.

Thermal parameters (B factors) were calculated for the mainchain and the sidechain atoms, an overall B value calculated for each protein and a plot of B factor versus residue number was made using the program BAVERAGE (version 2.12) (The CCP4 Suite of Programs for Protein Crystallography, 1994.) Plots of B factor versus residue number show that following deglycosylation there was an increase in the average B factor for the polypeptide loop around the site of glycosylation at Asn 137 compared with the rest of the domain. The average B factor of this loop was normalised for each protein using the ratio:

$$\frac{\text{Average B factor for residues 134-144}}{\text{Overall B factor for the protein}}$$

This figure was used to compare the flexibility of the glycosylation loop between one protein and another.

IV.3 RESULTS

IV.3.1 Preparation of Endoglycosidase

Analysis of endoglycosidase activity (using ovalbumin (Fig. IV.2)) showed the activity to be concentrated in the first two protein peaks (Fig. IV.3) eluting from the AcA_{44} gel filtration column used in the purification of the enzymes. Plummer *et al* (1984) have shown that the endoglycosidase prepared this way is in fact a mixture of two activities; endo-β-N-acetylglucosaminidase F (Endo F), and peptide:N-glycosidase F (PNGase F).

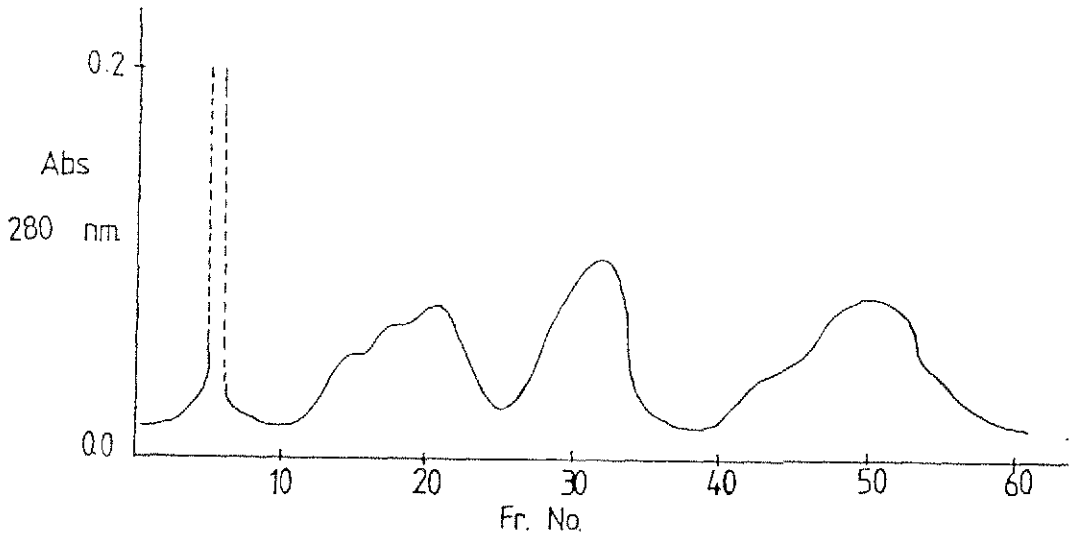


Figure IV.3 Elution profile for purification of endoglycosidase, using AcA_{44} gel filtration.

IV.3.2 Deglycosylation of Human Apolactoferrin

The deglycosylation of lactoferrin, when incubated with the endoglycosidase preparation, was quick, clean, specific and essentially complete. SDS polyacrylamide gel electrophoresis showed that after 16 hours incubation, lactoferrin was represented by only one band, with the molecular weight being reduced from ~80,000 Da for the native protein to 74,000 Da for the incubated protein (measured by peak maxima (Fig. IV.4).

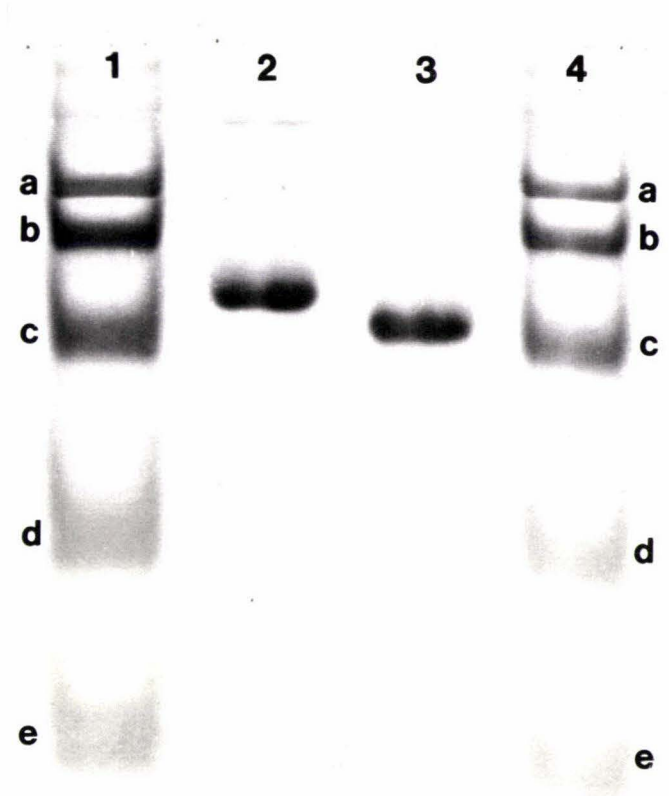


Figure IV.4 SDS gel monitoring the deglycosylation of human apolactoferrin. Lanes 1 and 4, molecular weight markers (Sigma M.W. SDS 200 kit: a 116 kDa, b 97.4 kDa, c 66 kDa, d 45 kDa, e 29 kDa), lane 2 native lactoferrin, lane 3 lactoferrin after 16 hours incubation with endoglycosidases.

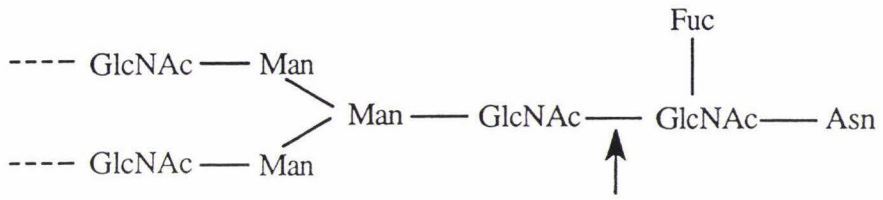
The purity of the deglycosylated protein was also assessed from the spectral ratios of $A_{280}/A_{466\ nm}$ and $A_{412}/A_{466\ nm}$. The values of 19-20 and 0.70-0.74 obtained are comparable to those of pure native lactoferrin. The fact that this deglycosylated protein has also been crystallised in its apo form (Norris *et al*, 1989) is another indication of its purity. Approximately 70% of the starting material was recovered as pure deglycosylated protein; a result similar to that obtained for other incubations (Tarentino *et al*, 1974).

Since the endoglycosidase prepared consisted of a mixture of Endo F and PNGase F (Plummer *et al*, 1984) with pH optima of 4-6 and 8.5 respectively (Tarentino & Plummer, 1987), the incubation was tried at several different pH values (pH 4.0, 6.0, 7.0 and 8.0). Deglycosylation occurred at each pH, but it was found to be most efficient at pH 6.0. Thus these results suggest that in this case it is the Endo F which is primarily responsible for the deglycosylation.

IV.3.3 **Confirmation of Deglycosylation**

SDS polyacrylamide gels stained with thymol-sulphuric acid showed a transient pink band appearing in the lanes with the control native lactoferrin, but none in the lanes with the endoglycosidase-incubated lactoferrin. This indicates that the latter contained little or no carbohydrate.

If Endo F is primarily responsible for the deglycosylation of lactoferrin then cleavage should occur as shown below.



This would leave one N-acetylglucosamine residue (GluNAc) remaining at each glycosylation site (Plummer *et al*, 1984), giving a total of two for deglycosylated human lactoferrin, whereas an average of nine would be expected for native lactoferrin (Spik *et al*, 1982). Thus an analysis of glucosamine peaks in an amino acid analysis of both native and deglycosylated lactoferrin should give a more quantitative estimate of the sugar content of the deglycosylated protein.

In the amino acid analysis, glucosamine appeared as a small peak eluting immediately after phenylalanine. The ratio of glucosamine peaks found for native and deglycosylated lactoferrin was 4:1 (after standardisation of the samples using methionine peaks, and integration of peak areas). This compares reasonably well with the predicted ratio of 9:2 and suggests that in the deglycosylated protein all except the two N-linked sugars have been removed.

The phenol-sulphuric acid test, which also gives a more quantitative analysis of the carbohydrate content gave similar results (Fig. IV.5).

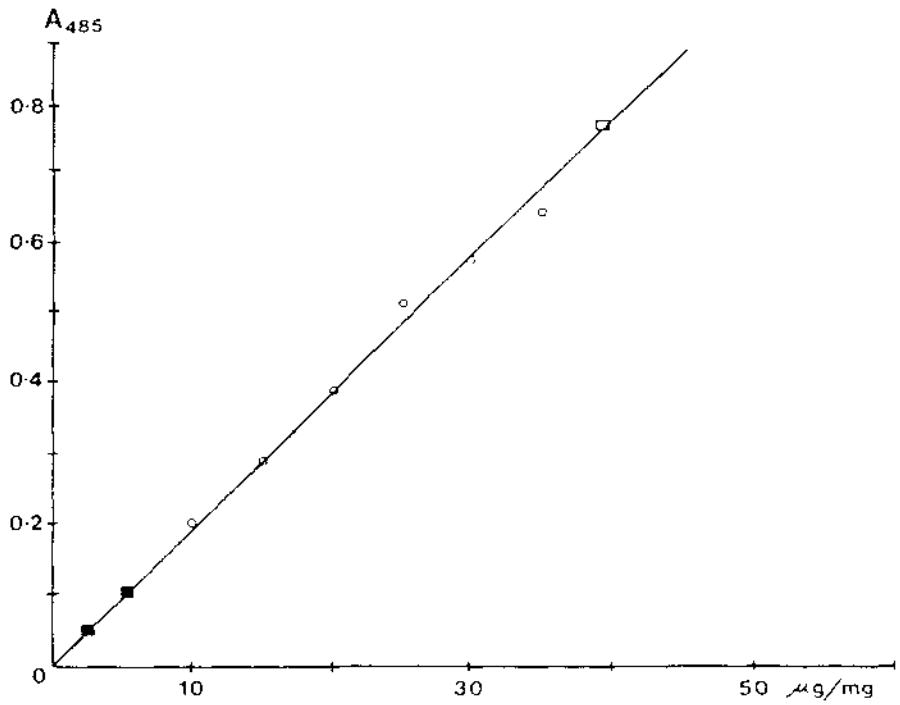


Figure IV.5 Sugar analysis by phenol-sulphuric acid test based on glucose standards. Plot of absorbance at 485 nm (A_{485}) versus sugar content in µg glucose ml^{-1} protein. Standard values shown O, native lactoferrin □, deglycosylated lactoferrin ■ (duplicates)

The standard plot of absorbance at 485 nm versus glucose concentration indicates that on our results native lactoferrin has a sugar content of 43 µg/mg compared with ~55 µg/mg calculated from the average number of reacting sugars in the primary structure (Spik *et al*, 1982). The extinction coefficients vary for different sugars, and some are inert to the test (Lee and Montgomery, 1961); thus the values obtained for a mixture of sugars, as in lactoferrin, will not be related directly to the extinction coefficient of glucose. The value is however indicative of the sugar content of the native protein. For deglycosylated lactoferrin, the value was only 9% of the original (4 µg/mg (deglycosylated lactoferrin) compared with 43 µg mg (native lactoferrin)).

These results are consistent with the removal of all but the two N-linked sugars and their attached fucose which should leave 7-8 µg/mg, ie 14% of the native lactoferrin, and agree well with the amino acid analyses. Similar results were found by Tarentino *et al* (1974).

IV.3.4 Circular Dichroism

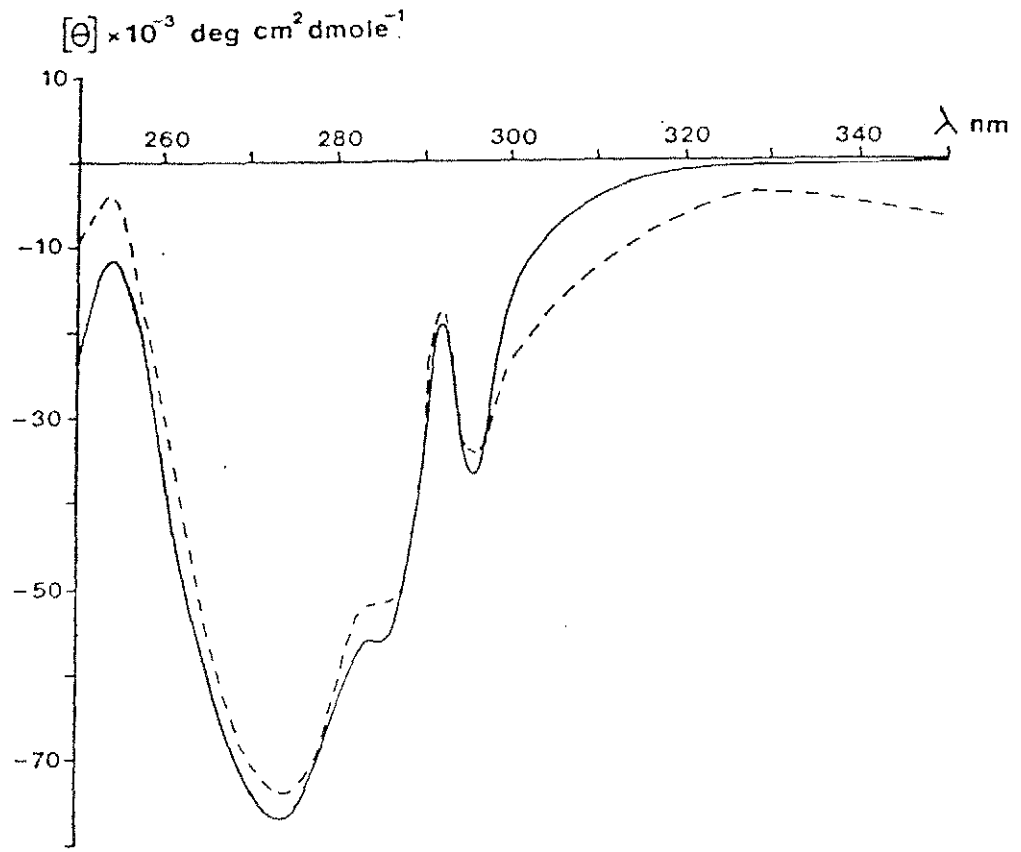


Figure IV.6 CD spectra in the range 250-350 nm. Broken line native lactoferrin, full line deglycosylated lactoferrin.

Fig. IV.6 shows the CD spectra for native and deglycosylated apolactoferrin in the range 250-350 nm. (The region below 250 nm could not be reliably accessed with the instrument used, while spectra in the region 350-700 nm were essentially featureless since the iron-free protein was used.) Three negative bands are seen for both the native and the deglycosylated lactoferrin; a broad band around 275 nm, and two smaller bands at 283 and 296 nm superimposed on the former. Their shapes and peak positions are extremely similar in the two spectra and similar also to the spectrum measured for apolactoferrin by Mazurier *et al* (1976). As these bands are due primarily to the aromatic residues in the protein (Trp, Tyr and Phe in order of intensity), this implies that there are no significant differences in the relative positions and environments of the aromatic sidechains between the native and deglycosylated proteins.

IV.3.5 Iron-binding Properties

Experiments here indicate that deglycosylated lactoferrin appears to have iron binding properties that are identical to those of the native protein. The UV-visible spectrum of the deglycosylated lactoferrin was the same as that of the native protein, except that the deglycosylated protein typically had a slightly lower ratio $A_{280}/A_{466\,nm}$ than the latter (19-21 compared with 20-22). This could have been due to the loss of

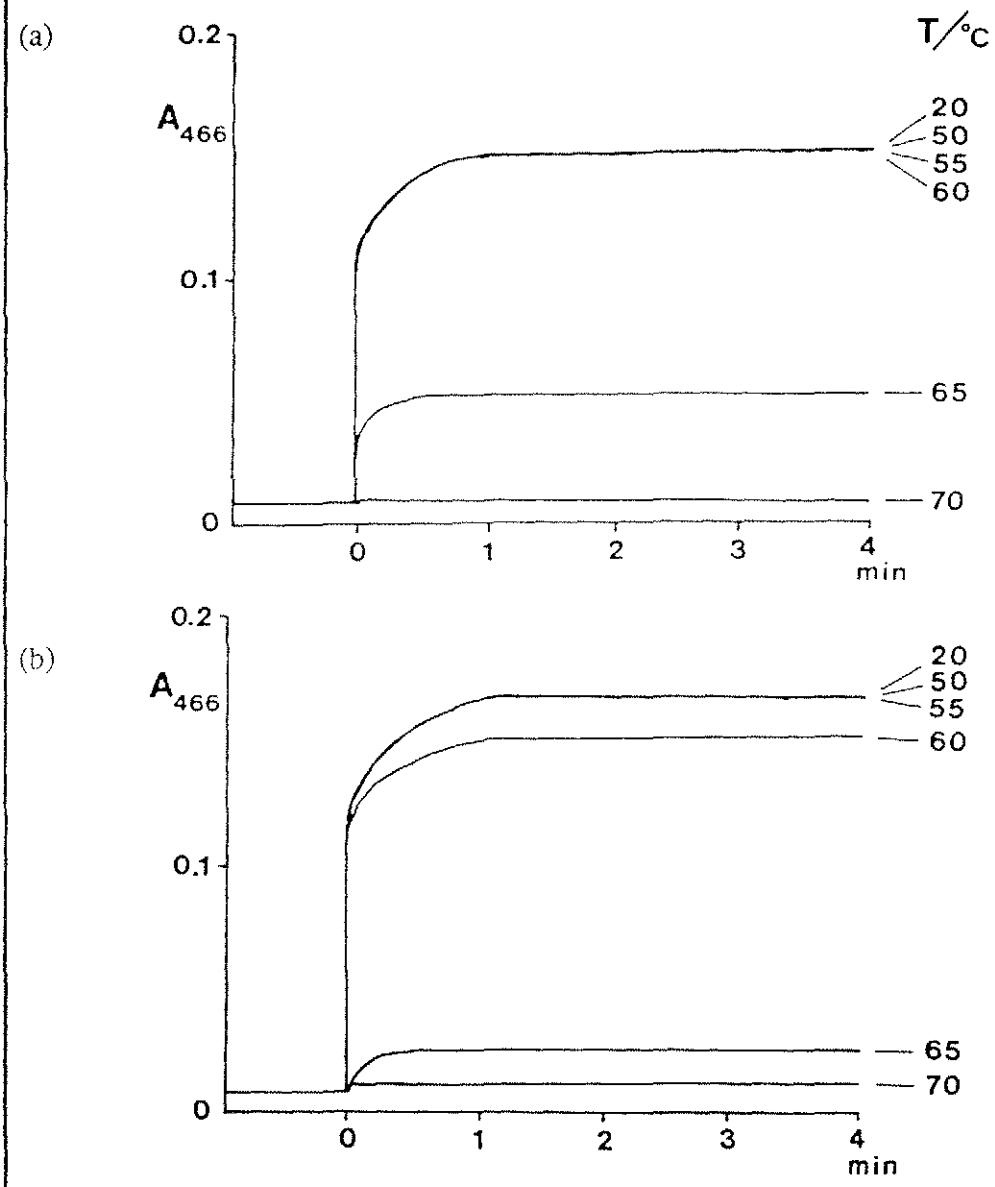


Figure IV.7 Iron binding by (a) native, and (b) deglycosylated apolactoferrin. The rate of iron binding is monitored by the increase in absorbance at 466 nm. The effects of pretreatment by heating at various temperatures are also shown, with the binding curves for samples previously heated for 1 hour at 50°, 55°, 60°, 65° and 70°C shown. In all cases the initial protein concentration (before heat treatment) was 3 mg/ml.

some absorption by carbohydrate at 280 nm. Experiments monitoring iron binding (as measured by the absorbance at 466 nm) as a function of time show that there is no apparent difference in the binding curves for the native and deglycosylated proteins and that in both cases binding is essentially complete within ~1 min (see Fig. IV.7).

