A mutational analysis of the hinge region of the N-lobe of lactoferrin.

A thesis presented in partial fulfilment of the requirements for the degree of Master in Science in Biochemistry at Massey University.

Steven Christopher Shewry

1996

ABSTRACT

Lactoferrin is an 80kDa iron binding glycoprotein that is found as a major component of human milk, as well as in many other exocrine solutions. Lactoferrin binds reversibly, and with high affinity, 2 Fe^{3+} ions with 2 synergistic CO_3^{2-} ions. The crystal structure shows that the polypeptide chain is folded into two similar lobes, each binding one Fe^{3+} and CO_3^{2-} ion. In the metal free state, the N-terminal lobe has been found to adopt an open structure, with rotation occurring around two residues found in separate beta strands located at the back of the binding site in a region referred to as the hinge. A sequence alignment of these two strands over the greater transferrin family shows a very high level of conservation particularly of the two residues at the centre of this rotation (Pro 251 and Thr 90).

The N-terminal half of human lactoferrin (Lf_N) has been constructed, expressed and the crystal structure determined.

In an attempt to understand the importance of the conservation of these two residues, and their effect on binding, a mutational analysis was initiated. Oligonucleotide site-directed mutagenesis has been used to construct mutants in the cDNA encoding for human lactoferrin using the M13 bacteriophage. The mutant cDNA was transformed into a mammalian expression vector (pNUT). After transfection of the pNUT vector into baby hamster kidney cells (BHK), the mutant proteins were expressed and purified from the culture medium using a CM-sephadex ion-exchange column.

Absorption maxima and pH-dependent iron-release experiments were carried out on the mutants. The data shows that the mutants behave essentially the same as Lf_N, the exception being P251G which appears to release the iron over a shorter pH range. The reason for this is not yet fully understood.

The crystal structure of P251A in the iron-bound form was solved by molecular replacement using the structure of LfN as the starting model. The structure of P251A was refined using data between 20.0 and 2.0 Å. The current model has good geometry and has an R-factor of 18.6 %. Analysis of the structure shows that it is essentially identical to that of the LfN structure.

Although the structure of the iron-free form has not been determined, it appears that changes to the hinge region of the N-lobe of lactoferrin do not affect the iron-binding or structural characteristics of the protein.

Acknowledgements.

This thesis was made possible by the contributions of a number of people, to whom I wish to offer my appreciation.

Firstly to my supervisors, Professor Ted Baker and Associate professor John Tweedie for their advice and guidance throughout the course of this work. I am particularly grateful for Ted who despite a busy schedule still found time to read this thesis in fast time.

To all members of the protein crystallography and Twilight zone groups (past and present) for helping me come to terms with the hazy aspects of this work. Special thanks go to Heather Baker, Drs Gillian Norris, Clyde Smith, Jakki Cooney, Stanley Moore, Rick Faber, Maria Bewley, and to Neil Peterson for their kind ear during the difficult times.

Special mention to Dr Hale Nicholson for his help in the initial stages of the DNA work and for helping me to enter this mysterious world of nucleotides, and to Dr Bryan Anderson for his help with the confusing (I mean computing) part of the structure refinement.

Special thanks also to Dr Simon Brown for his help in the mathematical part of the iron release curves. Coming to terms with the equations is not an activity I would wish on anyone.

I am especially grateful to my family (my wife Philippa and children Jackson, Nathan, and Esther) for standing by me during what must have been a black hole experience for them. They still don't know what I was doing all that time. In particular to the youngest two who know no other way of life. May it get more interesting for them from now on.

Table of Contents

List of Figures.

List of Tables

Abbreviations

A280	Absorbance at 280 nm
A454	Absorbance at 454 nm
ApoLf	Iron free native lactoferrin
ApoLf$_N$	Iron free Lf$_N$
ATP	adenosine triphosphate
BHK	baby hamster kidney
bp	base pair
BRL	Bethesda Research Laboratories
BSA	bovine serum albumin
cDNA	complementary DNA
CM-sephadex	carboxymethyl sephadex
C-terminal	carboxy terminal
Cu$_2$Lf	copper saturated native lactoferrin
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DG	deglycosylated
DGLf$_N$	deglycosylated Lf$_N$
dGTP	deoxyguanosine triphosphate
DHFR	dihydrofolate reductase
DMEM	Dulbeco's modification of Eagle's medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	mixture of dATP, dTTP, dCTP and dGTP
dsDNA	double stranded DNA
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
EDTA	ethylene diamine tetra-acetate
ELISA	enzyme linked immunoabsorbant assay
Endo F	endo β-N-acetylglucosamidase F1
F's	structure factors
Fcalc	calculated structure factors
Fobs	observed structure factors
F12	Hams-F12-medium
FCS	foetal calf serum
FeLf$_N$	iron saturated Lf$_N$
Fe$_2$Lf	iron saturated native lactoferrin

HEPES	N-2-hydroxethyl piperazine-N'-2-ethane sulfonic acid
hLf	human lactoferrin
hTr	human serum transferrin
I	intensities
IPA	isopropanol
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
LB	Luria-Bertani
Lf	lactoferrin
Lf_N	amino terminal half of human lactoferrin
MIR	multiple isomorphis replacement
MTX	methotrexate
NAG	N-acetyl glucosamine
NMR	nuclear magnetic resanance
NTA	nitrilotriacetic acid
N-terminal	amino terminal
PAGE	polyacrylamide gel electrophoresis
PBP	periplasmic-binding protein
PBS	phosphate buffered saline
PEG	polyethylene glycol
pfu	plaque forming units
PMN	polymorphonuclear
PNGase F	peptide-N4-(N-acetyl-β-D-glucosaminyl) asparagine amidase F
RNA	ribonucleic acid
RNAse	ribonuclease
SDS	sodium dodecylsulphate
SDS-PAGE	sodium dodecylsulphate polyacrylamide gel electrophoresis
ssDNA	single stranded DNA
TAE	Tris acetate EDTA
TE	10 mM Tris, 1 mM EDTA
TEMED	N,N,N',- Tetramethylethylenediamine
Tris	Tris-(hydroxymethyl) aminomethane
UV	ultraviolet
X-gal	5-bromo-4-chlor-3-indoyl β-D-galactopyranoside

A. INTRODUCTION

A.1. Introduction.

For all plants and animals, and for virtually all microbes, life without iron is impossible. In the cell, where essential iron-containing enzymes and proteins are found in species ranging from bacteria to man, the thermodynamically stable form of iron is Fe(III). Free Fe^{3+} is, however, rapidly hydrolysed in solution to form the insoluble $Fe(OH)_3$ complex. This insolubility problem is overcome by proteins involved in storage and transport of iron.

Transferrin is the name given to the family of proteins involved in the solubilisation, sequestration, and transport of ferric iron. The roles of transferrins in general is coupled to their tight binding of Fe^{3+}. These roles involve controlling the levels of free iron in body fluids, preventing precipitation of $Fe(OH)_3$ by keeping the concentration of free Fe^{3+} at less than 10^{-18} M, and preventing free radical damage (free radical production is catalysed by free iron). The name "Transferrin" is taken literally to mean "transport of iron" (Baker, 1994) and the transferrin family includes at least four iron-binding proteins, serum transferrin, ovotransferrin, lactoferrin, and melanotransferrin. Serum transferrin is the iron transport protein found in the blood of vertebrates and some invertebrates. Ovotransferrin, found in the white of eggs, is identical to the serum transferrin of birds except in the carbohydrate moiety attached to the protein (Thibodeau *et al,* 1978). Melanotransferrin is a membrane-associated protein found on all cells, but expressed at high levels on melanoma cells (Rose *et al,* 1986). Lactoferrin, the subject of this study, is the iron-binding protein most commonly found in milk, but also in other exocrine secretions including tears, saliva, seminal fluid, cervical mucous, gastric fluid, nasal exudate, bronchial mucous, and hepatic bile (Masson *et al,* 1969; Weinburg, 1984). Lactoferrin is also found in the specific granules of mature neutrophils (Baggolini *et al,* 1970).

The transferrins are glycoproteins with a molecular weight of about 78,000 Da. The number and positioning of the carbohydrate chains differs from one transferrin to another. For example, human serum transferrin has two chains attached to the C-lobe of the molecule, while human lactoferrin also has two chains, but one attached to each of the lobes.

Melanotransferrin has additional amino acids on the C terminal end of the molecule thought to be involved in anchoring the protein to the cell membrane (Rose *et al*, 1986), and responsible in part for its higher molecular weight of 97,000 Da. Other transferrins of differing size are the crab serum transferrin (Cancer magister) with a molecular weight of about 150,000 Da, and the transferrin of Pyura stolonifera with a molecular weight of about 40,000 Da (Baker, 1994).

The transferrins are generally folded into two lobes each of which reversibly binds a ferric ion concomitantly with a bicarbonate anion. Two members of the family, however, melanotransferrin and the transferrin of the hornworm (Manduca sexta), bind only one ferric ion (Baker, 1994) as a result of amino acid sequence changes in one of the two binding sites. Apart from these, the metal and anion-binding amino acid residues are identical in both lobes of the transferrins, and are conserved in all members of the transferrin family.

Lactoferrin is found in the milk of some, but not all, species, being completely absent in the milk of the rat, rabbit and dog. The milk of the rat and rabbit contains significant levels of transferrin, while the milk of the dog contains neither protein (Masson and Heremans, 1971). Lactoferrin levels are highest in slow-growing species such as the human and guinea pig, and are low or absent in fast-growing species such as rabbits, rats and cattle (Weinburg, 1984). Lactoferrin expression also varies during lactation and this suggests that the gene is hormonally regulated. This is supported by the work of Pentecost and Teng (1982), who found that the major oestrogen-stimulated protein synthesised in the uterine tissue of the mouse is lactoferrin.

A.2. Biological roles of lactoferrin

For many years, a considerable amount of effort has been directed at determining the biological role(s) of lactoferrin. Many different roles have been postulated, but as yet, no single, primary one has been identified. Among those that have been suggested are (i) iron-withholding in order to starve potential pathogens of iron, (ii) a bactericidal function, (iii) binding to receptors in order to perform various functions, (iv) modulation of the inflammatory response, (v) acting as a growth promoter, and (vi) as a source of iron for nutritional reasons.

Many other functions have been associated with lactoferrin, but the evidence is somewhat dubious. The following are the most widely accepted potential functions.

A.2.1. Iron withholding.

The egg is a good example of iron-withholding defence. The developing embryo is provided with a large amount of iron (1 μg in the egg yolk) for use in DNA synthesis, electron transport, and other important functions, while a porous shell is needed in order for air to be available. This porous shell, however, allows microbial invaders to enter the egg, and this problem of microbial invasion is overcome by placing no iron (an essential element for bacteria) in the egg white, while also including a powerful iron-binding protein. This iron-binding protein was discovered by Osborne and Campbell (1900), and is called ovotransferrin.

Schade and Caroline (1944) demonstrated that of the 10 vitamins and 31 elements tested, only iron overcame the bacteriostatic activity of egg white. The bacteriostatic factor (ovotransferrin) suppressed activity of gram negative bacteria, gram positive bacteria, and also fungi. Alexander in 1948 discovered that babies that were breast fed suffered less severe cases of gastric enteritis than did bottle fed babies. A factor in human milk was therefore responsible for this reduced level of infection. It was later discovered (Bullen *et al*, 1972) that samples of milk with unsaturated iron-binding capacities had a powerful bacteriostatic effect on *E.coli* 0111/B4. This effect was abolished if the proteins involved were saturated with iron.

In vivo work (Murray *et al*, 1978) shows that iron indeed enhances the incidence of infection. Somali nomads have a high incidence of iron deficiency, due to their all-milk diet, and also show a low level of intestinal parasites. Murray *et al* fed the nomads daily for 30 days, either with placebo (aluminium hydroxide) as a control, or with iron (900 mg $FeSO_4$). The subjects were observed for fever above 38°C and symptoms of infection. 7/67 of the placebo group and 36/71 of the iron group suffered infections, suggesting that the host defence against these infections was better with an iron deficiency than with iron present.

The factor accounting for these observations (lactoferrin), was discovered, purified and studied around 1960 (Johansson, 1960) An antibacterial role for lactoferrin involving iron sequestering is also supported indirectly by the observation that both human and bovine lactoferrin are secreted in the apo form (Lonnerdal, 1985).

It seems that vertebrate hosts inhibit microbial growth by withholding iron, and do so by the use of proteins of the transferrin family, serum transferrin, ovotransferrin, and lactoferrin.

A.2.2. Bactericidal activity.

Although the bacteriostatic activity of lactoferrin, based on iron sequestering, appeared well established, Arnold *et al* (1977) demonstrated that in some cases this activity was not reversed by the addition of iron. This pointed to an alternative mechanism that could be involved in some cases. In this work, immunofluorescence studies showed apolactoferrin binding to the surface of susceptible bacteria. A possible explanation for this was that inhibition was due to lactoferrin blocking sites that were required for the transport of an essential nutrient. If this was so, then inhibition should be reversed by the removal of any surface lactoferrin, and any lactoferrin-inhibited cells should then be able to continue to grow. Arnold *et al* (1982) removed the lactoferrin by techniques that retained the viability of control bacteria. The experimental bacteria (*Streptococcus mutans*) however failed to regain viability upon treatment to remove the lactoferrin. In effect, it showed the irreversible bactericidal effect of lactoferrin upon *S.mutans*. Other supporting evidence comes from Valenti *et al* (1985 and 1987) who showed, by use of a dialysis membrane, that the lactoferrin needed to be in direct contact with the bacterial surface to exert its bactericidal effect.

In a recent report (Tomita *et al,* 1994), lactoferrin was found to agglutinate the protoplasts of *Micrococcus lutens*. This agglutination was lost by chemical modification of the basic residues of lactoferrin, indicating that electrostatic action is involved. Using phase contrast microscopy, and spectroscopy, Tomita *et al*, (1994) found that both the apo and iron loaded forms of lactoferrin inhibited the growth of the bacteria. This points to an antibacterial mechanism in addition to the iron witholding bacteriostatic mechanism, as the latter should be effective only with the apo form of lactoferrin. Ovotransferrin and human transferrin had no effect and it was concluded that the cationic charge from lysine or arginine residues in lactoferrin are needed to agglutinate the bacterial cells.

Bactericidal activity is also found in proteolytically-derived and synthetic cationic peptides corresponding to the N terminal regions of lactoferrin (Bellamy *et al,* 1992). The precise bactericidal mechanism remains unclear. It may involve direct interruption of the membrane by this region, or the blocking of essential receptors on the bacterial outer membrane surface.

To understand the mechanism of action of one of these peptides, Bellamy *et al,* (1993) investigated the cell binding properties of lactoferricin B (a peptide corresponding to the N terminal 25 amino acids of bovine lactoferrin) with both gram positive (*Bacillus subtilis*) and gram negative (*E.coli*) bacteria, and compared these properties with the rate of irreversible death of these cells. Lactoferricin B is highly cationic with 8/25 basic residues. Other antimicrobial peptides with similar cationic

features include magainins from frog skin; cecropins from the haemolymph of insects; and defensins from mammalian neutrophils (Westerhoff *et al,* 1989), (Hill *et al,* 1991).

^{14}C-labelled lactoferricin B bound to the cell surface of both gram negative and gram positive bacteria and the rate of binding was consistent with the rapid rate of killing observed. Cell-binding activity was pH dependent implying that the cationic property of the peptide was important. The optimum pH however was dependent on the particular bacterial strain (pH 6.0 for *E.coli*, pH 7.5 for *B.subtilis*). With each strain, the killing effect was maximum near the optimum pH for cell binding implying that the bactericidal effect of lactoferricin B is dependent upon cell binding. This reflects the earlier work by Valenti *et al* (1987) with lactoferrin.

Naidu *et al* (1993) generated a peptide from lactoferrin by pepsin hydrolysis. This also showed antimicrobial activity, but failed to inhibit ^{125}I-labelled lactoferrin from binding to *E.Coli* and 10 other species of the *Enterobacteriaceae*. This indicated that lactoferrin and this peptide may bind by different mechanisms, and cast some doubt on the relevance of peptides in elucidating the function of lactoferrin.

An interesting twist to this bacteriostatic/bactericidal role is the work by Yu and Schryvers (1993) showing that some pathogenic bacteria (*Neisseria meningitidis*, *N.gonorrhoeae*, and *Moraxella catarrhalis*) may work by inhibiting lactoferrin and transferrin. They found by SDS PAGE that these bacteria bound lactoferrin to their membrane, and that the receptors that bound human lactoferrin were different from the ones that bound human transferrin. They propose that these bacteria bind lactoferrin in order to acquire iron from it, therefore overcoming its bacteriostatic effect. It does not rule out, however, a defensive role against the bactericidal effect of lactoferrin.

A.2.3. Receptor binding.

One of the intriguing characteristics of lactoferrin is its ability to bind to a wide variety of cell types. Several researchers have investigated this binding, and have related it to possible functions for lactoferrin.

A putative intestinal receptor for lactoferrin in humans was initially proposed by Cox *et al* (1979) and was isolated from rabbit brush border cells in 1989 (Mazurier *et al,* 1989). This receptor was found to be a 100 kDa protein.

Lactoferrin has also been found to bind to the brush border cells of rhesus monkeys (Davidsson and Lönnerdal, 1988) under conditions in which these cells can accumulate iron from lactoferrin. Only human lactoferrin released its iron to these cells, bovine lactoferrin, human transferrin, and ovotransferrin having no such ability

(Davidsson and Lönnerdal, 1989). Lactoferrin receptors have also been reported in mouse small intestinal brush border cells (Hu *et al,* 1988).

The binding constant and number of receptors per cell have been calculated for caco-2-cells, a human colon carcinoma cell line with characteristics of the brush border cells (Iyer *et al,* 1993). Human lactoferrin was labelled either with ^{59}Fe, or with ^{125}I. The results showed that human lactoferrin in both forms bound to the receptor in a saturable manner. Scatchard analysis indicated that it was a single binding site with a Kd of 1.7×10^{-6}, and about 6×10^{5} binding sites per cell. At 37°C, lactoferrin was taken up by the cell. ^{59}Fe taken up as ^{59}Fe-Lf was not transported across the monolayer, but when ^{125}I-Lf was used, there was a continuous transport of ^{125}I associated with a major protein later confirmed to be lactoferrin using gel electrophoresis and ELISA. In conclusion, Fe_2Lf is taken up by the cell, iron is then released and used by the cell, and Lf in part is transported across the cell.

This receptor has been cloned, sequenced and expressed in a baculovirus system, and confirmed to bind hLf by ligand blotting. The protein consists of 351 amino acids with four potential glycosylation sites. There is a strong hydrophobic region of 29 amino acids possibly acting to anchor the receptor to the membrane (Iyer *et al,* 1994).

Lactoferrin receptors have also been described in other cell types including lymphocytes, parenchyma liver cells, macrophages, and monocytes. With human peripheral lymphocytes, Mazurier *et al* (1989) found that in these cells during the resting state there were no lactoferrin receptors present. However upon stimulation by phytohemagglutinin, lactoferrin receptors were visualised using ^{125}I-labelled lactoferrin as a probe. The receptor gave rise to 2 bands of 100 and 110 kDa.

A high affinity receptor has also been discovered in monocytes with a Kd of 1 nM (Yuen *et al,* 1993). This work found that both monocytes and lymphocytes showed a dose dependent response in interleukin(IL)-1B and IL-6 to increasing levels of lactoferrin. It appears that lactoferrin binds to these cells and this results in increased IL-1B and IL-6 levels.

McAbee and Esbensen (1991) showed that hepatocytes also bind and internalise lactoferrin. 80% of the lactoferrin was internalised by these cells after 60 min at 37°C. The functional consequences of lactoferrin binding to these cells remains unclear, although lactoferrin has been found, in vitro and in vivo, to inhibit the binding and uptake of apoE-bearing lipoproteins by parenchyma liver cells (Huettinger *et al,* 1988; Van Dijk *et al,* 1991). Arginine residues in lactoferrin were shown to be crucial for recognition of lactoferrin by the liver cells (Ziere *et al,* 1992). More recent work by the same group (Ziere *et al,* 1993), showed, by removal of the first 14 amino acids, that it is the four arginine residues at the N terminus that are important.

As previously outlined, the pathogenic bacteria *Neisseria meningitidis*, *N.gonorrhoeae*, and *Moraxella catarrhalis*, overcome the iron sequestering effect of lactoferrin by using receptors in their membranes to bind the individual lobes of lactoferrin (Yu and Schryvers, 1993). In the same way, lactoferrin can also bind to certain bacteria from the Enterobacteriaceae family. Gado *et al* (1991) identified porins as the lactoferrin binding site on these bacteria. This binding was reversible and had a low affinity. Yu and Schryvers (1993) found that it was both the C- and N-lobes of lactoferrin that were responsible for binding to bacterial receptors. This contrasts with transferrin, in which the bacterial transferrin receptor only binds to the C-lobe of human transferrin (Alcantara *et al*, 1993). The region of human lactoferrin involved in binding to human lymphocytes however, has been localised to domain 1 of the N-lobe and there is no observable binding of the C-lobe (Rochard *et al*, 1989), (Legrand *et al*, 1992). This shows that the interaction with lactoferrin is different for mammalian receptors than for bacterial receptors.

A.2.4. Inflammation.

Another function proposed for lactoferrin is modulation of the inflammatory response. Lactoferrin is found in the specific granules of neutrophils (Baggiolini *et al*, 1970), and during the inflammatory response, levels of lactoferrin in the granules decline, while plasma levels of lactoferrin increase (Malmquist *et al*, 1978). It seems that the neutrophils release the lactoferrin, although it is not known how this occurs. Boxer *et al* (1982), using immunohistological techniques, found that lactoferrin released from neutrophils bound to the polymorphonuclear (PMN) membrane and altered its surface properties by reducing the surface charge. Transferrin did not do this. This work correlates with the observation that during inflammation there is an increased adherence of the PMN cells to the endothelial cells which results in an amplification of the inflammatory response (Oseas *et al*, 1981). By binding to the PMN membrane, lactoferrin helps the PMN cells to adhere to the endothelial cells and therefore amplify the inflammatory response.

The release of lactoferrin from neutrophils has also been suggested to have an antibacterial role (van Snick *et al*, 1974). It was proposed that lactoferrin released from the neutrophils bound free iron, possibly derived from transferrin, and once iron-saturated, lactoferrin bound to macrophages which were then removed by the reticuloendothelial system. This bacteriostatic role of lactoferrin in neutrophils is supported by the finding that patients whose neutrophils lack the specific granules suffer from recurrent infections (Sanchez *et al*, 1992), but has been questioned by Baynes and Bezuida (1992) after they could not detect significant levels of lactoferrin

in the plasma.

A.2.5. Growth factor activity.

Transferrin is known to be involved in cell proliferation (Casey *et al*, 1989), and this raises the possibility that lactoferrin may also have a role in growth-promoting activity. This was proposed when human lactoferrin was found to stimulate the growth of all the human lymphoid cell lines tested, but not mouse lymphocyte cell lines, a human epithelial cell line, or a human fibroblast cell line (Hashizume *et al*, 1983). Human lactoferrin in the iron loaded form has also been found to stimulate the growth of phytohaemagglutinin-stimulated human peripheral blood lymphocytes. Moreover, receptors for lactoferrin appeared on the surface of these cells upon stimulation by phytohaemagglutinin (Mazurier *et al*, 1989).

In addition to its possible growth-promoting activity, lactoferrin has also been shown to indirectly inhibit the growth of certain cell lines. Lactoferrin is responsible for regulating the proliferation of the granulocyte-macrophage progenitor cells by decreasing the production and release of a colony-stimulating activity found in macrophages (Broxmeyer *et al* 1978) . Lactoferrin is known to bind monocytes and macrophages, and this binding activity correlates with the inhibitory effect on these cells.

A.2.6. Iron nutrition.

Another suggested role for lactoferrin is in the uptake of iron for nutritional value. To perform this function, lactoferrin must be able to withstand proteolysis in the gastrointestinal tract, and evidence for this resistance was provided in the work of Spik *et al* (1982). A nutritional role for lactoferrin is also supported by the finding that receptors for lactoferrin have been identified in the small intestines of a variety of species (Iyer and Lonnerdal 1993). Lactoferrin binding then apparently leads to the retention of the iron by these cells (Iyer *et al*, 1993).

This function is questioned by Davidsson *et al* (1994) who separated out lactoferrin from human milk by treatment with heparin-sepharose, and then measured the incorporation of ^{58}Fe into red blood cells 14 days after feeding infants aged 2-10 months old. Their results showed that less ^{58}Fe was incorporated into red blood cells with lactoferrin-containing milk, than with lactoferrin-deficient milk.

A.2.7. Other possible roles.

As mentioned earlier, lactoferrin has many other postulated functions, one of which is to protect monocytes from oxidation (Brighton *et al,* 1991). It is proposed that lactoferrin binds to a receptor on the monocyte preventing hydroxyl damage to these and neighbouring cells.

Whatever the function(s) of lactoferrin, it is generally accepted that most are linked in some way to the binding of iron.

A.3. Structure of lactoferrin.

A.3.1. Primary structure.

At the present time, the amino acid sequences of 17 proteins of the transferrin family have been determined either directly or from cDNA sequences. These are listed in Table A.1.

All the members of the transferrin family demonstrate a high degree of sequence similarity. This includes approximately 65% amino acid identity within the lactoferrins, and approximately 55% between the transferrins of higher animals and lactoferrin (Mead, 1992). The sequences of the insect transferrins and the membrane-associated melanotransferrin are less conserved. Melanotransferrin has a 40% amino acid sequence identity with the mammalian transferrins, while the insect transferrins have 20 - 30% amino acid sequence identity with the higher transferrins (Baker, 1994).

In addition to the amino acid sequences, the complete genomic DNA sequences for human transferrin (Schaeffer *et al,* 1987), chicken ovotransferrin (Jeltsch *et al,* 1987), and mouse lactoferrin (Shirsat *et al,* 1992), have been reported. Comparisons of these sequences reveal that although the sizes of the genes vary, this is due to variation in the sizes of the introns and not the exons (coding regions). The positions of the intron/exon boundaries are conserved among transferrins. The position of one of the conserved exon/intron boundaries is between the N and C-lobes of the protein, supporting the theory of gene duplication put forth by Bowman and colleagues (1988).

Table. A.1. Primary sequences of the transferrin family known to date.

Protein	Reference
Human Serotransferrin	Park *et al*, 1985, MacGillivray *et al*, 1983
Horse Serotransferrin	Carpenter *et al*, 1993;
Pig Serotransferrin	Baldwin and Weinstock, 1988
Rabbit Serotransferrin	Banfeld *et al*, 1991
Rat Serotransferrin	Schreiber *et al*, 1979
Xenopus Serotransferrin	Moskaitis *et al*, 1991
Human Lactoferrin	Rado *et al*, 1987, Metz-Boutigue *et al*, 1984
Bovine Lactoferrin	Mead and Tweedie, 1990
Goat Lactoferrin	Le Provost *et al*, 1994
Mouse Lactoferrin	Pentecost and Teng, 1987
Pig Lactoferrin	Alexander *et al*, 1992
Human Neutrophil Lactoferrin	Rado *et al*, 1987
Human Melanotransferrin	Rose *et al*, 1986
Chicken Ovotransferrin	Jeltsch and Chambon,1982; Williams *et al*, 1982
Atlantic Salmon Transferrin	Kvingedal *et al*, 1994
Hornworm Transferrin	Bartfeld and Law,1990
Cockroach Transferrin	Jamroz *et al*, 1993

A.3.2. Three-dimensional structure.

A.3.2.1. Iron-loaded lactoferrin.

The first crystallographic studies on transferrins date back more than 20 years (Magdoff-Fairchild and Low, 1970), and the first low resolution analysis of a transferrin was published in 1979 (Gorinsky *et al*,1979). This involved rabbit serum transferrin, and demonstrated the bilobal nature of the molecule.

In 1987, Anderson *et al* determined the structure of human lactoferrin at 3.2 Å resolution and this structure was subsequently refined at 2.8 Å resolution (Anderson *et al*, 1989). The polypeptide folding of this structure is shown in Fig A.1.

Fig. A.1. Drawing created by Ribbons (Richardson, 1985; Priestle, 1988) showing the structure of the full-length lactoferrin molecule with iron bound in both lobes.

The polypeptide chain is folded into two lobes, representing the N- and C-terminal halves of the molecule (residues 1-332 and 345-691 respectively). These are joined by a short, 3-turn, alpha helix, (residues 333-344) that consists largely of glutamyl, arginyl and alanyl residues. Both lobes have very similar polypeptide folding, consistent with their high level of amino acid sequence identity (about 40%), and each lobe contains a single iron-binding site.

Inspection of the folding (Fig A.1) shows that each lobe is further subdivided into two similar sized domains (of about 160 residues), with a deep cleft between them, which houses the iron-binding site (Fig A.2). The domains in the N-lobe of human lactoferrin have been labelled N1 (residues 1-90 and 252-332) and N2 (residues 91-251), with the equivalent C-lobe domains being C1 (residues 345-433 and 596-691) and C2 (residues 434-595).

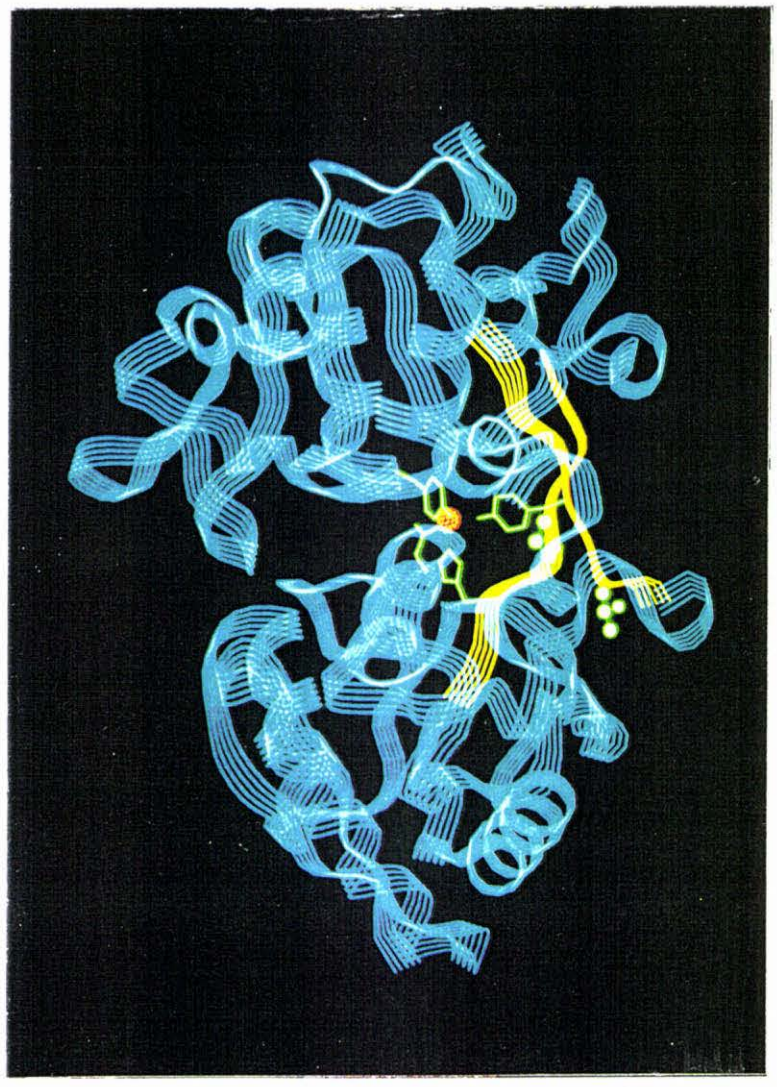


Fig. A.2. Drawing created by Ribbon (Richardson, 1985; Priestle, 1988) showing the structureof the half-length lactoferrin molecule (Lf_N) with iron bound. The side chains for the ligands are included, as are the side chain of two residues of this study, Thr 90 and Pro 251.

The two iron-binding sites are extremely similar. In each case the iron atom is coordinated by four protein sidechains, 2 tyrosines, 1 histidine, and 1 aspartate, coming from separate parts of the polypeptide chain. In the N-lobe these ligands are Asp 60, Tyr 92, Tyr 192, and His 253, while in the C-lobe they are Asp 395, Tyr 435, Tyr 528, and His 597. A unique feature of transferrins is that a carbonate (or bicarbonate) anion is bound with each metal ion, and neither metal ion nor anion is bound strongly by the protein in the absence of the other. The relationship is thus said to be synergistic. In the binding site the carbonate anion is found to bind to the metal in bidentate fashion, thus completing the metal coordination, and to also form a bridge between the metal ion and a positively charged region of the protein. The latter comprises an arginine sidechain

(Arg 121 in the N-lobe, Arg 465 in the C-lobe) and the N terminus of an alpha helix, these being presented by one wall of the second domain (N2 or C2). The binding site is shown schematically in Fig A.3.

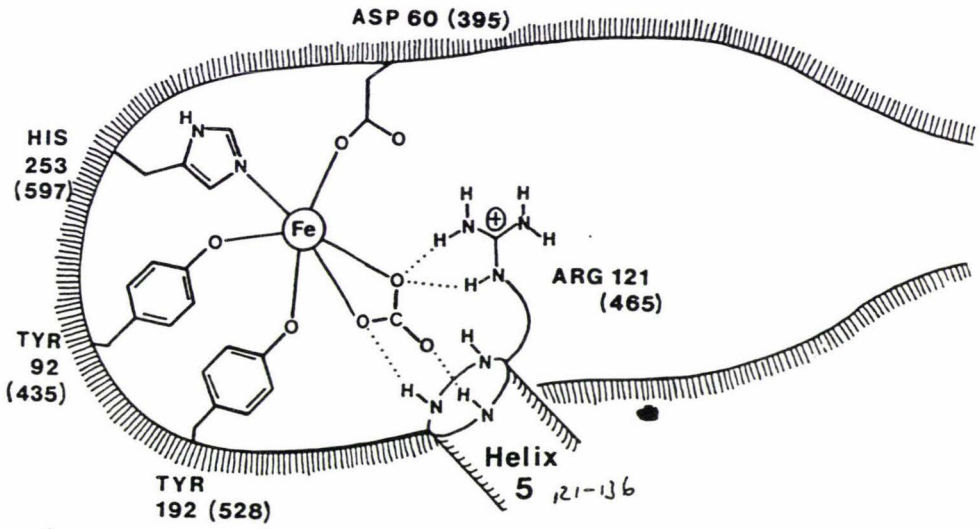


Fig A.3. Schematic diagram of the metal and anion binding site in lactoferrin. Residue numbers are for the N-lobe (with the C-lobe given in brackets). (From Baker, (1994) with permission).

With the knowledge of the lactoferrin structure, the structure of diferric rabbit serum transferrin has since been completed (Bailey *et al*, 1988). The overall folding pattern was similar to that of human lactoferrin, with similar domains, each based on a mixed beta sheet overlaid with alpha helices which pack against the two faces of the sheet. The main difference between the two structures was in the orientation of the two lobes. With the N-lobes superimposed, there is a 15 degree rotation of the C-lobe of transferrin compared to lactoferrin (Baker and Lindley ,1992).

Despite observed differences in function between rabbit serum transferrin and human lactoferrin, the iron-binding sites are similar. If the atoms of the iron-binding site are superimposed, the root mean square deviation between atomic positions for all the combinations of the N and C-lobes is less than 0.6 Å (Baker and Lindley, 1992). Therefore the functional differences between transferrin and lactoferrin are probably not concerned with the precise structure of the iron-binding sites. It has been suggested instead that the interactions between the two domains within a lobe may influence the functional differences between the transferrins and lactoferrins (Baker and Lindley, 1992).

The carbohydrate chains on transferrins are all N-linked through asparagine residues. Glycosylation sites vary in number from one in rabbit serum transferrin and chicken ovotransferrin, to four in bovine lactoferrin (human lactoferrin contains two sites). The sites are scattered over the surface of the molecule arguing against any direct functional role in iron-binding. The carbohydrate chains are heterogeneous with little defined structure observable by X-ray analysis. Except for certain species of fish (*Tinca tinca* and *Ctenopharybgodan idella*), all known vertebrate transferrins are glycoproteins (Stratil *et al,* 1983). The absence of carbohydrate chains in these species suggests that they are not important in the physiological functions of the proteins.