The release of iron as a function of pH also follows the same pattern for both native and deglycosylated lactoferrin (Fig. IV.8).

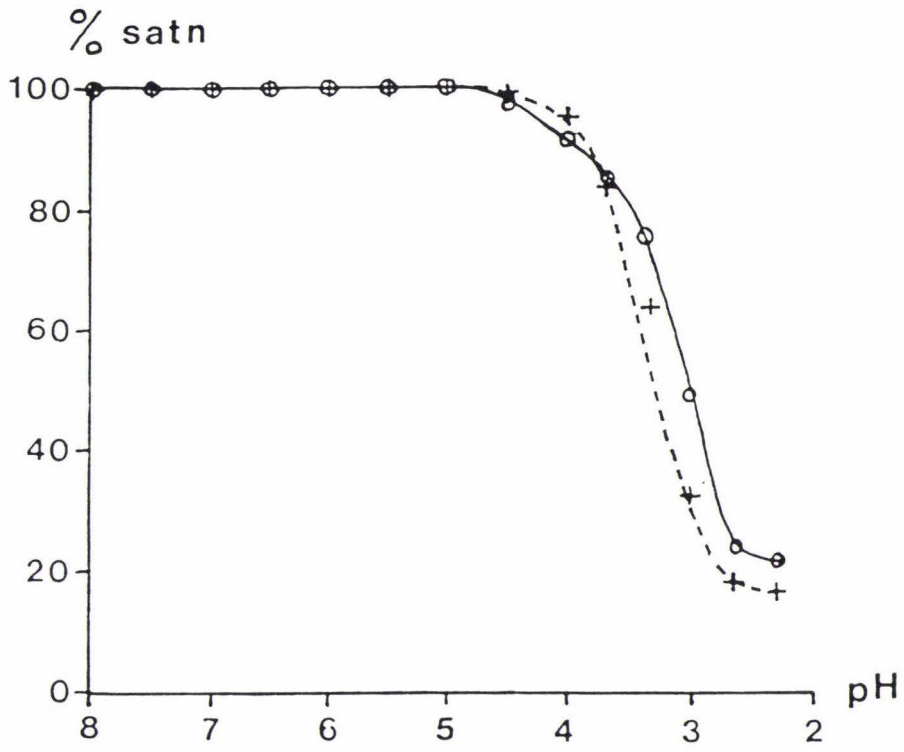


Figure IV.8 pH mediated iron release from native and deglycosylated lactoferrin. The percentage saturation with iron (calculated from the absorbance at 466 nm) is shown for native lactoferrin (dashed line) and deglycosylated lactoferrin (solid line), as the pH is reduced from 8 to 2.3. The differences between the two curves at lower iron saturations are probably not significant since the absorbance at 466 nm is then very low.

Both these curves are very similar to the results of Mazurier and Spik (1980) for native human lactoferrin at low ionic strength (I = 0.1), and emphasize the lack of any effect of the carbohydrate on iron binding and release *in vitro*. These results are in sharp contrast to those of Legrand *et al* (1990) which indicated a loss of 50% of iron-binding capacity of human lactoferrin when the carbohydrate was removed. These results will be

discussed further in Section V.2.2.

IV.3.6 **Protein Stability**

For some proteins, the carbohydrate has been shown to confer stability, eg by increasing resistance to proteolysis or thermal denaturation (Tashiro & Trevithick, 1977; Chu *et al*, 1978). As discussed in the introduction to this chapter, however, this property is by no means universal (Ploegh *et al*, 1981; Olden *et al*, 1982).

(a) *Resistance to proteolysis*

The effects of proteolysis by trypsin on native and deglycosylated lactoferrin are shown in the gel in Fig. IV.9. It can be seen that although some degradation occurs fairly quickly, even after 24 hours at 37°C most of the lactoferrin remains intact. The small amount that has been broken down has been cleaved into two major fragments corresponding to the 30 kDa and 50 kDa fragments characterised by Legrand *et al* (1984).

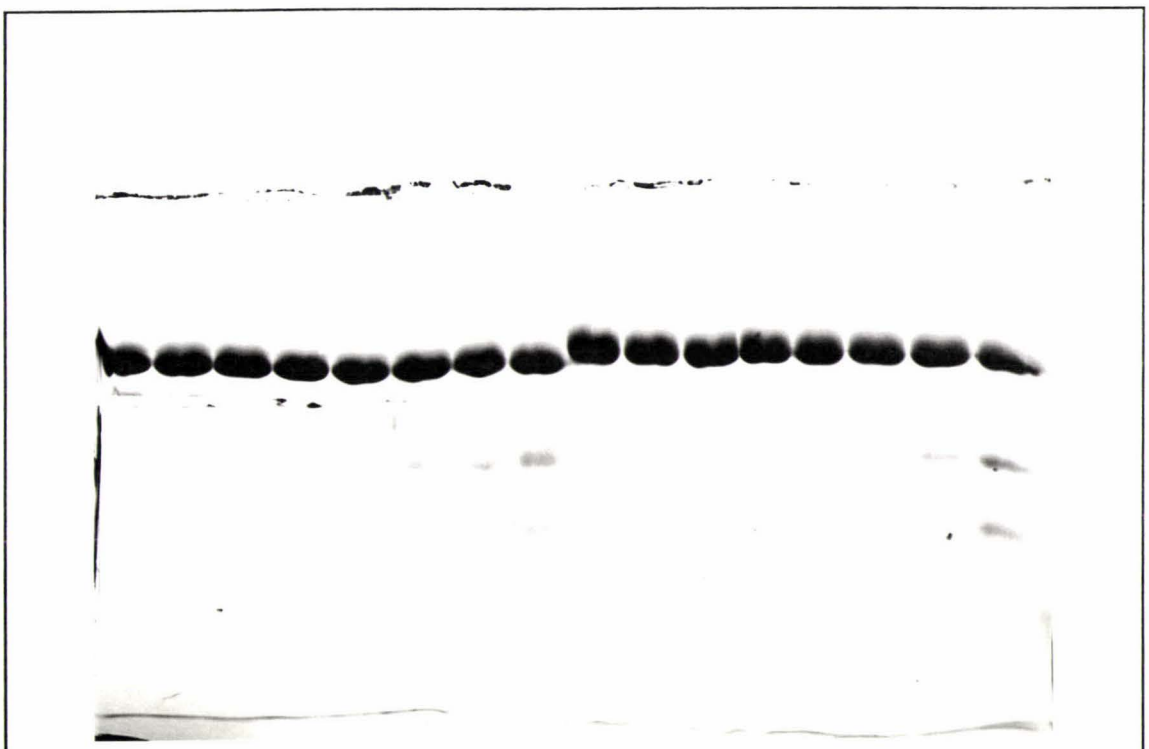


Figure IV.9 SDS polyacrylamide gel showing the incubation of native and deglycosylated lactoferrin with trypsin at 37°C. Lane 1 deglycosylated lactoferrin control, lanes 2-8 deglycosylated lactoferrin incubated with trypsin for 10, 30 min, 1, 2, 3, 4 and 24 hours; lane 9 native lactoferrin control, lanes 10-16 native lactoferrin incubated with trypsin for 10, 30 min, 1, 2, 3, 4 and 24 hours.

These gels are a striking demonstration of the similarity of the fragmentation patterns for the native and deglycosylated proteins, and are strong evidence that for lactoferrin the presence of carbohydrate does not confer any additional resistance to proteolytic attack, at least by trypsin. Results of experiments using an increased enzyme concentration (enzyme:substrate 1:25) were essentially identical. These results are also in sharp contrast to those reported by Spik *et al* (1985). These indicated that "partially deglycosylated lactoferrin" subjected to hydrolysis by neuraminidase, β galactosidase, and N-acetyl-β-glucosamidiase) was completely destroyed by trypsin in 4 hours in contrast to the native protein (ratio of trypsin:lactoferrin 1:50). (See Section V.2.2 for discussion of these results.)

(b) *Thermal stability*

No difference in the thermostability of the native and deglycosylated lactoferrin was apparent up to 55°C. Comparisons of the heat stabilities of the two proteins were based on measurements of their residual iron binding ability after heating for one hour at various temperatures, and for both native and deglycosylated apolactoferrin, samples heated at 50°C and 55°C showed identical binding curves (Fig. IV.7), and spectral ratios identical to those of unheated protein. A small amount of precipitation occurred at 60°C for the deglycosylated (but not native) protein. After centrifugation, however, the remaining soluble protein (~90% of original) retained full iron binding capacity; the binding curve (Fig. IV.7) was unchanged and the spectral ratio was similar to the unheated protein (although the protein concentration and absorbance at 466 nm were necessarily reduced). In spite of considerable precipitation of both glycosylated and non-glycosylated proteins at 65°C, the remaining soluble protein was still able to partially bind iron (spectral ratios of $A_{280}/A_{466\ nm}$ = 38 (native) and 62 (deglycosylated) corresponding to ~66% and ~40% binding respectively). At 70°C almost all protein had been precipitated for both native and deglycosylated lactoferrins, and no iron binding capacity was observed in the remaining soluble material.

These results show that the two forms of lactoferrin have very similar heat stabilities, although the deglycosylated protein appears to be slightly less stable. This has been presented graphically in Fig.. IV.10 where the percentage of iron binding capacity remaining after heating is plotted as a function of heating temperature.

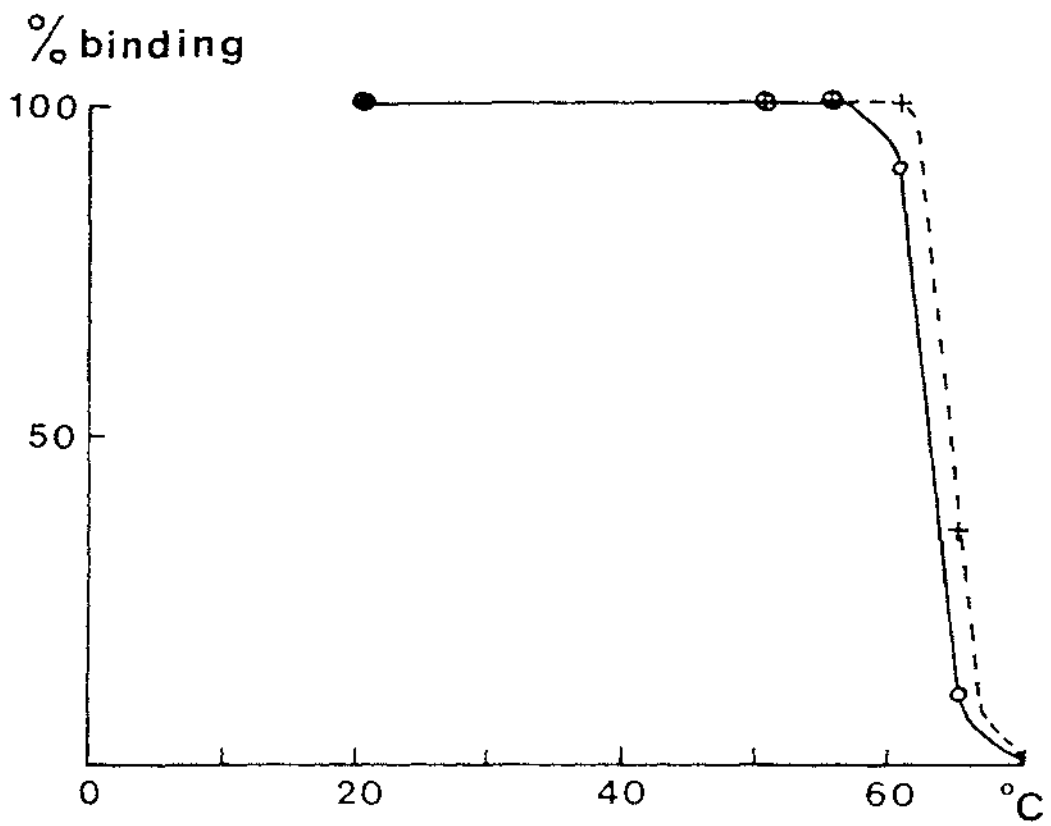


Figure IV.10 Effect of heating lactoferrin. % of iron binding by samples of native (dotted line) and deglycosylated (solid line) human lactoferrin are shown after prior heating for 1 hour at various temperatures. Levels of iron binding are calculated from the absorbance at 466 nm for the heated protein (after centrifugation and addition of FeNTA) compared with the unheated protein.

The melting temperature (estimated as the temperature at which 50% of iron binding capacity remained) of each protein, calculated from this graph is ~63°C for deglycosylated and 64°C for native apolactoferrin. It is interesting to note that in a previous study of heat stability of bovine lactoferrin (Baer *et al*, 1979), where neither the conditions of heating nor the assays were comparable to this work, complement fixation measurements showed a 50% loss of activity at ~60°C, - results similar to those reported here for human lactoferrin.

IV.3.7 Structural Comparisons of Deglycosylated and Native Lactoferrins

(a) *Comparison of domain orientations*

Comparison of $FeLf_N$ (deglycosylated) with the N-lobe of Fe_2Lf (glycosylated) show that with the exception of the C-terminal residues 313-327, the polypeptide chain

conformation is essentially identical. Superposition of residues 4-312 (ie. the whole N-lobe with the exception of the C-terminal peptide) gives an *rms* deviation in Cα atom positions of 0.41 Å indicating very little difference in the orientations of N1 and N2 domains between the two proteins. There is a major conformational difference involving the C-terminal residues of $FeLf_N$ 321-333 (Day *et al*, 1993). These residues form a helix and are part of the packing between the two lobes of Fe_2Lf, so it is reasonable to attribute this change in conformation to the absence of the C-lobe and not to the effect of deglycosylation.

(b) *Comparison of structure within a domain*

Superpositions of the glycosylated and deglycosylated N2 domains (residues 92-250) indicate that in all cases the overall conformations are essentially identical, with *rms* deviations varying between 0.56 Å and 0.29 Å (Table IV.2). In fact the extent of agreement between proteins does not appear to be associated with the state of glycosylation at all. Rather, the main trend shown in these superpositions in that the *rms* deviations associated with the N2 domain of ApoLf are consistently higher, suggesting that the absence of iron has a greater effect on the conformation within this domain than the presence or absence of the glycans.

Table IV.2 *RMS* deviation in Cα positions when N2 domains of Fe_2Lf, ApoLf, $Fe_2(C_2O_4)_2Lf$ and $FeLf_N$ are superimposed on each other

	ApoLf	$Fe_2(C_2O_4)_2Lf$	$FeLf_N$
$Fe_2(C_2O_4)_2Lf$	0.56 Å	-	-
$FeLf_N$	0.50 Å	0.35 Å	-
Fe_2Lf	0.48 Å	0.29 Å	0.29 Å

FeLf$_N$

ApoLf

$Fe_2(C_2O_4)_2Lf$

Fe_2Lf

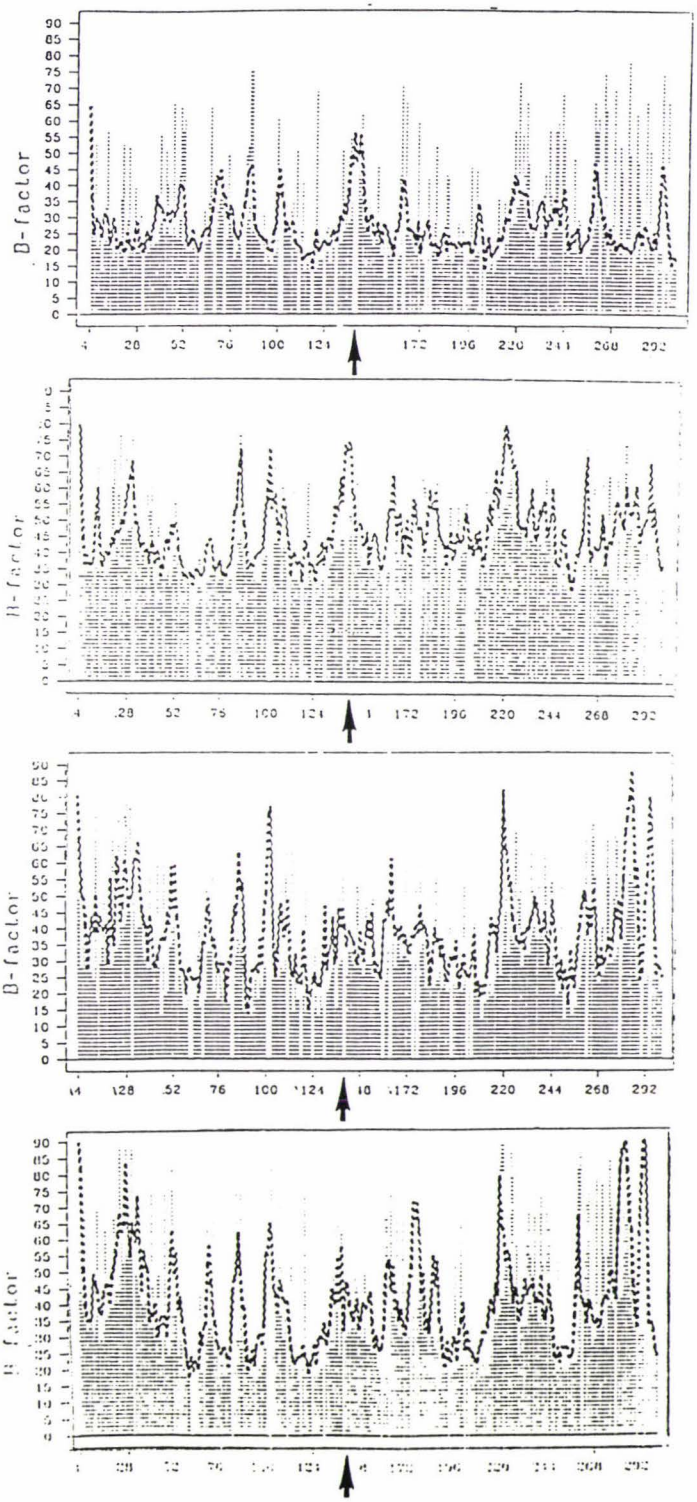


Fig. IV.11 Plots of B values against residue number for the N-lobe of (top to bottom) FeLf$_N$, ApoLf, $Fe_2(C_2O_4)_2Lf$ and Fe_2Lf. Plots show the increase in B values of residues around the glycosylation site 137 (arrowed) when the carbohydrate is removed.