One intriguing question that has never been unequivocally answered is that of whether there is any biological advantage in having a bilobal molecule, with two iron sites. The bilobal structure has clearly arisen by gene duplication and fusion from a small ancestral protein of half the size (40kDa, a single lobe) with a single iron-binding site (Bowman *et al,* 1988). One possible explanation for the duplication has been advanced by Williams *et al* (1982) who found that isolated lobes of serum transferrin were rapidly lost from the bloodstream via the kidney. In this case only species that had evolved this bilobal form would have survived. Williams however proposes that a single lobed ancestor is unlikely, and instead, the ancestral protein may have been a bilobal form, or a membrane associated protein. An alternative possibility is that the two lobes have become differentiated in their properties (see later).

A.3.2.2. Apolactoferrin structure.

A proper understanding of iron binding and release requires that the nature of the associated conformational changes be defined. Although X-ray analysis gives only a static picture, comparisons of the structures of iron-loaded and iron-free strucutres are an essential element in reaching such an understanding.

The apolactoferrin structure was solved from protein in which the carbohydrate had been removed by enzymatic digestion using peptide N-glycosidase F (PGNaseF) and endoglycosidase F (Endo F), both isolated from *Flavobacterium meningosepticum* (Elder and Alexander, 1982). This deglycosylated protein had identical properties of iron-binding and release, and identical spectroscopic parameters to the native form. The crystals diffracted to 2.0 Å resolution (Norris *et al,* 1989). The structure was determined by molecular replacement using diffractometer data to 2.8 Å resolution, and refined to an R factor of 0.213 for data between 10 and 2.8 Å.

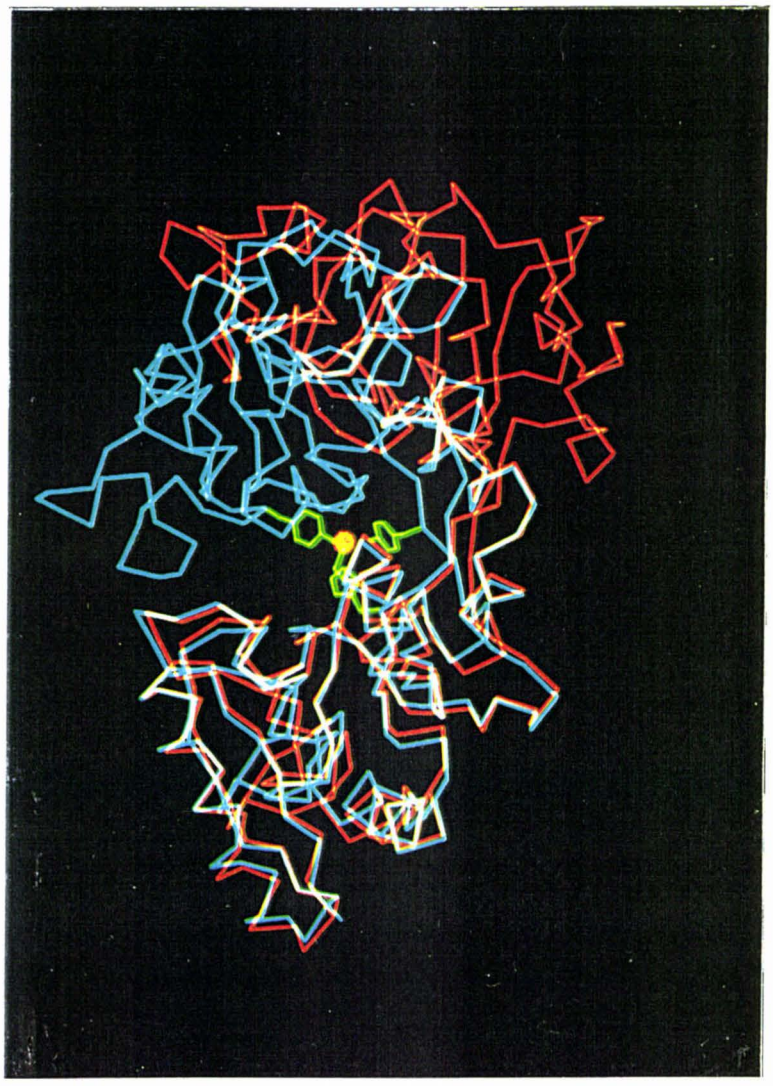


Fig. A.4. Superposition of the open and closed N-lobes in the full-length lactoferrin structure.

A striking feature of the X-ray analysis (Anderson *et al*, 1990), was the large conformational change seen for the N-lobe of the molecule (Fig A.4). One domain (N2) had rotated 54 degrees relative to the other domain (N1), resulting in a wide-open binding cleft. The equivalent movement was not seen in the C-lobe, however. Both sites (N and C-lobes) had lost their Fe^{3+} ion. This one-open, one-closed structure was unexpected and two theories to explain this were proposed, (i) that extra constraints in the C-lobe (specifically a disulphide bridge 483-677) inhibit opening of the C-lobe, and (ii) that an equilibrium exists between the open and closed forms in solution as seen for the periplasmic binding proteins (PBP's) (Oh *et al* 1993). Crystal packing could then have selected the closed form of the iron-free state. If this is so, then the energy difference between the open and closed forms of the iron free protein must be small. This second explanation was confirmed, firstly by solution X-ray scattering measurement (Grossmann *et al,* 1992), and secondly when Faber *et al* (1995) solved the

structure of a second crystal from the apoprotein in which both sites were open (Fig A.5). The C-lobe only opens by about 15 degrees, probably due to the extra constraints present in the C-lobe.

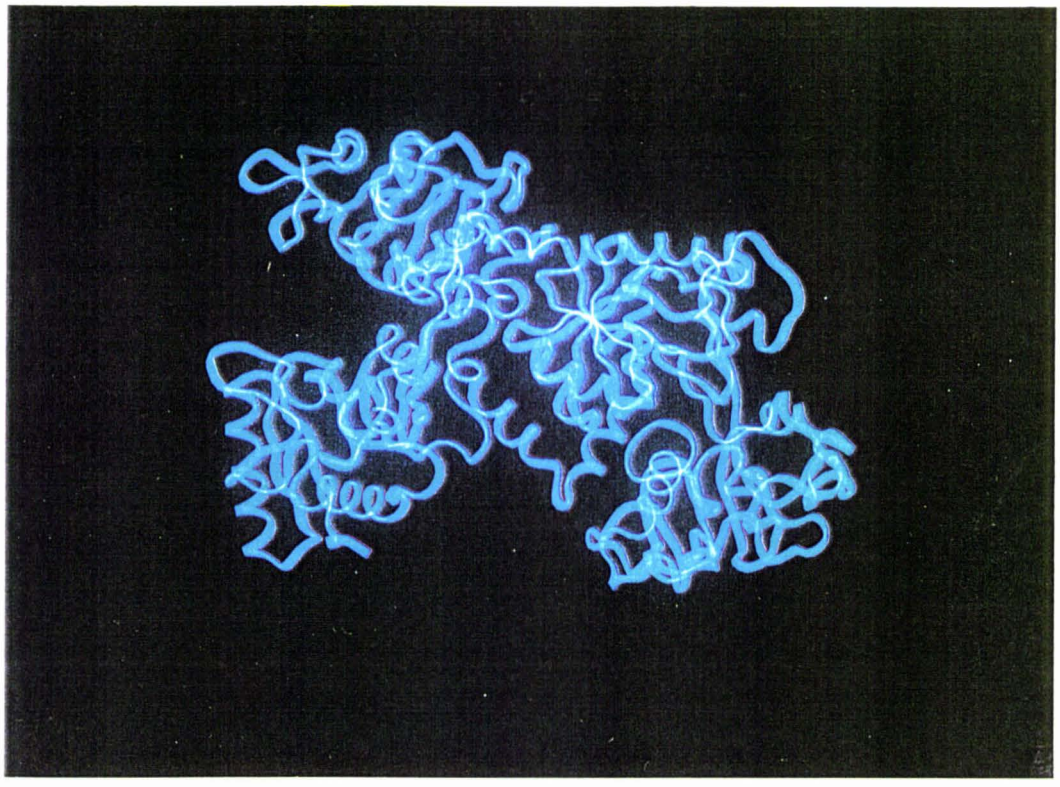


Fig. A.5. Structure of full-length lactoferrin with both the N- and C- lobes open.

In each lobe the conformational change that mediates the transition from closed to open form involves relative movement of the two domains. The axis of this movement is through a "hinge" in two backbone beta strands that connect the two domains and pass behind the iron-binding cleft (Fig A.6). The domains move as rigid bodies about this hinge, as is shown by superposition of the individual domains of Fe_2Lf on those of apoLf. When this is done, the rms deviation of the N2 domain (which rotates 54°) is only slightly greater that those for the other three domains (Anderson et al 1990); N1, 0.52 Å; N2, 0.65 Å; C1, 0.43Å; C2, 0.44Å. This means that the domain movement on iron release is a rigid body movement about a hinge. This rigid body movement for lactoferrin is the largest yet reported for any protein.

In serum transferrins, lactoferrins, and ovotransferrin, the C-lobe releases iron less readily, possibly due to its lesser flexibility, whereas in melanotransferrin and *M.sexta* transferrin, it does not bind iron at all. It may be that binding to the C-lobe has remained only in cases where a receptor mechanism exists to extract iron from the site and this idea again raises the question of the biological importance of the bilobal structure.

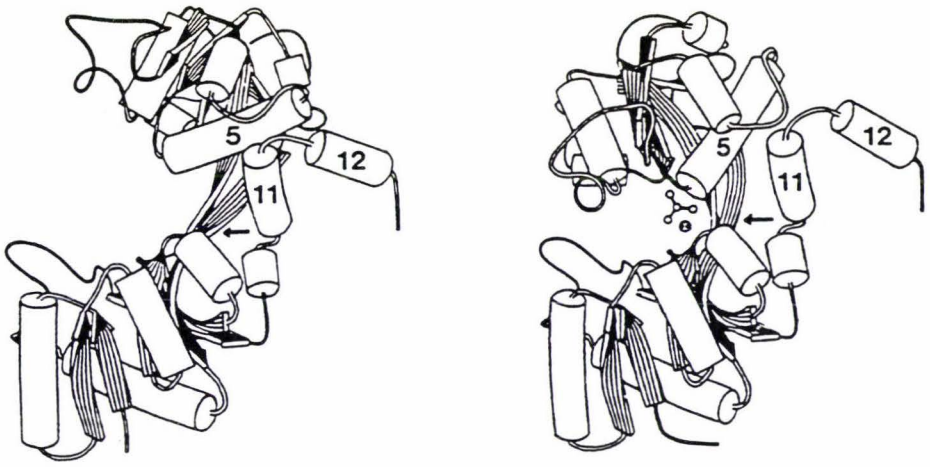


Fig A.6. Schematic diagram of the open (left) and closed (right) forms showing the action of the hinge in the N-lobe of lactoferrin. An arrow positions the part of the beta sheets involved in the "bending" of this region.

A.4. Iron-binding properties.

A.4.1. Introduction.

The orange-red color that develops when ferric iron is added to the transferrins is the main indicator of iron binding. The colour is due to an absorption band around 460nm, although the precise λmax varies among the transferrins. In addition to this band, there is another specific absorption near 295nm (Aisen *et al*, 1969).

One of the first studies to describe the iron-binding properties of lactoferrin was that of Aisen and Leibman (1972). These authors showed that the properties of transferrin and lactoferrin were very similar and suggested that the iron-binding ligands included 2 tyrosines and 1 nitrogen-containing amino acid, possibly histidine. These residues, and an additional aspartate residue, were confirmed when firstly the iron-loaded human lactoferrin structure (Anderson *et al*, 1987), and secondly the iron-loaded rabbit transferrin structure (Bailey *et al*, 1988) were determined.

A.4.2. Binding constant

Although the iron-binding ligands of the two proteins (transferrin and lactoferrin) are the same, there are differences of detail in their iron-binding properties. At pH 6.4, human lactoferrin (hLf) binds iron 300 times more strongly than transferrin (Tf) at pH 6.7 (Aisen and Leibman, 1972), and while hLf releases Fe^{3+} over pH range 4.0 to 2.5, transferrin releases it over the pH range 6.0 to 4.5 (Bullen *et al*, 1978).

The binding for both these proteins is represented by the equation

$$H_6\text{-Protein} + 2Fe^{3+} + 2HCO_3^- \text{ <---> } Fe_2\text{-Protein}(HCO_3^-)_2 + 6H^+$$

Because the binding of iron to any of the transferrins is fully reversible, it should be possible to determine an equilibrium binding constant. This is complicated, however, by the low solubility of Fe^{3+} in near-neutral solution (10^{-17}M at pH 7.0).

A method of equilibrium dialysis, using citrate as a competing iron complexing agent (Aasa *et al*, 1963), was used to overcome this low solubility problem, and therefore measure the equilibrium constant of the iron transferrin complex. This method analyses the amount of iron partitioned between two sides of a membrane with specific-sized pores. On one side only of the membrane is transferrin. Knowing the total iron present, and the amount of free iron (complexed to citrate) on each side, then the amount of iron bound to transferrin and hence the equilibrium constant can be calculated. Using this method, an equilibrium binding constant for lactoferrin at pH of 7.4, in air, was calculated to be 4.7×10^{22} M^{-1} (Harris and Aisen, 1989).

A.4.3. Anion binding.

A distinguishing feature of the transferrins is that an anion is bound with each metal ion. The anion found in vivo is CO_3^{2-} (or possibly HCO_3^-), although other anions than CO_3^{2-} are able to fulfil this role, a common feature of these synergistic anions being that they each contain a Lewis base close to a carboxylate group (Schlabach and Bates, 1975). Metal ion indicators, EPR experiments, NMR spectroscopy, and X-ray crystallography all indicate that the anion is directly attached to the metal ion in the transferrins (Harris and Aisen, 1989).

The ability of transferrin to bind two iron atoms per molecule, and the requirement for bicarbonate were first demonstrated by Schade *et al* (1949). The affinity for Fe^{3+} is weak in the absence of an anion (Bates and Schlabach, 1975), and that for the anion is weak in the absence of Fe^{3+}. Because of this feature, the relationship between the iron and anion is said to be synergistic.

A.4.4. Order of binding.

The next step in understanding the mechanism of iron-binding, was to determine the order of events when a transferrin binds iron. Whether metal or anion binding to transferrin occurs first is not completely clear, but using ^{13}C-NMR, Zweier et al (1981) observed that anions (oxalate and bicarbonate) bind weakly in the absence of metal ion, indicating that anion binding may precede metal binding.

Kinetic data also suggests that the anion binds first (Kojima and Bates, 1981). Using ferric chelate complexes, it has been proposed that there are five steps in the formation of the metal-anion-protein complex. These are (i) binding of the anion to the apoprotein, (ii) detachment of one or more ligands from the added ferric chelate, (iii) formation of a protein-anion-ferric-chelate complex, (iv) loss of the chelate ligand, and (v) conformational change involving the closing of the two domains over the ions. The closed conformation is locked together by the aspartate ligand, which plays a critical role in the metal-bound structure (Anderson *et al*, 1989). This sequence of events, which implies that the initial step involves binding of the anion to its site on the N2 (or C2) domain is shown schematically below (Fig A.7).

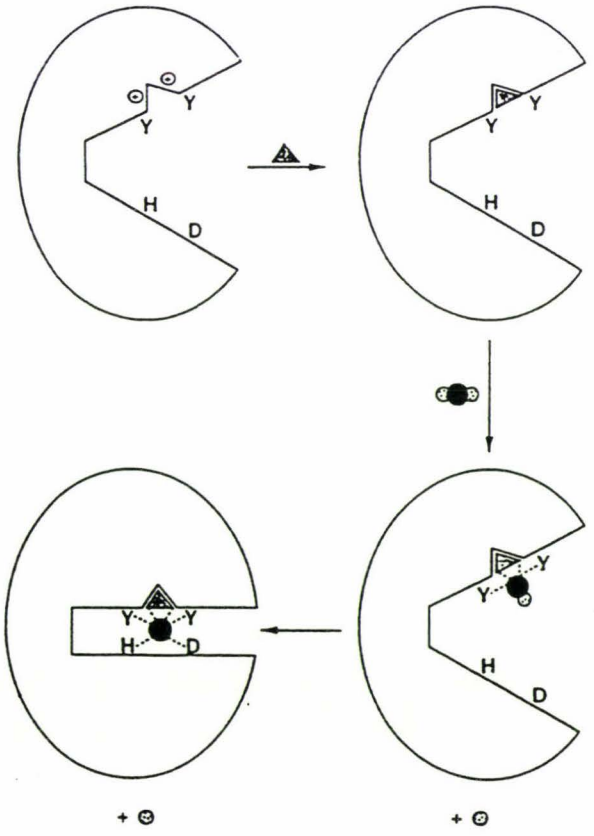


Fig. A.7. A schematic model of the steps involved in the uptake of iron by lactoferrin. (•) Iron; (▲) carbonate; Y, Tyr ligands; H, His ligand; D, Asp ligand; (O) chelate ligands. (From Baker (1994), with permission).

The role of the anion in binding is thought to be twofold. Firstly, it neutralises positive charges on the protein in the binding site which may otherwise repel the cation (Fe^{3+}), and secondly, by adding two more iron ligands (2 carbonate oxygens), it helps form the metal binding site.

A.4.5. Other metals.

While iron (as Fe^{3+}) is bound most strongly to the transferrins, other metal ions can also bind to lactoferrin (Ainscough *et al*, 1980). For one of these, Cu^{2+}, an X-ray crystallographic model of the complex with human lactoferrin has been determined (Smith *et al*, 1992). Although this shows a change in metal coordination from 6-coordinate to 5-coordinate in the N-lobe, the conformation of lactoferrin is still the same, suggesting that lactoferrin with metals other than Fe^{3+} bound, may still bind to receptors just as with Fe^{3+}.

A.4.6. Iron release.

In the mid 1970's, it was found that the pH dependence of iron dissociation from serum transferrin was biphasic, indicating that there was a difference between the two sites. $^{59}Fe^{3+}$ was found to be released completely from one site before the other (Princiotto and Zapolski, 1975), (Cannon and Chasteen, 1975). The N terminal site is acid labile, and the C terminal site is acid stable (Lestas ,1976), so that as the pH of human transferrin is lowered, iron is released from the N-lobe (pH 7 to 5.5) before it is released from the C-lobe (pH 5.5 to 4). The biphasic acid-induced iron release seen in serum transferrin is not seen in lactoferrin, however, in which both sites release iron together over the pH range 4.0-2.5.

The mechanism of iron release for the transferrins is not known, although it is thought to involve one or both of the two following proposals. In the first proposal, Kretchmar and Raymond (1986) showed that for serum transferrin iron was lost more readily from the N-terminal site and proposed that this may be due to the greater flexibility of this lobe. They suggest that conformational changes and electrostatic interactions may play a role in iron release. The interdomain interactions at the back of the binding pocket are known to differ between transferrin and lactoferrin (Baker and Lindley, 1992), and this may account for the difference in the pH at which dissociation occurs.

The second proposal for iron release in the transferrins suggests that a lowering of the pH would cause a protonation of the anion which would either disrupt the interaction between the anion and the anion-binding arginine residue (121 in the N-lobe of hLf, 465 in the C-lobe), or cause a change in anion coordination to the metal, from bidentate to the monodentate, as seen when copper binds to lactoferrin (Smith *et al*, 1992). Either mechanism could result in iron release. This second proposal involving the protonation of the anion, does not, however, explain why human serum transferrin and human lactoferrin release iron at different pH when they both contain the same ligands, Fe^{3+} and anion. Perhaps the environment around the binding site also plays an important role in this second scenario by altering the conditions under which protonation occurs.

Studies of half molecule fragments (one lobe only) suggest that the region at the back of the iron-binding site is important in iron release. The recombinant N-lobe of human lactoferrin differs in this region from the intact molecule, and it releases iron over the pH range 6 to 4.0 (Day *et al*, 1992). This is higher than that of the complete molecule which releases iron between 4.0 and 2.5 (Mazurier and Spik, 1980).

A.4.7. Differences between the lobes.

Given the existence of two similar binding sites on transferrins, four forms should exist, ie the apo form, two monoferric forms, and the diferric form. The existence of monoferric transferrin was first shown by electrophoresis (Aisen *et al*, 1966, Wenn and Williams, 1968) and routine separation is now made possible by electrophoresis of transferrin partially denatured by 6 M urea at pH 8.4. This method, due to Makey and Seal (1976), allows assesment of how much iron is bound to each lobe of the molecule under different conditions.

Using this same method, it was found that in fresh serum, the two sites of transferrin are not equally loaded with iron. Initially the N-lobe is preferentially loaded with iron, although continued incubation at 37°C distributes the iron evenly between the lobes. If the serum is stored at -15°C, then iron distribution is directed towards the C-lobe (Williams and Moreton, 1980).

In 1968, Fletcher and Huehns (1968) suggested that the two sites of iron-binding in transferrin may have different roles. They proposed that one site may be involved in the iron transport function often assigned to transferrin, while the other site may be involved in iron sequestering (antibacterial as for lactoferrin). This was followed by kinetic and thermodynamic work by Aisen *et al* (1978), in which they showed that the C terminal site of transferrin binds iron more strongly than the N terminal site. The

binding constant was 20 times greater. The release of iron also differed, being faster for the N terminal site (which is also more acid-labile, as noted earlier).

Studies of transferrin-receptor interactions (Bali and Aisen, 1991), show that the receptor specifically interacts with the C-lobe of transferrin to release the iron, whereas the N-lobe loses iron by a reduction in pH.

These differences in the two lobes of both human lactoferrin and human serum transferrin may be explained by the structure of lactoferrin (Anderson *et al,* 1990), where comparison of apolactoferrin with the iron-bound form shows that there is a difference in flexibility between the two lobes. It is thought that the greater flexibility of the N-lobe of lactoferrin may help in the binding and release of iron, and this would explain the greater thermodynamic stability and slower release of iron from the C-lobe compared with the N-lobe (Kretchmar and Raymond, 1986).

A.5. Conformational changes in lactoferrin.

A.5.1. Domain movements.

Lactoferrin shares a similar structure, topology, and binding site construction with the group II periplasmic binding proteins (Baker *et al,* 1987) (see Fig A.8). These proteins bind and transport small molecules (sugars, anions, amino acids) through the periplasmic space of gram negative bacteria, and interact with receptors in the bacterial cell wall. The structure of a number have been determined (Quiocho, 1990), and these studies show that they have about 300 residues in a two domain structure similar to a single lobe of the transferrins. In each domain, helices are packed on either face of a central sheet, and two similar extended polypeptides link the domains. Both domains provide ligands for the binding site, with one domain providing most of the groups thereby serving as the initial site of attachment (as seen in the transferrins). For the maltose binding protein (MBP), Spurlino *et al* (1991) have shown that maltose binds first to one domain in the open form.

The feature of lactoferrin structure and function that is addressed in this thesis involves the conformational change seen in the N-lobe of lactoferrin accompanying iron-binding and release. This movement is the largest yet seen in any such protein and involves a rigid body rotation of one domain relative to another, through an angle of 54° (Baker *et al,* 1991). It enables the protein to move between an open form (in which the binding cleft is wide open) and a closed form (in which the domains have closed over the bound metal ion).

Table A.2. lists all the proteins with domain motion for which both open and closed forms have been resolved by X-ray crystallography. Table A.3. lists proteins for which only one conformation is known, but in which a domain closure mechanism is thought to occur.

Table. A.2. Proteins with a hinge-type mechanism for domain closure for which open and closed structures have been determined.

Protein	Reference
Lactoferrin	Anderson *et al*, 1990
Tomato bushy stunt virus coat protein	Olson *et al*, 1983
Maltodextrin binding protein	Sharff *et al*, 1992
Lysine-arginine-ornithine binding protein	Oh *et al*, 1993
T4 lysozyme mutants	Faber and Mathews, 1991
Catabolite gene activator protein	Weber and Steitz, 1987
cAMP dependent protein kinase (catalytic domain)	Karlsson *et al*, 1993
Adenylate kinase	Schulz *et al*, 1990
Glutamate dehydrogenase	Stillman *et al*, 1993
Calmodulin	Meador *et al*, 1992, 1993

Table. A.3. Proteins with a hinge-type mechanism for which only one conformation has been determined.

Protein	Reference
Sulfate binding protein	Luecke and Quiocho, 1990
Phosphate binding protein	Pflugrath and Quiocho, 1988
Leucine binding protein	Gilliland and Quiocho, 1981
Galactose binding protein	Vyas *et al*, 1988, 1991
Arabinose binding protein	Sack *et al*, 1989
Transferrin	Sarra *et al*, 1990
Guanylate kinase	Stehle and Schulz, 1990
Porphobilinogen deaminase	Louie *et al*, 1992

An obvious question concerns why there is such a large domain movement, if the function of lactoferrin is only to sequester an atom of iron for bacteriostatic purposes. A possible answer for this could be the proposed receptor binding function of lactoferrin. Receptor binding is an important feature of the group II periplasmic binding proteins (Mowbray, 1992). For these proteins, the large conformational change is crucial for signal transduction in active transport (Jacobson *et al*, 1992), since the membrane bound receptors preferentially bind to the liganded, closed form of the maltodextrin binding protein (MBP). Furthermore, there are a number of mutations in the N and C domains of MBP (corresponding to the N1 and N2 domains of lactoferrin), at the opening to the cleft that affect the function of this protein.

Two of the periplasmic binding proteins have been analysed crystallographically in both the open and closed forms, ie the maltodextrin binding protein (Spurlino *et al*, 1991, Sharff *et al*, 1992), and the lysine-arginine-ornithine (LAO) binding protein (Oh *et al*, 1993). The nature of the domain movement is similar to that of lactoferrin. MBP has a 35° rotation of one domain relative to the other about an axis through the hinge region, while for LAOBP a movement of 52°, similar to that in the N-lobe of lactoferrin, is seen. This motion, as for lactoferrin (see below), involves only a few large torsion angle changes. In addition, two of the periplasmic binding proteins bind anions in a similar way to lactoferrin, ie the sulphate binding protein (Pflugrath and Quiocho, 1988), and the phosphate binding protein (Luecke and Quiocho, 1990). These two proteins have a similar fold to a single lobe of lactoferrin, and their anion binding sites coincide with the carbonate site in the transferrins. What is interesting is that there is a greater sequence similarity between the sulphate binding protein and lactoferrin

(15%), than between the sulphate and phosphate binding proteins (<10%) (Baker, 1994).

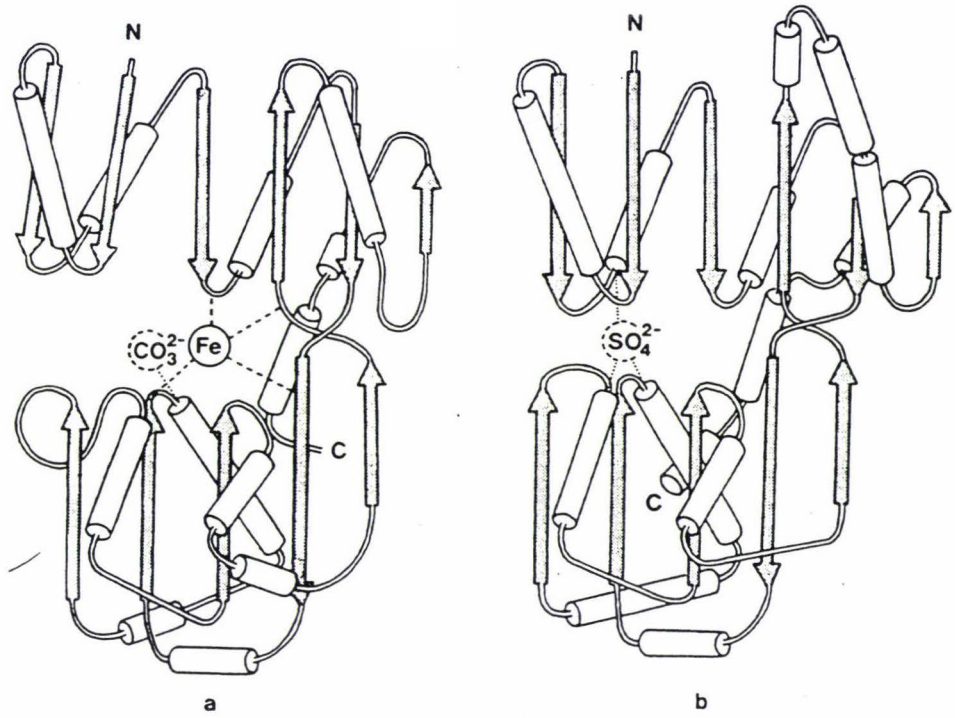


Fig. A.8. Comparison of the folding pattern for the N-lobe of lactoferrin (a) and for the sulphate binding protein (b). (From Baker (1994), with permission).

A.5.2. The hinge of lactoferrin.

The large rigid body motion in lactoferrin is made possible by two extended polypeptide strands that run behind the binding cleft connecting the two domains of each lobe.

In Fig A.9, the hinges in the N-lobe of human lactoferrin are indicated. Gerstein *et al* (1993) described the motion as a screw motion by fixing the origin so that translation was minimal upon domain movement. In this description, the N2 domain translates only 1.0Å with the rotation axis passing very close to Thr 90 and Pro 251, the two residues Anderson *et al* (1990) identified as the centre of the hinges; the axis is within 1.4Å of the Cα of the Thr 90, and 2.4Å of the Cα of Pro 251. Therefore the motion of N2 relative to N1 involves almost no translation, and is instead a pure rotation about the central residue in each hinge (Thr 90 and Pro 251). This is illustrated in Fig. A.10.

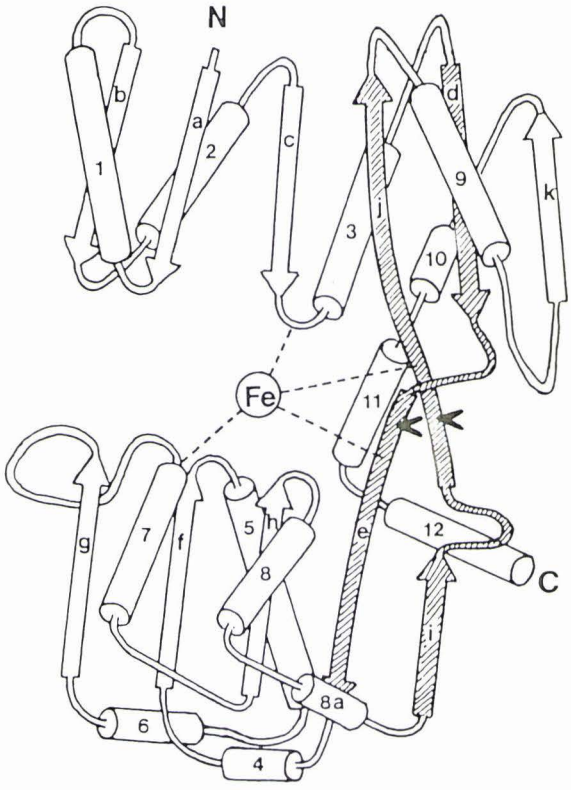


Fig. A.9. Polypeptide folding pattern for the N-lobe of lactoferrin indicating the two beta strands responsible for the position of the hinge. (From Baker (1994), with permission).

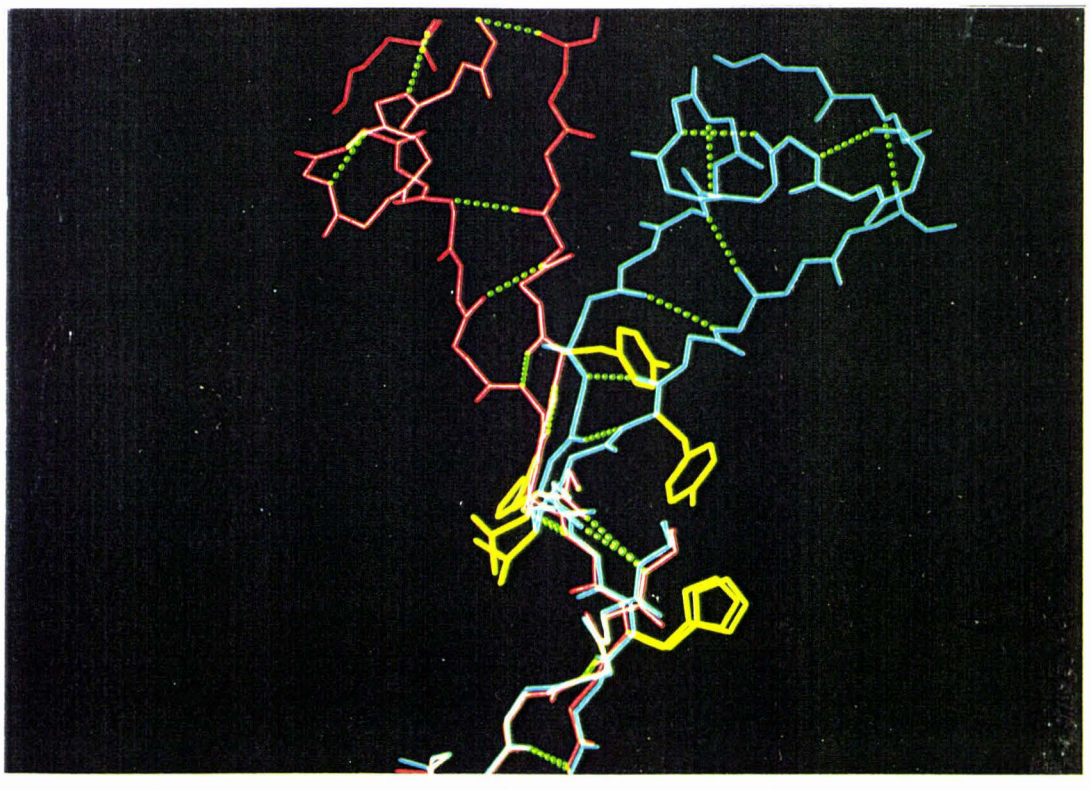


Fig. A.10. Main chain atoms only of the open (red) and closed (blue) structures of the two beta strands involved in the hinge of lactoferrin.

In the first hinge, the domain movement is associated with torsion angle changes in ψ (90) and ϕ (91), these changes being coupled across a peptide bond (Fig A.11). In total these produce a rotation of 81° , which, although greater than the overall 54° rotation seen in the N-lobe, is compensated by small opposing changes in the torsions of neighbouring residues (Gerstein *et al*, 1993).

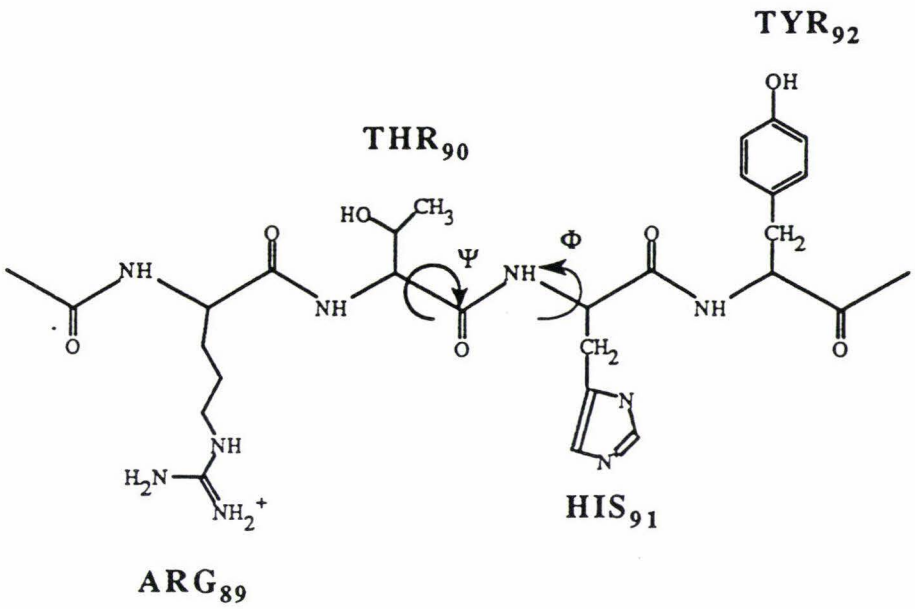


Fig. A.11. Residues 89 - 92 of the first beta strand involved in the hinge motion of the N-lobe of lactoferrin. The phi and psi angles responsible for the motion are indicated.

The second hinge is not so simple. Three torsion angles have changes greater that 20°; ie ϕ (250), ψ (250), and ϕ (251), with the central angle (ψ (250)) changing by 33 degrees. These angles are also coupled across a peptide bond to produce a rotation of 49° (Fig A.12).