(c) *Comparison of B values*

The distribution of B values along the polypeptide chain of the N-lobes of both glycosylated and deglycosylated lactoferrin is shown in Fig. IV.11. Normalised average B values around the glycosylation site are shown in Table IV.3.

Table IV.3 Comparison of normalised B values of residues around the glycosylation site 137 for glycosylated and deglycosylated lactoferrins

	Protein	$\bar{B}$* 134-144
Glycosylated	Fe_2Lf	38
Glycosylated	$Fe_2(C_2O_4)_2Lf$	39
Deglycosylated	ApoLf	51
Deglycosylated	$FeLf_N$	52

$$*\bar{B} = \frac{\text{Average B for residues 134-144}}{\text{Average B for the protein}}$$

Normalised average B values for the polypeptide loop 134-144 are ~39 when the protein is glycosylated, but this has increased to ~52 when the carbohydrate is removed (Table IV.3). Thus although deglycosylation does not appear to be associated with any change in the overall conformation from that of glycosylated lactoferrin, there is an increase in the general mobility of the polypeptide chain in the vicinity of the site of carbohydrate attachment (Asn 137) (see Fig. IV.11).

Discussion of these results and possible reasons for the conflict with earlier reports is continued in Section V.2.

CHAPTER V

DISCUSSION

This thesis has considered two aspects of lactoferrin structure and function. First is the anion-binding ability of lactoferrin. Although lactoferrin has not been as extensively investigated as transferrin, both proteins seem to show considerable versatility in their anion binding activities (Schlabach & Bates, 1975; Smith *et al*, 1991; Eaton *et al*, 1989). The structural basis of this forms the first part of the thesis. Second is the role of glycosylation of lactoferrin. All lactoferrins so far characterised have proved to be glycosylated, but no convincing function has yet been attributed to the attached glycans. The use of recombinant lactoferrin in which the glycosylation is either different or not present at all makes it essential that the effects of deglycosylation be investigated.

V.1 ANION BINDING BY LACTOFERRIN

The present crystal structure provides the first direct demonstration of the mode of binding of an anion other than carbonate in both lobes of a transferrin, and thus provides an experimental test of many of the spectroscopic studies that have been directed towards this question over the years. It has also provided direct evidence, for the first time, of the factors which lead to inequivalence between the two sites, and it leads to a fuller understanding of the extent to which the protein can adapt to the binding of different anions. This, then, allows some predictions to be made about other anions.

V.1.1 Anion Binding and the Schlabach Bates Model

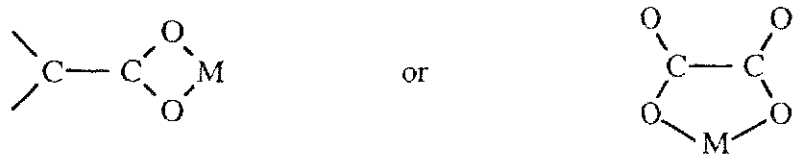
The early visible spectral studies of the binding of a wide variety of anions to transferrin by Schlabach and Bates (1975) were astutely interpreted to show that the anion acted as a bridge between a positively charged region of the protein and the metal. The subsequent crystal structure of human lactoferrin (Anderson *et al*, 1987, 1989) showed that this was correct. Nevertheless, the proposed interlocking-sites model also implied that dicarboxylic acids such as oxalate, malonate and maleate would coordinate to the metal ion in a monodentate or perhaps a 1,1-bidentate manner, ie. that one end interacted with the metal and the other end with the protein (see Fig. I.10.). It was suggested that other anions bound to the iron *via* the proximal electron donor group, while the carboxylate group interacted with a positively charged protein group. These ideas form one of the main propositions that this present work was intended to test.

V.1.2 Oxalate Binding to Transferrins

Many of the studies aimed at investigating the binding of synergistic anions to transferrins have been based on the use of oxalate. There appear to be two main reasons

for this, (i) that oxalate appears to bind more strongly than any of the other anions apart from carbonate (Schlabach & Bates, 1975), and (ii) that oxalate is the simplest of the other anions, and one with a well-known coordination chemistry. Complexes of oxalate and iron even occur naturally, eg as the complex mineral humboltine, $FeC_2O_4.2H_2O$.

With two carboxylate groups oxalate has the potential for considerable versatility in its coordination. Each carboxylate group has the potential to act either as a unidentate ligand or to act as a bidentate chelating ligand in a 1,1-bidentate or η^3 chelation mode as shown below. Alternatively, since the two carboxylate groups are close together, linked by a single C-C bond, oxalate can act as a chelate ligand using one oxygen from each of the COO^- groups, ie 1,2-bidentate coordination or η^4 chelation.



1,1-bidentate coordination or η^3 chelation

1,2-bidentate coordination or η^4 chelation

Fig. V.1 Two modes of oxalate coordination from small molecule structures.

Although there are examples of dibasic acids acting as monodentate ligands, as in the Schlabach-Bates model, in small molecule systems it is even more common to find them as chelating ligands (Oldham, 1987). Furthermore, although η^3 coordination is found for higher members of the dicarboxylate series, the four-membered ring introduces steric strain, and for the lower members, certainly oxalate and possibly malate and malonate, there is a strong tendency to form chelate rings with more than 4 atoms, eg by adopting (in the case of oxalate) the 1,2-bidentate mode.

Most of the studies of oxalate binding to transferrins have focused on the coordination mode of the anion. Some, such as the NMR study of ^{13}C oxalate binding to the transferrins, were interpreted as indicative of 1,1-bidentate coordination of oxalate to the metal (Bertini *et al*, 1986). ^{13}C NMR spectra of Al^{3+}, Ga^{3+} and Zn^{2+} derivatives of both human serum transferrin and ovotransferrin with ^{13}C-enriched oxalate, showed a strong inequivalence between the two carboxylate carbons. It was argued that this was not only supportive of the 'interlocking sites' model of Schlabach and Bates, but also indicative of monodentate oxalate binding to the metal. Later studies from the same group (eg Bertini *et al*, 1986) employing CD and ^1H NMR spectroscopy of cobalt derivatives of ovotransferrin, were interpreted also in terms of monodentate binding of oxalate. As late as 1990, after the crystal structure of human lactoferrin showed carbonate binding in a

bidentate mode, ^{113}Cd and ^{13}C NMR spectroscopy of the Cd^{2+} derivative of ovotransferrin containing oxalate were still interpreted as indicative of monodentate binding of oxalate to the metal (Sola, 1990).

In contrast to the above studies, magnetic measurements of electron spin echo envelope modulation (ESEEM) for human lactoferrin and transferrin were interpreted in terms of 1,2-bidentate oxalate coordination. These studies showed a deeper ^{13}C modulation for copper and vanadyl transferrin ^{13}C oxalate than vanadyl ^{13}C carbonate (Eaton *et al*, 1989). Since there is only one labelled carbon in carbonate, and the metal to ^{13}C distances are similar for coordinated carbonate and oxalate, the deeper modulation for the oxalate complex argued strongly for two interacting carbons. The interaction with two equivalent or nearly equivalent carbon nucleii indicated that the two ends of the oxalate were bound to the metal, ie a 1,2-bidentate coordination.

Recent EXAFS (extended X-ray absorption fine structure) measurements on oxalate-substituted diferric transferrin have also been interpreted in terms of 1,2-bidentate oxalate coordination (Mangani & Messori, 1992). In this case, however, knowledge of the 1,2-bidentate oxalate coordination seen crystallographically for the mono-oxalato copper-lactoferrin complex (Shongwe *et al*, 1992) played a significant part in the interpretation.

The present study shows that in both binding sites the oxalate binding mode is broadly the same, ie 1,2-bidentate coordination. These results, when combined with the similar EPR spectra for oxalate complexes of both lactoferrin and transferrin (Aisen & Leibman, 1972; Dubach *et al*, 1991) indicate that this is the normal binding mode for oxalate to transferrins. The non-equivalence of the two oxalate carbons in ^{13}C NMR spectra of ^{13}C-oxalate substituted Al^{3+}, Ga^{3+} and Zn^{2+} transferrin complexes (Bertini *et al*, 1986) can now be seen as having two possible origins. Either it reflects the different environments experienced by the oxalate carboxyl groups (in terms of their interactions with the protein) or it arises from different coordination in the two binding sites. From the crystal structure of $Fe_2(C_2O_4)_2Lf$ the former explanation seems the most likely since one carboxyl group interacts with the helix 5 N-terminus and the other with the arginine sidechain. On the other hand, the possibility cannot be excluded that when different metals are involved the anion coordination may also be affected.

V.1.3 **Metal-anion Relationships**

The behaviour observed when oxalate is substituted for carbonate while using Fe^{3+} as the associated metal ion contrasts with that observed for Cu^{2+}-lactoferrin complexes. In the latter case both the mono-oxalato, $Cu_2(CO_3)(C_2O_4)Lf$, and dioxalato, $Cu_2(C_2O_4)_2Lf$, lactoferrin complexes can be made by displacement of carbonate from $Cu_2(CO_3)_2Lf$ using 50-fold and 100-fold molar excess of oxalate respectively (Shongwe

et al, 1992). Furthermore the hybrid carbonate-oxalate complex is stable to atmospheric CO_2 for at least 3 months, and $Cu_2(C_2O_4)_2Lf$ only slowly converts to $Cu_2(CO_3)(C_2O_4)Lf$ over a period of a month. In contrast, stable oxalate complexes of Fe^{3+}-lactoferrin cannot be readily made by displacement of bound carbonate, and once made, the $Fe_2(C_2O_4)_2Lf$ complex is relatively unstable to atmospheric CO_2.

The different behaviour of Fe^{3+} and Cu^{2+} lactoferrin carbonate and oxalate complexes is also reflected in their visible spectra. Thus whilst the substitution of oxalate for carbonate shifts the absorption maximum of the charge transfer band to a lower energy (466 nm to 482 nm) when Fe^{3+} is the metal ion, if Cu^{2+} is the metal ion, the change is to a higher energy (434 nm to 420 nm) (Ainscough *et al*, 1979).

These contrasting results reflect both the different types of d acceptor orbitals involved and the greater ligand field strength of oxalate relative to carbonate (Gray, 1965; Moeller, 1982). The visible absorption bands in question have been assigned to a phenolate $\pi \rightarrow$metal d charge transfer transition (Gaber *et al*, 1974; Ainscough *et al*, 1980), but the lowest-energy available d orbital is of different type for the two metal ions; for d^5 Fe^{3+} it is d_{π^*}, whereas for d^9 Cu^{2+} it is d_{σ^*}. This is because the extra d electrons of Cu^{2+} lead to the filling of d_{π^*} (see Fig. V.2). The charge transfer transitions are thus phenolate $(\pi) \rightarrow Fe(d_{\pi^*})$ and phenolate $(\pi) \rightarrow Cu(d_{\sigma^*})$ respectively.

The difference between carbonate and oxalate is that oxalate is a stronger σ donor and weaker π donor than carbonate. Thus when the carbonate ligand is replaced by oxalate, the metal d_{σ^*} orbitals are increased in energy (because d_σ are lowered) and d_{π^*} are decreased in energy (because d_π are increased). The result is that a ligand$\rightarrow d_{\pi^*}$ transition (as for Fe^{3+}) is decreased in energy (longer wavelength) whereas a ligand$\rightarrow d_{\sigma^*}$ transition (as for Cu^{2+}) is increased in energy (shorter wavelength). These differences are summarised in Fig. V.2.

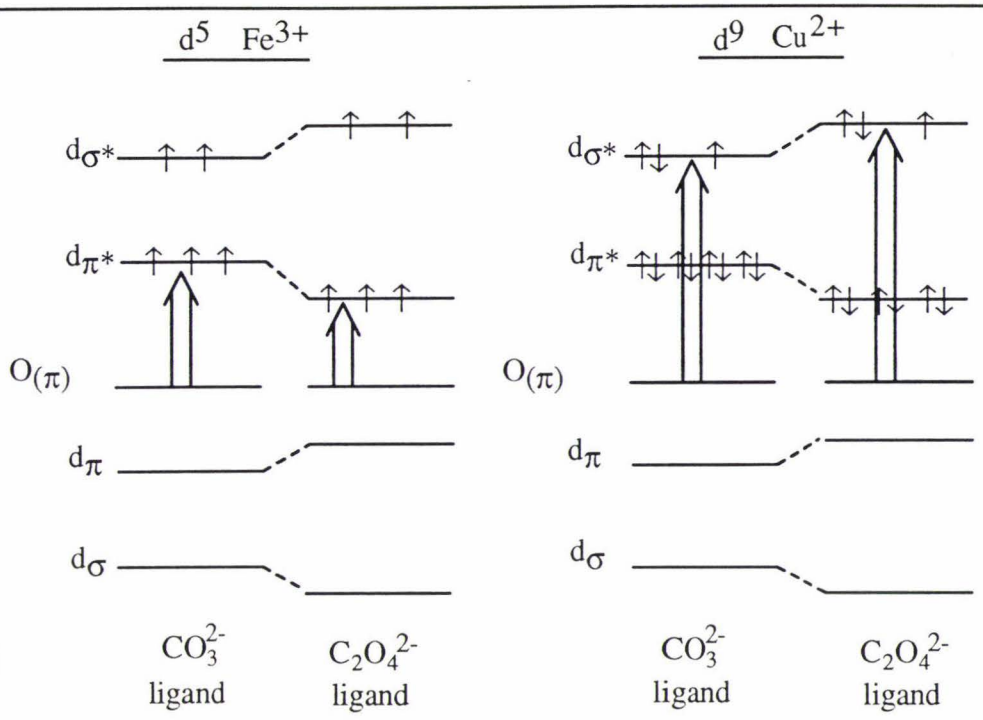


Fig. V.2. Schematic representation of energy levels for $Fe_2(CO_3)_2Lf$, $Fe_2(C_2O_4)_2Lf$, $Cu_2(CO_3)_2Lf$ and $Cu_2(C_2O_4)_2Lf$ to show changes that occur with different metal ions and anions.

The decrease by over 1000 mol^{-1} L cm^{-1} in the molar absorption coefficient that was noted from $Fe_2(CO_3)_2Lf$ to $Fe_2(C_2O_4)_2Lf$, (and from $Fe_2(CO_3)_2Tf$ to $Fe_2(C_2O_4)_2Tf$) can be attributed to weakened tyrosine-metal interactions. The intensity of ligand to metal charge transfer bands is proportional to the square of the overlap of the orbitals involved (Douglas and Hollingsworth, 1985) which, in these cases involved phenolate (π) and metal d types.

These observations provide experimental and theoretical support for two important aspects of transferrin chemistry. Firstly, the fact that both metal-anion and metal-tyrosine interactions are altered when oxalate is used in place of carbonate, with Fe^{3+} as the bound metal, is consistent with the idea that the sites are optimised for the binding of Fe^{3+} and CO_3^{2-}. These two ions together fit perfectly into the "closed" structure of the protein. Secondly, the weaker oxalate binding means that the precise mode of coordination of this anion is much more strongly influenced by the demands of the protein than is the case with carbonate. This is consistent with the repeated observations (see Section I.5) that differences between the sites become enhanced when ions other than Fe^{3+} and CO_3^{2-} are used, ie because "non-native" metal ions and anions bind more weakly, the minor differences in the two binding sites become more strongly expressed. The origin of these differences is discussed further in Section V.1.5.

V.1.4 Extension to Other Anions

A 'generalised' anion binding model was proposed by Shongwe *et al* (1992) as an extension of the Schlabach-Bates model, on the basis of oxalate binding to the C-lobe of copper lactoferrin. The present crystal structure shows that this model applies equally to both lobes of lactoferrin, and to the case where the physiologically-important Fe^{3+} is the bound metal ion. It is thus likely to prove valid for other metal ions and other anions.