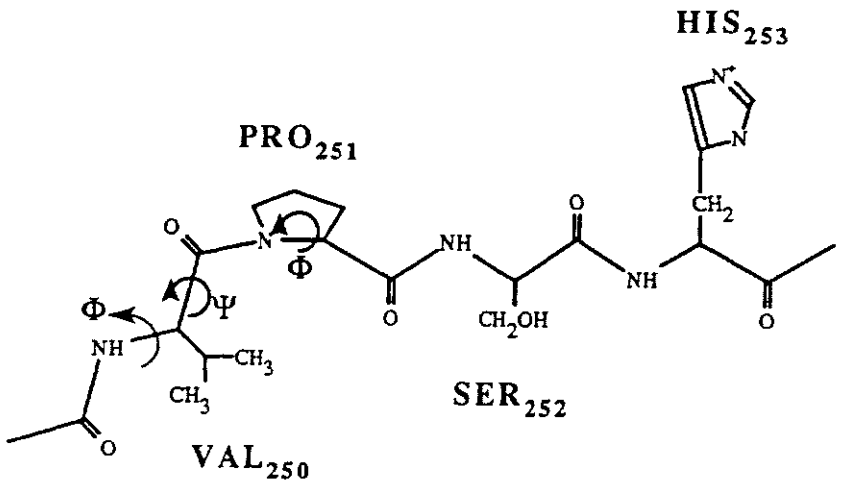


Fig. A.12. Residues 250 - 253 of the second beta strand involved in the hinge motion of the N-lobe of lactoferrin. The phi and psi angles involved in the motion are indicated.

The axis of the overall 54° rotation is closely aligned with the rotational axis of these principal torsion angle changes. This means that the local motion of the hinge reflects the overall domain movement of the molecule.

The principal torsion angle changes in the lactoferrin hinge occur in normally allowed regions of the Ramachandran plot (Fig A.13) and therefore involve only low energy transitions, indicating that the open and closed states can occur in dynamic equilibrium as proposed for the periplasmic binding proteins (Oh *et al*, 1993).

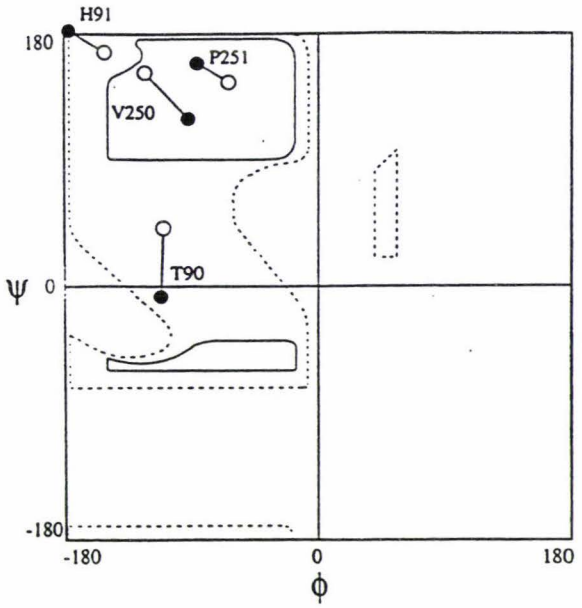


Fig. A.13. Ramachandran plot of the main chain torsion angle changes seen in the N-lobe of lactoferrin. Phi and psi angles are shown with open circles for the open conformation, and with filled circles for the closed conformation.

In the ribose and galactose binding proteins however, there is a residue (215 in RBP, 236 in GBP) that has an unfavourable conformation. This is one of only two Ramachandran violations in these proteins, and the conformation is necessary for the correct action of the hinge in these proteins (Mowbray, 1992). A buried aspartate helps maintain this conformation and therefore the correct action of the hinge. In human lactoferrin, the one Ramachandran violation is not associated with the hinges.

The location of the hinges with respect to the ligands seems likely to be important in lactoferrin. Tyr 92 (a ligand) is 2 residues away from Thr 90 (human lactoferrin numbering), and His 253 is two residues away from Pro 251. This allows the hinges to split these two ligands, with Tyr 92 moving with the N2 domain, and His 253 with the N1 domain. This provides the N2 domain with 4 of the 6 iron-binding groups (2 from the carbonate anion together with Tyr 92 and Tyr 192). The N2 domain is thought to serve as the initial site of attachment. This is supported by recent work by Lindley et al (1993) on the analysis of a proteolytic fragment of duck ovotransferrin corresponding to its N2 domain. In this one-domain molecule, the iron is bound to the carbonate ion and to two tyrosine residues just as in the intact molecule. Presumably

once the iron is bound to the N2 domain, domain closure occurs to allow the iron to complete its coordination to His 253 and Asp 60, and the aspartate helps lock the two domains together with an interdomain hydrogen bond (Anderson *et al*, 1989).

It has been proposed that the apo-protein is in a dynamic equilibrium, with the N1 and N2 domains "flexing" about the hinge, and that the iron then stabilises the closed form as the bound ligand does in the periplasmic binding proteins. The importance of the hinge in this mechanism would be in directing the two domains together into the correct position by restraining the number of conformations in the open state.

In most proteins, the main chain is buried beneath layers of side chains leaving little freedom for large torsion angle changes. In the hinge region of lactoferrin,however, the main chain is free from steric constraint, with the mainchain atoms of Thr 90, His 91, and Pro 251 making no contacts with the rest of the protein. The absence of main chain packing constraints with the rest of the protein, and the absence of internal hydrogen bonds is also found in hinged mechanisms in lactate dehydrogenase (Gerstein and Chothia, 1991) and in adenylate kinase (Gerstein *et al*, 1993) and appears to be a crucial structural requirement for hinged motion.

A.5.3. Amino acid sequences in the hinges.

The sequences of the hinge regions in transferrins are listed in Table A.4. and Table A.5. and display a remarkable degree of sequence conservation.

In the N-lobe, the hinges are located at Thr 90 in one "backbone" strand and Pro 251 in the other. Elsewhere in the two strands, a number of residues are conserved either because they are ligands (Tyr 92 and His 253) or because they contribute to the hydrophobic core of one domain or the other (the sequences AVAVV in the first strand, and AVV in the second strand). The most intriguing conservation, however, is that of Thr 90 and Pro 251.

Table A.4. Amino acid sequence in the two hinge strands of the N-terminal lobe, for all known transferrin sequence.

Protein	N-lobe	
	1st Hinge Strand	2nd Hinge Strand
	90	251
Human Serotransferrin	EDPQTFYYAVAVVKK	HLAQVPSHTVVAR
Horse Serotransferrin	TEPQTHYYAVAVVKK	YLASIPSHAVVAR
Pig Serotransferrin	DNPGTHYYAVAVVKK	YLAQVPSHAVVAR
Rabbit Serotransferrin	ENPKTFYYAVALVKK	HLARVPSHAVVAR
Rat Serotransferrin	EHRQTHYLAVAVVKK	YLARIPSHAVVAR
Xenopus Serotransferrin	TETDTCYYAVAVVKK	NLAKVPAHAVLTR
Human Lactoferrin	RQPRTHYYAVAVVKK	HLARVPSHAVVAR
Bovine Lactoferrin	ESPQTHYYAVAVVKK	HLAQVPSHAVVAR
Goat Lactoferrin	KSPQTHYYAVAVVKK	HLAQVPSHAVVAR
Mouse Lactoferrin	EQPRTHYYAVAVVKN	HLAQVPSHAVVSR
Pig Lactoferrin	ENPGTYYYAVAVVKK	HLARVPSHAVVAR
Human Melanotransferrin	QEVGTSYYAVAVVRR	HLARVPAHAVVVR
Chicken Ovotransferrin	EGSTTSYYAVAVVKK	NWARVAAHAVVAR
Atlantic Salmon Transferrin	EDSDTCYYAVAVAKK	HLARVPAHAVVSR
Hornworm Transferrin	PDAPFRYEAVIVVHK	SWAARPWQGLIGH
Cockroach Transferrin	PDEEFRYEAVCVIHK	IWAARPWQGYMAN

Thr 90 is totally conserved, except in the two insect transferrins. It is difficult to see why this should be so unless this residue is important for the function of transferrins in general. The obvious conclusion is that Thr 90 has some as yet unsuspected role in helping to define the conformational change, since it is at the exact location of one of the hinges. The only apparent interaction made by the sidechain of Thr 90 is a hydrogen bond with the C-terminal helix of the full-length molecule (90 OG1 - O=C 691), this being present in apolactoferrin but not in Fe_2Lf.

Pro 251 is also totally conserved, except in chicken ovotransferrin. Proline residues are known for their restraining role in the polypeptide conformation, and are present in the hinge regions of several bacterial binding proteins where their presence may modulate the motion of the hinge (Mowbray ,1992).

Table A.5. Amino acid sequence of the two hinge strands of the C-terminal lobe, for all known transferrin sequences.

Protein	C-lobe 1st Hinge Strand	2nd Hinge Strand 595
Human Serotransferrin	DTPEAGYFAVAVVKK	HLAMAPNHAVVSR
Horse Serotransferrin	DTPEEGYHAVAVVKS	YLARAPNHAVVSR
Pig Serotransferrin	NTPEKGYLAVAVVKK	HLARAPNHAVVAR
Rabbit Serotransferrin	KAPEEGYLSVAVVKK	HLAKAPNHAVVSR
Rat Serotransferrin	QSDVFPKGYYAVAVV	HLAQAPNHVVVSR
Xenopus Serotransferrin	SQAKGNYYAVAIVKK	NLAEVPAHAVVTL
Human Lactoferrin	DRPVEGYLAVAVVRR	HLAMAPNHAVVSR
Bovine Lactoferrin	LRPTEGYLAVAVVKK	HLAVAPNHAVVSR
Goat Lactoferrin	LRPTEGYLAVAVVKK	YLAVAPNHAVVSR
Mouse Lactoferrin	NRPVEGYLAVAAVRR	HLAIAPNHAVVSR
Pig Lactoferrin	HRPTGGYFAVAVVRK	HLAVAPSHAVVSR
Human Melanotransferrin	EDSSNSYYVVAVVRR	NLAQIPPHAVMVR
Chicken Ovotransferrin	ERPA SYFAVAVARK	NLAEVPTHAVVVR
Atlantic Salmon Transferrin	PGEASSYYAVAVAKK	HLAKVPAHAVITR
Hornworm Transferrin	GELKTPNYAVAVVKK	
Cockroach Transferrin	GEHGSLYYAVAVVRK	

In the C-lobe there is no conserved residue equivalent to Thr 90 in the first strand, although there is a totally conserved proline equivalent to Pro 251 in the second strand. In the C-lobe, however, the exact location of the hinge is not yet known.

Oh et al (1993) have suggested that the residues in the hinge do not play any part in the domain movement. The conservation of hinge residues seen in lactoferrin sequences argues against this view, however. In order to test this apparent contradiction, an obvious approach is to mutate the two conserved residues in the hinge (Thr 90 and Pro 251) and to characterise the resulting proteins by iron-binding studies and x-ray crystallography.

A.6. Aims of this project.

The aims of this project were to address the roles of the two "hinge" residues, Thr 90 and Pro 251, by site-directed mutagenesis and characterisation of the resultant mutant proteins.

The cDNA from human lactoferrin had previously been cloned into a suitable expression vector and expression had been shown to occur (Day, 1993). The complete structure for the half-length molecule had been solved by molecular replacement by Day (1993) using the Fe_2Lf structure as a starting model (Anderson *et al*, 1989). Bates (1994) had also cloned a segment of the human lactoferrin cDNA encoding amino acids 250 to 688 into the M13 bacteriophage.

In order to introduce amino acid changes into the cDNA for human lactoferrin, the method of Kunkel (1989) was chosen. Changes made in the M13 bacteriophage containing this segment were to be cloned into the pNUT vector containing the half-length lactoferrin cDNA (pNUT:Lf_N) constructed by Day (1993). This would then be used in tissue culture to produce the recombinant proteins. After purification of the recombinant protein, their iron-release properties would be examined, and the structural effects of mutation determined by x-ray crystallography.

The changes at Pro 251 were designed to allow varying degrees of flexibility into this region. For this reason, Gly, Ala and Val were chosen. Asp was also chosen as this is convenient using a degenerate oligonucleotide, and this change also introduces a negative charge into this region.

The changes at Thr 90 involved a change to alanine in order to determine the importance of the proposed hydrogen bond between the side chain hydroxyl group of Thr 90 and C-terminal end of the polypeptide chain.

B. MATERIALS AND METHODS.

B.1. Materials and reagents.

$MgCl_2$, Tris/HCl, Tetrachloroethanol, and various salts were supplied from Riedel de Haën.

All restriction endonucleases and enzyme reaction buffers, T4 polynucleotide kinase, T4 Ligase, T4 DNA Polymerase, Proteinase K, 1kb DNA ladder, DNA quantification standards, and ammonium persulphate came from Bethesda Research Laboratories (Maryland, USA).

All radiolabelled nucleotides were supplied from New England Research Products, (Boston, Massachusetts, USA).

E.coli XL-1 cells were obtained from Stratagene, (La Jolla, California, USA). The nutrients and agarose for plating were supplied from Gibro BRL.

All mutagenic oligonucleotides were obtained from Oligos Etc.Inc. (Connecticut, USA).

SDS (sodium dodecyl sulfate), phenol, chloroform, isoamylalcohol, methanol, ethanol, and HPLC grade acetonitrile were obtained from BDH Ltd (Poole, England). Acetic acid was obtained from Rhône Poulenc Chemicals, (Paris, France).

The pTZ:18U vector and M13KO7 helper phage, acrylamide, bis-acrylamide, and TEMED were from Biorad (California, USA).

The maxiprep and miniprep resin and plasticware came from Promega (Madison, Wisconsin, USA).

Glass wool, sodium chloride, and sodium acetate were obtained from AJAX chemicals. The dichloromethyl silan was from Merck (Germany).

Sequenase version 2.0 DNA sequencing kit came from United States Biochemical Corporation (Cleveland, Ohio, USA).

X-ray film for sequencing gels was from Fuji Photo Film Company Ltd (Japan). Polaroid film for DNA gels was from Polaroid Corporation (Cambridge, Massachusetts, USA).

Dimethyl sulphoxide (DMSO), Fetal Calf Serum, Dulbecco's modified Eagle's medium (DMEM), and Hams-F12 (F12) used for tissue culture, came from Gibco (Grand Island, New York, USA). All the tissue culture plasticware (50 ml and 300 ml bottles, and roller bottles), were supplied by Nunc (Denmark). Methotrexate was obtained from the Palmerston North Hospital. The antibiotic Baytril came from Bayer chemicals.

Sigma Chemical Company (Missouri, USA) provided the streptomycin and penicillin antibiotics for cell culture, all the dNTP's, ethidium bromide, agarose for DNA gels, glycerol, RNAse, X-gal, IPTG, DTT, PEG 8000, Coomassie brilliant blue R-250, protein molecular weight markers, CM-sephadex, and SP-sephadex resins.

The eukaryotic expression vector, pNUT (Palmiter et al, 1987) and the mammalian baby hamster kidney cells (BHK) were generously provided by Dr R.T.A. MacGillivray (Dept of Biochemistry, University of British Columbia, Vancouver, Canada).

The Cell-Phect kit for the transfection of the BHK cells came from Pharmacia LKB Biotechnology (Uppsala, Sweden). Cellulose acetate membrane filters (0.2 μm) came from Micro Filtration Systems (California, USA). All ultrafiltration membranes, Centricon and microcon devices were supplied by Amicon (Massachusetts, USA).

B.2. Methods used in handling DNA.

B.2.1. Maintenance and storage of bacterial strains.

E.coli strains JM101, CJ236, XL-1, and DH-1 were used. Stock cultures were maintained on Luria broth (LB) plates (1% tryptone, 0.5% yeast extract, 0.5% NaCl containing 1.5% agar) for up to 4 weeks before restreaking from a -70°C stock. The plates were supplemented with antibiotics according to the strain used. (DH-1, ampicillin); (XL-1, tetracycline); (CJ236, chloramphenicol). Clones of interest were stored frozen at -70°C in 20% glycerol.

B.2.2. Transformation of DNA.

DNA was transformed by the method of electroporation or the $CaCl_2$ method (Sambrook et al, 1989). Competent cells for electroporation were prepared by growing the cells to an A_{600} of 0.4-0.5 and spinning at 5000 rpm in an SS34 rotor, then rinsing twice with ice cold deionised water. An equal volume (compared with cell volume) of 10% glycerol was added to the cells and they were stored at -70°C. The DNA was electroporated at 200 ohms, 2.5 volts, and 25 microfarrads capacitance, using a Biorad gene pulser™ and 800 µl of SOC medium was added to the cells, which were then incubated at 37°C for 1 hr before plating on LB plates containing an appropriate antibiotic.

In the $CaCl_2$ method, the $CaCl_2$-competent cells were grown and collected in the same way as for electroporation except that instead of deionised water, 0.1M ice cold $CaCl_2$ was used. Transformation involved adding no more than 50 ng of DNA to 200 µl of $CaCl_2$ competent cells and storing on ice for 30 minutes. The cells were then transferred to a 40°C circulating water bath for 90 seconds, and rapidly transferred to an ice bath and chilled for a further 1-2 minutes. SOC medium was added, as for electroporation, and the reaction was incubated at 37°C for 1 hour prior to plating.

B.2.3. Preparation of DNA.

For rapid purification of small amounts of plasmid or double-stranded phage DNA to be used for restriction enzyme analysis, the rapid boil method of Sambrook et al (1989) was used. For a cleaner purification of small amounts of DNA to be used for sequencing, the magic miniprep method marketed by Promega was used. In this case up to 10 µg is obtained with 3 ml of culture medium.

For the large scale preparation of dsDNA, two methods were used, ie the alkaline lysis method (Sambrook et al, 1989), and the Promega maxiprep method. With the alkaline lysis method, the cells were grown overnight in a 200ml culture and harvested by spinning the cells down at 5000 rpm for 15 min in an SS34 rotor. The cells were then lysed and the DNA isolated. The DNA was made free of protein by phenol/chloroform extraction and precipitated by ethanol. Further purification was by RNAse incubation and PEG precipitation.

Phage ssDNA was prepared by growing the cells to an A_{600} of about 0.5 followed by inoculation of the culture with phage. This was allowed to continue to incubate at 37°C for 1-2 hours before spinning down the cells (12000 rpm, SS34) and collecting the supernatant. 15% polyethylene glycol 8000, 2.5M NaCl was added to the supernatant to make a final concentration of 3% PEG. This was stored at 4°C overnight, spun at 12000 rpm (SS34) and the pellet of phage collected. This pellet was purified by phenol/chloroform extractions and ethanol precipitation.

B.2.4. General precautions in handling DNA.

All solutions, glassware, and plasticware in direct contact with DNA were autoclaved at 15 psi for 15 min if not already sterile. Sterilisation by filtration through a 0.2 µm membrane filter was used for tissue culture medium in a laminar flow hood. Gloves were worn at all times to minimise contamination by nucleases.

B.2.5. Phenol/chloroform extraction of DNA.

Protein was removed from the DNA by 2 extractions with phenol/chloroform followed by an extraction with chloroform as described by Sambrook et al (1989). Phenol was obtained in a crystalline form and was made to 0.1% by dissolving in 8-hydroxyquinoline and then distilled at 180°C. This solution was then saturated with 50 mM Tris/HCl pH 8.0 before being used.

B.2.6. Ethanol precipitation of DNA.

DNA was precipitated from solution by adding 1/10 volume of 3 M sodium acetate, and then adding to this volume 2.5 volumes of ice cold 95% ethanol. After 15 min at -70°C, the samples were centrifuged at 12000 rpm for 10 min at 4°C to pellet the DNA. The pellet was then washed with 70% ethanol and dried in a vacuum concentrator.

B.2.7. Quantitation of DNA.

The concentration and purity of DNA were determined by spectrophotometric analysis. Pure DNA has an A_{260}/A_{280} ratio of about 1.8. Values higher than 1.8 are indicative of RNA contamination, while lower values suggest contamination by protein (Sambrook et al, 1989).

The concentration of DNA was estimated by the following conversions.

For dsDNA $A_{260} = 1.0$ corresponds to 50 $\mu g\ ml^{-1}$
For ssDNA $A_{260} = 1.0$ corresponds to = 33 $\mu g\ ml^{-1}$

Alternatively, small amounts of dsDNA can be quantitated by the use of agarose gels, using markers of known DNA concentration, and running these on an agarose gel.

B.2.8. Digestion of DNA with restriction endonucleases.

Restriction endonuclease digestions were generally carried out in a final volume of 25 µl for 2 hr at 37°C using React® buffers (Betheseda Research Laboratories, MD, USA). When using SmaI restriction enzyme, the reaction was carried out at 25°C.

Typically 1µl of each enzyme was used (2 µl total in double digests), and when using DNA prepared by the rapid boil method, 1µl of RNAse (10mg ml^{-1}, boiled for 15 min to inactivate DNAses) was added and further incubated for 30 min at 37°C.

B.2.9. Agarose gel electrophoresis of DNA.

1% agarose in 1 x TAE (0.04 M Tris/HCl, 0.02 M acetate, 1.0 mM EDTA, pH 8.0) was used to make the gel, and this was electrophoresed at 80 volts for 1-2 hours. Ethidium bromide was added to the gels before setting (5 µl of 10 mg ml^{-1} solution per 500 ml of agarose) to allow visualisation of the DNA bands using a UV transilluminator. In the case where quantitative markers were used to quantitate the DNA, then no ethidium was put into the gel Instead, the gels, after electrophoresis, were soaked for 20 min in a solution containing 0.5 $\mu g\ ml^{-1}$ of ethidium bromide before visualizing.

Prior to loading the DNA into the wells, 1 µl of loading dye (0.25% bromophenol blue, 40% w/v sucrose) was added to the DNA samples to increase the sample density (to ensure retention in the well), and to allow the progress of the migration to be monitored.

B.2.10. Isolation of DNA fragments from agarose gels.

When pieces of DNA were to be separated from each other for use in ligations, the method of Heery *et al* (1990) was used. In this method the samples were run on a normal 1% agarose gel and visualised under a UV light with a wavelength of 366 nm. The bands of interest were cut out of the gel and placed in a 0.5 ml microcentrifuge tube which had been packed half full of siliconised glass wool, with a pin hole placed in the bottom. This tube was placed into a 1.5 ml microcentrifuge tube and spun for 2 min at 6000 rpm. The DNA passed through into the large tube and could be used for further work without any other preparation.

The siliconised glass wool was prepared by soaking the glass wool for 1 hour in a solution containing 1% silane in tetrachloroethanol and leaving this to evaporate overnight in a fumehood.

B.2.11. DNA ligations.

All ligations were carried out using T4 DNA ligase (Bethesda Research Laboratories) with the 5 times stock buffer provided. All ligations involved an approximately 1:3 ratio of vector:insert unless otherwise specified, and the reaction was generally carried out overnight at about 10°C, followed by 1hr at 37°C unless otherwise specified.

B.3. Mutagenesis of DNA.

In all cases, the method of Kunkel et al (1989) was used. This uses a *dut*$^-$ *ung*$^-$ strain of bacteria (CJ236) to produce a template containing uracil. The *dut*$^-$ genotype lacks the dUTPase enzyme responsible for removing the dUTP from the cell. The *ung*$^-$ genotype lacks the enzyme uracil-N-glycosylase which initiates DNA repair, replacing uracil with thymine. Removing both enzymes results in an increase of uracil in the DNA.

A kit from Biorad® was used when working with M13. The uracil-containing template was elongated from an annealed oligonucleotide. Transformation of the uracil-containing DNA into a *dut*$^+$ *ung*$^+$ strain of bacteria resulted in degradation of the uracil-containing template strand, leaving only the mutated strand.

Two methods of mutagenesis were used. In the first, M13 bacteriophage was used for the template, and in the second, helper phage was used. These will be discussed separately.

B.3.1. Preparation of uracil-containing M13 phage.

A vector containing part of the lactoferrin gene was introduced into the phage M13 dsDNA. This phage was plated out on LB plates with a lawn of JM101 cells. One plaque was picked and put into 1 ml of 2xYT (Sambrook *et al*, 1989) and heated at 60°C for 5 min to kill the cells. This left phage only. At the same time, 5 ml of CJ236 were grown overnight.

Into 100 ml of 2xYT, was put 100 µl of phage, 100 µl of 30 mg ml^{-1} chloramphenicol, 50 µl of 5 mg ml^{-1} uracil and the 5 ml of CJ236 grown overnight. This was grown for 8 hr and then centrifuged at 5000 rpm for 30 min in a Sorvall SS-34 rotor. At this stage a sample was removed to titrate the amount of uracil-containing phage present. To the rest of this 100 ml supernatant, 25 ml of 15% PEG 8000 was added and allowed to precipitate the phage overnight. The pellet was redissolved into 1 ml of TES and incubated with 100 µg of RNAse at 37°C for 30 min. The protein left with the DNA was then extracted using phenol/chloroform, and precipitated by the method of ethanol precipitation.

The degree of incorporation of uracil into the DNA template was estimated by comparing the number of plaques obtained when the uracil-containing phage particles were introduced into JM101 (*dut*$^+$, *ung*$^+$) and CJ236 (*dut*$^-$, *ung*$^-$) strains of E.coli. A ratio of $1:1x10^5$ indicates that the level of incorporation of uracil is sufficient for the template to be used for mutagenesis.

B.3.2. Preparation of uracil-containing phagemids.

The phagemids were prepared according to the protocol of the Muta-Gene® Phagemid in vitro mutagenesis kit obtained from Biorad (Biorad Laboratories, Harbour Way South, Richmond).

A pTZ18U phagemid with the N-lobe of the lactoferrin gene in it (pTZ18U:Lf_N) was introduced into the cell line CJ236 by the $CaCl_2$ method of transformation. This was grown overnight and 1 ml of this overnight culture was used to infect 50 ml of 2xYT containing ampicillin (to select CJ236 with pTZ) This was grown to an A_{600} of 0.3, when helper phage (M13KO7) was added to obtain a

multiplicity of infection of around 20 helper phage per cell. This was incubated for another hour, at which time kanamycin was added (70 µl of 50 mg ml^{-1}) and incubated for 4-6 hours before harvesting the phage as for M13 (outlined above).

B.3.3. Phosphorylation of mutagenic oligonucleotide.

Prior to phosphorylation, the oligonucleotide was cleaned by passing it through a reverse phase SEP-PAK C_{18} column using the following protocol:-

(i) Wash the column with acetonitrile.
(ii) Wash the column with 50 mM NaCl.
(iii) Dissolve the oligonucleotide in 0.5 ml of 50 mM NaCl.
(iv) Load the oligonucleotide slowly on to the column.
(v) Wash the column with 10 ml of 50 mM NaCl.
(vi) Elute the oligonucleotide with 30% volume of acetonitrile, and 70% volume of 50 mM NaCl.

The concentration and purity were then determined, and the molar concentration recorded.

The oligonucleotide was phosphorylated using T4 polynucleotide kinase. 1 unit of kinase was added to 4 picomoles of oligonucleotide with kinase buffer and 1µl of 10 mM ATP present in a final volume of 30 µl. This was incubated at 37°C for 3 hr and then heated at 65°C for 10 min to stop the reaction.

B.3.4. Annealing and elongation of the mutagenic oligonucleotide.

The oligonucleotide (0.6 picomoles) was added to 0.2 picomole of template to make a final volume of 10 µl with a final buffer concentration of 20 mM Tris/HCl pH 7.4, 2 mM $MgCl_2$, and 50 mM NaCl. The tube containing the reaction was placed into 500 ml of water at 80°C, allowed to cool slowly to less than 35°C, and then placed in a 4°C room.

Elongation buffer (40 mM Tris/HCl-HCl pH 8.0, 4 mM DTT, 20 mM $MgCl_2$, 1 mM of each dNTP, 2 mM ATP) was added to the reaction, along with 1 unit of T4 DNA polymerase, and 1 unit of T4 DNA ligase. The reaction was placed on ice for 5 min, then at 25°C for 5 min, then at 37°C for 2 hours. The cool temperatures help the primer to base pair to the template before elongation occurs (37°C). The reaction is stopped by adding 2 µl of 0.25 M EDTA.

In addition to this reaction, a control reaction containing template only (no oligonucleotide) was carried out to evaluate the amount of exogenous priming occurring.

B.3.5. DNA sequencing.

Sequencing was carried out using both single-stranded and denatured double-stranded DNA as templates. Primers were annealed to the template DNA and extended using Sequenase version 2.0 according to the manufacturer's instructions.

Single-stranded templates were prepared as detailed in section B.2.3 for phage. Double-stranded templates were prepared by using the Promega Magic miniprep system. Typically 7 μg of dsDNA was used for sequencing. This was then denatured and precipitated according to the Sequenase version 2.0 protocol, before annealing of the primer prior to sequencing.

Sequencing gels were prepared from acrylamide made up for no more than 1 month, and electrophoresed as described in Sambrook et al (1989). The gels were then fixed in 20% methanol, 5% acetic acid for 30 min and dried for 1 hr in a Biorad® gel drier at 80°C. The gel was fully placed in a light-tight cassette adjacent to X-ray film (Fuji) for 36 hr before developing, either automatically with a Kodak X-Omat automatic film processor, or manually.

B.4. Expression of lactoferrin by tissue culture.

B.4.1. Maintenance of mammalian cells in tissue culture.

BHK cells were grown at 37°C in a humidified CO_2 incubator with a maintained atmosphere of 5% CO_2 in air. Transfections, passing of the cells, and feeding were carried out in a laminar flow hood. This was swabbed with 70% ethanol before every use and UV-sterilised periodically.

The BHK cells were grown in 50 ml or 300 ml flasks with the lids partially open to allow the CO_2/air into the flask. The cells were grown in a 1:1 mixture of Dulbecco's modified Eagle's medium (DMEM) and Hams-F12 (F12). This was supplemented with 10% foetal calf serum (FCS). The antibiotics used were changed regularly to minimise the development of antibiotic resistant strains of bacteria. The antibiotics used were penicillin (100 U ml^{-1}), streptomycin (100 U ml^{-1}), and enroflaxacin (Baytril)

(25mg l^{-1}).

All plasticware was supplied sterile, and glass pipettes were plugged with cotton wool and sterilised by dry heat at 160°C prior to use.

B.4.2. Preparation of tissue culture reagents.

DMEM/F12 was made by mixing equal volumes of DMEM and F12 which had been prepared according to the manufacturers' instructions and filter-sterilized through a 0.2 µm Acrocap filter into sterile bottles. A masterfile pump with a flow rate of about 50 ml min^{-1} was used to pump the medium. The sterile medium was stored at 4°C.

FCS was obtained in a heat-inactivated form and aliquoted into 20 ml aliquots. These were stored at -20°C.

Antibiotics were prepared in water, made up into 1.5ml aliquots and stored at -20°C.

Complete medium was made up prior to use. For flasks, 40 ml of FCS was mixed with 400 ml of DMEM/F12 and 1.5 ml of antibiotics. For roller bottles, 20 ml of FCS was mixed with 400 ml of DMEM/F12 with antibiotics, and 4 ml of 8 mM ZnSO4 was added to promote expression of recombinant protein using the pNUT expression vector.

B.4.3. Passage of cells.

Cells were grown until >80% confluent, which typically occurred 2 days after passage. To pass the cells, the medium was removed, and 4 ml of virsene (0.2 g Na_2EDTA, 8.0 g NaCl, 0.2 g KCl, 1.15 g Na_2HPO_4 per litre, pH 7.2) was added to cover the cells. This was left for 2 minutes, and then removed. The cells were left for a further 10 minutes in which time they became detached from the flask. 2.5 ml of medium was then added to the cells to suspend them before transfer into a larger flask. Sufficient medium was added and the flasks placed back into the incubator.

B.4.4. Freezing and thawing of cells.

Cells were grown in flasks until >80% confluence, and then "passed" (B.4.3). 2 ml fractions were collected from a 50 ml flask and resuspended into 2 ml of medium containing 20% DMSO and 80% FCS. This was aliquoted into four cryotubes which were then wrapped in paper towels and stored at -70°C for 24 hr to allow the cells to

freeze slowly. The tubes were then transferred into liquid nitrogen after freezing (24 hr).

Cells were thawed at room temperature in the laminar hood and transferred into 50 ml flasks. One tube was put into one 50 ml flask and 6 ml of media added. The media was changed 24 hr later to replace the DMSO.

B.4.5. Transfection and selection of BHK cells.

In order to introduce foreign DNA into the BHK cells for expression, the Cell Phect kit (Pharmacia LKB Biotechnology) was used and the protocol supplied with this was generally followed. 6 μg of DNA was used to transfect one flask, 70-80% confluent with cells.

The DNA to be transfected was dissolved in 120 μl of deionised water, and 120 μl of buffer A was added to this and incubated at room temperature for 10 min. 240 μl of buffer B was then added and immediately vortexed before a further incubation at room termperature for 15 min. The DNA solution (with the DNA now as a calcium phosphate precipitate) was added to the cells containing the medium, and incubated for 4-6 hours in the 37°C incubator. The cells were washed twice with medium, after which 15% glycerol in isotonic HEPES buffer pH 7.5 was added, and incubation was continued for 1.5 min. The glycerol in isotonic HEPES was removed and the cells rinsed two more times with medium prior to feeding them.

24 hours after transfection, the cells were fed with 0.25 mg ml^{-1} methotrexate (MTX) in media containing 10% FCS. The MTX was used to select for cells that had taken up the DNA, as the DNA expression vector (pNUT) contains a mutant form of dihydrofolate reductase (DHFR) which has a 270-fold lower affinity for MTX. This means that cells containing the vector express the mutant DHFR and become resistant to MTX.

B.4.6. Large scale growth of cells in roller bottles.

Cells were grown until four 300 ml flasks were 80% confluent, and then passed into roller bottles. These bottles allow large amounts of recombinant protein to be grown. 80-100 ml of medium containing $ZnSO_4$ were placed into these bottles and these were fed every 48 hr. From this stage onwards, there was no methotrexate present in the media. The bottles were placed in a 37°C room on a rotating machine, and the medium in the bottles was changed and collected every 48 hr until extensive cell death occurred. This was typically after 1 month.

B.5. Protein purification and analysis.

B.5.1. SDS-polyacrylamide gel electrophoresis of proteins.

In general the protocols in the Hoefer Mighty Small II instruction manual were used. The resolving gel consisted of 10% acrylamide, 0.9% bisacrylamide, 0.375 M Tris/HCl-HCl pH 8.8, 0.1% SDS, 0.1% w/v ammonium persulphate, and 0.07% v/v TEMED. Stacking gels were always used and consisted of 4% acrylamide, 0.36% bisacrylamide, 0.2 M Tris/HCl-HCl pH 6.8, 0.1% SDS, 0.4% w/v ammonium persulphate, and 0.05% v/v TEMED. The tank buffer was 0.02 M Tris/HCl-HCl pH 6.8, 0.192M glycine, 0.1% SDS.

Prior to loading, the samples were diluted 1:1 with 2x treatment buffer (0.125 M Tris/HCl-Cl pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol containing a drop of Bromophenol Blue) and placed in a boiling water bath for two minutes.

The gels were run at 15 mA until the dye band was through the stacking gel, and then at 20-25 mA through the resolving gel.

B.5.2. Staining and destaining of polyacrylamide gels.

Gels were stained using Coomassie brilliant blue R-250 (0.125% Coomassie blue R-250, 50% methanol, 10% acetic acid) for 1 hr or more at room temperature. The stain was tipped off and strong destainer (50% methanol, 10% acetic acid) was used for two changes. Further destaining was in a weak destainer containing 7% acetic acid and 5% methanol until there was no background colour. The gels were photographed at this stage, and then dried on to blotting paper using a Biorad® gel dryer at 80°C for 30 minutes.

B.5.3. Immunoprecipitation of human lactoferrin in tissue culture.

The cells were shown to express the recombinant human lactoferrin by immunoprecipitation with antibodies specific to human lactoferrin. These antibiotics had been raised in New Zealand white rabbits, purified, and demonstrated to be specific for human lactoferrin (Stowell, 1990).

Excess media (0.5 ml) was mixed with 30 µl of antibody solution and incubated at 37°C for 1 hr, before being centrifuged at 12000 rpm for 10 min at 4°C. The pellet was washed with phosphate buffered saline containing 1% Triton-X-100, and spun at 12000 rpm for 10 min. This washing step was repeated and the pellet was then resuspended in 10 µl of SDS loading dye ready for SDS analysis.

B.5.4. Purification of recombinant lactoferrin.

Media containing expressed protein were centrifuged in GSA bottles at 8000 rpm and 4°C for 60 min to remove any cell debris. Before loading on to an ion exchange column, the pH and conductivity of the supernatant was checked, and adjusted to the loading conditions if required. It was then loaded on to a CM-Sephadex C-50 column under gravity and rinsed with buffer A (10 mM HEPES pH 7.8, 0.2 M NaCl). The protein was eluted using an Econosystem pump with a flow rate of 1.0 ml min^{-1}. A gradient was used starting with 100% buffer A and increasing to 100% Buffer B (10 mM HEPES pH 7.8, 1.0 M NaCl) after 60 mins. This was continued at 100% buffer B for a further 30min or until the A_{280} peak was seen.

After deglycosylation (see section B.5.7), the proteins were purified either by the above method, or by using a SP-sephadex column, loading the protein in 20 mM Tris/HCl-HCl pH 7.0, 0.2 M NaCl, and eluting with 20 mM Tris/HCl-HCl pH 7.0 1.0 M NaCl.

B.5.5. Determination of protein concentration.

The concentrations of the proteins were estimated from the A_{280} absorbance using the following calculations.

For iron loaded lactoferrin, conc (mg ml^{-1}) = A_{280} x 1.2
For apolactoferrin, conc (mg ml^{-1}) = A_{280} x 1.09

B.5.6. Concentration of protein samples.

The samples were concentrated using a Centricon-10 concentrator for the half length molecules, and a Centricon-30 concentrator for the full length molecules.

The total protein content before and after concentration was checked and typically about 90% of the protein was accounted for in the concentrated sample. The A_{280} of the eluates was checked to make sure no protein had leaked through the ultrafiltration membranes.

B.5.7. Deglycosylation of lactoferrin.

Deglycosylation of lactoferrin was carried out prior to crystallisation, as previous work (Norris *et al*, 1989) had shown that crystals can be of better quality if the protein had been deglycosylated. The proteins were deglycosylated after being purified on the CM-Sephadex column using a preparation comprising two enzymes, endoglycosidase F (Endo F), and peptide N-glycosidase (PGNase F), both isolated from *Flavobacterium meningosepticum* (Elder and Alexander, 1982).

Endo F/PNGase F (kindly prepared by G.E.Norris) was incubated with the glycosylated lactoferrin at 37°C for 48 hr until deglycosylation as judged by SDS-PAGE was complete. The lactoferrin was separated from the deglycosylating enzyme on a CM-Sephadex column, or a SP-Sephadex column (section B.5.4).

B.5.8. Iron release from lactoferrin.

The effect of pH on iron release from lactoferrin was determined by iron release experiments in which the amount of iron bound to the lactoferrin was measured under varying pH conditions. The buffers used for the various pH conditions were pH 7.0 - 5.0, 50 mM MES; pH 5.5 - 4.0, sodium acetate (I=0.05); pH 3.5 - 2.0, 100 mM glycine HCl. In addition, all the buffers contained 0.2 M NaCl.

Samples were dialysed for 48 hr at each particular pH, with separate protein samples being used for different buffers. At each pH value, the absorbtion spectrum from 250-700 nm was taken using a Cary1 spectrometer (Varian). The absorbances were measured at 280 nm, λ_{max} for the iron absorption (usually around 454 nm) and 700 nm. The A_{700} value was subtracted from the A_{280} and iron peak absorbances to subtract any background absorbance. The iron peak absorbance was then "normalised" by multiplying its value by $1/A_{280}$.

C. RESULTS - DNA MANIPULATION.

C.1. Site directed mutagenesis by in vitro oligonucleotide extension.

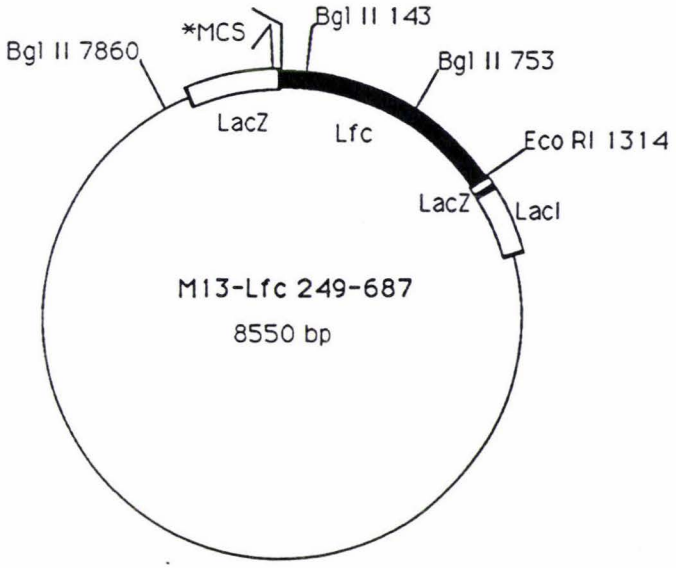
Oligonucleotide-directed in vitro mutagenesis is a widely used procedure for introducing specific point mutations into the cDNA encoding for a protein of interest. The segment of DNA to be mutated is cloned into a vector in which the DNA can exist in both single- and double-stranded forms (as in M13 bacteriophage). An oligonucleotide complementary to the region to be altered, except for the base mismatches to introduce the mutation, is annealed to the single stranded form of the vector. The complementary strand that includes the mismatch(s) is then synthesised by T4 DNA polymerase using the oligonucleotide as the primer. T4 ligase is used to join the 3' end of the new strand to the 5' end of the oligonucleotide, producing covalently closed double-stranded DNA. The double stranded DNA is then transformed into E.*coli*, resulting in two classes of progeny, the original wild type, and those carrying the mutation.

A method described by Kunkel (1989) provides a strong selection against the original wild type strand. Before elongation, the single strand of the vector is synthesised in a *dut*$^-$ *ung*$^-$ double mutant bacterium, CJ236. This causes the DNA to contain a high proportion of uracil(s) in place of thymine. The uracil-containing strand is then used as the template for the in vitro synthesis of the complementary strand that contains the mutation of interest. The resulting double-stranded DNA is transformed into an E.*coli* cell that is *dut*$^+$ *ung*$^+$ as in JM101 or XL-1, and the uracil containing strand is inactivated leaving the non-uracil-containing strand with the mutation in it to survive and form clones.

C.2. Mutation of Pro 251 in LfN.

C.2.1 Preparation of uracil-containing single-stranded DNA template for mutagenesis.

A template for mutagenesis was constructed by inserting the 1314 bp Sma I/Eco RI fragment encoding residues Val 250 to Glu 688 of the human lactoferrin cDNA, into the bacteriophage M13 genome. This work was carried out by Alan Bates, and the resulting construct is illustrated below (Fig. C.1).



* MCS = multiple cloning site. (Hind III; Sph I; Sal I; Acc I; HInc II; Xba I; Bam HI; SmaI)

Fig. C.1. M13:Lfc template used for mutagenesis of Pro 251.

A ssDNA uracil containing template was produced from this vector as outlined above and described in detail in section B.3.1.

C.2.2. Design of the mutagenic oligonucleotide.

An oligonucleotide with mutations in the codon for Pro 251 was designed (Fig. C.2.) in order to introduce 4 different amino acids in place of proline. Glycine and alanine were chosen to introduce flexibility into the peptide chain at this point, while valine and aspartate were introduced in order to add an unconstrained hydrophobic group and a charged hydrophilic group respectively.

The sequence to be mutated lies at one end of the fragment of the lactoferrin gene cloned into M13, close to the polylinker region. The mutagenic oligonucleotide was designed to overlap the polylinker site in M13 and to introduce a change in sequence which would ablate the Bam H1 site in the multiple cloning site, as well as introducing the required change(s) into the lactoferrin cDNA fragment. This would allow the initial selection of mutants by digesting the elongation products with Bam H1. DNA molecules which do not have the Bam H1 site will survive.

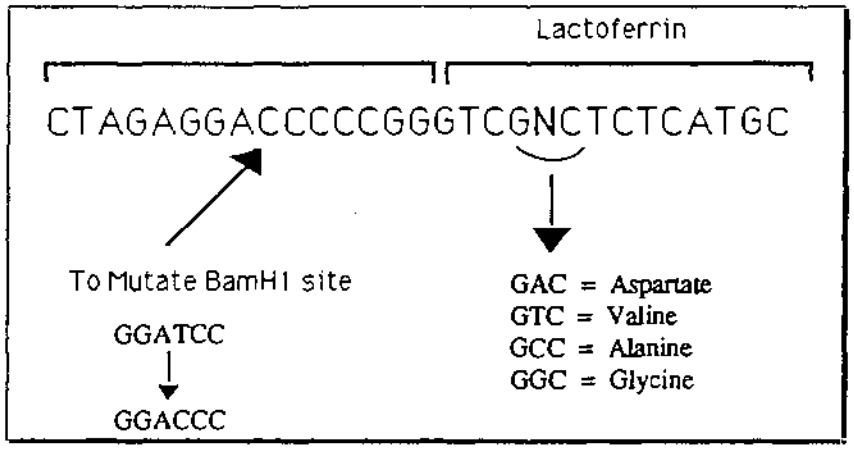


Fig. C.2. Mutagenic oligonucleotide to introduce changes at Pro 251.

Table C.1. shows the wild type sequence coding for Pro 251, along with the changes in sequence required to introduce the changes (P251V, P251A, P251D and P251G) in the cDNA. As seen in Table C.1, the mutagenic oligonucleotide contains an extra change to the sequence (T -> C). This is included in order to obtain an equal ratio of all four mutants, as annealing would favour the oligonucleotide with the least number of mismatches (alanine GCT). By including the extra nucleotide change, the number of mismatches on the oligonucleotides is made more equivalent for each changed amino acid.

Table. C.1. Mutagenic oligonucleotide sequence changes.

Protein	Sequence	With additional T to C mutation
Wild type (Proline)	CCT	
Valine	GTT	GTC
Alanine	GCT	GCC
Aspartic acid	GAT	GAC
Glycine	GGT	GGC

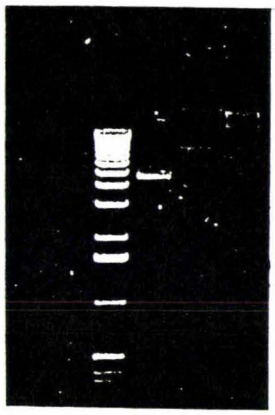
Another way to try and obtain a more even representation of all four mutants was to use a low ratio of oligonucleotide to template which would favour all the mutagenic oligonucleotides annealing to the template. A ratio of 0.05 picomoles of oligonucleotide to 0.1 picomoles of template was used for this. This ratio would not normally be considered favourable as it would allow any contaminating fragments to anneal to the excess template resulting in unwanted clones. However as outlined earlier,

this problem can be overcome by cutting the resulting elongation products with Bam H1 and in theory only mutagenic clones should remain.

C.2.3. Elongation of the mutagenic oligonucleotide.

The mutagenic oligonucleotide was prepared by passing it through a SepPac™ column and adding 5'-phosphate groups as described in section B.3.3.

The oligonucleotide was annealed to the template and elongated according to section B.3.4. A control reaction containing template only (no oligonucleotide) was run alongside the reaction to test for any excessive endogenous priming that may have occurred.



L 1 2 3

Fig. C.3. DNA gel of the products of elongation from the mutagenic oligonucleotide. Lane 1 contains the ssDNA template before elongation. Lanes 2 and 3 contains the products of the elongation reaction with (Lane 3) and without (Lane 2) the oligonucleotide.

Both the control (template only) and the mutagenic (template and oligonucleotide) elongation reactions show products suggesting that elongation had occurred. This was evident by the presence of a band on the DNA gel (Fig. C.3) corresponding to the double stranded form of the M13 vector (8.5 kb). A greater degree of elongation, however, had occurred in the mutagenic reaction (lane 3), seen by a more intense band.

A puzzling discovery observed on the gel (Fig. C.3) was the presence of extra bands following the elongation reaction in the presence of the mutagenic oligonucleotide. Whereas the control reaction had 2 bands corresponding to two forms of the double stranded product, the mutagenic reaction had 2 sets of 2 bands, the top band of each set corresponding to one of the 2 bands seen in the control reaction (see Fig. C.3). The extra bands were unexpected, and may be due to a deletion caused by the DNA bending back on itself leading to deletion of a segment of the DNA. At this stage, this looked to be a major concern and further separation of these two elongation products by gel purification was contemplated.

C.2.4. Transformation of E.*coli* with the products of the mutagenic reactions.

A small sample was taken out from the mutagenic reaction for comparison when plating. The two reaction products (control and mutagenic) were then dialysed against restriction buffer 4 (for Bam H1) and digested with the Bam H1 restriction enzyme for 2 hours. As outlined earlier (C.1.2), the mutagenic oligonucleotide contained a mutation in the Bam H1 restriction enzyme site (Fig.2). In theory therefore, all clones containing the required mutation(s) should have lost the Bam H1 site, rendering them resistant to Bam H1 digestion.

After cutting with Bam H1, the reaction mixtures (template only, and template with oligonucleotide) were dialysed against 10% glycerol in water and introduced into JM101 cells by electroporation, (prepared as in section B.2.2), plated onto a lawn of JM101 cells on LB plates, and incubated overnight at 37°C.

The results of these transformations are summarised in Table 2. The plate of the control reaction (cut with Bam H1) had 15 plaques. The transformation from the elongation-ligation reaction involving the mutagenic oligonucleotide resulted in confluent plaques with no dilution, while a 1/25 dilution yielded approximately 350 plaques per plate. The comparative transformation with the reaction products which had not been treated with Bam H1 yielded a confluent plate at 1/25 dilution. The difference in transformation efficiency following Bam H1 digestion demonstrates the benefit of including the mutation to the Bam H1 site in the mutagenic oligonucleotide. From the BamHI-cut 1/25 dilution plate, 20 plaques were chosen and added to 5 mls of 2xYT containing 5 μl ml^{-1} of overnight JM101 and prepared for sequencing.

Table. C.2. Number of plaques after transformation into *E.coli* JM101 cells.

Reaction	no of plaques on plate	no of plaques on 1/25 diluted plate
Control cut with Bam HI	15	N/A
Mutagenic	Confluent	Confluent
Mutagenic cut with Bam HI	Confluent	350

C.2.5. Identification of mutants by DNA sequencing.

For ten of the clones, single-stranded DNA was prepared as outlined in section B.2.3 for DNA sequencing by the dideoxy method of Sanger *et al* (1977). A sample of the DNA in each case was run on a gel to check purity, and this showed only one band present. This means that the previous problem of the two bands seen in the elongation reaction had vanished. Restriction enzyme analysis of a double-stranded DNA preparation showed that the band that remained was the larger band (the desired product). A possible explanation for the loss of the other band may have been that this deletion was in a region that is an essential part of the M13 genome and loss of this resulted in death of the bacteriophage.

The clones were sequenced, and the results showed that one clone (no 3) was wild type, but clones 1-2, and 4-10 all contained a band that lay across the four lanes of the sequencing gel at the mutated site (N of GNC). Initially this was thought to be due to inadequate isolation of the clones, but further streaking and sequencing produced the same result.

An analysis of the sequence around this region, carried out by Dr Hale Nicholson, using the programmes "Fold" and "Squiggles" from the GCG package, showed that the mutation site was at a critical point on a hairpin. This is shown diagrammatically in Fig. C.4.

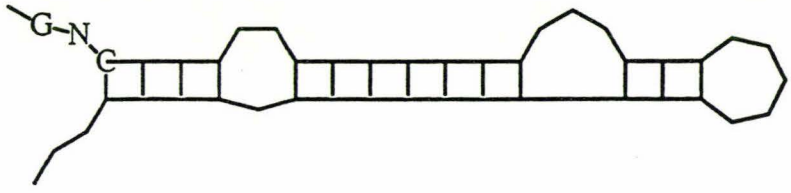


Fig. C.4. Hairpin of sequence around the mutation site of Pro 251.

Several strategies to overcome this problem are outlined in the instruction manual for the Sequenase 2 kit. Several of these strategies were tested, with dilution of the labelling mix, or addition of manganese to the reaction mixture producing the best results. The first 9 clones selected produced 3 of the 4 desired mutants, and after picking another 12 plaques, 2 clones of each mutant were isolated.

For each of these mutants, a double stranded DNA preparation (alkaline lysis and PEG precipitation) was carried out, and the purity and concentration of the samples established. The A_{260}/A_{280} ratios ranged from 1.83 to 1.91 and the yield ranged from 270 μg to 385 μg. This was sufficient DNA for any subsequent analysis and cloning.

C.3. Cloning into pNUT.

Protein expression requires that the DNA be transfected into baby hamster kidney (BHK) cells. As M13 is not suitable for protein expression in these cells, it is necessary to move the mutation from M13 to a suitable expression vector, pNUT. The vector pNUT:Lf_N (shown in Fig. C.5) contains the cDNA for the N terminal lobe of lactoferrin, along with a metallothionein promoter, and a mutated dihydrofolate reductase gene that allows for the selection of BHK cells containing pNUT (see section C.4.1).

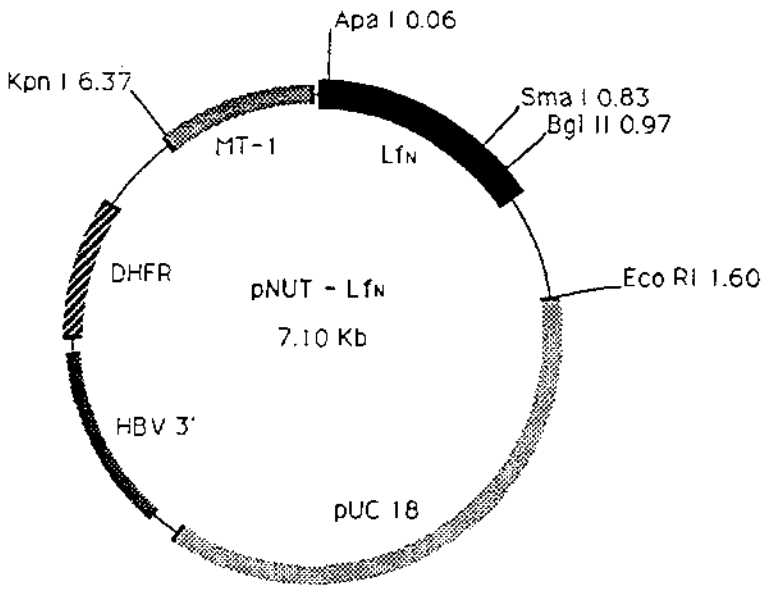


Fig. C.5. The vector pNUT:Lf_N used for expression of lactoferrin in Baby Hamster Kidney cells.

Ligations were generally carried out using gel-purified fragments. This method of gel purification is explained in section B.2.10. In this particular case, however, an altered pNUT:Lf$_N$ clone had been constructed by Dr Hale Nicholson. PNUT:Lf$_N$ was digested with the restriction enzymes Sma I and Bgl II to release a small fragment (145 bp) from bases 829 to 974 of the Lf$_N$ cDNA (including Pro 251), leaving a large 7.0 kb construct (Fig.C.6) that was isolated.

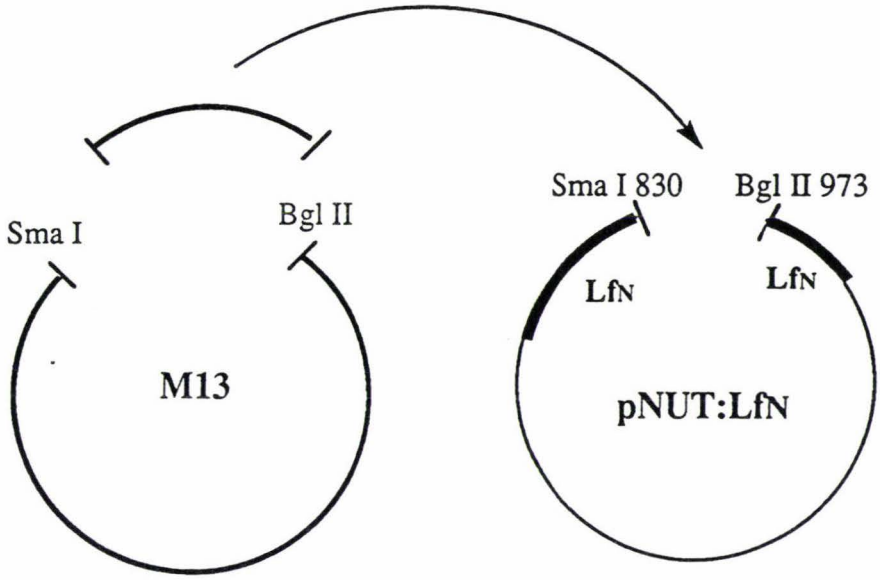


Fig. C.6. Construction of pNUT:Lf$_N$:M13 from pNUT:Lf$_N$.

In a separate experiment, the M13 genome was also digested with Sma I and Bgl II to release a fragment of 693 bp. This fragment was isolated and inserted into the large fragment of pNUT:Lf$_N$ between the Sma I and Bgl II sites as illustrated in Fig. C.6 above.

C.3.1. Ligation into pNUT.

Restriction digests were carried out on both the DNA from the M13 clones containing the Pro 251 mutations, and the altered mammalian expression vector pNUT:Lf$_N$:M13. A Sma I : Bgl II cut was chosen as this involved a small, easily-sequenced piece of DNA which could then be easily moved.

For each of the Pro 251 mutations, 5 μg of M13:Lf$_N$ in a total volume of 80 μl was cut at 25^{o}C (Sma I conditions) for 1hr in buffer 4 with both restriction enzymes present. After 1hr the temperature was raised to 37^{o}C (Bgl II conditions) for a further

1hr incubation. The reaction was stopped by phenol/choroform extraction and ethanol precipitation. The altered pNUT:Lf_N:M13 vector (Fig. C.6) was also cut with Sma I and Bgl II restriction enzymes.

Products of each of the two restriction enzyme digestions (M13:Lf_N and pNUT:Lf_N:M13, each cut with Sma I/Bgl II) were ligated together in an M13 insert:pNUT vector ratio of 2:1. 0.25 µl of T4 ligase enzyme was added, and the reaction was carried out at 16°C overnight in a total volume of 15 µl.

The products of this ligation were electroporated into DH1 cells as these cells are *rec*$^-$ and therefore the vector DNA should not recombine. The cells were plated out on to LB plates containing ampicillin to select for cells containing the pNUT vector (pNUT is *amp*^{r}). Several colonies were picked and the clones were checked by restriction enzyme analysis using Sma I and Eco RI restriction enzymes. The pNUT:Lf_N:M13 contains a 1329bp Sma I/Eco RI fragment when run on an agarose gel. The pNUT:Lf_N vector that contains the Sma I/Bgl II lactoferrin sequence (with the Pro 251 mutations) gave a 920bp Sma I/Eco RI fragment when run on an agarose gel. This is the ligation that was required.

For each of the Pro 251 mutations, the DNA gel showed that most samples were cut only once. This may have been because of poor restriction enzyme conditions, or because of a blunt-ended piece of DNA ligating to the blunt-ended Sma I site. About 20% of the clones showed the pattern expected if the mutagenised cDNA fragment had been inserted into the Sma I/Eco RI-cut pNUT:Lf_N.

C.3.2. Sequencing of the DNA insert in pNUT.

A double-stranded preparation (alkaline lysis and PEG precipitation) was carried out on clones which showed the correct size fragment after digestion with Sma I/EcoR1. The A_{260}/A_{280} ratio of these preparations ranged from 1.6 to 1.88, and the yield was about 1mg of DNA in each case. 15 µg of DNA was used for double-stranded DNA sequencing according to the Sequenase* version 2.0 procedure. The CLD5 oligonucleotide was used as the sequencing primer (see appendix). In addition to normal sequencing methods, the P251G and P251A clones were sequenced with a reaction mixture that included Mn^{2+}, in order to avoid the earlier problem of the band lying across all four lanes of the sequencing gel (section C.1.5).

The autoradiograph from the sequencing gel showed that all four clones had the expected sequence between Sma I and Bgl II. The Mn^{2+} was not needed in this case as the earlier problem of a band across all four lanes did not occur.

C.4. Expression of recombinant lactoferrin in baby hamster kidney cells.

C.4.1. Selection method.

The vector pNUT, incorporating the lactoferrin cDNA was introduced into BHK cells with the aid of a method of selection that makes use of the pathway for thymine biosynthesis in mammals. Tetrahydrofolate is involved in the addition of methyl groups to dUMP to make dTMP. One step in the pathway catalysed by the enzyme dihydrofolate reductase (DHFR), regenerates dihydrofolate and NADPH to tetrahydrofolate and $NADP^+$. An analogue of dihydrofolate, methotrexate, is a potent competitive inhibitor ($K_i < 10^{-9}M$) of DHFR. By binding to the active site of DHFR, methotrexate prevents the regeneration of tetrahydrofolate, resulting in death of the cell.

This characteristic of methotrexate is used in the selection of cells that have taken up the lactoferrin cDNA in the form of the vector pNUT. It involves introducing a second DHFR cDNA into the cell as part of the introduced pNUT vector. This has a point mutation in it that decreases the binding constant for methotrexate, but does not dramatically alter the binding constant for dihydrofolate. Cells that have taken up pNUT therefore, express the mutated form of DHFR and are able to grow in the presence of elevated levels of MTX. Non-transformed cells synthesise only the wild-type form of DHFR and consequently are sensitive to elevated levels of MTX.

C.4.2. Transfection into BHK cells.

BHK cells were transfected with the pNUT cDNA (containing the mutations) according to the procedure outlined in section B.4.5. The cells were transfected with 4 μg of DNA from each mutant (4 cell lines in total). Following transfection, the cells were incubated at 37°C, 5% CO_2, for 24 hr before selection with methotrexate was commenced.

Cells with the pNUT:Lf_N vector stably integrated into their genome were cultured under methotrexate selection for a period of 2 - 3 weeks. Following MTX selection, an aliquot of the growth medium was tested for lactoferrin expression by immunoprecipitation using antibodies against human lactoferrin. Expression of the N-lobe of lactoferrin was confirmed by SDS analysis of the immunoprecipitate product. A gel of the immunoprecipitate is shown in Fig. C.7.

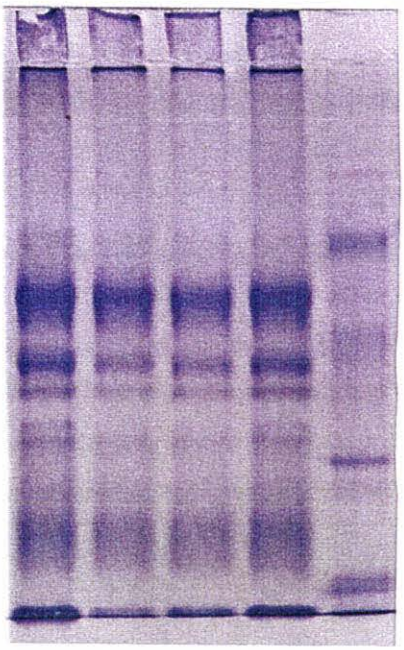


Fig. C.7. SDS gel of the immunoprecipitate of all four proteins. Lanes A (P251A), D (P251D), G (P251G), and V (P251V) refer to the recombinant proteins, L refers to a low to middle range standard ladder.

Once confluent cultures of methotrexate-resistant BHK cells had been obtained, frozen stocks of the cells were prepared as described in section B.4.4.

C.5. Production of the mutant T90A.

C.5.1. Construction of a uracil-containing template for mutagenesis.

An oligonucleotide corresponding to the region around Thr 90 in the lactoferrin cDNA, but encoding alanine in place of threonine, was passed through a SepPak™ column. The 5' end of the oligonucleotide was then phosphorylated as described in section B.3.3, before annealing to uracil-containing ssDNA isolated from an M13 clone containing the lactoferrin cDNA. This oligonucleotide was elongated (section B.3.4) and transfected by electroporation (200ohms, 25 microFarrads, 2.5 Volts) into *E.Coli* JM101 cells, and plated onto luria broth plates along with extra JM101 cells. Few plaques were formed however.

The failure to obtain plaques was initially attributed to an inability of the 3' end of the oligonucleotide to anneal correctly to the template. This was checked by

sequencing another clone using the oligonucleotide as the sequencing primer. This produced a correct, readable, sequence indicating that elongation could occur from the 3' end of the oligonucleotide, and that this end was correctly annealed.

Another possibility was that the oligonucleotide may be fragmented and therefore only a small proportion of the oligonucleotide was annealing properly and elongating. This was checked by taking some of the oligonucleotide that had not been previously phosphorylated, and phosphorylating it using γATP. It was then compared with another oligonucleotide of the same length, also phosphorytlated with γATP, on a 20% acrylamide gel (Fig. C.8). The autoradiograph contained one band at the correct position on the gel, indicating that fragmentation had not occurred (seen by a ladder of bands).

Fig. C.8. Autoradiograph of the polyacrylamide gel showing the T90 oligonucleotide as one band of approximately 20 base pairs. (Lane 1, An oligonucleotide of 26 base pairs, Lane 2, The T90 oligonucleotide, Lane 3, Another oligonucleotide of 20 base pairs.)

Given the difficulty of obtaining plaques with M13 clones, an alternative strategy, using phagemids, was explored.

C.5.2. In vitro mutagenesis using phagemids.

Vectors used for in vitro mutagenesis are often derivatives of the *E.coli* bacteriophage M13 as the life cycle of M13 includes both single- and double-stranded DNA stages.

The Muta-Gene phagemid kit from Biorad, however, makes use of phagemid vectors which are derived from the plasmids pUC18 and pUC19, but with the incorporation of the single-stranded replication origin of the phage f1. These phagemids are called pTZ18U and pTZ19U respectively.

The phagemid replicates as a double-stranded circular plasmid until the host cell is infected with a helper phage derived from f1 or M13. Proteins coded by the helper phage act at the single strand origin in the phagemid causing replication, packaging and export from the cell, of the single-stranded phagemid.

C.5.3. Preparation of uracil containing template for pTZ18U:Lf_N.

The cDNA for the N-lobe of lactoferrin was cloned into the phagemid vector pTZ18U by Dr C.Day, to give pTZ18U:Lf_N. This is illustrated in Fig. C.9.

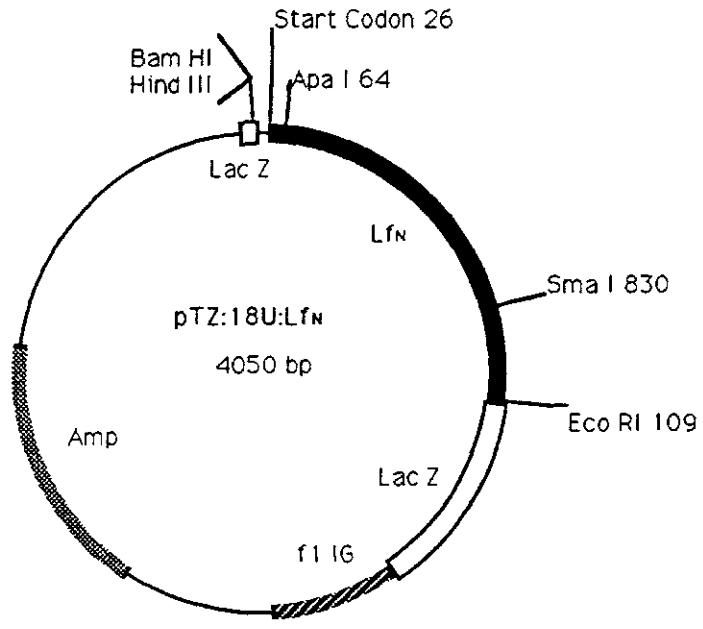


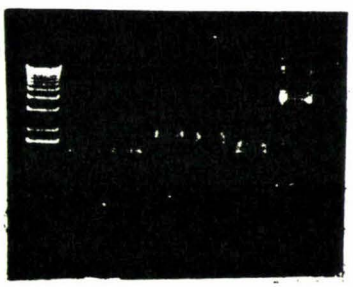
Fig. C.9. The plasmid pTZ:18U:Lf_N used to make mutants of Thr 90.

The double-stranded DNA form of this vector was transformed into *E.coli* CJ236 by the calcium chloride method described in section B.2.2. The single-stranded DNA from the phagemid was produced by infection of the host cells with the helper phage MK13KO7. The single-stranded template DNA was purified and then run on a gel to check for purity.

C.5.4 Production of the T90A mutation in pTZ18U:Lf_N.

The mutagenic oligonucleotide (T90A) was annealed on to the uracil-containing pTZ:18U:Lf_N template and elongated under various conditions, as outlined below, and illustrated in Fig. C.10.

Lane 1 - Ladder
Lane 2 - ssDNA template marker
Lane 3 - Template only }
Lane 4 - Template and 1.0 µl oligonucleotide } @ 15°C, Overnight.
Lane 5 - Template and 2.0 µl oligonucletide }
Lane 6 - Template and 2.0 µl p251 oligonucleotide }
Lane 7 - Template and 2.0 µl oligonucleotide @ 37°C, 2 hr



1 2 3 4 5 6 7

Fig. C.10. DNA gel of the products of elongation of template DNA with the various oligonucleotides and conditions as outlined above.

As can be seen, the best result came from elongation at 37°C for 2 hr, conditions more suited to the polymerase enzyme than to the ligase enzyme. The resulting elongation products were transformed into JM101 cells by the $CaCl_2$ method and plated. The ratio of template + oligo to template only was 10:1, and colonies were picked from the template + oligo plate for sequencing.

A point to note, however, was that the control reaction was incubated at 15°C overnight, and not at 37°C for 2 hr. Under these conditions the ratio is academic and sequencing will determine the true ratio of mutant to wild type clones.

C.5.5. Identification of the T90A mutant by DNA sequencing.

Both the single-stranded and double-stranded DNA were prepared as outlined in section B.2.3. DNA sequencing was carried out as outlined in section B.3.5.

The initial attempts at sequencing the ssDNA with the CLD3 primer produced a smear on the DNA sequencing gel. A control M13 sequence, using the control DNA in the sequenase kit, produced a perfect sequence indicating that the problem lay in the preparation of single-stranded DNA, not in the sequencing method. Double-stranded DNA sequencing of the phagemid DNA was subsequently used.

12 samples were sequenced by dsDNA sequencing, but only 3 of these produced readable sequences. The reason for this was not known at this stage, but is explained later in section D.3.3. Of the three readable sequences, 2 were wildtype (threonine), and the other was the required mutant (alanine).

C.6. Cloning of the segment containing T90A into pNUT.

The pNUT vector with a full length lactoferrin insert (pNUT:hLf) was used for introduction of the mutation, as it is in the context of the whole molecule that any functional effects of this sidechain are likely to be felt. The 3D structure of human lactoferrin had showed a possible interaction between the side chain hydroxyl group of Thr90, and the carboxylate group at the C-terminal end (residue 691) of the full length lactoferrin molecule (B.F.Anderson, personal communication). This interaction is shown in Fig.C.11.

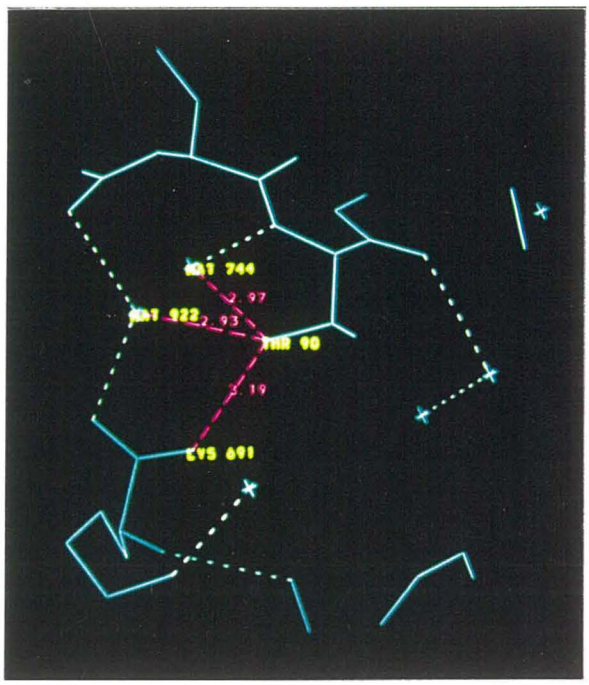


Fig. C.11. Possible interaction between the hydroxyl group of Thr90 and the carboxy group of the main chain at Lys 691 (terminus of the C-lobe main-chain).

C.6.1. Gel Purification of the vector and insert.