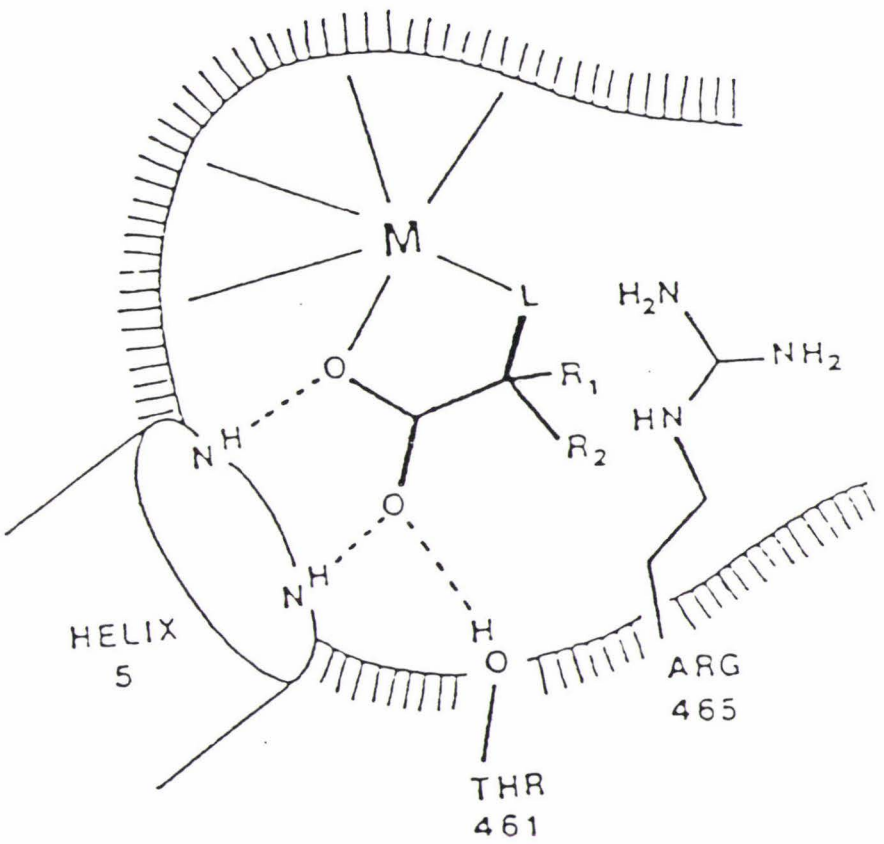


Fig. V.3 A generalised model for anion binding to transferrins, based on crystal structure analyses of carbonate and oxalate substituted lactoferrin. Binding of a general anion (thick lines) is compared to carbonate (thin lines). The anion interacts with the protein primarily through the carboxylate group (carbon 1). The interactions with the arginine sidechain depend on the nature of the substituents R_1 and R_2. (From Shongwe *et al*, 1992.)

The extension to other, larger, anions, with different substituents, can be made on the basis of the comparisons of oxalate and carbonate binding modes, from the crystal structures of $Fe_2(CO_3)_2$Lf (Anderson *et al*, 1989), $Cu_2(CO_3)(C_2O_4)$Lf (Smith *et al*, 1994) and $Fe_2(C_2O_4)_2$Lf (this work).

In the case of oxalate, one carboxylate group binds (like carbonate) to the positively-

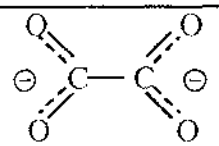
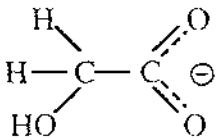
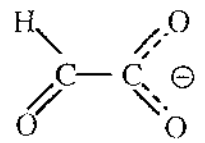
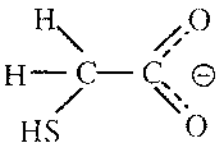
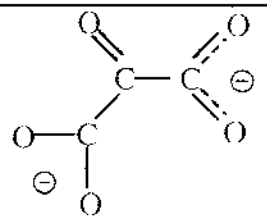
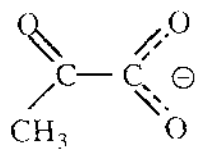
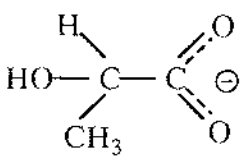
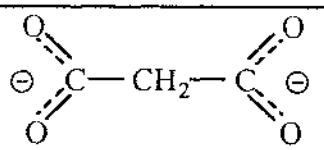
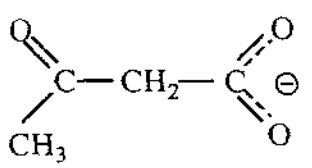
charged N-terminus of helix 5, and to the ferric ion. To accommodate the second carboxylate group, the arginine sidechain, which forms the rest of the carbonate site, is displaced away from the metal. This is necessary because of the larger size of the oxalate ion, relative to carbonate, but is able to occur because the interdomain cleft contains a large solvent-filled cavity next to the binding site which allows room for movement of the arginine and some other neighbouring sidechains (see Section III.3.3).

The common feature of all synergistic anions is the presence of at least one carboxylate group, and this would be expected to bind to the N-terminus of helix 5 and to the metal. To accommodate the rest of the anion, and to allow coordination of the second donor group, L, to the metal, the arginine sidechain must be displaced (as seen with oxalate). The binding of anions larger than carbonate is thus likely to be influenced by the extent to which substituents on the anion alter hydrogen-bonding interactions with the arginine and between the latter and other protein groups.

This model fits nicely with the results of Campbell and Chasteen (1977). In their study of VO^{2+}-substituted transferrin, using 16 different anions, it was found that the EPR spectra could be grouped in two classes, A and B, which were anion dependent. Those anions which gave both class A and B spectra comprised carbonate and the dicarboxylic acids such as oxalate and malonate. On the other hand, anions with only one carboxylate and a uncharged electron donor group on the proximal carbon, gave only one class of spectrum, Class B. Anions in the second group, which include lactate, glycolate, thioglycolate, glyoxylate and salicylate, are incapable of hydrogen-bonding to the arginine, but those in the first group with two or more carboxylate groups are capable of hydrogen-bonding. In the light of the crystal structures of $Fe_2(C_2O_4)_2$Lf and $Cu_2(CO_3)(C_2O_4)$Lf, it is possible to ascribe the differences seen in the EPR spectra observed by Campbell and Chasteen, to the nature of the interaction (if any) between the anion and the arginine sidechain.

The above interpretation is consistent with the observed pH dependence of the VO^{2+} EPR spectra. For the dicarboxylate anions, transition from the A spectrum to the B spectrum is associated with the deprotonation of a protein group of $pK_a \sim 10$. It is reasonable to conclude that this group corresponds to the arginine sidechain, and that its ionisation state influences its interaction with the anion, and in turn the metal-anion interaction and EPR spectrum. The A spectrum corresponds to the case where the arginine is protonated, positively charged and a better hydrogen bonding group, whereas the B spectrum is associated with a deprotonated arginine, which would be a poorer hydrogen bonding group and would no longer have a charge-charge interaction with the anion.

Table V.1: Comparison of hydrogen-bonding potential of synergistic anions to the 'essential' arginine with the relative stability[a] of the related anion-transferrin complex

Anion	Structure	Able to H-bond to Arginine?	% Fe^{3+} removed by citrate[a]
Group 1			
oxalate		Yes	1%
glycolate		No	32%
glyoxylate		No	30%
thioglycolate		No	37%
Group 2			
ketomalonate		Yes	20%
pyruvate		No	47%
lactate		No	45%
Group 3			
malonate		Yes	30%
acetoacetate		No	48%

[a]Figures of the % of iron removed from the ternary complex by citrate under CO_2-free conditions are taken from Schlabach and Bates, 1975. Note, citrate used here as a competitor for iron-binding by transferrin, has been shown to be a non-synergistic anion.

A corollary of this argument is that for synergistic anions of a similar size, the stability of the complexes formed with the transferrin will be dependent not only on the interaction with the metal and the N-terminus of Helix 5, but also on the anion's ability to interact with the 'essential' arginine. Table V.1 lists synergistic anions in groupings of similar size, and compares their potential ability to hydrogen-bond to the arginine with the stability of the complex formed with transferrin

Table V.1 shows that the competition of citrate with iron-binding by the transferrin complex is clearly dependent on the nature of the synergistic anion. Furthermore, there is a high correlation between the potential hydrogen bonding ability of the synergistic anion to the 'essential' arginine, and the relative stability of the Fe^{3+}-anion-transferrin complex. In Group 1, which comprises synergistic anions with a two-carbon backbone, oxalate, which is able to hydrogen-bond to the arginine, forms a ternary complex with transferrin that donates only 1% of its iron to the completing chelate. In contrast, those anions in Group 1 which are unable to hydrogen-bond to the arginine, - glycolate, glyoxylate and thioglycolate, - all yield transferrin complexes which lose between 30 and 37% of their iron to citrate. A similar effect is seen in both Groups 2 and 3 where the carbon backbone is lengthened to 3 carbons. In all three groups an extra 20-30% of the iron is lost from the transferrin complex to the competing chelate if the anion does not have the additional potential to hydrogen bond to the arginine.

An increase in the size of the synergistic anion from two- to a three-carbon backbone also results in a pronounced loss of stability of the Fe^{3+}-anion-transferrin complex. This would be expected from the crystal structure of $Fe_2(C_2O_4)_2$Lf which shows that not only would the arginine sidechain have to be displaced, but there also must be a consequent rearrangement of residues beyond the arginine to make room for the larger anion. The N-lobe of $Fe_2(C_2O_4)_2$Lf shows that constraints by sidechains in the vicinity of the arginine may restrict the conformation of the anion and so possibly reduce the stability of the Fe^{3+}-anion-transferrin complex. In $Fe_2(C_2O_4)_2$Lf the average anion-Fe bond is 1.99 Å in the C-lobe, but this has lengthened to 2.21 Å in the N-lobe, indicating reduced iron stability in the N-lobe.

Anions such as malonate and acetoacetate imply a 1,3-bidentate coordination to the iron. Such coordination is clearly less stable than the 1,2-bidentate coordination for dicarboxylic acids (loss of iron from the oxalate-Tf complex, 1%; malonate-Tf complex, 30%) but it appears to have little effect on the stability of the ketone derivatives, pyruvate (1,2-coordination) and acetoacetate(1,3-coordination) where both complexes lose the same amount of iron to citrate (~48%).

V.1.5 **Inequivalence of the Sites**

An unexpected result of the structure analysis of $Fe_2(C_2O_4)_2$Lf has been the discovery that the oxalate coordination is different in the two binding sites. In the N-lobe

the oxalate ion is asymmetrically bound, with Fe-O bonds of 1.87 Å and 2.55 Å, whereas in the C-lobe site it is much more symmetric, with bond lengths of 1.91 Å and 2.07 Å. Although the differences are not great, in terms of the likely coordinate error in the structure, they have been well tested during the refinement of the structure. All attempts to orient the anions differently were counteracted both by the appearance of the electron density and by the movements that took place during least squares refinement. Most importantly, the differences demonstrate, for the first time, how sequence differences beyond the immediate binding site can modulate binding.

The close similarity between the sites when carbonate is the anion can be explained if it is assumed that the site is optimised for carbonate binding. When a larger anion is substituted, the arginine sidechain which helps form the binding site is forced to move, and it is in this movement that the effects of neighbouring residues come into play. There is then a subtle interplay between the interactions the arginine makes with neighbouring sidechains and with the anion, and the binding mode of the anion to the metal. As discussed in Section III.3.7, and summarised in Table V.2, residues Ser 191, Phe 183, Phe 63 and Phe 190 in the N-lobe, and their C-lobe equivalents Gly 527, Pro 519, Tyr 398 and Tyr 526 exert different effects on the displaced arginine sidechain. These include both steric effects (eg Phe 183 restricting the movement of Arg 121) and stabilising effects (Tyr 398 and Tyr 526 stabilising the position of Arg 465 once it has moved).

The result is that in the C-lobe the anion-binding arginine can swing further away from the anion site, allowing the oxalate both to remain symmetrically coordinated to the iron atom and to maintain its hydrogen bonding to the arginine. In the N-lobe, however, the arginine is restricted in its movement and this in turn restricts the configuration of the bound oxalate. Assuming that the preferred binding mode for the oxalate ion is symmetrical 1,2-bidentate, this explains why oxalate appears to bind preferentially in the C-lobe site of lactoferrin (which supports symmetrical binding) rather than the N-lobe.

Table V.2 Residues that may modulate anion binding in human lactoferrin

N-lobe		C-lobe	
Ser 191	Oγ hydrogen bonds to Arg 121; helps hold Arg 121 in place	**Gly 527**	no hydrogen bond can be made
Phe 183	Large hydrophobic residue which sterically hinders movement of Arg 121	**Pro 519**	Sited further from Arg 465; no steric hindrance to Arg 465
Phe 63) **Phe 190**)	Cannot make hydrogen bonds to Arg 121	**Tyr 398**) **Tyr 526**)	Provide stabilising H-bonds when Arg 465 moves away from anion site.

V.1.6 **Extension to Other Transferrins**

Attempts to extrapolate the present results, for the binding of oxalate to human lactoferrin, to the binding of oxalate or other anions to different members of the transferrin family must be treated with caution. Consideration must be given to both steric and stability factors, for example the extent to which the anion-binding arginine can move, and the question of whether it can find a stable configuration once it has been displaced. Moreover, an unknown factor is the extent to which the domains are closed over each binding site in the different transferrins; this is unlikely to be radically different, but variations of 3-5° in domain closure would not be unreasonable, and could influence the movements that are possible in the binding cleft.

Some conclusions can be drawn, however, from comparisons of the residues that modulate anion binding in human lactoferrin. These are summarised in Table V.3.

Table V.3 Comparison of residues that may modulate anion binding in transferrins

	N-lobe				C-lobe			
Residue no.	63	183	190	191	398	519	526	527
human Lf	Phe	Phe	Phe	Ser	Tyr	Pro	Tyr	Gly
bovine Lf	Met	Cys	Phe	Gly	Tyr	Pro	Tyr	Gly
murine Lf	Thr	Ser	Ser	Gly	Tyr	Pro	Gln	Gly
serum Tf	Leu	Cys	Phe	Gly	Phe	Pro	Tyr	Gly
ovo Tf	Gln	Arg	Ser	Gly	Leu	Ala	Phe	Gly

It is apparent that with respect to anion substitution the N-lobe of human lactoferrin is the most sterically constrained lobe of all the transferrins. Three phenylalanines, large bulky residues, appear to surround or occupy the space to which the arginine may swing to make more room for the oxalate (see Fig. III.15). Not only is there steric hindrance, especially from Phe 183, but also none of these residues can provide stabilising hydrogen bonds to the arginine were it able to swing further away from the oxalate. Furthermore, amongst all the transferrins, it is only in the N-lobe of human lactoferrin that the arginine is held back by a hydrogen bond to Oγ of Ser 191. In all other transferrins, and in the C-lobe of lactoferrin, the equivalent residue is a glycine, and unable to make such a stabilising hydrogen bond.

On the other hand, murine and bovine lactoferrin and serum transferrin appear to be somewhat less restricted around the N-lobe site. All have a small residue, either serine or cysteine, at position 183, so the arginine would have the freedom to move in this direction. Murine lactoferrin in particular appears to have quite a spacious binding cleft,

with small residues, threonine and serine at positions 63 and 190. Such residues, although theoretically potential hydrogen bonding partners for the arginine are unlikely to be able to reach far enough to make such stabilising bonds as seen in the C-lobe of human $Fe_2(C_2O_4)_2$Lf. Bovine lactoferrin and serum transferrin both have phenylalanine at 190, and another bulky residue, methionine or leucine at 63, none of which is a potential hydrogen bond partner for the arginine. Both methionine and leucine are quite flexible residues however, and whether or not they sterically hinder the movement of Arg 121 would depend on their conformation. Thus, in contrast to the situation in human lactoferrin, it is probable that Arg 121 in murine lactoferrin could move sufficiently freely to allow symmetrical 1,2-bidentate oxalate coordination to the iron. Whether such symmetrical coordination occurs in bovine lactoferrin and serum transferrin depends to some extent on the conformation of the methionine and leucine residues at position 63.

The situation in the N-lobe of ovotransferrin is interesting. Here the residue corresponding to Phe 183 is an arginine, which is positively charged. This could limit the movement of Arg 121 through charge-charge repulsion. On the other hand, Gln 63 could provide a stabilising hydrogen bond for Arg 121. These uncertainties make it very difficult to predict the likely effects of anion substitution in this site.

The C-lobe sites of transferrins show a much greater similarity with each other than do the N-lobe sites. In human lactoferrin symmetrical 1,2-bidentate oxalate binding is associated with a small residue (Pro) at position 519 and hydrogen bonding residues (Tyr) at positions 398 and 526, to stabilise the displaced Arg 465. The equivalent residues in bovine lactoferrin are identical to human lactoferrin, and in murine lactoferrin only Tyr 526 is replaced, by Gln, a residue which may also be able to reach far enough to hydrogen bond with Arg 465. Thus it could be confidently predicted on this basis, that in the C-lobe of both bovine and murine lactoferrin, oxalate would bind in a symmetric 1,2 bidentate fashion.

On the other hand, the ovotransferrin C-lobe site appears somewhat more constrained, with bulky hydrophobic residues Leu 398 and Phe 526 unable to provide hydrogen bonds to Arg 465, but a small residue, alanine, at position 519. This configuration of residues is very similar to that seen in the N-lobe of serum transferrin. It is interesting to note that when copper is the metal, oxalate binds preferentially to the C-lobe for ovotransferrin and the N-lobe for serum transferrin.

Oxalate binding in the C-lobe of serum transferrin presents a puzzle. Only one of the modulating residues is changed from human lactoferrin (Tyr 398→Phe). This suggests that transferrin should bind oxalate well in its C-lobe site, although the displaced arginine would lack one of the stabilising interactions seen in lactoferrin. Diferric transferrin does indeed bind oxalate strongly, in both sites (Schlabach & Bates, 1975). On the other hand, it has been shown that the C-lobe site of human serum transferrin is unable to bind oxalate at all when copper is the metal. It is difficult to see why this should

be so, especially when the C-lobe site is the preferred site for oxalate binding in lactoferrin when copper is the associated metal ion (Shongwe *et al*, 1992). This emphasizes the complexities that arise when the final structure depends on the competing requirements of protein, metal ion and anion.

V.1.6 **Anion Binding and Domain Closure**

The versatility shown by lactoferrin in the binding of oxalate in place of carbonate illustrates an important feature of transferrin structure. Not only does the molecule have the large-scale flexibility required for binding and release, with domain movements that provide the necessary framework for tight reversible binding, but it also has scope for internal adjustment because of the large solvent filled cavity beyond the arginine. Thus carbonate is not the only anion which can support a closed structure. The crystal structure of $Fe_2(C_2O_4)_2$Lf shows, in the case of oxalate at least, lactoferrin can form a closed structure in both N- and C-lobes. Whether this can be extended to even larger anions is not clear, but anions as large as phenlylactate and even xylenol orange have been reported to act as synergistic anions (Schlabach & Bates, 1975; Harris & Gelb, 1979).