The pNUT:hLf vector was cut with Sma I and Apa I restriction enzymes and the fragments were run on a normal 1% agarose gel. A large 7.5 kb fragment corresponding to the vector without the 766 bp Sma I - Apa I fragment was isolated as in section B.2.10.

The pTZ:18U:Lf_N clone containing the T90A mutation was also grown up and the DNA was prepared by using the magic miniprep marketed by Promega. The DNA was cut with Sma I and Apa I restriction enzymes and the766 bp Sma I/Apa I fragment containing the T90A mutation was also isolated by gel electrophoresis (section B.2.10). Both the 7.5 kb pNUT:hLf vector (minus the Sma I/Apa I fragment), and the Sma I/Apa I fragment from pTZ:18U:Lf_N containing the T90A mutation, were quantified by comparison to known amounts of DNA from quantitation standards run in parallel lanes.

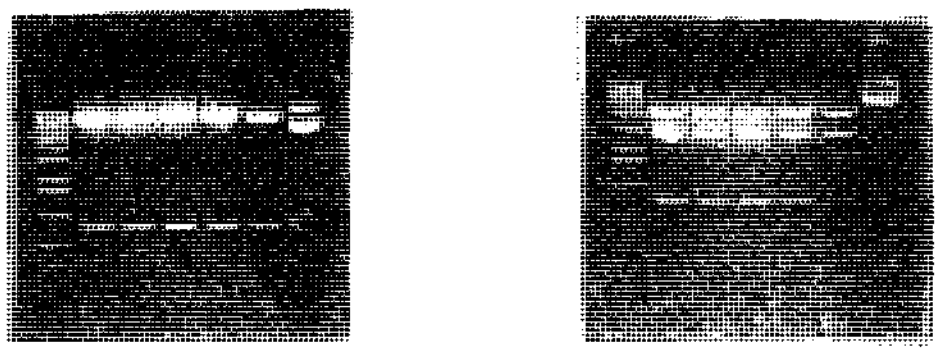
C.6.2. Ligation of the T90A fragment into pNUT:hLf.

The 766 bp T90A Sma I/Apa I fragment and the 7.5 kb pNUT:hLf fragment were ligated using a ratio of 3:1 insert to vector, at 12°C for 24 hr in a total volume of 15 µl. The reaction was electroporated into XL-1 cells and plated onto TET/AMP plates to provide a selection in which only cells with the vector present grow in the presence of these two antibiotics. A control reaction containing cut vector but without insert was also electroporated and plated.

The results showed 50 times as many colonies on the plate with the insert included, than on the plates with vector only (see Table.3). Four of these colonies from the vector + insert plate were grown up and magic minipreps carried out. DNA from these clones was cut with Kpn 1 and Eco RI restriction enzymes, or with Sma I and Apa I restriction enzymes and run on a gel to check for correct fragments (Fig. C.12).

Table. C.3. Number of colonies on LB plates after ligation.

Ligation	No of Colonies
Vector only	3
Vector + Insert	~ 150



Gel A. Gel B.

Fig. C.12. DNA gels of (Gel A) Sma I / Apa I restriction digest.
(Gel B) Kpn I / EcoR1 restriction digest.

DNA from all four clones showed the expected fragments indicating that correct ligation had occurred. DNA from clones 1-4 were then quantitated on a gel and prepared for sequencing.

C.6.3. Sequencing of pNUT:hLf (T90A) clones.

Because previous attempts to sequence this part of the lactoferrin gene had produced few readable lanes, the following tests were carried out.

1. A different template (previously used for successful sequencing), but using the CLD3 primer, was sequenced. This was to test the capacity of the primer.
2. The clones were sequenced using different primers for this region of the lactoferrin cDNA, to check the integrity of the template.

The following reactions were prepared;

(i) M13 control - to test the sequencing kit and the method.
(ii) Clone 2 with CLD4 primer - using another primer, previously proven.
(iii) Clone 3 with new CLD3 primer - using a new aliquot of CLD3.
(iv) Clone 4 with CLD13 primer - using another previously proven primer.
(v) Clean DNA hLf-5 with CLD4 primer - using proven DNA and proven primer. (a positive control).
(vi) Clean DNA hLf-5 with CLD3 primer - to test the primer using proven DNA.
(vii) M13 control - another positive control also used to define the radiograph.

Table. C.4. Results of sequencing the pNUT:hLf clones.

Primer	Template	Testing	Readable sequence
Universal primer	M13	Method	√
CLD4	Clone 2	Template 2	√
new CLD3	Clone 3	Template 3	√
CLD13	Clone 4	Template 4	√
CLD4	pNUT:hLf-5[a]	Primer +ve control	√
CLD3	pNUT:hLf-5[a]	Primer	X
Universal primer	M13	Flank the sequencing gel.	√

a = Clone made by Dr K. Stowell, previously used in DNA sequencing.

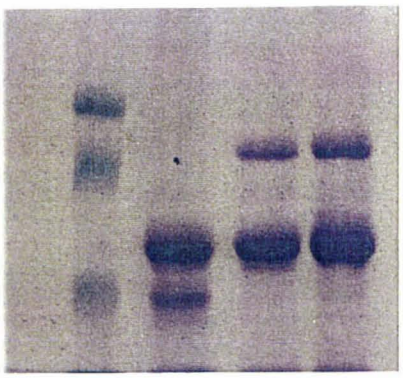
The results (Table.C.4) showed that all the reactions except lane 6 produced a good readable sequence. As this involved CLD3 with clean DNA, it suggested that this batch of CLD3 primer was at fault. This explains why sequencing of pTZ:18U:Lf_N mutants produced such a poor result (3/12 readable sequences), and possibly why attempts at sequencing the single-stranded DNA produced a smear. Lane 3, using a new batch of CLD3, produced a sequence which when analysed proved to have the required mutation (T90A).

This clone (3) was grown up and a maxiprep carried out. The dsDNA was sequenced from bases 66 to 829 (Apa I to Sma I) and found to have the correct sequence. This DNA was used to transfect BHK cells.

C.7. Expression of T90A in BHK cells.

The cells were transfected with 10 μg of DNA containing the T90A mutation, and were then selected with methotrexate as outlined earlier (sections C.3.1 and C.3.2).

Immunoprecipitates prepared from 100 μl, and 200 μl, of T90A media were run alongside an immunoprecipitate prepared from a culture known to be expressing Lf_N:P251V. This provided a positive control for the method of immunoprecipitation. The gel showed that the cells were expressing a protein immunologically related to lactoferrin and of about the expected molecular weight (Fig.C.13.).



1 2 3 4

Fig. C.13. Immunoprecipitate of (Lane 2) P251V acting as a control, and (Lanes 3 and 4) T90A showing expression of this protein in BHK cells. (Lane 1 is a ladder)

BHK cells from a newly confluent culture of this cell line were harvested and frozen under liquid nitrogen as described in section B.4.4.

C.8. Discussion

The method used to express the recombinant protein is expensive, as it involves mammalian tissue culture. In addition to this expense, cultured mammalian cells selected in methotrexate can become resistant to the inhibitory action of methotrexate due to defects in the transport system of the cell, or due to point mutations in the active site of the dihydrofolate reductase gene (Stryer, 1988). The use of a negative control when selecting the cells that have taken up the pNUT vector (no pNUT cDNA), will prevent wrong diagnosis due to this rare resistance of BHK cells to methotrexate.

One group (Ward *et al,* 1992) have recently expressed recombinant human lactoferrin in species of yeast. Ward *et al* (1992) have expressed lactoferrin in *Aspergillus oryzae* with expression levels up to 25 $mg.l^{-1}$. The recombinant lactoferrin is indistinguishable from human milk lactoferrin with respect to size, immunoreactivity, iron binding capacity (Ward *et al,* 1992), and to pH-dependent iron release properties (see protein results section). Lactoferrin produced in this way has recently been crystallized and the crystal structure determined, and a detailed comparison with native lactoferrin will soon be possible. If this proves successful, production of lactoferrin mutants in *Aspergillus oryzae* may prove to be a more cost-effective method.

D. PURIFICATION, DEGLYCOSYLATION AND CHARACTERISATION.

D.1. Purification of the proteins.

BHK cells that had been stably transfected with the cDNA encoding pNUT:Lf_N with the specific mutations (P251V, P251A, P251D, and P251G) or pNUT:hLf with the mutation T90A, were grown in roller bottles (section B.4.6). The medium was collected every 48 hours and stored at 4°C. Prior to loading the medium on to a CM-sephadex column for purification, the cell debris was removed by centrifugation at 8500 rpm for 60 minutes in a Sorvall GS3 rotor.

The lactoferrin half-length molecules (Lf_N) and full-length molecules (hLf) were purified on a CM-sephadex C50 column but different conditions proved to be required in the two cases. The half-length molecule was loaded on to the column in 10mM Hepes at pH 7.8 and 0.2M NaCl. Under the same conditions, the full-length molecule did not completely bind to the column and some of it came straight through. Increasing the buffer concentration in the media to 25 mM Hepes, and lowering the salt concentration in the media from 0.2 M to 0.12 M by dilution, increased the binding of this protein (hLf) to the column. In all cases (P251V, P251A, P251D, P251G, and T90A), the protein could be seen on the column as an orange band, indicating that the proteins had bound iron. With all proteins (half and full-length molecules), elution was acheived with a buffer concentration of 10 mM, and an increase in the salt concentration to 1.0 M NaCl in a gradient over a period of 60mins. The full-length molecule (T90A) eluted at about 0.4M NaCl, whereas the half-length molecules (P251V, P251A, P251D, and P251G) eluted at about 0.8M NaCl.

The fractions within the elution peak were collected and run on a denaturing sodium dodecyl sulphate (SDS) gel to check for purity (Fig. D.1). The gel in Fig. D.1 shows that the molecular weight of the protein in this case (P251A), was 40kDa, consistent with the glycosylated form of Lf_N. A minor band of around 37kDa was also present corresponding to the non-glycosylated form. The fractions were pooled together and typically the yield was 15mg of lactoferrin per litre of medium.

Fig D.1. SDS Gel showing a typical elution profile after elution off a CM-sephadex column (the lanes represent successive fractions of the protein peak).

D.2. Deglycosylation.

For iron release experiments, no further preparation of the protein was required since the glycosylated protein behaved in an equivalent manner to the non-glycosylated protein when used in such studies (Baker 1995). However, for proteins to be used in X-ray crystallographic analysis, Norris *et al* (1989) found that removal of the carbohydrate chains produced crystals that diffracted better.

The endoglycosidase preperation that was used to remove the carbohydrate chains from the lactoferrin was isolated from *Flavobacterium meningosepticum*. It actually contained two activities, Endo F (which cleaves after the first NAG residue) and PNGase F (which cleaves at the point of attachment, converting Asn to Asp). Previous studies have shown, however that this mixture is very effective for deglycosylating lactoferrin (Baker *et al*, 1994) and it was therefore used in this form, as supplied by Dr. G.E. Norris.

D.2.1. Deglycosylation with endoglycosidase.

In order to deglycosylate the lactoferrin, the endoglycosidase preparation was added to the lactoferrin and the mixture incubated for a period of time. The temperature and incubation time were decided by a series of 10-fold dilutions of the enzymes and incubation at 37°C or 25°C for 24 hrs. The reaction mixtures were then run on an SDS gel to compare treatments. Complete deglycosylation was seen by the complete disappearance of the 40kD band, and an increase in the 37kD band. This is shown in Fig. D.2

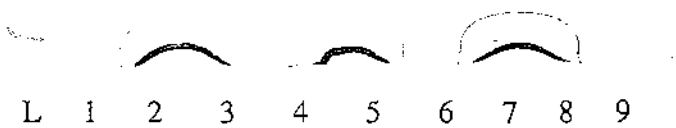


Fig D.2. SDS gel showing the P251A protein before (lanes 1 and 9) and after enzymatic deglycosylation at 37°C using 0.01 ul (lane 2), 0.1 ul (lane 3) or 1 ul (lane 4) of the endoglycosidase, or at 25°C using 0.01 ul (lane 8), 0.1 ul (lane 7),1 ul (lane 6), or 10 ul (lane 5) of the endoglycosidase. Deglycosylation can be seen by the disappearance of the 40 kDa band, and an increase in the 37 kDa band.

Comparing lane 3 and lane 5 of Fig D.2., it can be seen that incubating the lactoferrin at 37°C is 100 times as efficient as incubation at 25°C, and degradation of lactoferrin had not occurred. The conditions that resulted in complete deglycosylation (37°C, 24hrs, 1 µl endoglycosidase per 3 mg of lactoferrin), were used for deglycosylating the two proteins P251V and P251A prior to growing crystals. The endoglycosidase however, needed to be removed again before crystallisation.

The endoglycosidase was loaded on to an SP-sephadex column (20mM Tris/HCl-Cl pH 7.0, 0.2M NaCl) and fractions that came straight through the column were collected. Incubation with lactoferrin showed that these fractions retained the endoglycosidase activity of the original sample. This gave a further purification of the endoglycosidase, on a similar SP-sephadex column as is used for loading lactoferrin, and ensured that the endoglycosidase used for deglycosylation would be easily separated in subsequent steps. Lactoferrin was then incubated with the endoglycosidase for 24 hr before being separated from the latter on another SP-sephadex column as described above; the lactoferrin bound to the column, but the endoglycosidase came straight through. The lactoferrin was finally eluted using 1 M NaCl.

D.3. Characterisation of the Proteins.

The proteins were characterised by two methods, ie by X-ray crystallography, which will be covered in a later section, and by analysing the pH dependent release of iron.

D.3.1. pH dependent release of iron.

Iron bound to lactoferrin has a characteristic absorption band in the visible spectrum with a maximum at around 464nm for the full-length lactoferrin molecule, and around 450nm for the half-length molecule when measured at pH 8.0 in 0.2M NaCl. This absorption maximum does however vary with pH (as seen in the present studies) and with ionic strength (Kretchmar et al 1988).

As the pH of the lactoferrin is lowered in a suitable buffer (section B.5.8), the iron is released from the protein, and this is seen by a decrease in the absorbance at 450nm (or 464nm for the full-length molecule). An absorption maximum at 280nm is also found and is characteristic of all proteins with tyrosine or tryptophan residues present. By measuring the absorbance at both peaks (A280 and A450 (or A464)), a ratio of A450/A280 (or A464/A280) could be calculated and used to indicate the percentage of the protein with iron bound. An absorbance reading at around 700nm was also taken to allow subtraction of any background absorbance from the measurements.

The calculations used to establish the amount of protein with iron bound for the half-length lactoferrin molecule for example is as follows:-

$$\frac{A450 - A700}{A280 - A700} = A450/A280 \text{ (net)}$$

The higher the absorbance at 450 nm, the more iron is bound. A ratio of about 0.04 to 0.05 indicates fully iron-loaded lactoferrin (Day *et al*, 1992).

D.4. Pro 251 mutants

D.4.1. Absorbance maxima.

The Pro 251 mutants were initially characterised with respect to their UV/visible absorption spectra and the wavelength of visible absorption maximum. These were measured after dialysing the protein at pH 8.0 and 0.2 M NaCl. The five proteins all have a similar absorption maximum as shown in Table D.1.

Table D.1. Absorption maxima for Lf_N and the Pro 251 mutants

Protein	Abs^{max}
Lf_N (Day *et al*, 1992)	454 nm
Lf_N (this work)	451 nm
P251V	450 nm
P251A	451 nm
P251D	451 nm
P251G	452 nm

D.4.2. pH dependent iron release.

Previous studies on wild-type Lf_N found that the absorbance peak for iron was at 454nm, and that iron removal begun at about pH 6.0, and was essentially complete at pH 4.0 (Day *et al*, 1992).

The pH dependence of iron release from each of the four Pro 251 mutants was compared with the native half-length molecule. The extent of iron saturation, calculated as described earlier, was measured over the pH range 6.0 to 3.2.

The curves on the following pages show the pH dependence of iron release for the wild type half-length molecule (Figs. D.3 and D.4) together with the four mutant proteins (Figs. D.5 - D.8).

The corrected absorbance ratio was plotted as a function of pH for all the proteins. A line of best fit was calculated using the Marquardt Algorithm (Marquardt, 1963). The error of the measurements can be seen by the scatter of points around the best fit curve. The pKa for each protein was calculated, and is included in the inset table, together with the least square errors and the steepness of the curve. The latter is related to the number of protons involved in the release (n). The Lf_N data refers to the data collected by Dr C.Day (Day *et al*, 1992).

Table D.2. Summary of the iron release parameters for Lf_N and its Pro 251 mutants.

Protein	no of data points	pka	*Sum of squared errors	n
WT N lobe (Day, 1993)	22	4.88 ± 0.07	1.392 e-3	0.9 ± 0.1
Wild Type N lobe	34	4.34 ± 0.03	6.772 e-4	1.9 ± 0.2
P251V	30	4.55 ± 0.02	4.896 e-4	2.4 ± 0.1
P251A	37	4.24 ± 0.01	1.973 e-4	2.8 ± 0.3
P251D	43	4.51 ± 0.04	1.780 e-3	1.8 ± 0.1
P251G	23	4.70 ± 0.02	3.71 e-4	#(5)

*Errors are standard errors. # P251G was constrained to an n value of 5 in the least squares analysis.

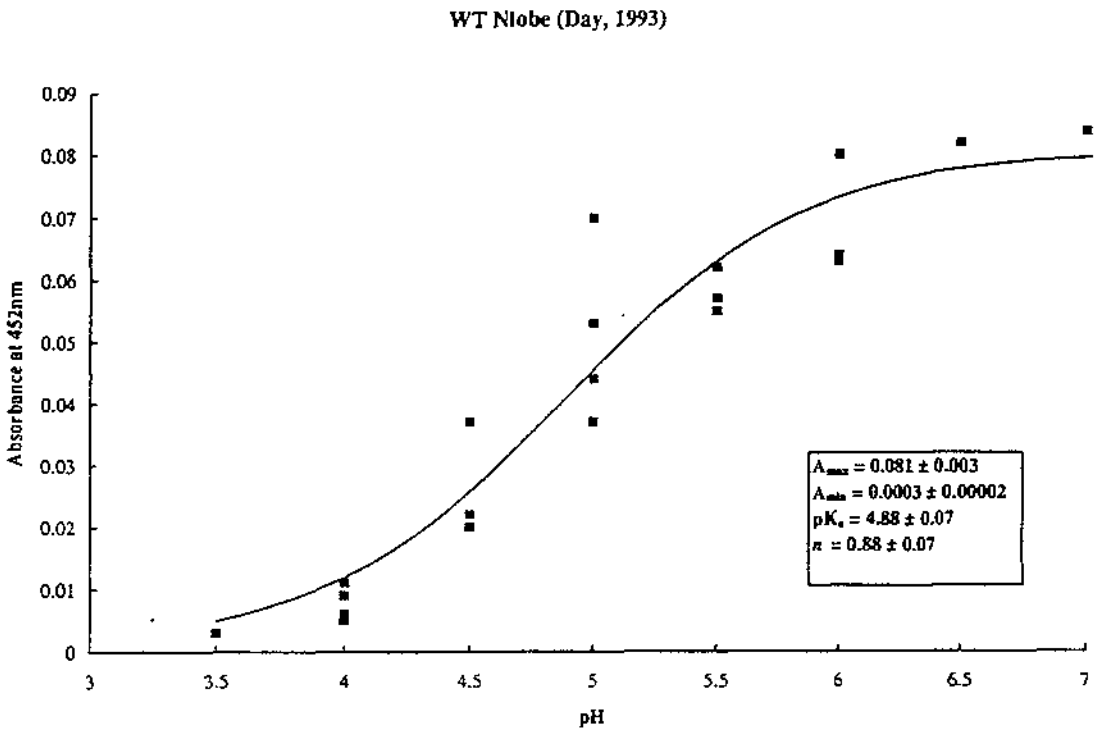


Fig. D.3. pH-dependant iron release curve for the N-lobe of lactoferrin as calculated by Day (1993). The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

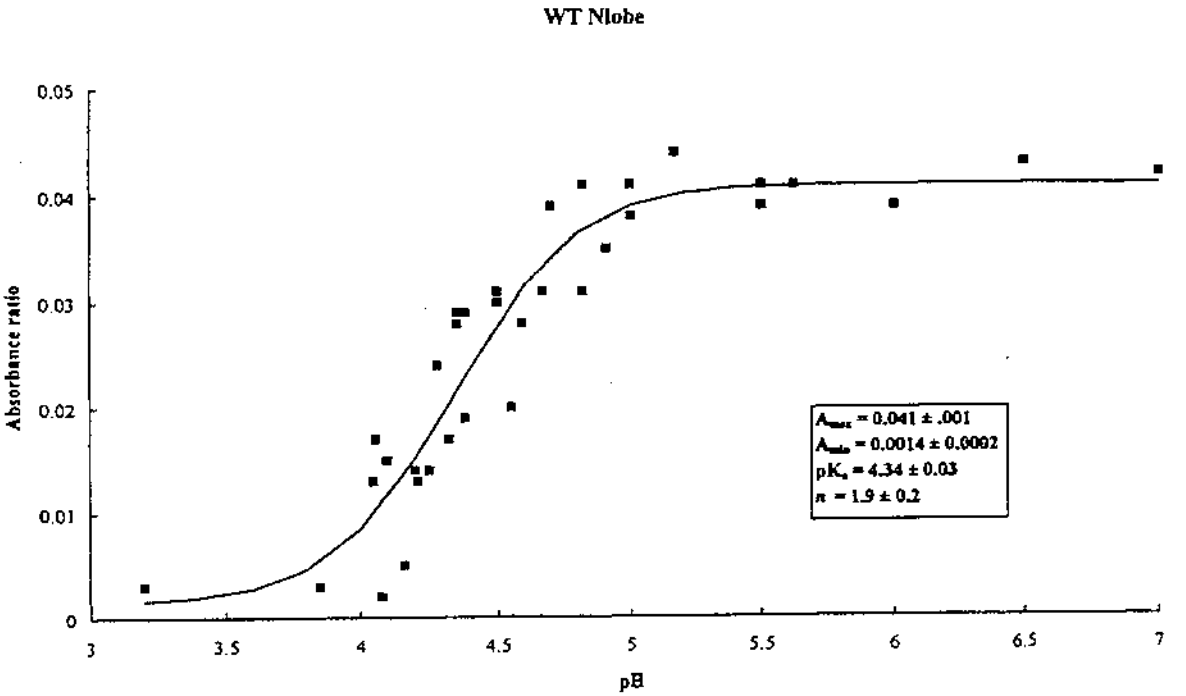


Fig. D.4. pH-dependant iron release curve for the N-lobe of lactoferrin as calculated in this study. The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

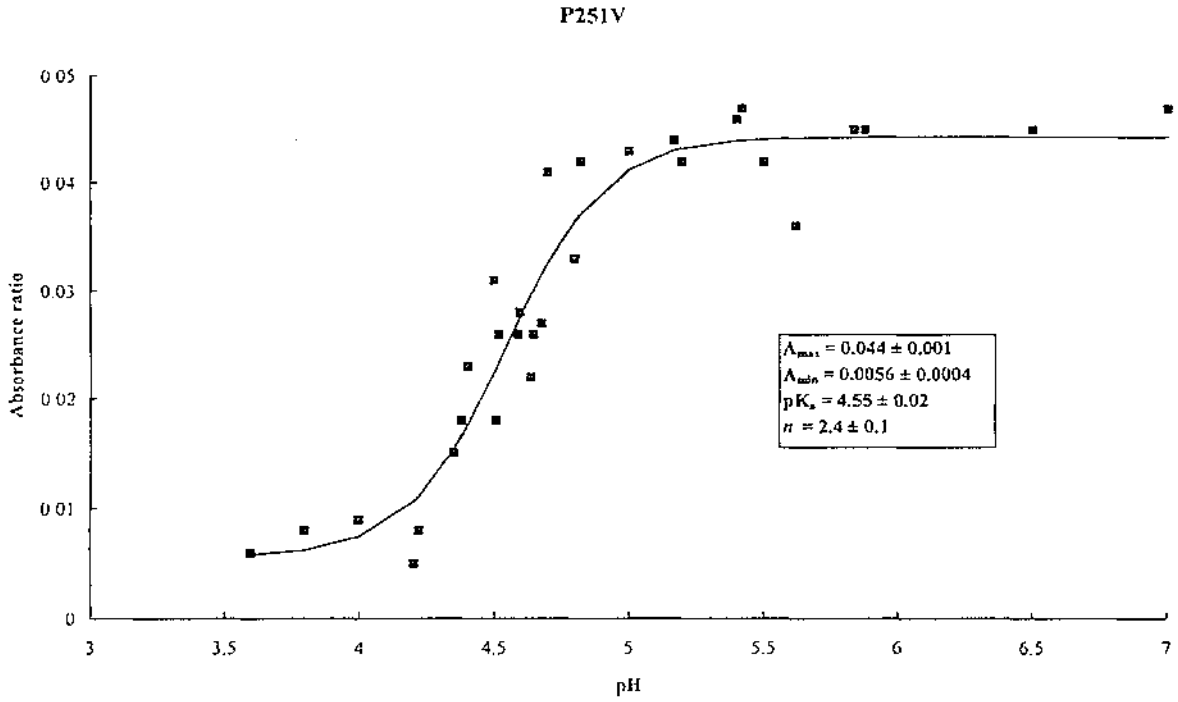


Fig. D.5. pH-dependant iron release curve for P251V as calculated in this study. The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

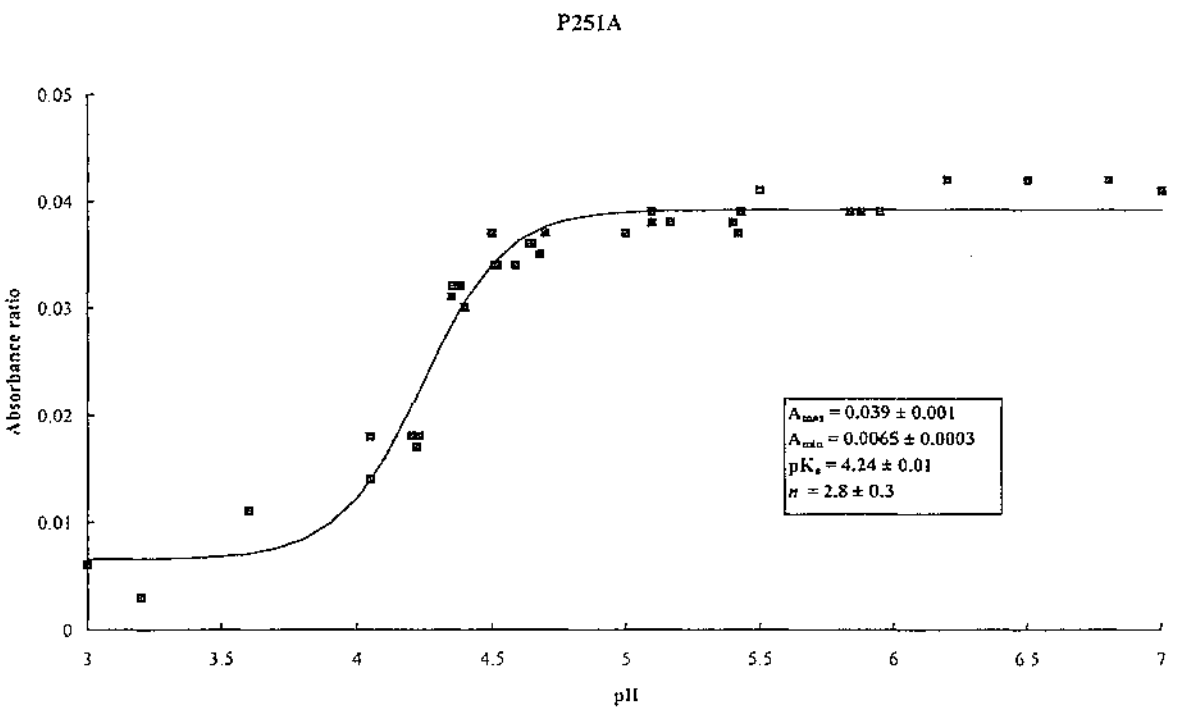


Fig. D.6. pH-dependant iron release curve for P251A as calculated in this study. The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

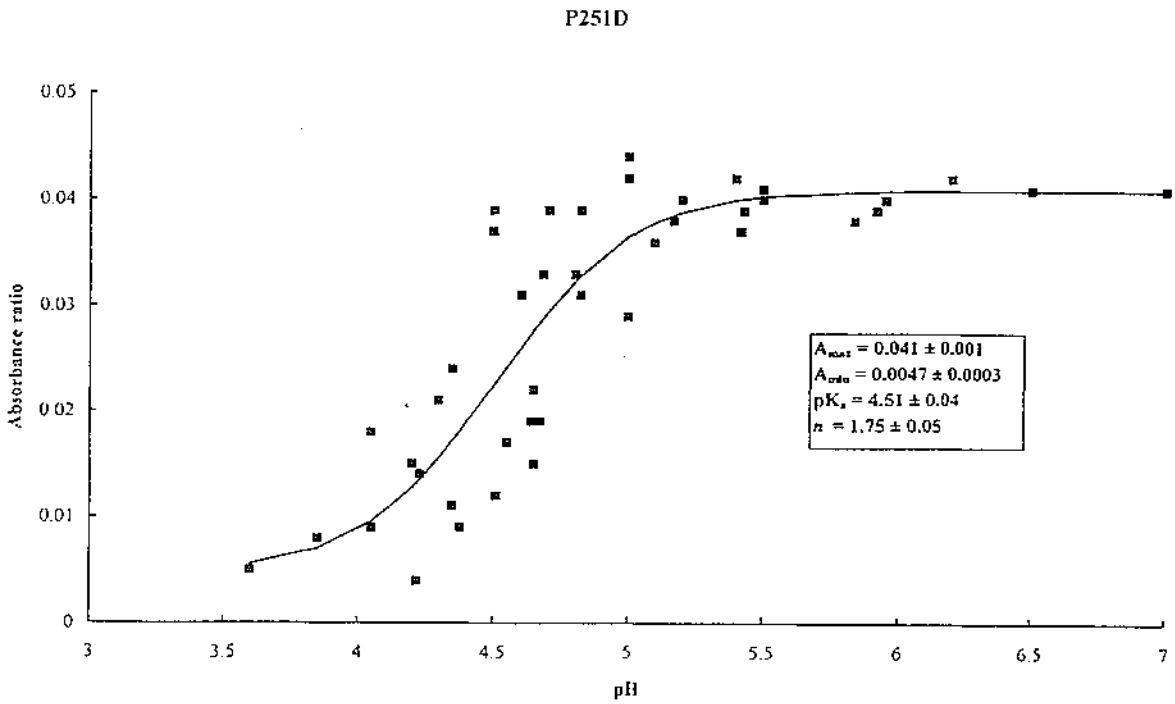


Fig. D.7. pH-dependant iron release curve for P251D as calculated in this study. The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

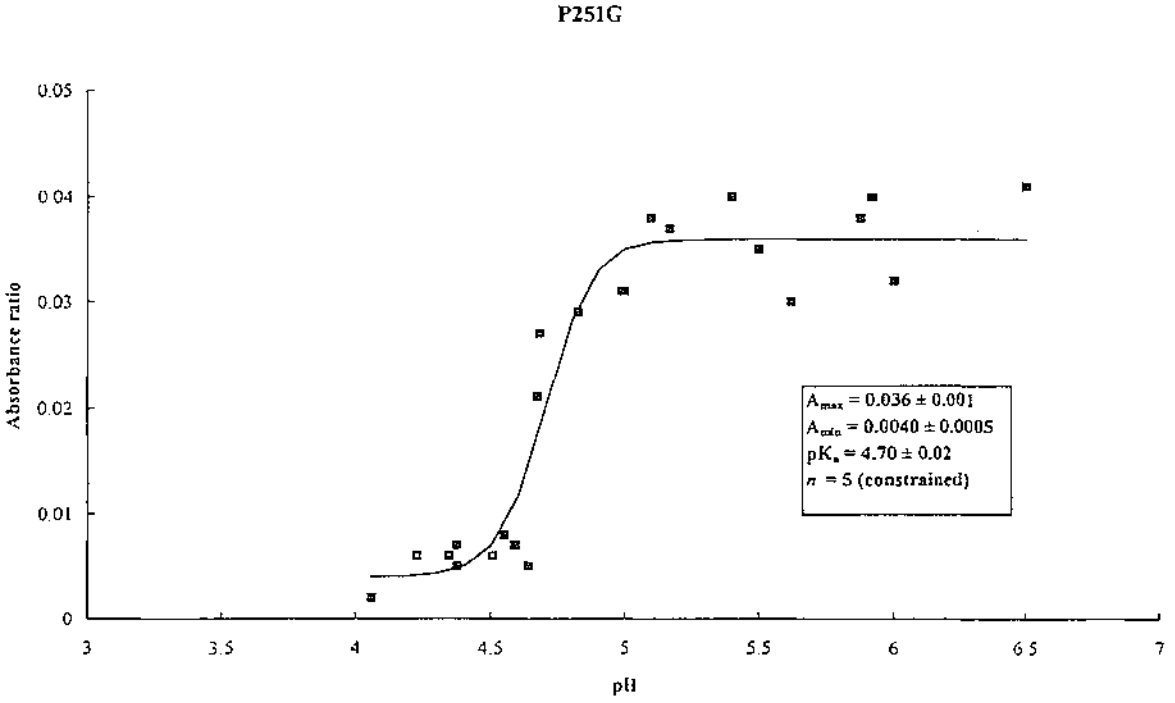


Fig. D.8. pH-dependant iron release curve for P251G as calculated in this study. The points represent the absorbance at 452 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

D.4.3. Discussion.

Some aspects of the method used to measure the pH dependence of iron release from lactoferrin are worth commenting on. Firstly, the low solubility constant of $Fe(OH)_3$ ($K_{sp} = 4 \times 10^{-38}$)means that even when iron release is minimal (eg at high pH) any iron that is released from lactoferrin reacts with water to form the insoluble $Fe(OH)_3$ complex. The equilibrium is thus shifted as iron is removed from solution. This means that the visible absorption spectrum should decrease even in the absence of any change in pH. This does not seem to occur significantly in practice however as the A452/A280 ratio of a solution of iron loaded lactoferrin does not decrease over a period of a few weeks, indicating that there has been no significant loss of iron bound to lactoferrin.

Secondly, time must be allowed for equilibrium to be reached at each pH during the pH titration. Previous iron release studies have used a time of either 24 hr (Stowell 1990) or 48 hr (Day, 1993). In this study, to be sure that equilibrium was reached, the solution was maintained for 48 hr at each pH. The validity of this was checked by equilibrating a sample for 24 hr, measuring the absorbance, then equilibrating for another 24 hr at the same pH before further absorbance measurements were made. The calculations at the two times (24 hr and 48 hr) suggested that the A452/A280 ratio for the 48 hr dialysis time was lower than for the 24 hr period. The decrease was not significant, however when compared with experimental error (seen by the scatter of points on the pH titration curve). A possible problem with this length of time (48 hr) however, is that during longer periods at low pH the reduced stability of lactoferrin may result in protein precipitation.

Thirdly, at low pH proteins tend to precipitate due to denaturation. During iron release titrations this causes problems especially if the protein that precipitates is predominantly in one form (iron-loaded or apo) as this would affect the ratio of iron bound/total protein irrespective of any release due to pH. Before every absorbance measurement the protein was filtered in an 0.2 μm spin filter and a colloidal material was removed. It is not known whether this colloidal material was protein, and/or $Fe(OH)_3$, and if protein, whether it was iron-saturated or iron-free, or both.

Fourthly, the molar absorptivity at 280 nm differs between iron-loaded (1.2) and apo lactoferrin (1.09), meaning that there is a decrease in the A280 absorbance when the protein loses iron. This should increase the final ratio by a decrease in the A280 absorbance. This was not allowed for, but it does not appear to have much effect on the overall pH curve as experiments with the wild type N-lobe protein using different plots (A452, A452-A700, and A452/A280, each graphed against pH) all produced a similar pH titration curve and pKa value.

The similarity in absorbance maxima and pKa values between the wild type N-lobe and the four mutant proteins of this study indicates that the pH of iron release is not dependent on the conservation of Pro 251. This point is more fully discussed in section F.

A feature of interest, however, is the shape of the pH titration curves for the five proteins here relative to the wild type N-lobe data of Dr C Day. Page 74 shows the two curves for the wild type N-lobe, one measured by C.Day, and the other measured in this study. In both cases the lactoferrin appears to have fully lost its iron at around pH 3.5 to 4.0. What is different is the *n* value between the graphs. The *n* value as outlined earlier refers to the slope of the graph and should be indicative of the number of protons involved in the mechanism of iron release.