The ability of transferrins to bind other species also highlights a potential use, that of drug delivery. The $Fe_2(C_2O_4)_2$Lf structural analysis has shown that when oxalate is substituted for the smaller carbonate, the overall protein conformation and the extent of closure of the domains over each site is unchanged. If the closed conformation of lactoferrin is important for receptor binding or other activities, then lactoferrin carrying anions or metals other than the physiological CO_3^{2-} and Fe^{3+} could equally well be recognised, provided the closed configuration is formed. Thus a possible scenario for drug delivery is that the drug binds to lactoferrin, and once bound induces the closed conformation so that the drug remains safely locked inside the lactoferrin molecule until some external factor such as receptor binding or pH destabilises the complex.

Interestingly, aspirin, perhaps the most common drug of all, could well bind to lactoferrin. Salicylate is a synergistic anion (Schlabach & Bates, 1975) and is small enough that it could well induce the closed state. Ruthenium complexes with antitumour activity; *trans*-imidazolium[tetrachlorobis(imidazole)ruthenateIII], *trans*-indazolium[tetrachlorobis(indazole)ruthenate(III)] and indazolium[pentachloro(indazole)ruthenate(III)], $Ru(Im)_2$, $Ru(Ind)_2$ and Ru(Ind),respectively, bind to transferrin and lactoferrin. Crystallographic studies show that these complexes bind with high affinity to the binding cleft of lactoferrin (Smith *et al* , in press). These studies were carried out using crystals of apolactoferrin in which the molecules are restrained by their crystal environment. They did not show, therefore, whether the protein could close over the Ru(III) complex. Modelling studies, however, suggest that there is enough space for the drugs to bind and still permit closure, thus suggesting a possible model for the delivery of these ruthenium complexes, incorporated

into transferrin, to target cancer cells.

The question of whether an open or closed structure is adopted is a difficult one to address. Methods used to follow binding, such as UV difference spectra and visible charge transfer spectra monitor the binding of the tyrosine ligands to the metal ion. Because the tyrosine ligands are only associated with domain 2, these techniques reflect only binding to this domain and do not show whether an open or closed structure is adopted for the particular anion. Binding of the metal to the ligands on the other domain, and thus the formation of a closed structure, may be indicated by techniques sensitive to the histidine ligand, eg. NMR or EPR. Thus answers to important questions as to whether a particular anion supports an open or closed transferrin structure, and the implications of such a structure for the transport of such a species are best addressed by a combination of techniques. Such techniques may include not only UV/visible spectroscopy but also EPR, NMR, low angle electron scattering and X-ray crystallography.

It can be seen that the role of spectroscopy in these studies is changing. No longer is it necessary to determine the basic structure; something that spectroscopy is not good at doing, but at which X-ray crystallography has been very successful. Rather, spectroscopy is invaluable now as a monitor of small structural and electronic changes accompanying binding experiments and mutagenesis. The physicochemical data gained from these measurements can now be far more accurately interpreted/reinterpreted in the light of the three-dimensional structures of the transferrins.

The crystal structure of $Fe_2(C_2O_4)_2$Lf does, however, highlight a problem in interpreting spectra of a protein with two similar but distinct binding sites. Spectra such as UV difference, visible and EPR present an average of the two sites; small differences in coordination from one site to the other are revealed only by crystallography. Progress in the next few years will be heavily dependent on the powerful combination of crystallography, molecular biology and spectroscopy.

V.2 THE ROLE OF GLYCOSYLATION OF LACTOFERRIN

The discovery of endoglycosidases has now made it possible to enzymatically cleave an entire carbohydrate chain from a protein in one step. This method of removal of carbohydrate formed the basis of the present work, with the rapid and complete deglycosylation of the whole molecule of human lactoferrin being achieved by the use of the endoglycosidases PNGase and EndoF. This has provided, for the first time, sufficient protein to enable an investigation of some physicochemical properties of deglycosylated human lactoferrin, and allows some predictions to be made about recombinant human lactoferrin.

V.2.1 **Deglycosylation of Human Lactoferrin**

The deglycosylation of human lactoferrin was rapid and complete. This was not the case, however, when deglycosylation was attempted with either bovine lactoferrin or rabbit serum transferrin. With bovine lactoferrin some deglycosylation occurred immediately but even after prolonged incubation (>5 days) with increased concentrations of PNGase and EndoF, some carbohydrate remained attached to the protein. PNGase and EndoF did not appear to remove any carbohydrate at all from human serum transferrin. In all three proteins the carbohydrate is N-linked, and should be susceptible to cleavage by these endoglycosidases. The probable explanation for the difference in susceptibility lies in the three-dimensional structures of the proteins. The crystal structure of Fe_2Lf shows that the glycosylation sites of human lactoferrin are exposed, allowing easy access by the deglycosylating enzymes. Sites of glycosylation in other proteins where the Asn-glycan linkages are not so exposed may not be susceptible to the enzyme. This is certainly seen in the three dimensional structure of bovine lactoferrin (S. Moore, personal communication), in which one glycosylation site, Asn 545, has the Asn-glycan linkage deeply buried in a pocket of the protein (Fig. V.4).

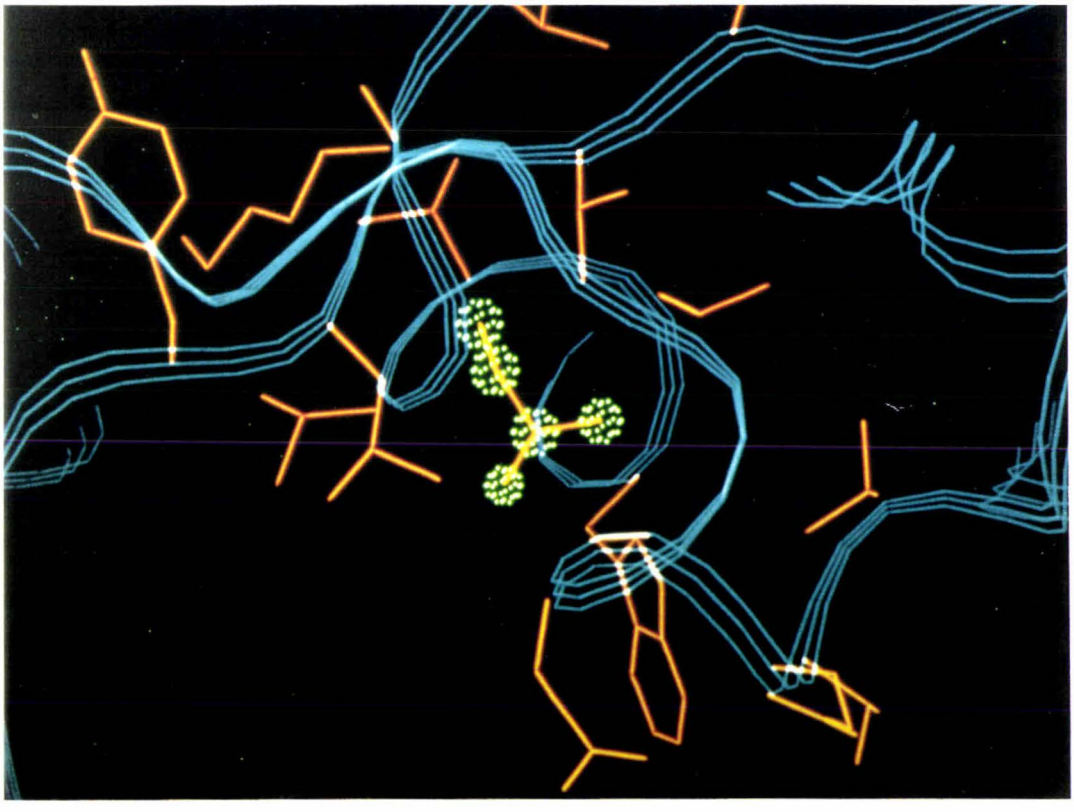


Fig. V.4 Structure of bovine lactoferrin around the glycosylation site Asn 545. Asn 545 (highlighted with green stippled spheres) is seen buried in a pocket in the protein. The mainchain of the protein is shown as a blue ribbon, sidechains in orange.

V.2.2 **Role of Glycosylation of Lactoferrin**

A variety of functions have been suggested for the carbohydrate groups attached to proteins (see Section IV.1.1 for more detail), and a number of these have been tested as part of this study. Results here have shown that native and deglycosylated human lactoferrin have:

(i) the same iron-binding and release properties,

(ii) the same resistance to limited proteolysis by trypsin,

(iii) very similar thermal stabilities,

(iv) the same overall three dimensional structure, but different flexibility of a small polypeptide loop around one of the carbohydrate attachment sites.

These results refute several of the possible roles for the glycans of lactoferrin, in particular the maintenance of protein conformation through specific protein-carbohydrate interaction, the protection against proteolysis and an *in vitro* role in either iron binding or iron release. There is an indication that the native protein has a slightly greater thermostability than its deglycosylated counterpart, but even in this respect the two forms are very similar.

That deglycosylation of human lactoferrin does not affect either the conformation of the protein or its *in vitro* iron binding properties (as far as experiments here can judge) is not surprising in light of the crystallographic results of the native Fe_2Lf structure (Haridas *et al*, 1995). This shows the glycosylation sites to be distant from the iron sites; 26 Å in the N-lobe and 20 Å in the C-lobe. In each lobe, the glycosylation site is on the surface of the molecule, while in contrast, the iron site is found deeply buried in a cleft between the two domains (Fig. V.5). The high resolution (2.2 Å), well refined X-ray structure of Fe_2Lf (Haridas *et al*, 1995) reveals no evidence of any significant carbohydrate-protein interaction. Indeed the little electron density visible beyond the first one to two sugar residues implies either disorder or flexibility in the glycans, and that the carbohydrate does not play any structural role. Instead the glycans appear to extend from the polypeptide surface offering neither protection from proteolysis, stabilising hydrogen bonds back to the polypeptide, nor any interaction whatsoever with the iron binding site.

Comparison of the native Fe_2Lf (Anderson *et al*, 1989), native apolactoferrin (Faber, H.R., personal communication), and deglycosylated apolactoferrin (Anderson *et al*, 1990) structures adds weight to this analysis. The domain structure is almost identical in these species. It appears that, in solution, there is a dynamic opening and closing of both lobes of apolactoferrin, and that conformations which include either only the N-lobe or C-lobe open, both lobes open, or both lobes closed are possible. It is clear that the different conformations seen in the different crystal forms result from crystal packing forces, and not the glycan interaction with the polypeptide.

Examination of glycosylation sites of other lactoferrins further supports these conclusions and suggests that they are likely to apply also to other members of the lactoferrin family. Table V.4 lists the potential glycosylation sites of various lactoferrins. The glycosylation sites were predicted from the presence of an Asn-Xaa-Ser/Thr motif (where Xaa is any amino acid other than proline) in the primary sequence of the lactoferrin. These glycosylation sites are shown in Fig. V.5.

Table V.4 Potential glycosylation sites of lactoferrins

	Residue number of potential glycosylation site						
Human lactoferrin	**137**					**478**	
Bovine lactoferrin		233	281	**368**		**478**	**547**
Mouse lactoferrin	100					478	
Porcine lactoferrin					370	478	

Sites known to be glycosylated are in bold

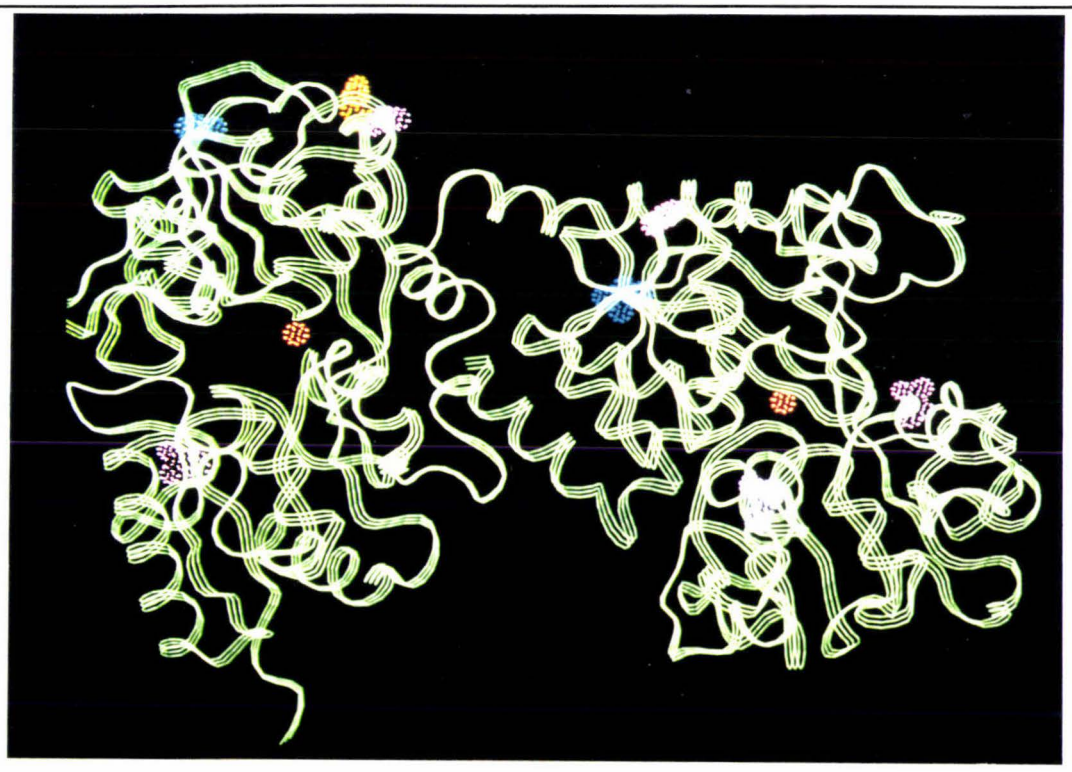


Fig. V.5 Ribbon diagram of human iron-lactoferrin showing potential glycosylation sites of different lactoferrins (human, orange; bovine, purple; mouse, light blue; procine, dark blue). N-lobe on the left, C-lobe on the right. Iron atoms are shown as red spheres.

It can be seen that in general the glycosylation sites are scattered over the whole surface of the molecule, often far removed from the iron-sites and binding clefts, and vary in position from one lobe to the other.

Extension of this analysis to members of the wider transferrin family shows that these conclusions are likely to apply also to members of the wider transferrin family (see Fig. V.6).

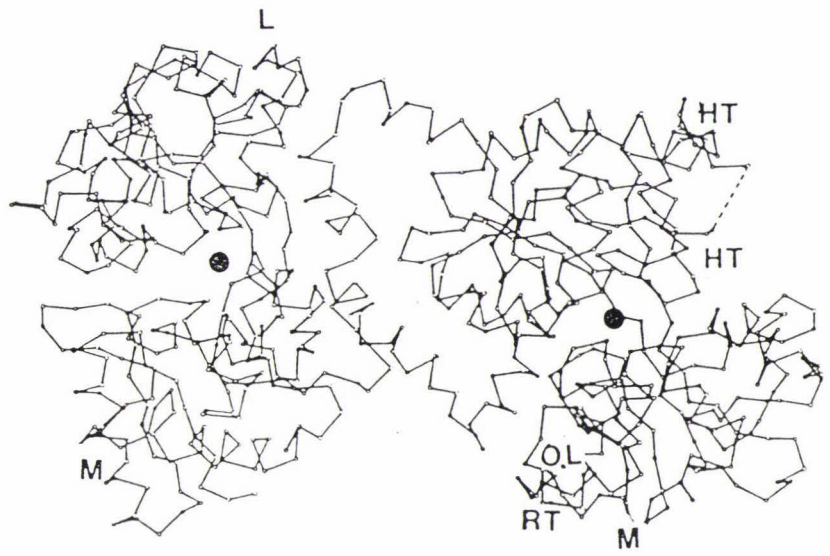


Fig. V.6 Glycosylation sites of the transferrin family. C_α plot of human iron-lactoferrin showing iron atoms in their binding clefts. N-lobe at left, C-lobe at right. Glycosylation sites are shown for lactoferrin (L), human serum transferrin (HT) and chicken ovotransferrin (O), rabbit serum transferrin (RT) and potential sites for melanotransferrin (M).

In all cases the protein structures are very similar, as judged by the similarity of transferrin to lactoferrin (Baker & Lindley, 1992) and by their amino acid sequence homology (Metz-Boutigue *et al*, 1984), but in contrast their glycosylation sites are very different, varying both in number and in location over the molecular surface (Baker *et al*, 1987) (Fig. V.6), and again often far removed from the iron-binding clefts. It seems unlikely, therefore, that such a variation in glycosylation amongst the transferrins would be correlated to properties such as an iron-binding function or structure that are invariant within this family.

The conflict of the present results with those of earlier reports can be traced back to the experimental methods employed. One study (Spik *et al*, 1985) that indicated an increased susceptibility to proteolysis of deglycosylated lactoferrin used quite prolonged methods of stepwise hydrolysis by a number of exoglycosidases to remove the carbohydrate. This may have resulted in partial denaturation of the protein and increased

its vulnerability to proteolysis. The fact that much of the protein irreversibly precipitated during this deglycosylation procedure (G. Spik, personal communication) supports this hypothesis. Care was taken in this present work to duplicate as closely as possible the conditions of proteolysis by trypsin, so that an accurate comparison of results could be made.

A more recent study of iron binding (Legrand *et al*, 1990) that used PNGase (as in this study) to remove the carbohydrate, based conclusions about iron binding on the analysis of deglycosylated N-terminal 20 kDa and 30 kDa N-tryptic fragments of lactoferrin. Such fragments are neither representative of a complete lobe (40 kDa) of lactoferrin nor the whole molecule. These fragments lack the helix behind the iron site and the loop 280-330. It could be thus expected that this fragment would release iron more easily, and that the conformation change would be different irrespective of the presence of carbohydrate. There may in fact be a different interaction of the glycans with this fragment because of the missing 280-330 polypeptide loop. In contrast, the present results are based on the well characterised whole molecule lactoferrin where the deglycosylation was rapidly and effectively carried out using highly active PNGase and Endo F.

There remains the strong possibility that the glycans of lactoferrin may be important in its adhesion to cells, and interaction with specific receptors, although evidence so far is contradictory.

Native and deglycosylated lactoferrins seem to have similar immunological properties; polyclonal antibodies raised against the native protein have been shown to cross react also with the deglycosylated form (Norris, G.E., personal communication). Although antibodies are known to occur against carbohydrate epitopes (Tarentino *et al*, 1974; Feizi, 1985) they do not appear to be essential in the antigenic reactions of human lactoferrin. This is in common with many other glycoproteins (eg Bose *et al*, 1976; Parham *et al*, 1977).

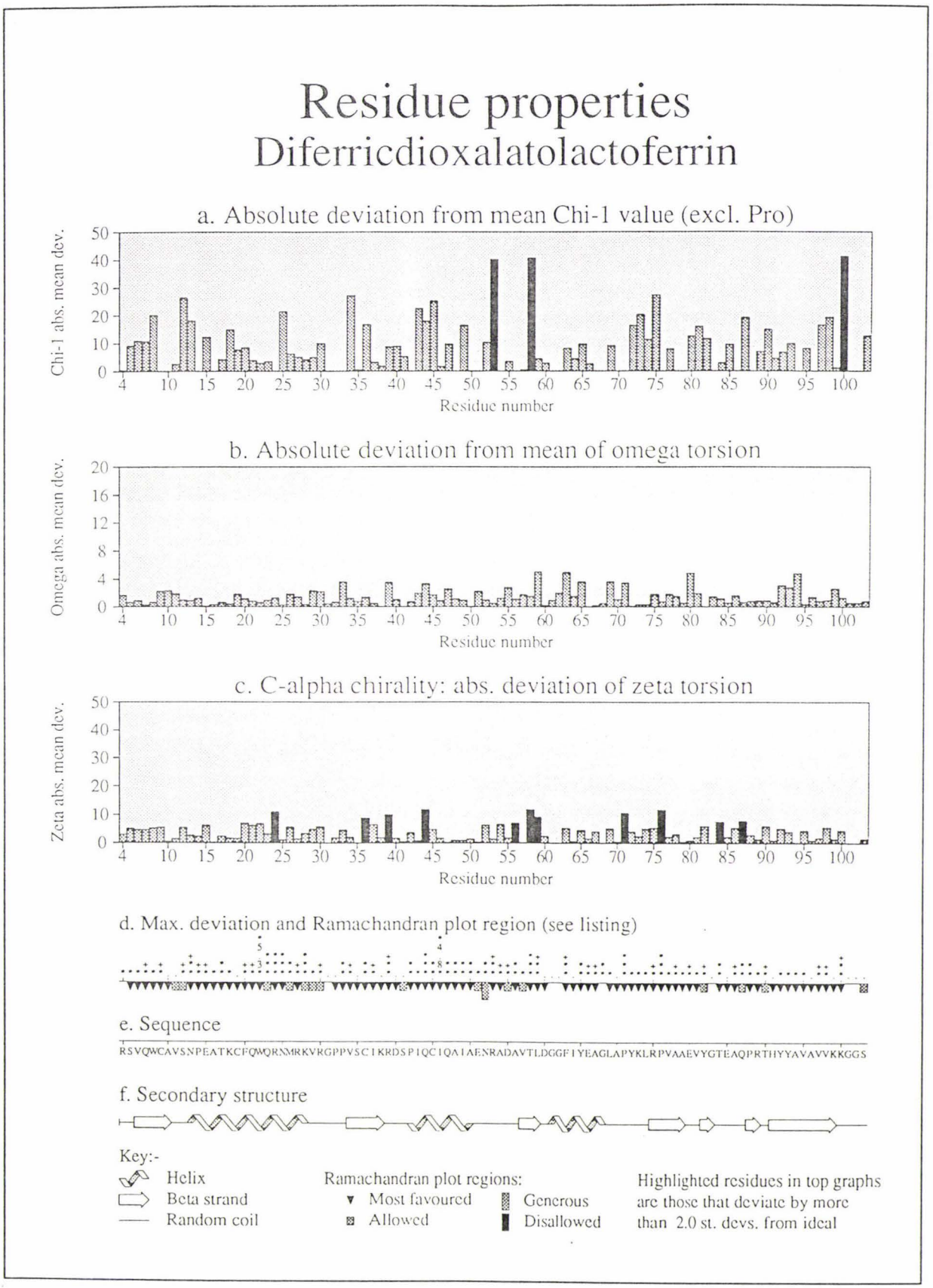
Fucose residues of the carbohydrate have been indicated as being essential for the binding of lactoferrin to specific brush border membrane receptors (Davidson & Lönnerdal, 1988) and to macrophages (Spik *et al*, 1985). These results, however, are in disagreement with a study of defucosylated lactoferrin (Imber & Pizzo, 1983). The carbohydrate of transferrin has been implicated in signalling the clearance of this protein from plasma (Regoeczi *et al*, 1974). Similarly lactoferrin clearance by liver is reported to be mediated by an interaction with a hepatocyte membrane protein sensitive to a fucosyl α-1-3-N-acetyl glucosaminyl group of the glycan chain (Prieels *et al*, 1978), but this has since been questioned (Imber & Pizzo, 1983).

One of the most extensively studied and widely accepted roles of lactoferrin is as an antibacterial agent. It does appear that some of the antibacterial properties of lactoferrin and other transferrins may result from actual contact with bacteria rather than a simple

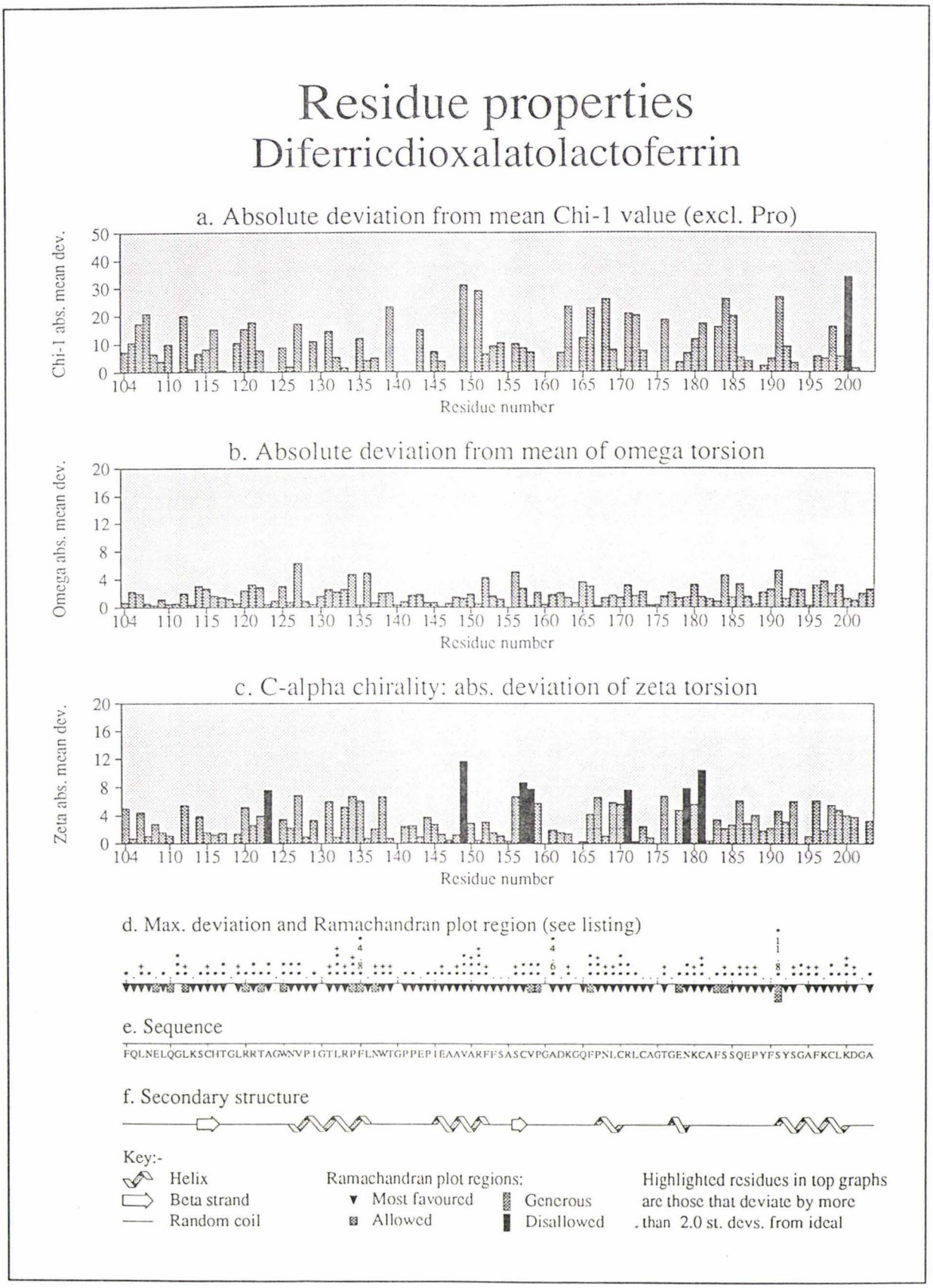
withholding of iron (Valenti *et al*, 1987; Dalmastri *et al*, 1988). It is also known that a peptide loop of 18 amino acids (20-37), derived from human lactoferrin, has potent bactericidal properties which are independent of iron, but require contact with bacteria (Bellamy *et al*, 1992). Such contact could be facilitated by the presence of the glycans on native lactoferrin. It is interesting to note that an examination of the potential glycosylation sites shows that the glycosylation site at position 478 is conserved in all lactoferrins but absent in all transferrins (Table V.4, Fig. V.5 & Fig. V.6). This raises the speculation that the conservation of glycosylation at this point indicates a possible importance of these glycans in an interaction that is unique to the lactoferrins.

It will be of great interest to determine whether deglycosylated lactoferrin has any of these activities, particularly in view of the recent founding of a company to produce and market recombinant lactoferrin as an antibacterial agent. The recombinant lactoferrin, produced in *Aspergillus*, will have different glycosylation from native human lactoferrin; thus the role of the carbohydrate (if any) is not only of scientific, but also therapeutic and commercial interest. The results presented here indicate that this recombinant lactoferrin is likely to have the same overall three dimensional structure, the same thermal stability and resistance to proteolysis, and the same properties of iron-binding and release. Maintenance of these properties is essential if this lactoferrin is to act effectively as an antibacterial agent.

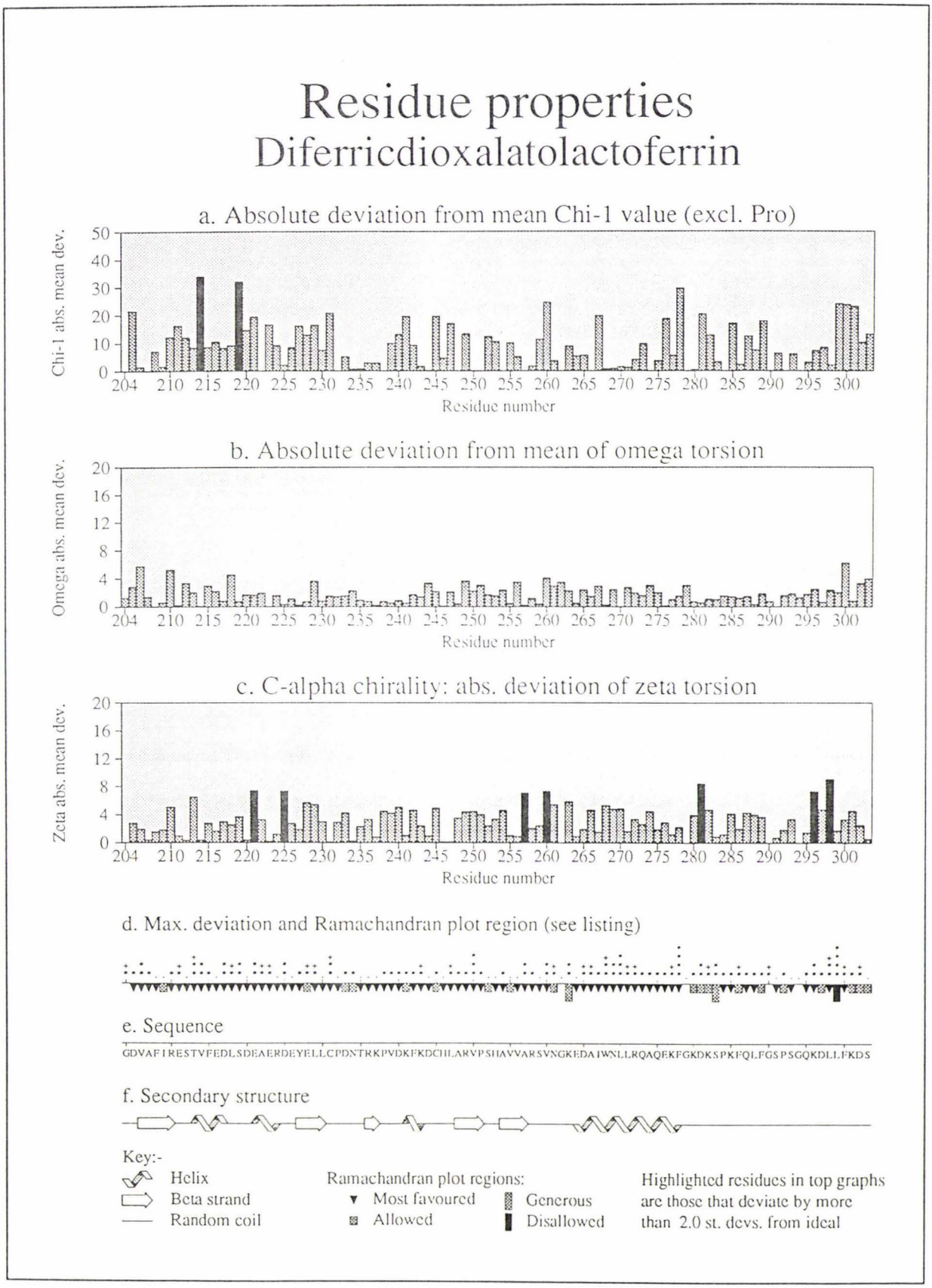
APPENDIX I. Residue properties of diferricdioxalatolactoferrin