In this work, the wild type N-lobe has an *n* value of around 2 and this is also seen in three of the four mutants. The exception is P251G which had to be constrained during the least squares fitting to a value of 5, (the least squares process was unstable in this case, and the value of *n* tended to increase to unrealistic values). In the earlier studies of C.Day the *n* value clearly is one. Constraining it to 2 gave a poorer fit and a greater sum of squared errors indicating that the true value is indeed 1. This points to a difference in the methods used. The conditions (buffer concentration, pH, titration time) were all the same. The calculations were different in that Dr C.Day used the A452 data, while in this study, the A280 and A700 measurements were taken into account as well. When only the A452 data was used on the present wild type data however, it did not significantly alter the shape of the graph, and the *n* value did not change at all.

This difference has not been resolved in this study. Further work with the N-lobe of lactoferrin may resolve this difference.

D.5. Thr 90 mutant (T90A).

D.5.1. Absorbance maxima.

Because this mutation was cloned into the full-length lactoferrin cDNA, the spectral measurements and pH titrations were measured alongside and compared with the native full-length molecule. The visible absorption maxima, measured at pH 8.0 in 0.2 M NaCl, are tabulated below.

Table D.3. Absorption maxima for the full-length lactoferrin molecules.

Protein	Abs^{max}
Native lactoferrin	464 nm
*Asp lactoferrin	466 nm
T90A	466 nm

* Asp lactoferrin refers to the recombinant form of human lactoferrin grown up in *Aspergillus oryzae*.

The similarity in these results shows that changing Thr 90 to Ala does not alter the iron site, at least as far as the Fe-Tyr binding is concerned.

D.5.2. pH dependent iron release.

Previous work has shown that for full-length human lactoferrin iron removal begins at about pH 4.0 - 3.5, and is complete at about pH 2.5 (Stowell, 1990).

The iron release measurements in this study were made from pH 6.0 to pH 4.0, to establish a base line prior to iron release, and from pH 4.0 to pH 2.0 for the titration of the iron release.

A new recombinant form of wild type lactoferrin, expressed in *Aspergillus oryzae* (Ward *et al*, 1992) was also used in these experiments, to see if the altered glycosylation resulting from this different expression system affected the iron release properties of the molecule.

The pH titration curves on the following page refer to the work done in this study on recombinant human lactoferrin (Asp hLf) and on its Thr 90 mutant (T90A).

The results were graphed and are shown on the following pages. These have been summarised in Table D.4.

Table D.4. Summary of the iron release parameters for the two full length molecules of this study.

Protein	No of data points	pKa	*Sum of sqared errors	*n*
asp hLf	18	2.85 ± 0.01	$1.613\ e^{-4}$	4.7 ± 0.3
T90A	32	2.73 ± 0.01	$5.425\ e^{-4}$	#(5)

*Errors are standard errors. # T90A was constrained to an *n* value of 5 in the least squares analysis.

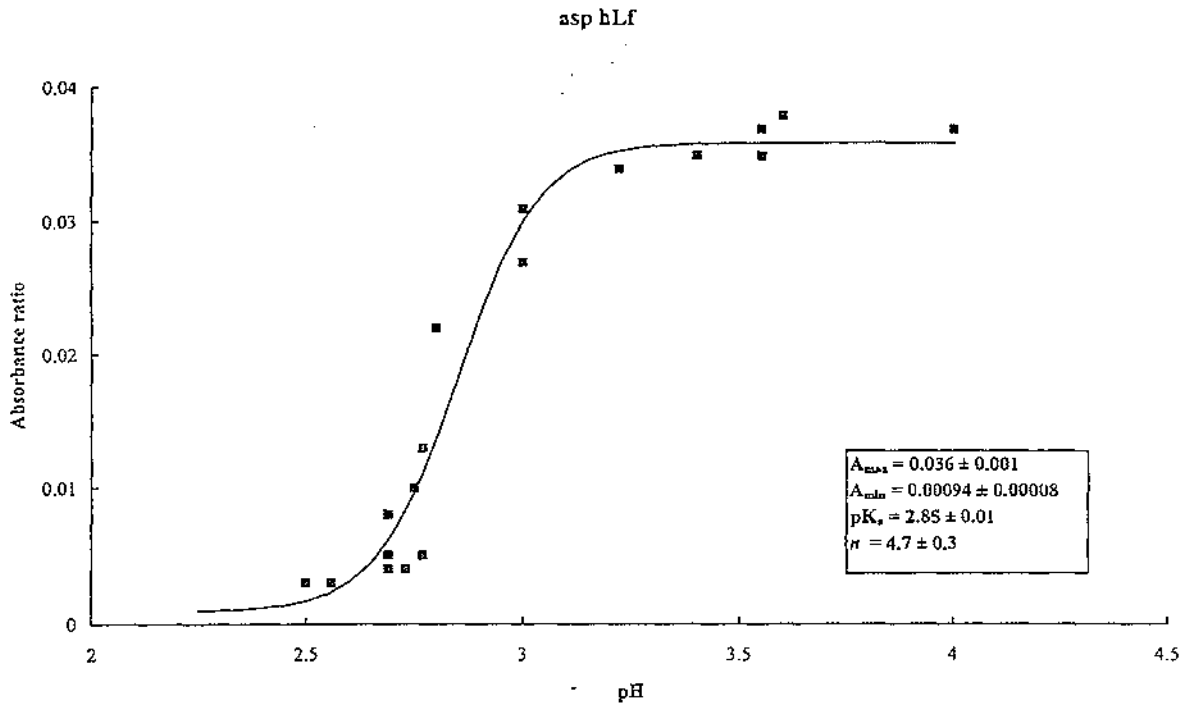


Fig. D.9. pH-dependant iron release curve for asp hLf as calculated in this study. The points represent the absorbance at 464 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

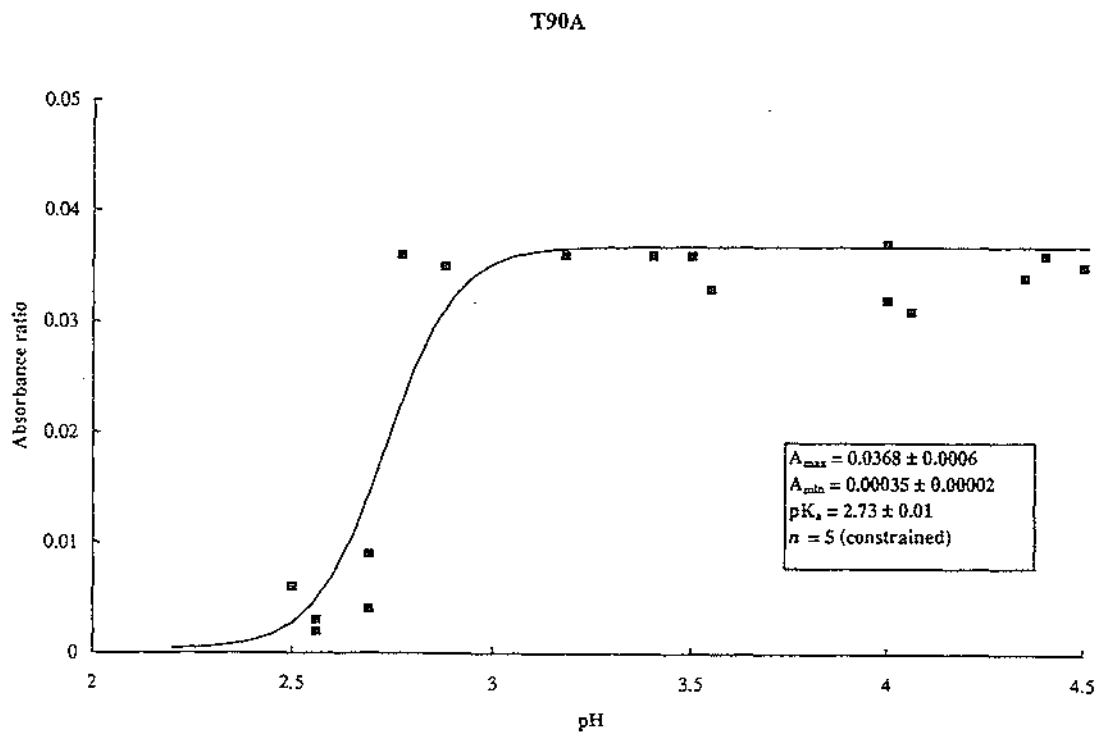


Fig. D.10. pH-dependant iron release curve for T90A as calculated in this study. The points represent the absorbance at 464 nm for various pH. A line of best fit was calculated using the Marquardt Algorithm.

D.5.3. Discussion.

The results for the full-length proteins (Asp hLf and T90A) show that mutating Thr 90 in human lactoferrin does not affect the absorption maxima. Comparing the iron release curves (Figs. D.9 and D.10), their pKa values are similar (2.73 and 2.85) and in each case iron removal is complete at about pH 2.5.

The other point worth noting is that the recombinant form of human lactoferrin grown in *Aspergillus oryzae* behaves in a similar way to the recombinant lactoferrin grown in baby hamster kidney cells (Stowell 1992) with similar pKa values (Asp hLf 2.85, BHK expressed hLf ~ 3.0) and both proteins having lost the Fe^{3+} fully by pH 2.5.

E. CRYSTALLOGRAPHY.

The first lactoferrin structure determined by X-ray crystallography was that of iron-saturated human lactoferrin (Fe_2Lf) which was solved by multiple isomorphous replacement (Anderson *et al* 1987). Subsequent structures of lactoferrin solved include bovine lactoferrin (Moore *et al* 1996), human apolactoferrin (Anderson *et al* 1989; Faber *et al* 1995), and copper saturated lactoferrin, Cu_2Lf (Smith *et al* 1992), which were all solved by molecular replacement using the structure of Fe_2Lf as the starting model.

The steps involved in the present work were crystal growth, data collection, refinement, and analysis of the final structure.

E.1. Crystal growth.

Various methods to grow crystals have been used over the years, including the hanging drop, sitting drop, and microdialysis methods (McPherson, 1982). As previous crystallizations of lactoferrin have all depended on the use of dialysis, the method of microdialysis was used in this work. Microdialysis involves placing the protein solution into a microdialysis button and covering with dialysis membrane (Fig E.1). Protein molecules cannot pass through the membrane but small molecules such as salt and buffer can. This button is placed in a larger volume of an external solution of different ionic strength, so that the solution inside the button equilibrates with the solution outside the button by dialysis through the membrane. This may lead to crystallization of the protein.

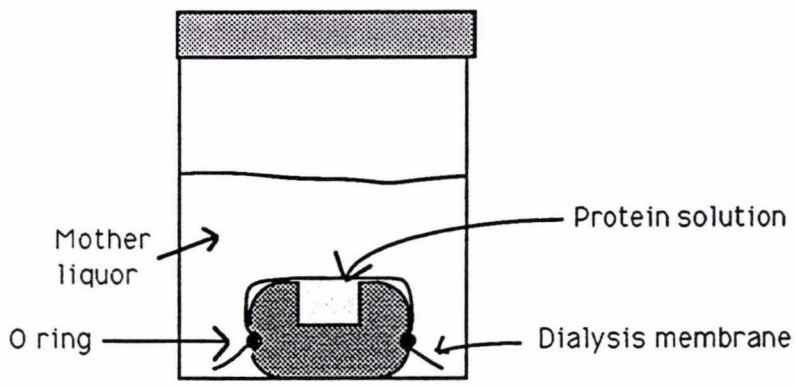


Fig. E.1. Schematic diagram of a microdialyis button set up for crystallization.

The protein was prepared for crystallization by equilibrating it against starting buffer (10mM Tris/HCl pH 8.0, 0.5M NaCl) that had been filtered in a 2 μm filter and autoclaved. The protein was then concentrated to a final concentration of 80mg ml^{-1} in a centricon 10 concentrator.

Table E.1. summarises the conditions used previously in crystallising other lactoferrin species. In all cases, successful crystallizations employed an external solution of low ionic strength, with small amounts of added alcohol (typically ethanol, methanol or isopropanol at levels of 10 - 12%).

Table. E.1. Previous conditions used for crystallization of various lactoferrins.

Protein	Concentration	Crystallization conditions
Native human Fe_2Lf (Baker and Rumball, 1977)	40 - 80 mg ml^{-1}	0.01 M sodium phosphate pH 7.8 10% v/v ethanol, 4°C.
Native human apoLf (Norris *et al.* 1989)	20 - 30 mg ml^{-1}	0.01 M sodium phosphate pH 7.8 10% v/v methanol, 4°C.
Deglycos human apoLf (Norris *et al* 1989)	12 mg ml^{-1}	0.05 M Tris/HCl, pH 7.8, 4°C 5% v/v MPD, 5% v/v ethanol
Native bovine Lf (G.Norris, *pers. comm*)	100 - 150 mg ml^{-1}	0.01 M Tris/HCl, pH 8.25, 4°C 10% IPA or 10% ethanol.
N lobe human Lf (Day *et al* 1992)	80 mg ml^{-1}	0.01 M Tris/HCl, pH 8.1, 4°C 10 - 12% IPA, 4 mM NaCl.

MPD (2-methyl-2,4-pentandiol), IPA (isopropanol), Deglycos (enzymatically deglycosylated).

As the two proteins to be crystallised (P251A and P251V) were both half-length molecules, the crystallization conditions previously used for the half-length wild-type molecule (Lf_N) were initially tried. These conditions (10 mM Tris/HCl pH 8.0; 4 mM NaCl; 12% IPA), when used for the mutants, produced large numbers of small needles within one week, but no crystals of size or quality suitable for X-ray diffraction.

Under these conditions, IPA is used to precipitate the protein, and as these conditions produced crystals quickly, a possible solution to obtaining larger crystals would be to slow the process down by decreasing the IPA content.

Two trials with 10% IPA (10 mM Tris/HCl pH 8.0; 4 mM NaCl; 10% IPA) and 8% IPA (10 mM Tris/HCl pH 8.0; 4 mM NaCl; 8% IPA) were set up, both producing crystals of diffraction quality. What was significant was that as the IPA concentration

was lowered (12% to 10% to 8%), the number of crystals decreased, but the size of them increased.

From the 8% IPA crystallization, two crystals were obtained for each protein (P251A and P251V).

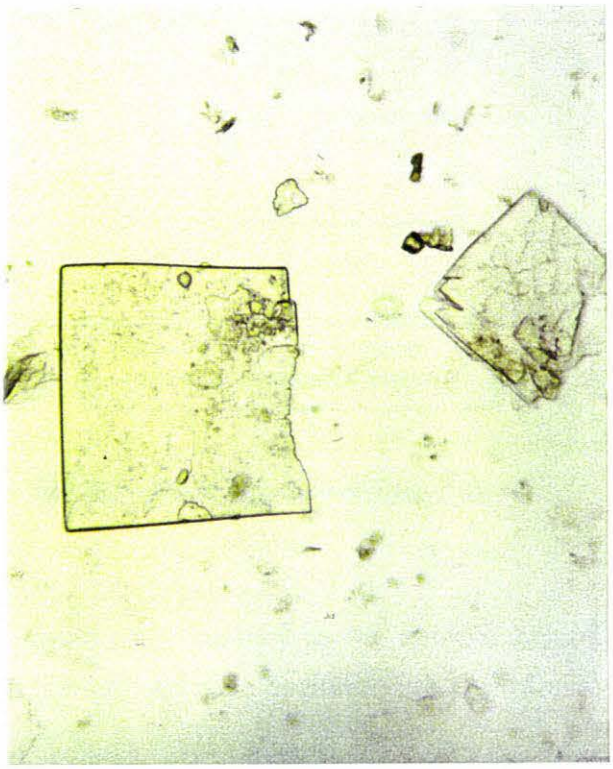


Fig. E.2. Photograph of the P251A crystal.

The general shape of the crystals (Fig. E.2) suggested that they might be tetragonal and therefore isomorphous with one of the forms of the native half-length molecule analysed by Dr Catherine Day (Day, 1993). Subsequent work proved this to be the case.

E.2. Data collection and processing

The best crystal of the P251A mutant was a thin square plate with dimensions 0.4 mm x 0.4 mm, and 0.04 mm thick. This crystal was used for collection of X-ray diffraction data. Data collection followed the flow diagram below.

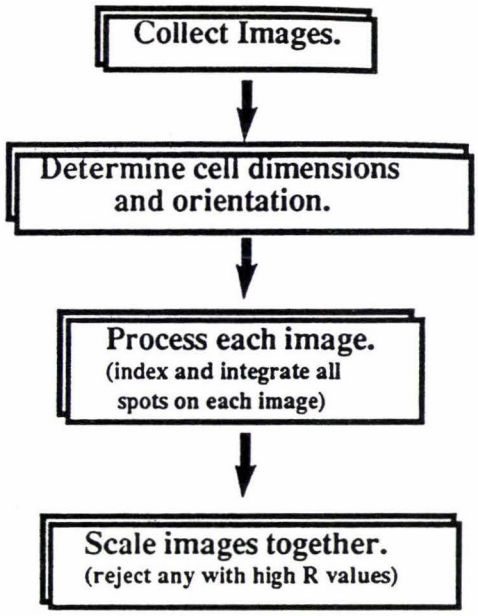


Fig. E.3. Flow diagram of data collection.

The crystal was mounted in a thin-walled capillary tube and optically aligned with the c-axis approximately parallel to the spindle axis. Diffraction data were collected on a Rigaku R-axis IIc area detector with the following instrument settings (Table E.2).

Table. E.2. Instrument settings and conditions for data collection.

Instrument	Rigaku R-axis IIc
Radiation	Cu-Kα
Wavelength	1.5418Å
Power	50kV; 100mA
Detector offset	-8.995°
Crystal to film distance	130mm
Phi angle of stills	45°, 90°
Δ phi range	1.5°
Exposure time per frame	90mins
Number of frames	30

Two still photographs were taken and showed a tetragonal cell with a=b=58.5 Å and c= 218.5 Å. The diffraction data was seen to extend to at least 1.9 Å. The long c-axis meant that the detector had to be set well back. With the detector well back, in order to get high resolution data, the detector angle needed to be a non-zero value. For this crystal an angle of -9.0° was used.

The crystal was mounted approximately around the c-axis such that only 45° of data needed to be collected. 30 oscillation images of 1.5° were collected.

The data were processed using the software provided with the R-axis detector. Initial cell dimensions were taken from the still photographs and these values were refined from the oscillation photographs to give the following cell dimensions; a=b= 58.5; c= 218.56 Å; $\alpha=\beta=\gamma=90^{o}$. The crystal system was tetragonal, with space group either $p4_12_12$ or $p4_32_12$.

The data processing involved the following steps:-

(i) Prediction and integration of the spots on each of the 30 oscillation images.

(ii) Calculation of preliminary scales between the images.

(iii) Post refinement to improve the missetting angles for each image to account for any crystal movement that may have occurred during data collection.

(iv) Rescaling and merging of the data. Partial reflections were included, but the last 3 oscillation images, which had a high R_{sym} (~15 %), were discarded. The R_{merge} for all data was 0.0782 (R_{merge} for full reflections = 0.075, R_{merge} for partial reflections = 0.089) for data between 20.0 and 1.9 Å. There was no sigma cut off.

(v) After applying Lorenz and polarisation corrections and converting to structure amplitudes, a final data set comprising 19480 reflections was obtained.

Table E.3. shows the results of the data processing.

			Data per resolution shell			Cumulative data			
Resolution	<FSQ>	<SIG>	Measured	Theoretical	Completeness (%)	Measured	Theoretical	Completeness (%)	FSQ/SIG
15.00	523.67	34.15	75	93	80.6	75	93	80.6	15.22
10.00	577.96	35.43	181	183	98.9	256	276	92.8	15.70
7.00	333.88	20.39	463	459	100.9	719	735	97.8	15.09
3.50	252.22	17.21	4444	4492	98.9	5163	5227	98.8	13.05
3.00	100.30	10.77	2681	2924	91.7	7844	8151	96.2	7.92
2.80	48.69	8.17	1397	1801	77.6	9241	9952	92.9	5.32
2.50	31.59	6.83	2757	3874	71.2	11998	13826	86.8	4.16
2.20	18.25	5.88	3662	6235	58.7	15660	20061	78.1	2.95
2.00	11.12	5.68	2933	6438	45.6	18593	26499	70.2	1.96
1.90	6.31	5.20	1247	4290	29.1	19840	30789	64.4	1.25

Table E.3. Summary of the details of data processing.

E.2.1. Assessment of the data

Once the data had been collected and processed, it was assessed by plotting it as a function of resolution using different sigma cut off levels. These plots are shown in Fig. E.4.

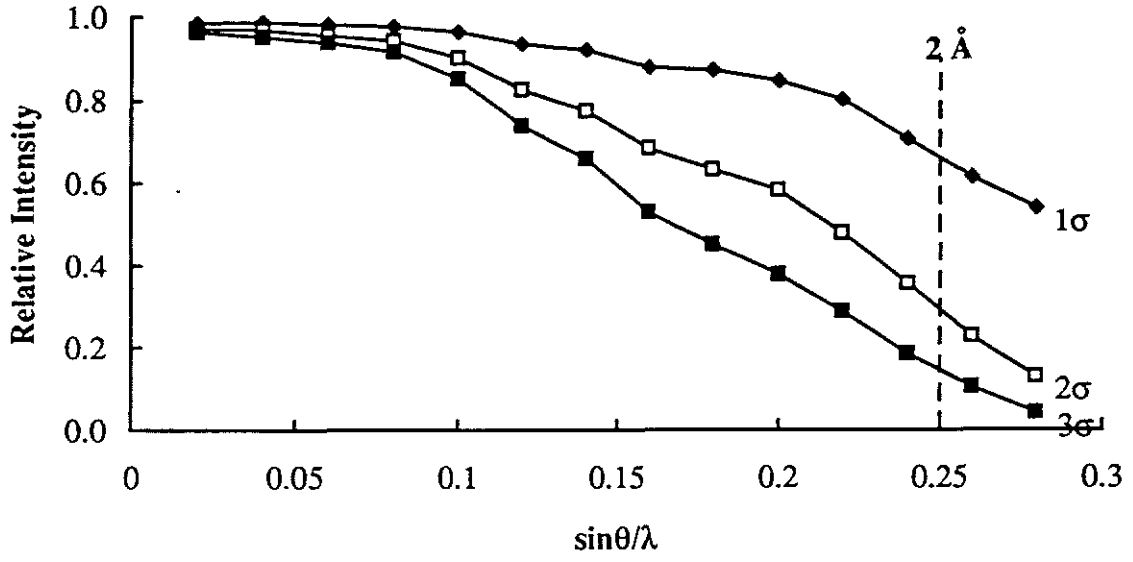


Fig. E.4. Plots showing the fraction of data with intensities greater than 1σ, 2σ and 3σ as a function of resolution.

Both Table E.3. and Fig. E.4. show that the data were weak at high resolution. In the resolution range 2.0 - 1.9 Å the mean I/σ ratio is only 1.25. Moreover, the completeness is also low in this shell (29%) probably as a result of the data collection geometry; at a crystal-to-detector distance of 130mm not all the 1.9 Å data is accessible, even with θ set at a non-zero angle.

Although data from the 1.9 - 2.0 Å shell were used during much of the refinement, the effective resolution is only 2.0 Å and the final refinement included only the data to 2.0 Å resolution. At this resolution, the I/σ ratio is 1.96, and the outermost shell is 50% complete.

E.3. Structure determination and refinement.

E.3.1. Introduction.

The structure amplitudes can be obtained from the measured intensities, but the phases cannot. The latter therefore need to be determined, and would normally be derived by MIR or molecular replacement methods. In the present case, however, the wild-type half length molecule crystallises in an isomorphous cell ($p4_12_12$; a = b = 58.35 Å, c= 215.2 Å; $\alpha = \beta = \gamma = 90.0^o$) and its structure has been solved. Since the wild-type molecule differs from the mutant only at one residue (Pro 251), the atomic positions for the former could be used as a starting model for the refinement of the P251A structure, ie the initial phases could come from the known wild-type structure.

The aim of refinement is to minimise the difference between the observed and calculated structure amplitudes, ie. minimise |Fobs| - |Fcalc| by refining the atomic positions (x,y,z), B values (thermal parameters), and occupancies. Restraints on bond lengths, bond angles, planar groups, and distances between non-bonded atoms are used to aid the refinement process. In this study, the restrained least squares refinement programme TNT was used (Tronrud *et al* 1987).

Throughout the process each round of refinement was followed by checks on bond lengths, bond angles, and contacts, using the programme PROCHECK (Laskowski *et al* 1993). Electron density maps were then calculated and in conjunction with the analysis by PROCHECK, rebuilding was carried out with the aid of the program FRODO (Jones, 1978). Two types of maps were calculated, ie with coefficients 2|Fo| - |Fc|, and |Fo| - |Fc| respectively. The maps were calculated using the programmes from the TNT package (Tronrud *et al* 1987), and converted into a form readable by the program FRODO.

Contour levels corresponding to 1σ and 0.6σ of the 2|Fo| - |Fc| map, and 2σ and -2σ of the Fo-Fc map were used to make adjustments in the early stages of refinement, while 2σ (2|Fo| - |Fc|) and 3σ or -2σ (|Fo| - |Fc|) were used in the latter stages of refinement.

E.3.2. Refinement process.

The progress of the refinement is summarised in Table E.4. Initially a round of refinement was carried out in which the molecule was treated as one rigid body. This was to ensure that the model was correctly oriented in the unit cell. The Fe^{3+} and CO_3^{2-} ions, and all solvent molecules were omitted from this starting model, and the side chain of residue 251 was also omitted.. Three cycles of refinement saw the R-factor drop from 33.7% to 32.9%.

The next round treated the molecule as the two domains in order to determine whether the relative orientations of the two domains differed from the wild-type molecule. Once again the Pro 251 side chain was omitted, and there were no iron or carbonate or solvent molecules present. In both rounds, the resolution range was from 10.0 to 2.5 Å. In this second case, the R-factor dropped only marginally from 32.9% to 32.8% indicating that the molecule was in essentially the same conformation as in the wild-type structure.

With the next round of refinement (R1) the resolution range was increased to include data between 20.0 and 1.9 Å. The Fe^{3+} and CO_3^{2-} ions, solvent molecules, and residue 251 side chain were still omitted, as were the iron ligands to prevent bias of these regions. The various residues not complete in the starting model were also left out at this stage. After six rounds of refinement, the R-factor decreased to 26.2%.

An electron density map now showed that the density of the binding site (Fe^{3+}, CO_3^{2-}, and ligands) was clear and very well defined, particularly the carbonate anion. The density around residue 251 identified this residue as an alanine and not a proline. Model building at this stage involved placing the alanine 251 side chain into density from the 2|Fo| - |Fc| map and repositioning any side chains not currently in the density. The ligands, Fe^{3+} and CO_3^{2-} ions were also put in. This model was refined (R2) for 12 cycles to give a final R-factor of 22.6%.

Side chains missing from the original model were added where density could be seen in the 2|Fo| - |Fc| map. Residues 313 to 320, were also added at the C-terminus. The B values had been fixed at 25 $Å^2$ prior to refinement. These were now allowed to vary, and 40 cycles of refinement (R3) saw the R-factor decrease from 26.1% to 21.8%.

At this stage, 2|Fo| - |Fc| and |Fo| - |Fc| maps were again calculated. From the density of the |Fo| - |Fc| map, 68 water molecules were added where density could be seen as a sphere when contoured at 2σ. Asparagine 137 (the site deglycosylated by PNGase) was changed to an aspartate residue. The B values were constrained to lie between 10.0 and 60.0 $Å^2$, and a further 18 cycles of refinement (R4) saw the R-factor drop from 29.7% to 20.1%.

The fifth round of refinement was carried out with 104 water molecules present. The geometry terms were tightened, and 18 cycles of refinement (R5) decreased the R-factor further to 19.6%

A final check of the model was carried out using 2|Fo| - |Fc| and |Fo| - |Fc| electron density maps. Little change to the protein structure was necessary. The solvent model was thoroughly examined, and only those water molecules which had density at the 2σ level in the 2|Fo| - |Fc| map and density at the 3σ level in the |Fo| - |Fc| map were retained. This led to the removal of 36 water molecules and the addition of 36 giving a total of 104 water molecules in the final model. The B-values were allowed to lie between 10.0 and 80.0 $Å^2$ during the final 30 refinement cycles (R6) which also imposed higher weighting on the geometry terms. The final R-factor was 18.6%. Full statistics of the final refined model are given in Table E.5.

Table E.4. Summary of the refinement process.

Round of Refinement	Number of cycles.	Resolution limits	R-factor (%)	rms bond lengths (Å)	rms bond angles (°)
Rigid body 1	3	10.0 - 2.5	33.7 - 32.9	0.021	2.931
Rigid body 2	3	10.0 - 2.5	32.9 - 32.8	0.021	2.933
R1	6	20.0 - 1.9	33.6 - 26.2	0.021 - 0.022	2.932 - 3.187
R2	12	20.0 - 1.9	28.6 - 22.6	0.042 - 0.020	3.731 - 1.457
R3	40	20.0 - 1.9	26.1 - 21.8	0.093 - 0.012	3.956 - 1.668
R4	18	20.0 - 1.9	29.7 - 20.1	0.024 - 0.017	2.623 - 2.061
R5	18	20.0 - 1.9	20.1 - 19.6	0.017 - 0.013	1.576 - 1.563
R6	30	20.0 - 2.0	22.4 - 18.6	0.017 - 0.013	1.829 - 1.481

E.4. Accuracy of the structure.

The structure was determined and refined using data between 20.0 and 2.0 Å resolution.and has a final R-factor of 18.6 %. The final model includes residues 4 - 320 including 2393 protein atoms, 1 ferric ion, 1 carbonate ion, and 104 solvent molecules

Table E.5. Geometry values for the final P251A structure.

Final Refinement Details	
Resolution data	20.0 - 2.0 Å
No of reflections	19840
R-factor	18.6 %
RMS deviation from standard values	
Bond lengths (Å)	0.013
Bond angles (deg)	1.481
Torsion angles (deg)	17.956
Trigonal planes (Å)	0.012
General planes (Å)	0.015
Bad contacts (Å)	0.298
B-value correlations ($Å^2$)	4.365

E.4.1. Agreement of the model with the data.

E.4.1.1. Luzzati plot.

The maximum average error in the model is estimated from the variation in R-factor with resolution as described by Luzzati (1952). The Luzzati plot is shown in Fig. E.6. with error lines indicated at 0.20 Å, 0.25 Å, 0.30 Å, and 0.35 Å. Disregarding the outermost shell, the average coordinate error in this structure is approximately 0.25 Å.

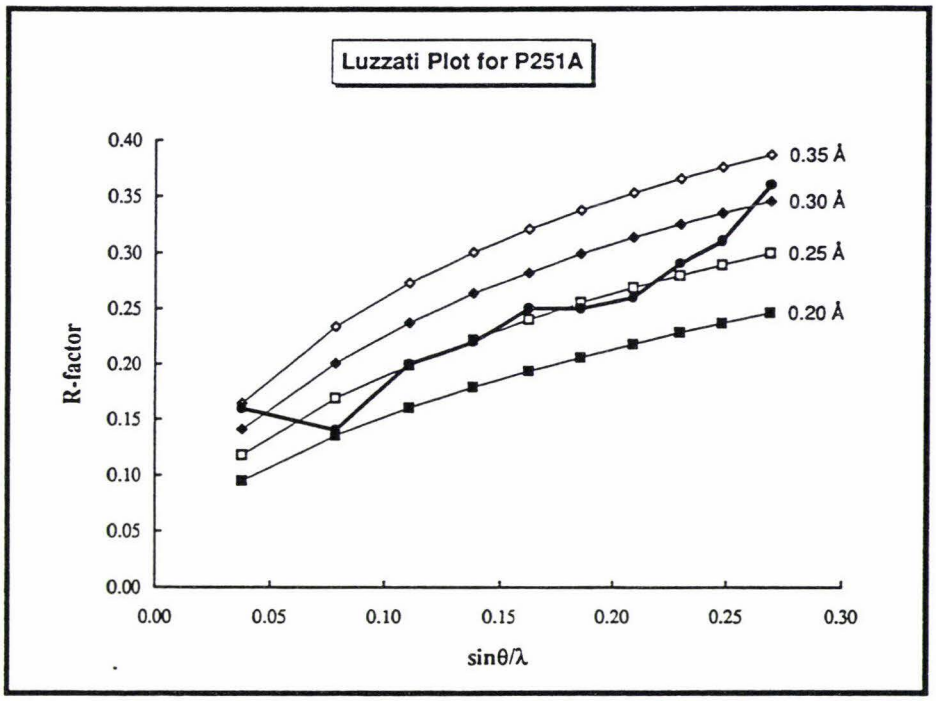


Fig. E.5. Luzzati plot for the P251A data.

E.4.1.2. Real space correlation coefficient.

The real space correlation coefficient (RSCC) is a measure of the fit of the coordinates to the final 2Fo-Fc electron density map. The RSCC for the individual residues was calculated using the real space fit part of the map building programme O (Jones et al 1991).

Fig. E.6. shows that in general, the correlation is high with most of the residues having correlation coefficients greater than 80%. The overall average coefficient for the model is 82.7%, while taking out the residues around 141 gave a value of 83.5%. The regions with notably low coefficients include the N-terminal region (Arg 4); Leu 74; prolines 141 and 142; and Ile 145. The absence of good real space correlation coefficents for the region around Pro142 correlates with large B-values found for this region and is discussed in the section on B-values (section E.4.3). Well defined regions include the iron and carbonate binding site, with alanine 251 also having a good correlation of 85.8%.

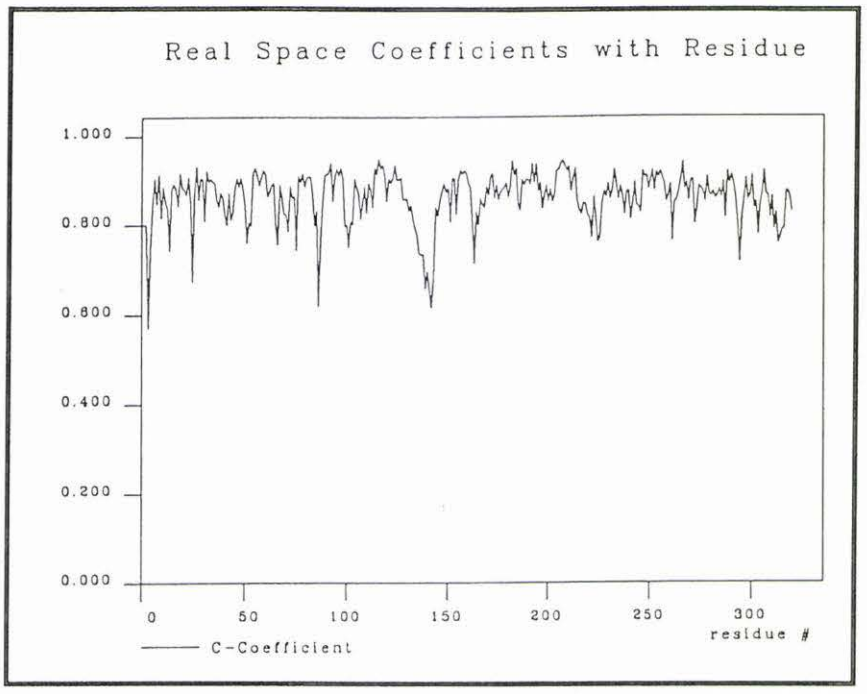


Fig. E.6. Graph showing the real space correlation coefficients for all atoms of P251A

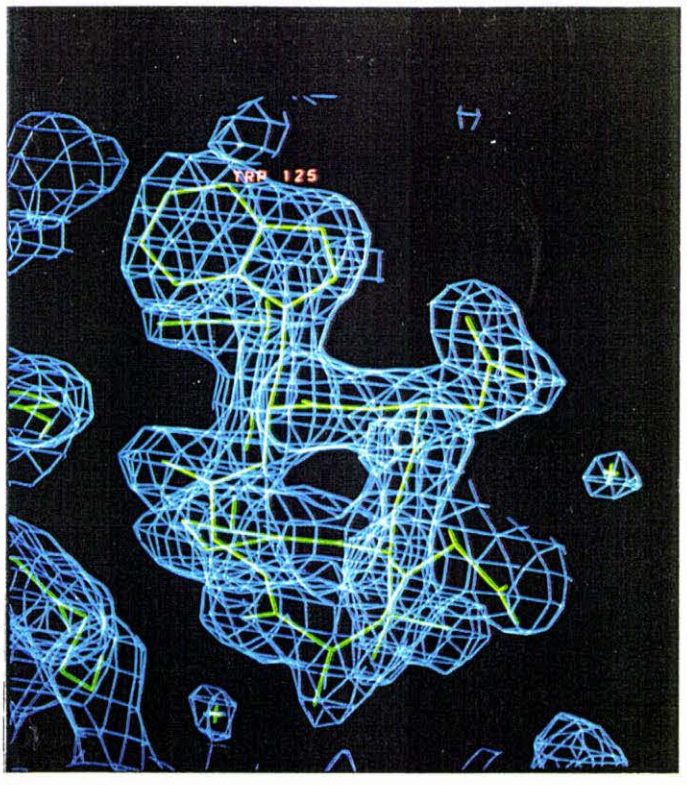


Fig. E.7. Photo of good density showing an end on view of an alpha helix around Trp 125

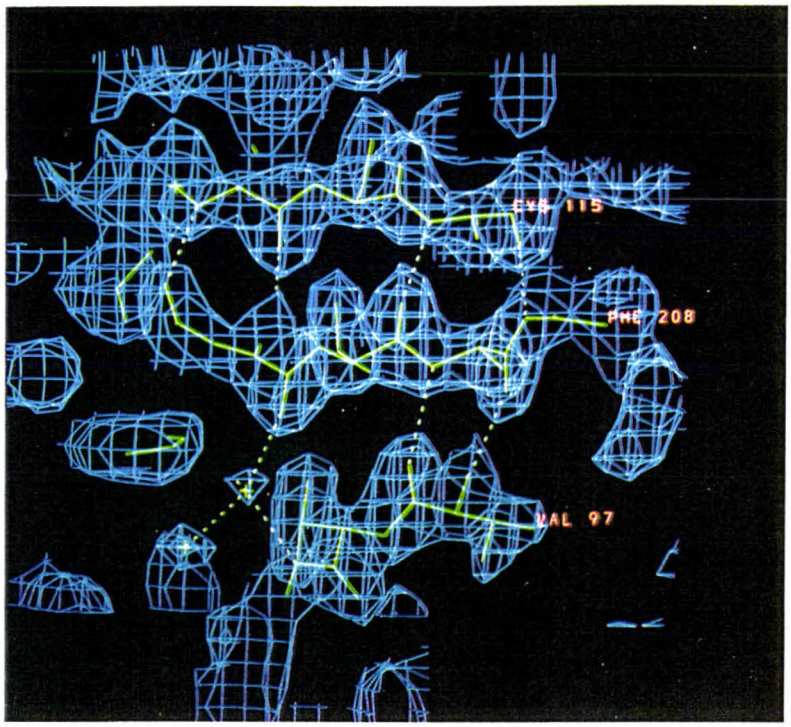


Fig. E.8. Good density showing a beta sheet in the internal part of the molecule. The 2 |Fo| - |Fc| map was contoured at 2σ.

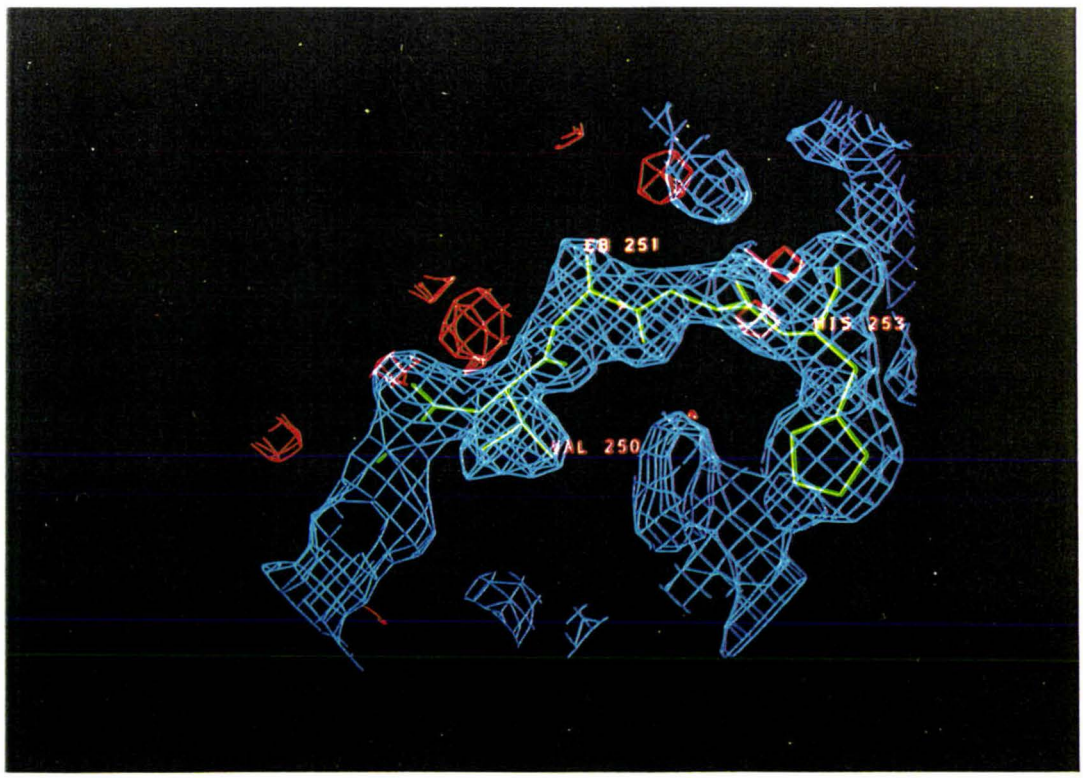


Fig. E.9. Photo showing density around Ala 251. The blue density corresponds to the 2 |Fo| - |Fc| map contoured at 2σ, while the red density is a |Fo| - |Fc| map contoured at 2σ to show that the proline had been replaced by alanine.

E.4.2. Agreement of the model with accepted geometry values.

E.4.2.1. Dihedral angles.

The correctness of the model was also assessed with regard to the distribution of dihedral angles (ϕ and ψ). Fig. E.10 shows the Ramachandran plot (Ramakrishnan and Ramachandran 1965) of the main chain torsion angles obtained using the program PROCHECK (Laskowski et al 1993). The Ramachandran plot shows that 98.9 % of the residues are found in favoured or additional allowed regions, with one residue (Ser 191) in the generously allowed region, and two residues (Arg 4 and Leu 299) in the disallowed regions.

Serine 191, although in a well ordered part of the molecule (in the binding site) has unusual mainchain torsion angles of phi = 69 and psi = -177, designated an ε conformation (Sibanda et al 1989). Serine 191 is involved in hydrogen bonding to Arg 121 presumably to help stabilise Arg 121 in the correct position for anion binding. This ε conformation has been found before in proteins such as alcohol dehydrogenase (Eklund *et al* 1976) and the fungal triglyceride lipase from *Rhizomucor miehei* (Derewenda and Derewenda 1991).

Leu 299 is found as the central residue of a γ-turn (298 O - N 300). This type of turn is characterised by mainchain torsion angles of $\phi = 70^{o}$; $\psi = -50^{o}$ (Baker and Hubbard 1984)and Leu 299, with angles of phi = 71^{o} and psi = -48^{o}, conforms closely to these values. The side chain of Arg 4 has been omitted from the final model, and although the main chain follows the density in this region, the density is weak and this residue has an uncertain conformation.

E.4.2.2 Main and side chain parameters.

The program PROCHECK (Laskowski et al 1993) also assesses the geometry of the model by comparison with 118 well defined structures. Comparisons include the quality of the Ramachandran plot (Fig. E.10.), and a series of plots (Figs E.11. and E.12.) comparing different geometric parameters. The result of this assessment is illustrated in Figs E.10, E.11. and E.12. on the following pages.

At 2.0 Å resolution, the quality of the structure is better than average in the omega angle standard deviation, and in the overall G-factor. The G-factor is an overall "goodness factor" defined by the programme PROCHECK which takes into account all the factors in the assessment. With respect to the rest of the parameters, the structure is within acceptable limits.

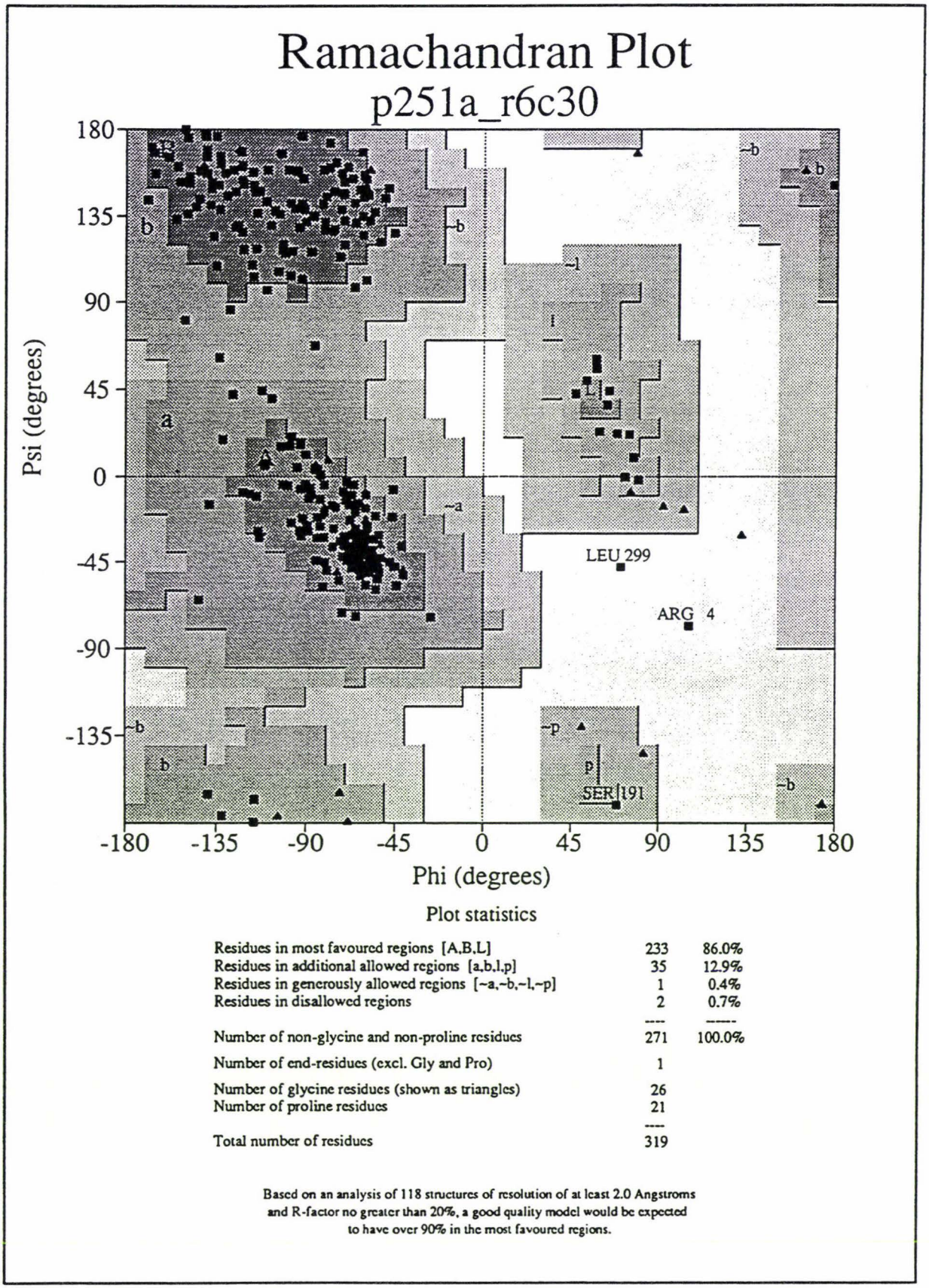
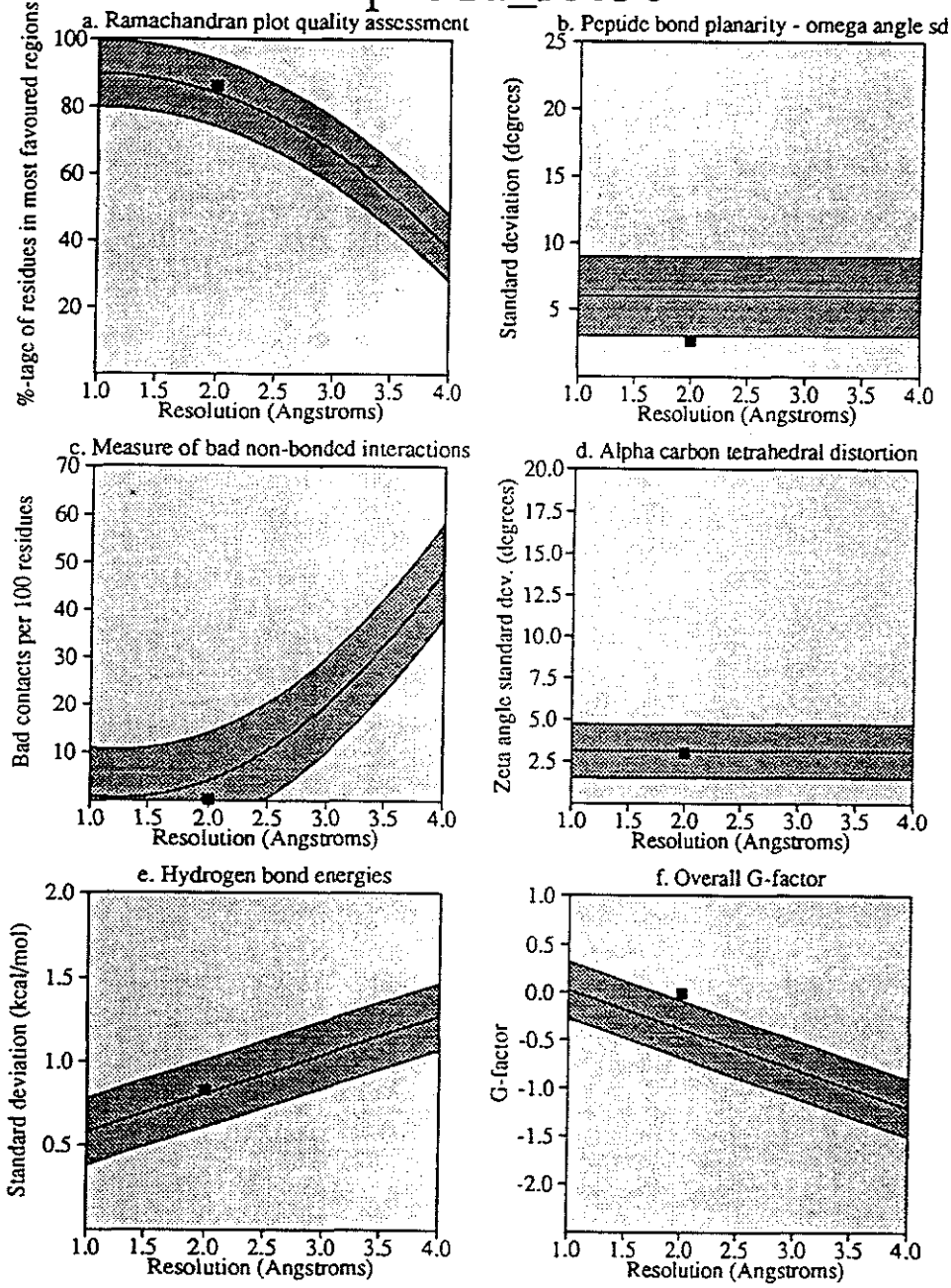


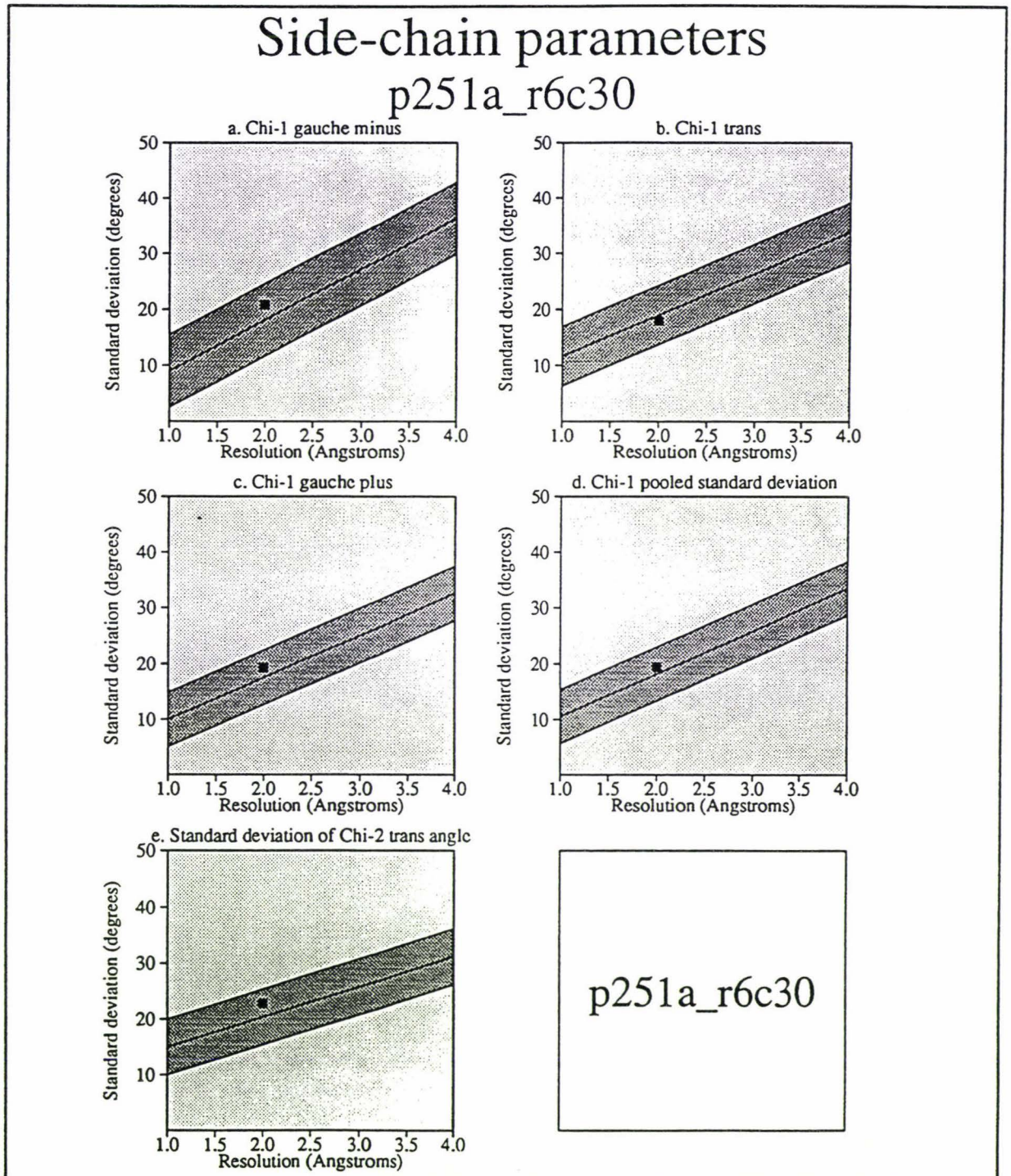
Fig. E.10. Ramachandran plot for the final P251A structure.



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	271	86.0	83.8	10.0	0.2	Inside
b. Omega angle st dev	315	2.6	6.0	3.0	-1.1	BETTER
c. Bad contacts / 100 residues	1	0.3	4.2	10.0	-0.4	Inside
d. Zeta angle st dev	288	3.0	3.1	1.6	-0.1	Inside
e. H-bond energy st dev	193	0.8	0.8	0.2	0.1	Inside
f. Overall G-factor	319	0.0	-0.4	0.3	1.3	BETTER

Fig. E.11. Comparrison of the main chain geometric parameters for P251A as assessed by PROCHECK (Lasdowski *et al*, 1993).



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	40	20.9	18.1	6.5	0.4	Inside
b. Chi-1 trans st dev	65	18.0	19.0	5.3	-0.2	Inside
c. Chi-1 gauche plus st dev	127	19.3	17.5	4.9	0.4	Inside
d. Chi-1 pooled st dev	232	19.4	18.2	4.8	0.3	Inside
e. Chi-2 trans st dev	67	22.8	20.4	5.0	0.5	Inside

Fig. E.12. Comparrison of the side chain geometric parameters for P251A as assessed by PROCHECK (Lasdowski *et al*, 1993).

E.4.2.3. Geometrical deviations.

The quality of the structure can be seen in Fig. E.13. by the low number of deviations from expected geometry. The structure was assesed for all 317 residues, and out of a total of 920 bonds, only 7 had lengths differing by > 0.05 Å from standard values. Five mainchain bond angles (out of a total of 638) were found to have differences greater than 10.0°, while only 3 planar groups were outside the acceptable levels assessed by the program PROCHECK (Laskowski et al 1993).

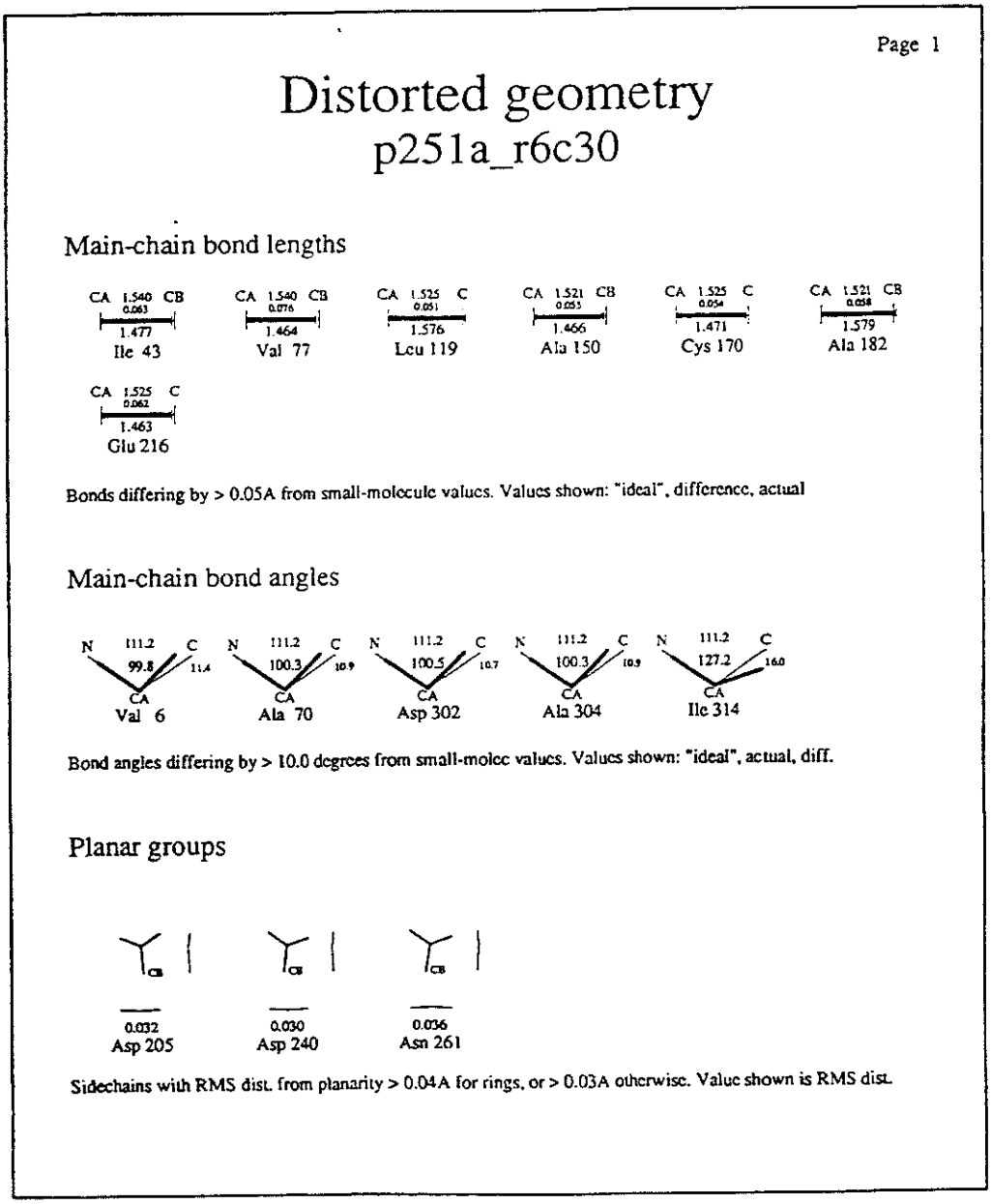


Fig. E.13. Distorted geometry found in the final structure of P251A by the programme PROCHECK (Laskowski *et al*, 1993).

E.4.3. Temperature factors.

Fig. E.14. shows the average mainchain B-values as a function of residue number. The average mainchain B-value of 34.1 $Å^2$ is consistent with the weakness of the high resolution data set, and the fact that all the data were used in the least squares refinement process (with no sigma cutoff). Large B-values are caused by fitting calculated structure factors to weak observed data. The average B-value for the side chains only is 39.9 $Å^2$, while over all the atoms it is 37.0 $Å^2$.

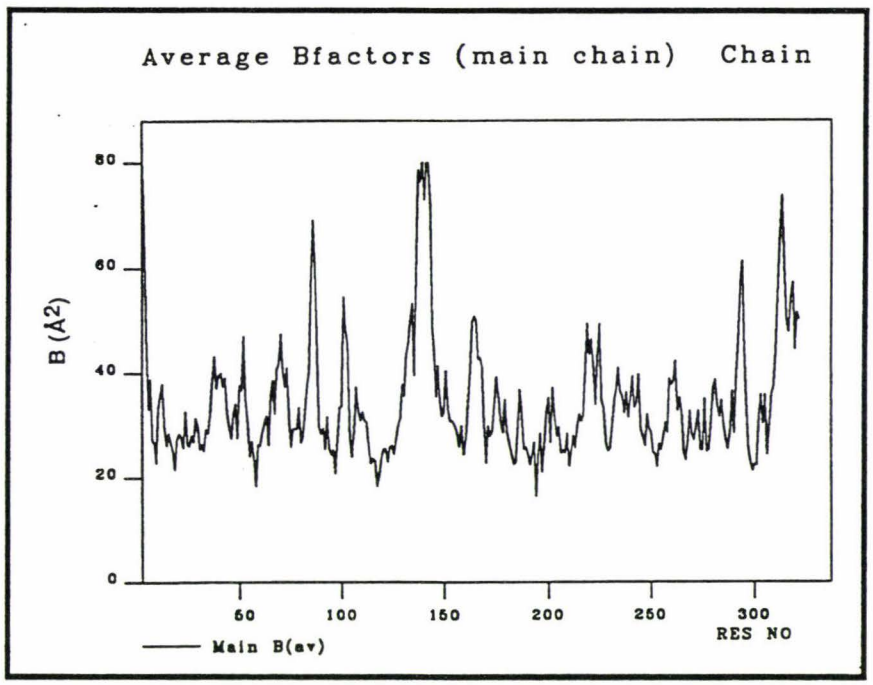


Fig. E.14. Plot of average mainchain B-value as a function of residue number for P251A.

From the figure (Fig. E.14.) it can be seen that the low real space correlation coefficients seen for Pro 141 and Pro 142 are reflected in high B-values for these residues. The reason Pro 141 and 142 are poorly defined is unknown, although it has been found with other mutants of lactoferrin that this feature of high B-values and low real space correlation coefficients occurs in those proteins in which the glycan chain from asparagine 137 has been removed (as in this case). This implies that the glycan chain is in some way stabilising the loop containing Pro 141 and Pro 142, although this cannot be confirmed until a structure showing the structure of the glycan chain is solved.

E.5. Structure description.

E.5.1. General topology.

The secondary and tertiary structure of P251A are as described previously for the N-lobe of Fe_2Lf (Anderson *et al*, 1987) and for Lf_N (Day *et al*, 1992).Secondary structure comprises about 25% beta sheet, and 38% helices with the remainder as turns or regions of extended chain or loops.

The polypeptide is folded into 2 similar domains with the Fe^{3+} and CO_3^{2-} ions bound in between. A ribbon diagram of the structure of P251A generated by the program Molscript (Kraulis, 1991) is illustrated in Fig.E.16, while a shematic representation is shown in Fig. E.15.

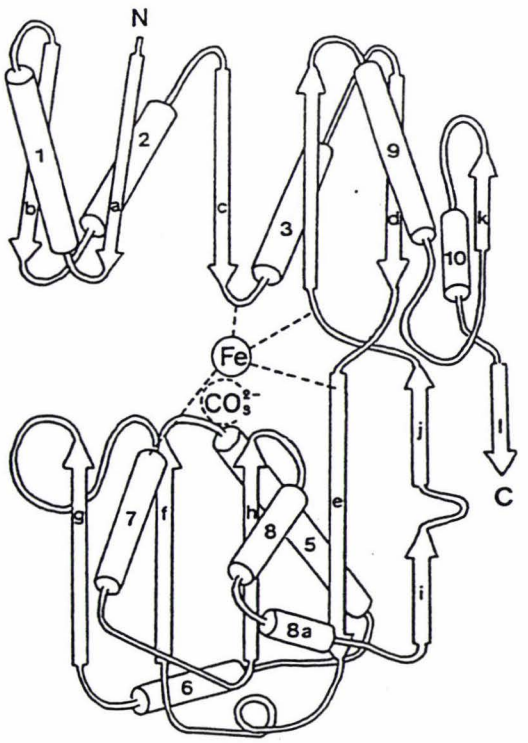


Fig. E.15. Schematic representation of the N-lobe of lactoferrin. Helices are shown as cylinders, while β-sheets are shown as arrows. The position of the protein ligands, the iron atom and the anion are indicated.

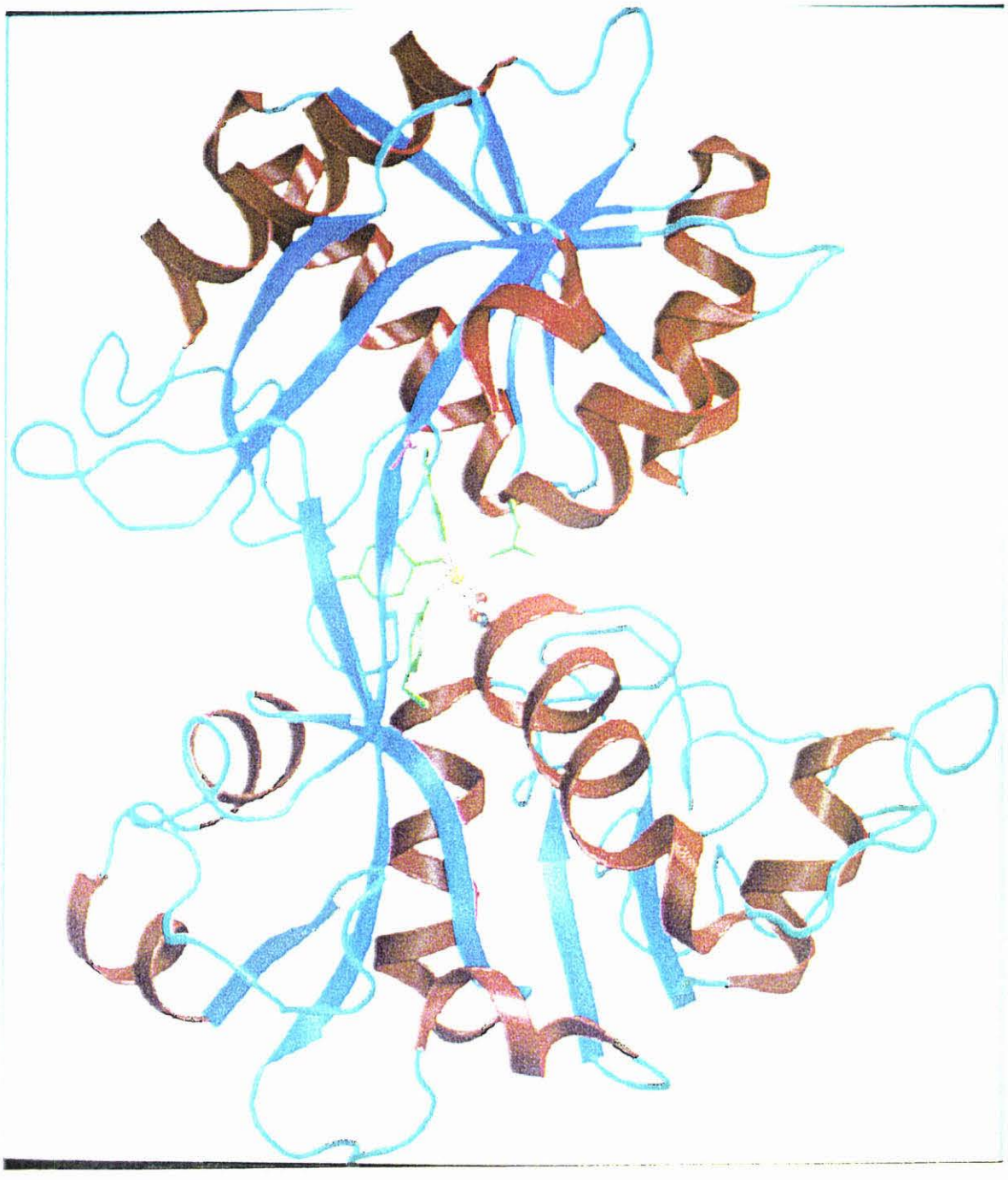


Fig E.16. Molscript image of the final P251A structure. The helices are shown in red, while the β-sheets are shown as dark blue. The turns are light blue. The four protein ligands are shown (in green) coordinated to the iron atom (orange), which is coordinated in turn to the carbonate anion (red). The side chain of Ala 251 is shown in purple where it is part of a β-sheet.

The beta sheet in domain N1 (Fig. E.15), consists of six strands in the order b-a-c-j2-d-k. with 4 parallel strands (a, b, c, and d) pointing towards the cleft of the binding site. These four strands are connected to one another by α-helices 1, 2, and 3. In the N1 domain, the four β-strands with their connecting α-helices thus form an α/β-type structure (Levitt and Chothia, 1976). The N2 domain is similar in folding to the N1 domain with the same α/β-type strucutre with the four beta stands pointing into the binding cleft, but in this case, extra helices are inserted between β-strands f and g, and between β-strands h and i. The other two β-strands are parallel to each other, but antiparallel to the other 4 strands (a, b, c, and d), and therefore point away from the binding site cleft.

The structure of the N-lobe found in Fe_2Lf, Lf_N, and P251A consists of three common structural elements. Firstly, as described above, the two α/β-type structures form a major part of each of the domains (N1 and N2), and these cause the positively charged N-terminal ends of the helices to point towards the binding site. This has been suggested by Anderson et al (1987) to have a role in attracting the anion into the binding site as it is thought that the anion binds first before the iron.

Secondly, the N-lobe has a pair of antiparallel beta strands which run behind the metal binding site, and which make up the two strands involved in the "hinge" mechanism (Gerstein *et al*, 1993) which incorporates Pro 251. These also are common to all N-lobe structures.

A feature which varies in the N-lobe structures is the C-terminal portion of the polypeptide, residues 312 - 333. In Fe_2Lf, residues 315 - 321 form a short helix (number 10), and residues 322 - 333 form another helix (helix 11). In the recombinant N-terminal half-molecule, Lf_N, however, helix 11 is unwound such that residues 322 - 327 have an extended conformation (no density was seen beyond residue 327) (Day, 1993). The density of the P251A structure is not complete beyond residue 320 however, possibly due to the different crystal packing found between Lf_N (C2) and P251A ($p4_12_12$).

E.5.2. Iron binding and the mutation site.

For the initial round of refinement, the metal ion, ligands and carbonate anion were not included in order not to bias this region of the structure. These atoms were added to the model by fitting to a subsequent 2|Fo| - |Fc| map and were then refined with the rest of the structure. The iron atom makes the same six bonds with Asp 60 OD1, Tyr 92 OH, Tyr 192 OH, His 253 NE2, and the two carbonate oxygens O1 and O2, as seen in Fe_2Lf and Lf_N. The carbonate is hydrogen bonded to Arg 121, Thr 117, and to the main chain amide nitrogens of two residues at the N-terminus of helix 5 (Ala 123 and Gly 124). These contacts are illustrated in Figs. E.17. and E.18. below.