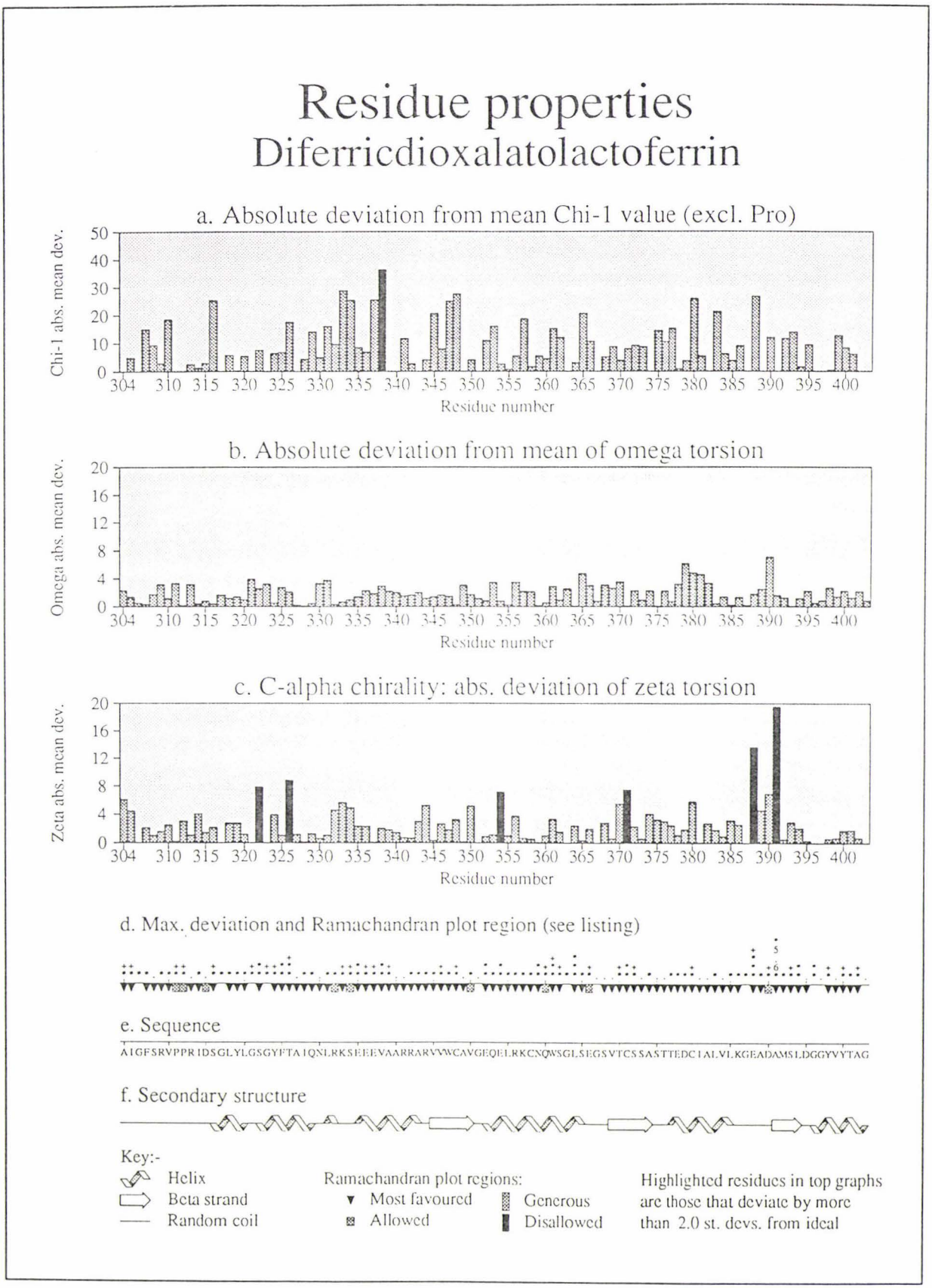
APPENDIX I continued



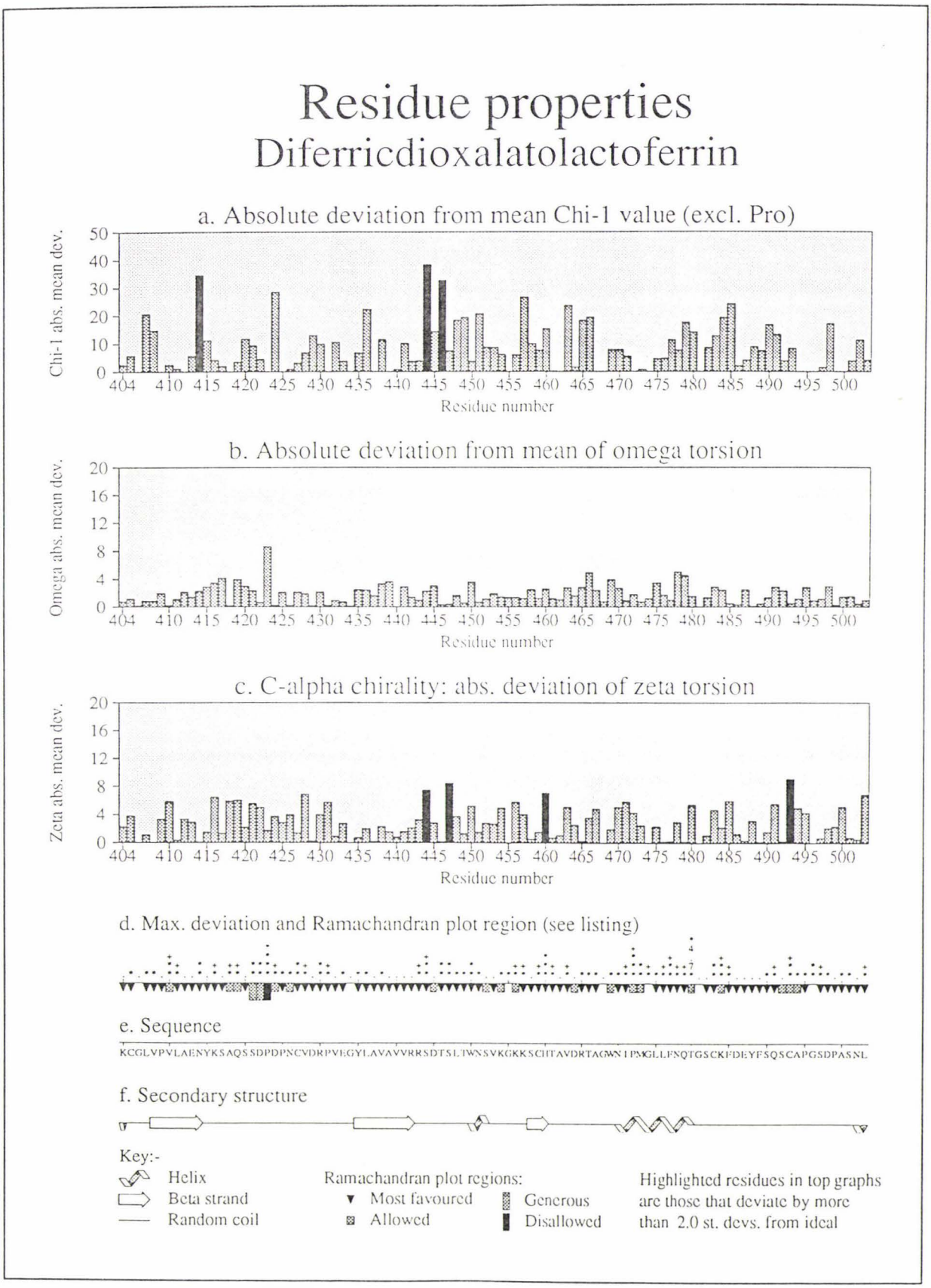
APPENDIX I continued



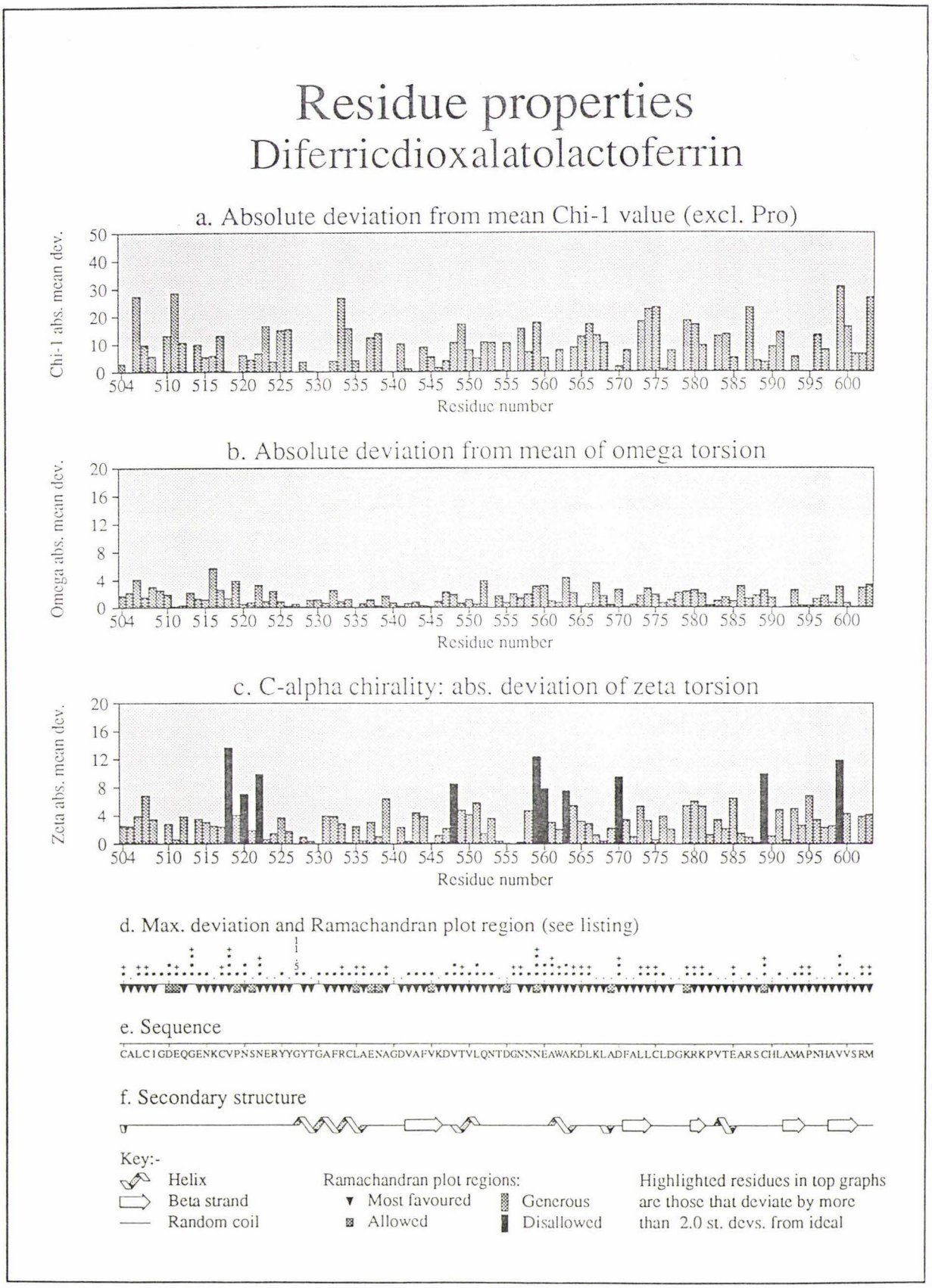
APPENDIX I continued



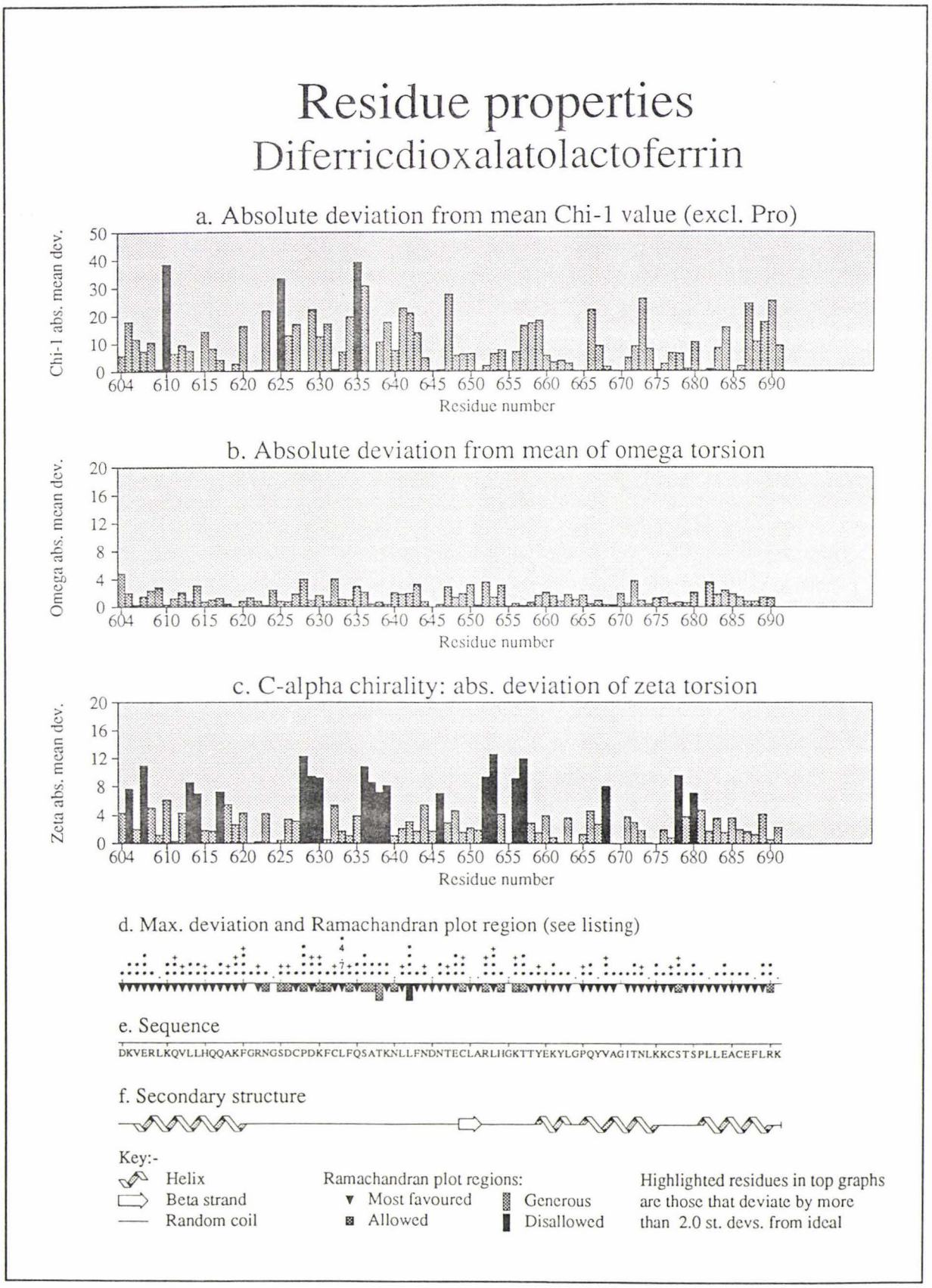
APPENDIX I continued



APPENDIX I continued



APPENDIX I continued



REFERENCES

Aasa, R., Malmström, B.G., Saltman, P. and Vänngard, T. (1963). *Biochim. Biophys. Acta* **75**, 203-223.

Abola, J.E., Wood, M.K., Church, A., Abraham, D. and Pulsinella, P.D. (1982). In *The Biochemistry and Physiology of Iron*, pp. 27-34 (Saltman, P. and Hegenauer, J. (eds)), Elsevier, New York.

Ainscough, E.W., Brodie, A.M. and Plowman, J.E. (1979). *Inorg. Chim. Acta* **33**, 149-153.

Ainscough, E.W., Brodie, A.M., McLachlan, S.J. and Ritchie, V.S. (1983). *J. Inorg. Biochem.* **18**, 103-112.

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

Aisen, P. and Harris, D.C. (1989). In *Iron Carriers and Iron Proteins* (T.M. Loehr, ed), pp. 241-351, VCH Publishers, New York.

Aisen, P. and Leibman, A. (1972). *Biochim. Biophys. Acta* **257**, 314-323.

Aisen, P. and Listowsky, I. (1980). *Ann. Rev. Biochem.* **49**, 357-393.

Aisen, P., Aasa, R. and Redfeild, A.G. (1969). *J. Biol. Chem.* **244**, 4628-4633.

Aisen, P., Aasa, R., Bo, G., Malmström, B. and Vänngärd, T (1967). *J. Biol. Chem.* **242**, 2484-2490.

Aisen, P. (1989). In *Iron Carriers and Iron Proteins*, (Loehr, T.M. (ed)), pp. 355-371, VCH Publishers, New York.

Aisen, P., Leibman, A. and Zweier, J. (1978). *J. Biol. Chem.* **253**, 1930-1937.

Al-Hilal, D., Baker, E., Carlisle, C.H., Gorinsky, B., Horsburgh, R.C., Lindley, P.F., Moss, D.S., Schneider, H. and Stimpson, R. (1976). *J. Mol. Biol.* **108**, 255-257.

Alsaadi, B.M., Williams, R.J.P. and Woodworth, R.C. (1981). *J. Inorg. Biochem.* **15**, 1-10.

Anderson, B.F., Baker, H.M., Dodson, E.J., Norris, G.E., Rumball, S.V., Waters, J.M. and Baker, E.N. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 1769-1773.

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

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

Arnold, R.R., Brewer, M. and Gauthier, J.J. (1980). *Infection and Immunity* **28**, 893-898.

Arnold, R.R., Cole, M.F. and McGhee, J.R. (1977). *Science* **197**, 263-265.

Ashwell, G. (1966). *Methods in Enzymol.* **VIII**, 93-95.

Ashwell, G. and Harford, J. (1982). *Ann. Rev. Biochem.* **51**, 531-554.

Baer, A., Oroz, M. and Blanc, B. (1979). *J. Dairy Res.* **46**, 83-93.

Bailey, S., Evans, R.W., Garratt, R.C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti,

H., Lindley, P.F., Mydin, A., Sarra, R. and Watson, J.L. (1988). *Biochemistry* **27**, 5804-5812.

Baker, E., Shaw, D.C. and Morgan, E.H. (1968). *Biochemistry* **7**, 1371-1378.

Baker, E.N. (1993). *Perspectives in Bioinorganic Chemistry* **2**, 161-205.

Baker, E.N. (1994). *Advances in Inorganic Chemistry* **41**, 389-463.

Baker, E.N. and Hubbard, R.E. (1984). *Progr. Biophys. Mol. Biol.* **44**, 97-179.

Baker, E.N. and Lindley, P.F. (1992). *J. Inorg. Biochem.* **47**, 147-160.

Baker, E.N. and Rumball, S.V. (1977). *J. Mol. Biol.* **111**, 207-210.

Baker, E.N., Anderson, B.F., Baker, H.M., Haridas, M., Jameson, G.B., Norris, G.E., Rumball, S.V. and Smith, C.A. (1991). *Int. J. Biol. Macromol.* **13**, 122-129.

Baker, E.N., Baker, H.M., Smith, C.A., Stebbins, M.R., Kahn, M. Hellström, K.E. and Hellström, I. (1992). *FEBS Lett.* **298**, 215-218.

Baker, E.N., Rumball, S.V. and Anderson, B.F. (1987). *Trends in Biochem. Sci.* **12**, 350-353.

Baldwin, G.S. and Weinstock, J. (1988). *Nucleic Acids Res.* **16**, 8720.

Bali, P.K. and Aisen, P. (1991). *Biochemistry* **30**, 9947-9952.

Bali, P.K. and Aisen, P. (1992). *Biochemistry* **31**, 3963-3967.

Bartfeld, N.S. and Law, J.H. (1990). *J. Biol. Chem.* **265**, 21684-21691.

Bellamy, W., Takase, M., Yamauchi, K., Wakabayashi, H., Kawase, K. and Tomita, M. (1992). *Biochim. Biophys. Acta* **1121**, 130-136.

Bertini, I., Luchinat, C., Messori, L., Scozzafava, A., Pellacani, G. and Sola, M. (1986a). *Inorg. Chem.* **25**, 1782-1786.

Bezkorovainy, A. (1980). In *Biochemistry of Non-heme Iron*, pp. 127-204, Plenum, New York.

Bishop, J.G., Schanbacher, F.L., Ferguson, L.C. and Smith, K.L. (1976). *Infection and Immunity* **14**, 911-918.

Bose, S., Gurari-Rotman, D., Ruegg, U.Th., Corley, L. and Anfinsen, C.B. (1976). *J. Biol. Chem.* **251**, 1659-1662.

Britten, J.R. and Koldovsky, O. (1989). *Early Human Development* **19**, 127-135.

Brock, J.H. (1985). In *Metalloproteins*, Vol. 2 (Harrison, P.M. (ed)), pp. 183-505, MacMillan, London.

Brock, J.H., Arzabe, F., Lampreave, F. and Pineiro, A. (1976). *Biochim. Biophys. Acta* **446**, 214-225.

Broxmeyer, H.E. (1978). *Blood* **51**, 889-901.

Broxmeyer, H.E., de Sousa, M., Smithyman, A., Ralph, P., Hamilton, J., Kurland, J.I. and Bognacki, J. (1980). *Blood* **55**, 324-333.

Broxmeyer, H.E., Smithyman, A., Eger, R.R., Meyers, P.A. and de Sousa, M. (1978). *J. Exp. Med.* **148**, 1052-1067.

Bullen, J.J., Rogers, H.J. and Leigh, L. (1972). *Brit. Med. J.* **3**, 69-75.

Campbell, R.F. and Chasteen, N.D. (1977). *J. Biol. Chem.* **252**, 5996-6001.

Charpin, C., Lachard, A., Pourreau-Schneider, N., Jacquemeir, J., Lavout, M.N., Andonian, C., Martin, P.M. and Toga, M. (1985). *Cancer* **55**, 2612-2617.

Chasteen, N.D., White, L.K. and Campbell, R.F. (1977). *Biochemistry* **16**, 363-368.

Chu, F.K., Trimble, R.B. and Maley, F. (1978). *J. Biol. Chem.* **253**, 8691-8693.

Cowart, R.E., Kojima, N. and Bates, G.W. (1982). *J. Biol. Chem.* **257**, 7560-7565.

Cox, T.M., Mazurier, J., Spik, G., Montreuil, J. and Peters, T.J. (1979). *Biochim. Biophys. Acta* **588**, 120-128.

Dalmastri, C., Valenti, P., Visca, D., Vittorioso, P. and Orsi, N. (1988). *Microbiologica* **11**, 225-230.

Davidson, L.A. and Lonnerdal, B. (1989). In *Milk Proteins* (Barth, C.A. and Schimne, E., eds), pp. 76-82.

Davidson, L.A. and Lönnerdal, B (1988). *Am. J. Physiol.* **254**, G580-G585.

Day, C.L., Anderson, B.F., Tweedie, J.W. and Baker, E.N. (1993). *J. Mol. Biol.* **232**, 1084-1100.

Day, C.L., Norris, G.E., Tweedie, J.W. and Baker, E.N. (1992). *J. Mol. Biol.* **228**, 973-974.

DeLucas, L.J., Suddath, F.L., Gams, R.A. and Bugg, C.E. (1978). *J. Mol. Biol.* **123**, 285-286.

Dewan, J.C., Mikami, B., Hirose, M. and Sacchenttini, J.C. (1993). *Biochemistry* **32**, 11963-11968.

Douglas, B.E. and Hollingsworth, C.R. (1985). Symmetry in Bonding and Spectra, p.254, Academic Press, New York.

Dubach, J., Gaffney, B.J., More, K., Eaton, G.R. and Eaton, S.S. (1991). *Biophys. J.* **59**, 1091-1100.

Eaton, S.S., Dubach, J., Eaton, G.R., Thurman, G. and Ambruso, D.R. (1990). *J. Biol. Chem.* **265**, 7138-7141.

Eaton, S.S., Dubach, J., More, K.M., Eaton, G.R., Thurman, G. and Ambruso, D.R. (1989). *J. Biol. Chem.* **264**, 4776-4781.

Elder, J.H. and Alexander, S. (1982). *Proc. Natl. Acad. Sci. USA* **79**, 4540-4544.

Feeney, R.E. and Komatsu, S.K. (1966). *Struct. Bonding* **1**, 149-206.

Feizi, T. (1985). *Nature* **314**, 53-57.

Finkelstein, R.A., Sciortino, C.V. and McIntosh, M.A. (1983). *Reviews of Infectious Diseases* **5**, S759-S766.

Fransson, G.-B. and Lönnerdal, B. (1980). *J. Pediat.* **96**, 380-384.

Fransson, G.-B., Keen, C.L. and Lönnerdal, B. (1983b). *Journal of Pediatric Gastroenterology and Nutrition* **2**, 693-700.

Fransson, G.-B., Thoren-Tolling, K., Jones, B., Hambraeus, L. and Lönnerdal, B. (1983a). *Nutrition Research* **3**, 373-384.

Funk, W.D., MacGillivray, R.T.A., Mason, A.B., Brown, S.A. and Woodworth, R.C. (1990). *Biochemistry* **29**, 1654-1660.

Gaber, B.P., Miskowski, V. and Spiro, T.G. (1974). *J. Am. Chem. Soc.* **96**, 6868-6873.

Garratt, R.C., Evans, R.W., Hasnain, S.S., Lindley, P.F. and Sarra, R. (1991). *Biochem. J.* **280**, 151-155.

Gelb, M.H. and Harris, D.C. (1980). *Arch. Biochem. Biophys* **200**, 93-98.

Gilliland, G.L. and Quiocho, F.A. (1981). *J. Mol. Biol.* **146**, 341-362.

Gorinsky, B., Horsburgh, C., Lindley, P.F., Moss, D.S., Parkar, M. and Watson, J.L. (1979). *Nature (London)* **281**, 157-158.

Gray, H.B. (1965). Electrons and Chemical Bonding, pp. 198-199, W.A. Benjamin, New York.

Groves, M.L. (1960). *Acta Chemica Scandinavica* **14**, 510-512.

Gutteridge, J.M.C., Halliwell, B., Treffry, A., Harrison, P.M. and Blake, D. (1983). *Biochem. J.* **209**, 557-560.

Haridas, M., Anderson, B.F., Baker, E.N. (1994). *Acta Cryst. section D*, in press.

Harris, D.C. and Gelb, M.H. (1980). *Biochim. Biophys. Acta* **623**, 1-9.

Harris, W.R. (1986). *Inorg. Chem.* **25**, 2041-2045.

Harris, W.R., Carrano, C.J., Pecoraro, V.L. and Raymond, K.N. (1981). *J. Amer. Chem. Soc.* **103**, 2231-2237.

Hendrickson, W.A. and Konnert, J.H. (1980). In *Biomolecular Structure, Function Conformation and Evolution* (R. Srinivasan, Ed) Vol. 1, pp. 43-57, Pergamon Press, Oxford.

Higashi, T. (1991). *J. Appl. Cryst.* **22**, 9-18.

Hol, W.G.J., van Duijnen, P.T. and Berendsen, H.J.C. (1978). *Nature (London)* **273**, 443-446.

Hu, W.-L., Mazurier, J., Sawatzki, G., Montreuil, J. and Spik, G. (1988). *Biochem. J.* **249**, 435-441.

Ikeda, H., Nabuchi, Y., Nakazato, K., Tanaka, Y. and Satake, K. (1985). *FEBS Lett.* **182**, 305-309.

Imber, M.J. and Pizzo, S.V. (1983). *Biochem. J.* **212**, 249-257.

Jamroz, R.C., Gasdaska, J.R., Bradfield, J.Y. and Law, J.H. (1993). *Proc. Natl. Acad. Sci. USA* **90**, 1320-1324.

Jeltsch, J.M. and Chambon, P. (1982). *Eur. J. Biochem.* **122**, 291-295.

Jhoti, H., Gorinsky, B., Garratt, R.C., Lindley, P.F., Walton, A.R. and Evans, R.W. (1988). *J. Mol. Biol.* **200**, 423-425.

Joao, H.C., Scragg, I.G. and Dwek, R.A. (1992). *FEBS Lett.* **307**, 343-346.

Johanson, B. (1960). *Acta Chem. Scand.* **14**, 510-512.

Jones, T.A. (1978). *J. Appl. Cryst.* **11**, 268-272.

Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991). *Acta Cryst.* **A47**,

110-119.

Kilar, F. and Simon, I. (1985). *Biophys. J.* **48**, 799-802.

Kojima, N. and Bates, G.W. (1981). *J. Biol. Chem.* **256**, 12034-12039.

Kornfeld, S. (1992). *Ann. Rev. Biochem.* **61**, 307-330.

Koscielak, J. (1986). *Glycoconjugate J.* **3**, 95-108.

Kretchmar, S.A. and Raymond, K.N. (1986). *J. Amer. Chem. Soc.* **108**, 6212-6218.

Laemmli, U.K. (1970). *Nature (London)* **227**, 680-685.

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

Lee, Y.C. and Montgomery, R. (1961). *Arch. Biochem. Biophys.* **93**, 292-296.

Legrand, D., Mazurier, J., Colavizza, D., Montreuil, J. and Spik, G. (1990). *Biochem. J.* **266**, 575-581.

Legrand, D., Mazurier, J., Metz-Boutigue, M.-H., Jolles, J., Jolles, P., Montreuil, J. and Spik, G. (1984). *Biochim. Biophys. Acta* **787**, 90-96.

Levitt, M. and Chothia, C. (1976). *Nature (London)* **261**, 552-558.

Lis, H. and Sharon, N. (1993). *Eur. J. Biochem.* **218**, 1-27.

Lönnerdal, B. (1984). In "*Iron Nutrition in Infancy and Childhood*", Stekel, A. (ed), Vevy/Raven, New York, pp. 95-117.

Lönnerdal, B., Forsum, E. and Hambraeus, L. (1976). *Am. J. Clin. Nutr.* **29**, 1127-1133.

Luecke, H. and Quiocho, F.A. (1990). *Nature (London)* **347**, 402-406.

Luk, C.K. (1971). *Biochemistry* **10**, 2838-2843.

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

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

MacGillivray, R.T.A., Mendez, E., Shewale, J.G., Sinha, S.K., Lineback-Zins, J. and Brew, K. (1983). *J. Biol. Chem.* **258**, 3543-3553.

Mangani, S. and Messori, L. (1992). *Inorg. Biochem.* **46**, 1-6.

Martin, A.W., Huebers, E., Huebers, H., Webb, J. and Finch, C.A. (1984). *Blood* **64**, 1047-1052.

Mason, S. (1962). *Arch. Dis. Child* **37**, 387-391.

Matthews, B.W. (1968). *J. Mol. Biol.* **33**, 491-497.

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

Mazurier, J. and Spik, G. (1980). *Biochim. Biophys. Acta* **629**, 399-408.

Mazurier, J., Aubert, J.-P., Loucheux-Lefevre, M.-H. and Spik, G. (1976). *FEBS Lett.* **66**, 238-242.

Mazurier, J., Legrand, D., Hu, W.L., Montreuil, J. and Spik, G. (1989). *Eur. J. Biochem.* **179**, 481-487.

Mazurier, J., Metz-Boutigue, M.-H., Jollés, J., Spik, G., Montreuil, J. and Jollés, P.

(1983). *Experientia* **39**, 135-141.

Mead, P.E. and Tweedie, J.W. (1990). *Nucleic Acids Research* **18**, 7167.

Mead, P.E., Thesis, PhD, Massey University, 1992.

Metz-Boutigue, M.-H., Jollés, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. and Jollés, P. (1984). *Eur. J. Biochem.* **145**, 659-676.

Moeller, T. (1982). In *Inorganic Chemistry*, pp. 460 and 470, John Wiley, New York.

Molloy, A.L. and Winterbourn, C.C. (1990). *Blood* **75**, 984-989.

Monteiro, H.P. and Winterbourn, C.C. (1988). *Biochem. J.* **256**, 923-928.

Moskaitis, J.E., Pastori, R.L. and Schoenberg, D.R. (1991). *Nucleic Acids Res.* **18**, 6135.

Nemethy, G. and Printz, M.P. (1972). *Macromolecules* **5**, 755-758.

Norris, G.E., Anderson, B.F., Baker, E.N., Baker, H.M., Gartner, A.L., Ward, J. and Rumball, S.V. (1986). *J. Mol. Biol.* **191**, 143-145.

Norris, G.E., Baker, H.M. and Baker, E.N. (1989). *J. Mol. Biol.* **209**, 329-331.

Olden, K., Parent, J.B. and White, S.L. (1982). *Biochim. Biophys. Acta* **650**, 209-232.

Oldham, C. (1987). In *Comprehensive Coordination Chemistry, Vol. 2* (G. Wilkinson, R.D. Gillard and J.A. McCleverty, eds), pp. 435-459. Pergamon Press, Oxford.

Olsen, O. and Thomsen, K.K. (1991). *J. Gen. Microbiol.* **137**, 579-585.

Parekh, R.B., Dwek, R.A., Rudd, P.M., Thomas, J.R., Rademacher, T.W., Warren, T. Wun, T.C., Reitz, H.B., Palmier, M., Ramabhadran, T. and Tiemeier (1989). *Biochemistry* **28**, 7670-7679.

Parham, P., Alpert, B.N., Orr, H.T. and Strominger, J.L. (1977). *J. Biol. Chem.* **252**, 7555-7567.

Pentecost, B.T. and Teng, C.T. (1987). *J. Biol. Chem.* **262**, 10134-10139.

Pflugrath, J.W. and Quiocho, F.A. (1985). *Nature (London)* **314**, 257-260.

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

Ploegh, H.L., Orr, H.T. and Strominger, J.L. (1981). *J. Immunol.* **126**, 270-275.

Plowman, J.E., Thesis, PhD, Massey University, 1979.

Plummer, T.H., Elder, J.H., Alexander, S., Phelan, A.W. and Tarentino, A.L. (1984). *J. Biol. Chem.* **259**, 10700-10704.

Prentice, A., Ewing, G., Roberts, S.B., Lucas, A., MacCarthy, A., Jarjou, L.M.A. and Whitehead, R.G. (1987). *Acta Paediatr. Scand.* **76**, 592-598.

Prieels, J.-P., Pizzo, S.V., Glasgow, L.R., Paulson, J.C. and Hill, R.L. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 2215-2219.

Priestle, J.P. (1988). *J. Appl. Crystallogr.* **21**, 572-576.

Princiotto, J.V. and Zapolski, E.J. (1975). *Nature* **255**, 87-88.

Rado, T.A., Wei, X. and Benz, E.J. (1987). *Blood* **70**, 989-993.

Ramakrishnan, G. and Ramachandran, G.N. (1965). *Biophys. J.* **5**, 909-933.

Regoeczi, E., Hatton, M.W.C. and Wong, K.-L. (1974). *Can. J. Biochem.* **52**, 155-161.

Rehse, P.H. and Zubak, V. (1992). High Energy Physics Laboratory Manual.

Reiter, B. (1983). *Int. J. Tiss. Reac.* **1**, 87-96.

Richardson, J.S. (1985). *Methods in Enzymology* **115**, 359-380.

Rogers, T.B., Borresen, T. and Feeney, R.E. (1978). *Biochemistry* **17**, 1105-1109.

Rose, T.M., Plowman, G.D., Teplow, D.B., Dreyer, W.J., Hellström, K.E. and Brown, J.P. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 1261-1265.

Rossenau-Montreff, M.Y.F., Soetewey, R., Lamote, R. and Peeters, H. (1971). *Biopolymers* **10**, 1039-1048.

Saarinen, U.M., Simes, M.A. and Dallman, P.R. (1977). *J. Pediatrics* **91**, 36-39.

Sakabe, N. (1991). *Nuclear Instrum. and Methods in Physics Research* **A303**, 448-463.

Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. and Lindley, P. (1990). *Acta Crystallogr.* **B46,** 763-771.

Schlabach, M.R. and Bates, G.W. (1975). J. Biol. Chem. **250,** 2182-2188.

Schneider, D.J., Roe, A.L., Mayer, R.J. and Que, L. (1984). *J. Biol. Chem.* **259**, 9699-9703.

Shongwe, M.S., Smith, C.A., Ainscough, E.W., Baker, H.M., Brodie, A.M. and Baker, E.N. (1992). *Biochemistry* **31**, 4451-4458.

Sibanda, B.L., Blundell, T.L. and Thornton, J.M. (1989). *J. Mol. Biol.* **206**, 759-777.

Smith, C.A. , Thesis, PhD, Massey University, 1993.

Smith, C.A., Ainscough, E.W., Baker, H.M., Brodie, A.M. and Baker, E.N. (1994b). *J. Am. Chem. Soc.* **116**, 7889-7890.

Smith, C.A., Anderson, B.F., Baker, H.M. and Baker, E.N. (1992). *Biochemistry* **31,** 4527-4533.

Smith, C.A., Anderson, B.F., Baker, H.M. and Baker, E.N. (1994). *Acta Cryst.* **D50**, 302-316.

Smith, C.A., Baker, H.M. and Baker, E.N. (1991). *J. Mol. Biol.* **219**, 155-159.

Sola, M. (1990). *Eur. J. Biochem.* **194**, 349-353.

Sorenson, M. and Sorenson, S.P.L. (1939). *Comptes Rendus des Travaux du Laboratorie Carlsberg* **23**, 55-59.

Spik, G., Coddeville, B. and Montreuil, J. (1988). *Biochimie* **70**, 1459-1469.

Spik, G., Coddeville, B., Legrand, D., Mazurier, J., Leger, D., Goavec, M. and Montreuil, J. (1985). In Proteins of Iron Storage and Transport (Spik, G., Montreuil, J., Crichton, R.R. and Mazurier, J., eds) pp. 47-51, Elsevier Press.

Spik, G., Strecker, G., Fournet, B., Bouquelet, S., Montreuil, J., Dorland, L., van Halbeek, H. and Vliegenthart, J.F.G. (1982). *Eur. J. Biochem.* **121**, 413-419.

Springer, T.A. (1990). *Nature* **346**, 425-434.

Stowell, K.M., Rado, T.A., Funk, W.D. and Tweedie, J.W. (1991). *Biochem. J.* **276**, 349-355.

Stratil, A., Bobak, P. and Valenta, M. (1983). *Comp. Biochem. Physiol.* **74B**, 603-610.

Tan, A.T. and Woodworth, R.C. (1969). *Biochemistry* **8**, 3711-3716.

Tarentino, A.L. and Plummer, T.H. (1987). *Methods in Enzymol.* **138**, 770-778.

Tarentino, A.L., Plummer, T.H. Jr. and Maley, F. (1974). *J. Biol. Chem.* **249**, 818-824.

Tashiro, Y. and Trevithick, J.R. (1977). *Can. J. Biochem.* **55**, 249-256.

The CCP4 Suite: Programs for Protein Crystallography 1994. *Acta Cryst.* **D50**, 760-763.

Tomita, M., Takase, M., Wakabayashi, H. and Bellamy, W. (1994). *Advances in Experimental Medicine and Biology* **357**, 209-218.

Valenti, P., Visca, P., Antonini, G., Orsi, N. and Antonini, E. (1987). *Med. Microbiol. Immunol.* **176**, 123-130.

van Bockxmeer, F.M. and Morgan, E.H. (1982). *Comp. Biochem. Physiol.* **71A**, 211-218.

Vyas, N.K., Vyas, M.N. and Quiocho, F.A. (1983). *Proc. Natl. Acad. Sci. USA* **80**, 1792-1796.

Warner, R.C. and Weber, I. (1953). *J. Am. Chem. Soc.* **75**, 5094-5101.

Wassaram, P.M. (1990). *Development* **108**, 1-17.

Wharton, B.A., Balmer, S.E. and Scott, P.M. (1994). *Advances in Experimental Medicine and Biology* **357**, 91-98.

Williams, J. (1962). *Biochem. J.* **83**, 355-364.

Williams, J. (1974). *Biochem. J.* **141**, 745-752.

Williams, J. (1975). *Biochem. J.* **149**, 237-244.

Williams, J. and Moreton, K. (1988). *Biochem. J.* **251**, 849-855.

Williams, J., Elleman, T.C., Kingston, I.B., Wilkins, A.G. and Kuhn, K.A. (1982b). *Eur. J. Biochem.* **122**, 297-303.

Wilson, A.J.C. (1942). *Nature* **150**, 152.

Winton, E.F., Kinkade, J.M., Vogler, W.R., Parker, M.B. and Barnes, K.C. (1981). *Blood* **57**, 574-578.

Zak, O. and Aisen, P. (1988). *Biochemistry* **27**, 1075-1080.

Zweier, J.L. (1980). *J. Biol. Chem.* **255**, 2782-2789.

Zweier, J.L. and Aisen, P. (1977). *J. Biol. Chem.* **252**, 6090-6096.

Zweier, J.L., Aisen, P., Peisach, J. and Mims, W.B. (1979). *J. Biol. Chem.* **254**, 3152-3515.

Zweier, J.L., Peisach, J. and Mims, W.B. (1982). *J. Biol. Chem.* **257**, 10314-10316.

Zweier, J.L., Wooten, J.B. and Cohen, J.S. (1981). *Biochemistry* **20**, 3505-3510.