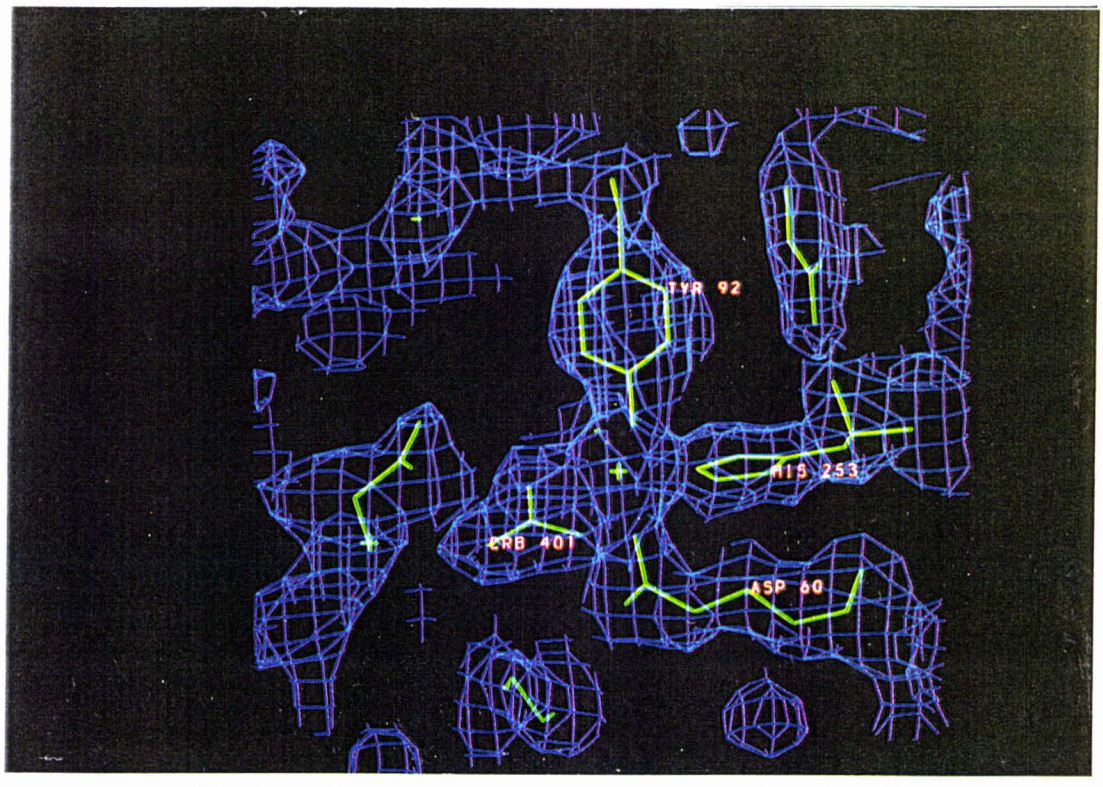


Fig. E.17. Density around the iron binding site of P251A.

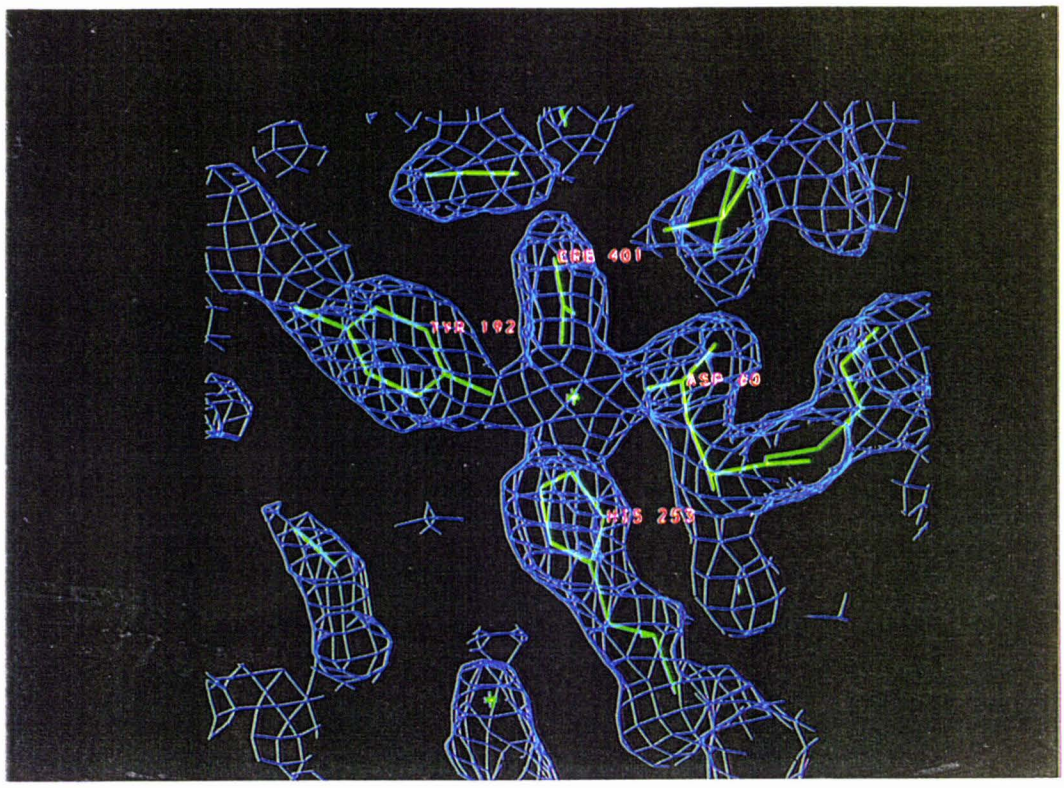


Fig. E.18. Density around the iron binding site of P251A.

The bond lengths and angles for the residues involved in Fe^{3+} and anion binding are indistinguishable within experimental error to those found for both the full-length (Fe_2Lf) and half-length (Lf_N) molecules (Table E.7).

The effect of changing the proline to alanine in position 251 thus has little effect either on the binding site or on the overall structure.

E.5.3. Water molecules.

As outlined earlier, there are 104 water molecules present in the final model for P251A. This is less than the 186 found for the wild-type structure (Lf_N). Water molecules were only included where potential hydrogen bonds could be made with 3.5 Å being the maximum length for the hydrogen bond. This lower number of water molecules present in the P251A model is the result of more conservative criteria being used in the P251A analysis when identifying potential water molecules in 2|Fo| - |Fc| and |Fo| - |Fc| maps. The water molecules common to the two structures are well defined. This is shown by the lower B-values of the water molecules found to be conserved between the two structures, compared with higher B-values of the water molecules unique to this structure.

Fig. E.19 shows a Cα-trace of P251A with the associated water molecules. The pink coloured water molecules refer to those conserved between P251A and the N-lobe of Fe2Lf, while the yellow coloured water molecules are those not conserved.

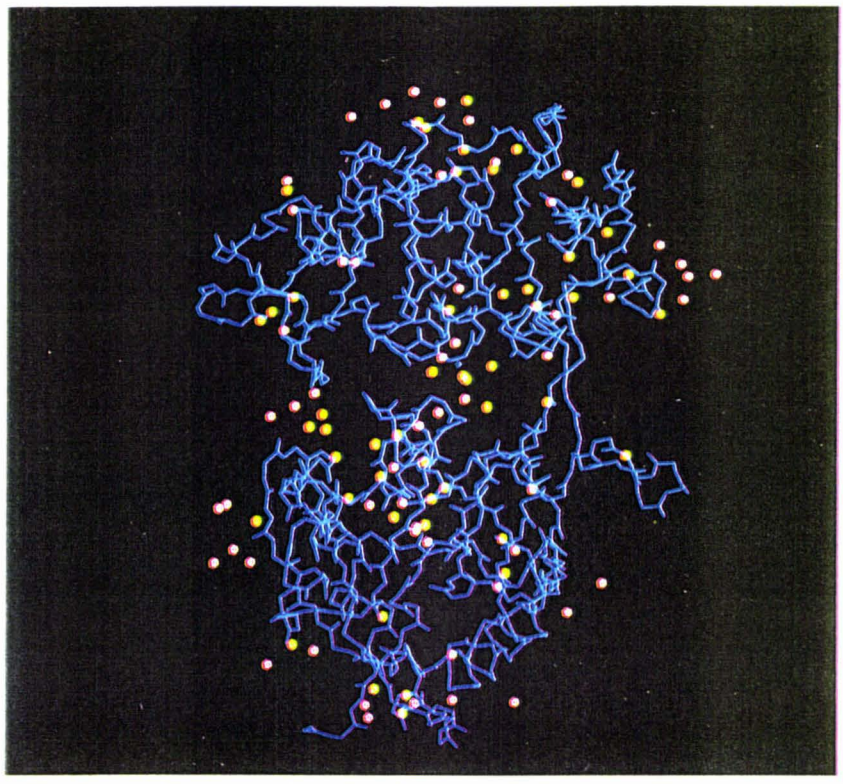


Fig. E.19. Water molecules found in P251A and Fe_2Lf (pink) and those found only in P251A (yellow).

E.6. Structural comparison between NLf and P251A.

E.6.1. Rms differences between the two polypeptide chains.

The rms deviations between the two structures (wild- type N-lobe and P251A) were determined using the programme LSQKAB from the CCP4 suite of programmes (The CCP4 suite, 1994).

For mainchain atoms from residue 5 to 320 (1261 atoms), the rms deviation is 0.38 Å, whereas for side-chain atoms (1093 atoms) the rms deviation is 1.06 Å. For all the atoms (mainchain and sidechain) the difference is 0.78 Å.

To determine the effect the mutation had on the extent of closure, the P251A structure was superimposed on to the Lf_N structure as a full molecule firstly, and then as two separate domains. The difference in domain closure was determined by first rotating P251A onto Lf_N using domain N2, and then further superimposing this rotated P251A onto Lf_N using domain N1. Using C_α atoms only, the rms deviation for the full molecule is 0.33 Å (residues 5 - 320) or 0.28 Å (residues 5 - 312), for the N2 domain (residues 92 - 250) the deviation is 0.24 Å, and for the N1 domain (residues 5 - 91, and 251 - 312), it is 0.26 Å. The difference in domain closure between Lf_N and P251A was only 0.9^o, - essentially identical structures.

Fig. E.20. shows the C_α trace of the mainchains for Lf_N and P251A overlaid onto each other.

For residues 248 - 254 from strand j of the hinge, ie. the residues immediately surrounding the mutation site, there is even closer correspondence; the rms difference for the main-chain atoms is 0.22 Å (Ala251 is 0.19 Å), and for the side-chain atoms 0.35 Å. As seen in Fig. E.21, there is little change in this region indicating that the mutation of proline to alanine at residue 251 does not perturb this region in any way from the wild-type protein.

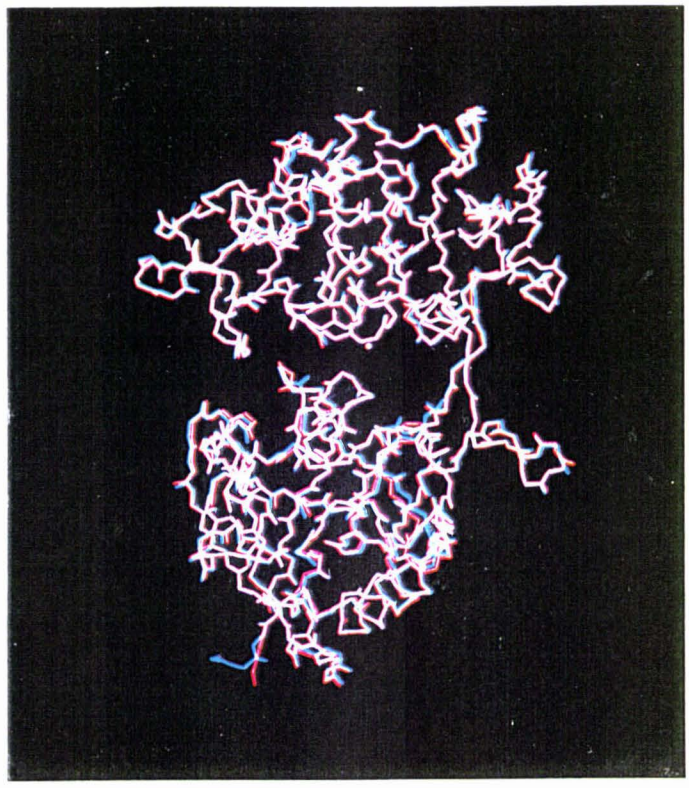


Fig. E.20. C_α-trace of the polypeptide chain of Lf_N (red) and P251A (blue). The white trace shows where the two structures superimpose.

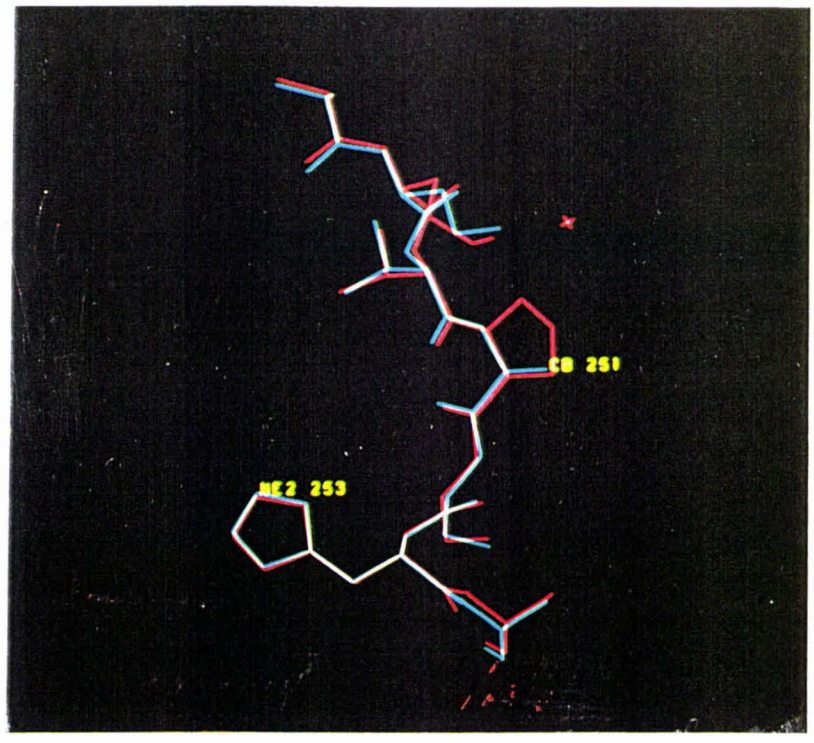


Fig. E.21. Diagram showing the overlay of residues 248 - 254, between the structures Lf_N (red) and P251A (blue). The white trace is where the two structures superimpose.

The only regions where the main chain differs by more than 1.0 Å between the two structures, are the N- and C-terminal regions, and residues Thr 139, Ser 293 and Asp 302. These residues with their rms differences are tabulated in Table E.6.

Table E.6. Residues where the rms difference between Lf_N and P251A is greater than 1.0 Å.

Residue	rms difference
Arg 4	1.93 Å
Thr 139	1.34 Å
Ser 293	1.65 Å
Asp 302	1.50 Å
Pro 312	1.24 Å
Arg 313	2.32 Å

The differences at the N- and C-terminal ends of the molecule can be explained by the flexibility of the molecule in these regions. Thr 139 is found in a loop which appears to be affected by deglycosylation, which has resulted in greater flexibility in this region. The conformation therefore may be variable, explaining the differences between the structures. Residues 293 and 302 are also at sites which show some variability in different lactoferrin structures; Ser 293 is in a loop with high B-values, and Asp 302 is in a β-turn in which the peptide 302 - 303 can "flip" in different structures (Faber *et al*, 1996).

E.6.2. Differences in the binding cleft.

Fig. E.22. superimposes the iron binding sites of Lf_N and P251A. The bond-lengths and -angles for the P251A structure are tabulated in Table E.7. These are essentially the same as found for Lf_N.

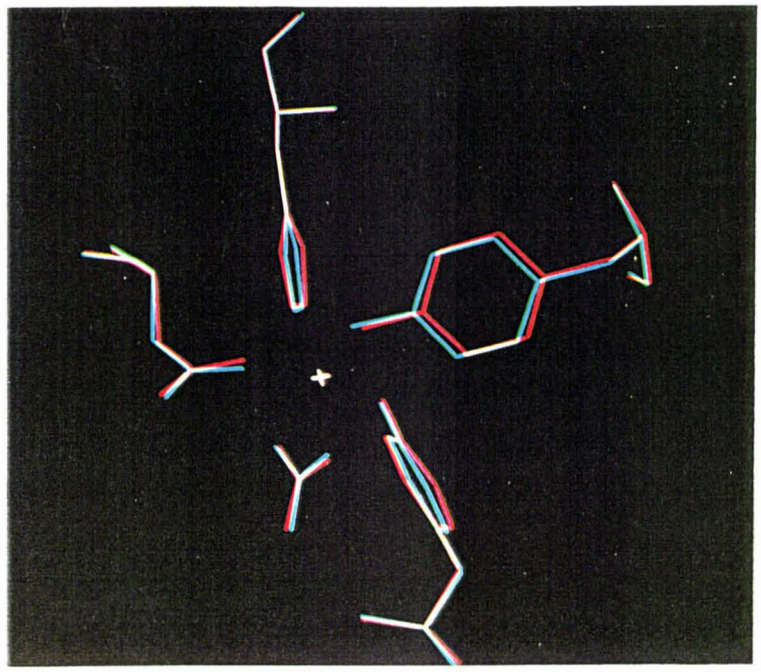


Fig. E.22. Superposition of the iron binding sites between P251A and Lf_N.

Table E.7. Bond-lengths and -angles at the iron site. Numbers in brackets refer to the values found for the wild-type structure (Day, 1993).

Ligand	Iron to ligand distance (Å)	Angle formed (°)		
Asp 60 OD1	1.91 (1.99)	60 OD1-Fe-OH 92	91	(90)
		60 OD1-Fe-OH 192	157	(164)
		60 OD1-Fe-NE2 253	80	(87)
		60 OD1-Fe-O1	77	(87)
		60 OD1-Fe-O2	88	(94)
Tyr 92 OH	1.85 (2.00)	92 OH-Fe-NE2 253	91	(96)
		92 OH-Fe-O1	157	(160)
		92 OH-Fe-O2	96	(97)
Tyr 192 OH	1.90 (1.92)	192 OH-Fe-OH 92	112	(106)
		192 OH-Fe-NE2 253	92	(86)
His 253 NE2	2.03 (2.15)	253 NE2-Fe-O1	107	(104)
		253 NE2-Fe-O2	171	(167)
Carbonate O1	2.07 (2.04)	O1-Fe-OH 192	82	(80)
		O1-Fe-O2	65	(63)
Carbonate O2	2.22 (2.10)	O2-Fe-OH 192	90	(97)

E.6.3. Crystal packing of half-length molecules.

Although the two structures of Lf_N and P251A are closed and iron-bound, the crystal packing for them is different (C2 space group for Lf_N, $P4_12_12$ for P251A). It has been suggested (Day, 1993) that crystal packing of Lf_N may be responsible for the altered conformation of residues 321 - 333. When compared with native, full-length lactoferrin, an α-helix, 321 - 332, becomes unwound in Lf_N and forms a piece of extended chain which is disordered beyond residue 327. In the Lf_N structure, the molecules pack back to back such that the residues 313 - 333 are able to make contacts with the same region of another molecule and it was thought possible that this packing was responsible for the unwinding. For P251A ($p4_12_12$), the molecule packs differently such that these contacts are not present. In spite of this different crystal packing there is still no helix 321 - 332, and the chain is completely disordered from residue 321 onwards. The conformation is, however, essentially the same for residues 313 - 320, which is as far as the chain can be traced.

F. DISCUSSION.

The aim of this thesis was to examine the importance of residues in the hinge region of human lactoferrin through mutagenesis and subsequent characterisation of the mutants. The N-lobe of lactoferrin was chosen as the vehicle for the mutagenesis studies (a) because it possesses only a single well defined hinge, (b) because several highly conserved residues in the hinge offered an obvious target, and (c) constructs were already available which could be used as a basis for construction of mutants. In the event, four mutants of Pro 251 were successfully constructed, expressed and purified, together with a single mutant of Thr 90. The time available did not permit a full functional analysis of the mutants, but the effect of mutagenesis on domain closure was examined crystographically for one mutant (P251A), and all of the mutants were characterised with respect to their iron-release properties and their absorption spectra.

F.1. Mutagenesis

The N-lobe of human lactoferrin has previously been expressed in BHK cells by integration of the pNUT:Lf_N construct (illustrated in Fig. C.5.) into the DNA of the cell, The expressed protein was subsequently shown to be folded correctly by X-ray crystallographic studies (Day, 1992).

In order to express protein with a single amino acid change, the part of the cDNA around the mutation site was inserted into the bacteriophage M13 and by the use of oligonucleotide site-directed mutagenesis, various mutations were introduced into different cDNA clones. The segments were moved from M13 into pNUT:Lf_N, replacing the wild-type sequence with the mutation. This process was aided by the presence of two unique restriction enzyme sites in both M13 and pNUT:Lf_N that were not found in other parts of the two constructs. The mammalian cell culture system used for expression is, however, both expensive and time consuming. An alternative system has recently been developed in the fungus *Aspergillus oryzae* (Ward *et al*, 1992) and lactoferrin expressed in this system has been found to have the same structure and properties as the native protein. This promises to provide a cheaper and easier expression system for future mutagenesis studies.

The method of purification of the recombinant protein was modified from the method used by Day (1992). The method developed removed the first step in Tris/HCl buffer and involved only one step on a CM column with 10 mM HEPES and a salt gradient from 0.2 M to 1.0M. The eluted protein gives two bands of 40 kDa and 37 kDa when run on an SDS gel. The explanation for these two bands was explained (Day *et al*, 1992) where the 40 kDa band correspond to the glycosylated form of the protein, and the 37 kDa band the non-glycosylated form.

In order to obtain homogeneous protein for subsequent x-ray analysis, deglycosylation with endoglycosidase is required (Norris *et al*, 1989). As the endoglycosidase is added to the protein mixture, this needs to be removed again before the protein is ready for growing crystals. The endoglycosidase was passed through a SP-sephadex column and the fractions collected and checked for activity with a sample of protein. As the lactoferrin binds to this column under the same conditions, this allows the use of these fractions of endoglycosidase to be used to deglycosylate the proteins prior to growing crystals.

F.2. Effect of mutation on protein structure.

Lactoferrin has two distinct structural forms, an open, iron-free state, and a closed, iron-bound structure. Mutation of an amino acid residue could affect either the open form, or the closed form (or both), or could perturb the transition between the two forms. In the present study the three dimensional structure of the closed, iron-bound form of one of the mutants, P251A, was determined, although it is clearly desirable that the open, iron-free structure should also be determined in future work.

Comparison of the three dimensional structure of P251A with Lf_N shows that neither the polypeptide chain conformation nor the iron site have been affected by the mutation. In particular, the domain closure is essentially identical to that seen in Lf_N. When the entire half-molecule is superimposed as a single unit, the rms deviation is only 0.28 Å. The two individual domains match only slightly more closely - when two N1 domains are superimposed, deviation is 0.26 Å and when two N2 domains are superimposed, rms deviation is 0.24 Å. This translates to a difference of only 0.9° in the closure of the two domains.

It appears that neither changing the metal ion (as in Fe^{3+} to Cu^{2+} (Smith *et al* 1992)), nor one of the ligands (H253M, (Nicholson *et al*, mannuscript in prep)), nor the central residue of the hinge (P251A) affects the overall structure of the N-lobe of lactoferrin, as stabilising contacts maintain the structure.

Gerstein *et al* (1993), have proposed that the closure of the domains in

lactoferrin depends on the matching of two interfaces (a large interface, which includes most of the residues in the binding cleft, and a small interface, which includes residues at the back of the hinge). Gerstein *et al* (1993) also suggest that the exposure of the hinge residues and their lack of interaction with the rest of the protein means that they do not significantly influence domain closure. The present results support this view. Proline can have a relatively strong influence on polypeptide chain conformation, because of the angular constraints imposed on the N-Cα bond. Mutation to alanine has left the domain closure unaffected, however, implying that it is other residues (eg in the interfaces) that are much more important. On the other hand, mutation of the aspartate ligand, in the D60S mutant, has been shown to have substantial effect on domain closure. Asp 60 forms important interactions in the interdomain cleft, and makes up part of the interface, - its mutation to serine causes a difference of 7° in domain closure (Faber *et al*, 1996) as the domains collapse further together. Mutagenesis of other residues in the large interface would make it possible to examine further the effects these contacts have on the mechanism of binding, and to see whether this moved the equilibrium between the open and closed forms to a predominantly more open state by preventing these contacts from initially stabilising the closed form.

Lack of change in the iron site is not unexpected. The rms deviation is 0.190 Å. This similarity is consistent with the similarity in absorption spectra. The visible absorption maximum is very similar to that of the wild type, indicating that there are similar interactions of tyrosine residues, since it is the Fe-Tyr interaction that gives rise to the spectral band.

F.3. Iron release.

One of the most significant elements of transferrin chemistry is the pH dependence of iron release. In fact this is one of the principal differences between lactoferrin and serum transferrin. In serum transferrin, iron is released below pH 6.5, whereas in lactoferrin, release occurs 2 pH units lower. The origin of this pH dependence is unknown though several suggestions have been made.

The iron release experiments with the lactoferrin mutants of this study proved difficult to carry out experimentally, because of the low solubility of free iron, and the low pH at which lactoferrin releases iron. These two effects meant that the process is not reversible because as the pH is lowered, iron is released where it forms insoluble $Fe(OH)_3$. The lactoferrin is less stable to pH in the iron-free form, resulting in denaturation of the protein. This means that upon lowering the pH, both the iron and protein end up as a precipitate when the titration is taken to its conclusion (at pH < 3).

Iron release experiments did, however, show that none of the mutants showed any significant difference in the pH of iron release, either from each other or from the wild-type Lf_N. The effect of pH presumably depends on protonation of one or more functional groups. What the iron release experiments on the mutants show, therefore, is that these groups are not affected by mutations at the hinge. Two suggestions have been advanced. The first is that residues behind the iron site are responsible for these effects, notably several cationic groups. In serum transferrin, a pair of lysine residues have been suggested to act as a trigger (Dewan *et al*, 1993). In lactoferrin the equivalent residues are an arginine and a lysine, and the implication is that these are possible sites of protonation. The pKa values for lysine and arginine (normally between 10 and 12) do not however correspond to the observed pH at which iron is released, and the pKa values would have to be very much depressed by their environment. A second suggestion is that the CO_3^{2-} ion is protonated and that protonation is the first step in the break-up of the iron site. The pKa value for free HCO_3^- is high (10.25) but its environment, with an arginine closely associated should depress it greatly. Also in favour of this suggestion is the fact that HCO_3^- ions tend to bond to metals as monodentate ligands. Conversion of bidentate CO_3^{2-} to monodentate HCO_3^- could allow water to bind to the iron atom and stimulate its release.

The studies on these mutants do not obviously distinguish between these two possibilities. However the lysine and arginine residues are behind the iron site very close to the hinge and might be expected to be affected by mutations at the hinge (particularily by introducing a negative charge as in P251D). The lack of change of the pH of iron release would tend to suggest that it is protonation of CO_3^{2-} that is more likely to be important.

F.4. Conservation of hinge residues.

Given that the mutation of the hinge residues does not affect the three dimensional structure (at least in the closed state), nor iron release, why are proline and threonine residues at the hinge so highly conserved?

It could be that this is purely coincidental (the next protein sequenced may show differences here). A point to make is that the periplasmic-binding proteins which have a similar hinge mechanism, have no similarity in the hinge residues between them.

It may be that the residues are conserved to help define the open state. As there is no open structure yet available for wild-type Lf_N or any mutant, this question cannot be answered.

A more likely explanation is that these residues in some way influence the flexibility of the hinge. This may be particularly true of proline. The three dimensional structure of one crystal form of iron-free lactoferrin shows that even in the absence of iron, the closed state can be stabilised. The same is true for the periplasmic-binding proteins. Closed unliganded forms have been characterised for both the glucose/galactose-binding protein (Flocco and Mowbray, 1994), and for the arabinose-binding protein (Quiocho, *person commun*). The implication is that for effective binding, the open and closed states need to be close in energy, and probably are in dynamic equilibrium in the ligand-free state. The hinge residues may play an important role in this equilibrium.

A final consideration is that kinetic factors may play a part. A most important target for future characterisation of these mutants is to analyse their kinetic properties. This can be carried out using a method developped by Egan *et al* (1993) who use a spectroflurometric method with an iron-sequestering agent present. This has the advantage of using small amounts of protein, and has been used to study the effect of mutations of human serum transferrin (Zak *et al*, 1995).

F.5. Hinges in periplasmic-binding proteins.

A family of proteins structurally similar to the transferrins, are the periplasmic-binding proteins which are involved in active transport of growth essential nutrients across the periplasmic space of gram negative bacteria. From this family, two members have structures solved in both open and closed forms. These are the lysine-arginine-ornithine-binding protein (LAOBP) (Oh *et al*, 1993), and the maltodextrin-binding protein (MBP) (Sharff *et al*, 1992). For the LAOBP, the mechanism of domain movement appears to be similar to that in lactoferrin with the large torsion angle changes occurring in a region structurally equivalent to that of the lactoferrin hinge. For MBP, two of the hinges are also structurally equivalent to those of the lactoferrin hinge. The central residues involved in the hinge motion for the two proteins are Ala90 and Asp193 for the LAOBP, and Glu111, Val261 and Ala312 for the MBP. A sequence alignment on several members of this periplasmic-binding protein family was carried out using the programme CLUSTAL W (ref) to determine whether these residues were found in a conserved region, but sequence identity between members of this family generally ranged between 10 and 15% identity, indicating poor sequence conservation.

Finally, a newly found member of the periplasmic-binding protein family, the ferric iron-binding protein (FBP) is a periplasmic-binding protein found in pathogenic Neisseira involved in the active transport of growth-essential iron by these bacteria. This protein is localised in the periplasmic space (as are the other PBP's) and reversibly binds a single ferric iron per molecule.

FBP has functional and structural characteristics similar to both the PBP's and the transferrin family. It is a 36 kDa protein, similar to a single lobe of the transferrins. Experimental results (Nowalk *et al*, 1994) provides evidence that FBP binds iron in a similar way to the transferrins. This evidence includes spectrometric, chemical modification and flurometric data.

The ligands responsible for iron-binding in FBP appear to be two tyrosines, one histidine, a synergistic bicarbonate anion, and one other ligand. FBP has an absorbance maximum of around 460 - 480 nm, and a wide range of transition cations that can bind.

Although the structure of this protein has not yet been solved, FBP is the first periplasmic-binding protein that binds iron, and possibly is the protein that bridges the evolutionary gap between the iron-binding bilobal transferrins, and the single-lobed PBP's.

References

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

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

Aisen, P., Leibman, A. and Reich, H.A. (1966) *J. Biol. Chem.* **241,** 1666-1671.

Aisen, P., Aasa, R., Malmstrom, B.G. and Vanngard, T. (1967) J.Biol.chem. **242,** 2484-2490.

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

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

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

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

Alcantara, J., Yu, R.H. and Schryvers, A.B. (1993) *Molecular microbiology* **8,** 1135-1143.

Alexander, L.J., Levine, W.B., Terry, C.T. and Beattie, C.W. (1992) *Anim. Genet.* **23,** 251.

Alexander, M.B. (1948) *British Medical Journal* **4,** 973-978.

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* **344,** 784-787.

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

Arnold, R.R., Russell, J.E., Champion, W.J., Brewer, M. and Gauthier, J.J. (1982) *Infection and Immunity* **35,** 792-799.

Baggiolini, M., de Duve, C., Masson, P.L. and Heremans, J.F. (1970) *J.Exp.Med.* **131,** 559-570.

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, H.M. (1995) phD thesis.

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

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

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

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

Baker, E.N., Anderson, B.F., Baker, H.M., Haridas,M., Jameson, G.R., Norris, G.E., Rumball, S.V. and Smith, C.A. (1991) *Int.J.Biol.Macromol.* **13,** 121-130.

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

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

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

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

Banfeld, D.K., Chow, B.K-C., Funk, W.D., Robertson, K.A., Umelas, T.M., Woodworth, R.C. and MacGillivray, R.T.A. (1991) *Biochim. Biophys Acta* **1089,** 262-265.

Bates, A. Thesis, PhD, Massey University, 1994.

Bates, G.W. and Schlabach, M.R. (1975) *J.Biol.Chem.* **250,** 2177-2181.

Baynes, R.D. and Bezwoda, W.R. (1992) in Proceedings of the First International Symposium on Lactoferrin Structure and Function.

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

Bellamy, W.R., Wakabayashi, H., Takase, M., Kawase, K., Shimamura, S. and Tomita, M. (1993) *J.Appl.Bact.* **75,** 478-484.

Boxer, L.A., Haak, R.A., Yang, H-H., Wolach, J.B., Whitcomb, J.A., Butterick, C.J. and Baehner, R.L. (1982) *J.Clin.Invest.* **70,** 1049-1057.

Bowman, B.H., Yang, F. and Adrian, G.S. (1988) *Advances in Genetics* **25,** 1-38.

Britigan, B.E., Serody, J.S., Hayek, M.B., Charniga, L.M. and Cohen, M.S. (1991) *J. Immunol* **147,** 4271-4277.

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 Griffiths, E. (1978) *Curr.Top.Microbiol.Immunol.* **80,** 1-35.

Bullen, J.J., Rogers, H.J. and Leigh, L. (1972) *Brit.Med.J.* **1,** 69-75.

Cannon, J.C. and Chasteen, N.D. (1975) *Biochemistry* **14,** 4573-4577.

Carpenter, M.A. and Broad, T.E. (1993) *Biochim. Biophys. Acta* 1173, 230

Casey, J.L., Koeller, D.M., Ramin, V.C., Klausner, D. and Harford, J.B. (1989) *EMBO* **8,** 3693-3699.

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

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

Davidson, L.A. and Lonnerdal, B. (1989) in *Milk Proteins* (Barth, C.A. and Schlimne, E. eds) p76-82.

Davidsson, L., Kastenmayer, P., Yuen, M., Lonnerdal, B. and Hurrell, R.F. (1994) *Pediatric Research* **35,** 117-124.

Day, C.L. Thesis, PhD, Massey University, 1993.

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

Derewenda, Z.S. and Derewenda, U. (1991) *Biochem. Cell Biol.* **69,** 842-850.

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

Egan, T.J, Zak, O. and Aisen, P. (1993) *Biochemistry* **32,** 8162-8167.

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

Ellison, R.T., Giehl, T.J. and Laforce, F.M. (1988) *Infection and Immunity* **56,** 2774-2781.

Ellison, R.T., Larforce, F.M., Giehl, T.J., Boose, D.S. and Dunn, B.E. (1990) *J.Gen.Microbiol.* **136,** 1437-1446.

Ellison, R.T. and Giehl, T.J. (1991) *J.Clin.Invest.* **88,** 1080-1091.

Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B-O., Taipu, O., Branden, C-I. and Akeson, A. (1976) *J. Mol. Biol.* **104,** 27-59.

Faber, H.R., Anderson, B.F., Baker, H.M. and Baker, E.N. (1996) manuscipt in preparation.

Faber, H.R., Bland, T., Day, C.L., Norris, G.L., Tweedie, J.W. and Baker, E.N. (1995) *J. Mol. Biol.* in press.

Faber, H.R. and Mathews, B.W. (1990) *Nature* **348,** 263-266.

Fletcher, J. and Huehns, E.R. (1968) *Nature* **218,** 1211-1214.

Flocco, M.M. and Mowbray, S.L., (1994) *J. Biol. Chem.* **269,** 8931-8936.

Gado, I., Erdei, J., Laszlo, V.G., Paszti, J., Czirok, E., Kontrohr, T., Toth, I., Forsgren, A. and Naidu, A.S. (1991) *Antimicrob. Agents Chemother.* **35,** 2538-2543.

Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M. and Chothia, C. (1993) *J.Mol.Biol.* **234,** 357-372.
Gerstein, M., Lesk, A.M. and Chothia, C. (1994) *Biochemistry* **33,** 6739-6749.

Gerstein, M., and Clothia , C.H. (1991) *J. Mol. Biol* . **220,** 133

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* **281,** 157-158.

Grossmann, J.G., Neu, M., Pantos, E., Schwab, F.J., Evans, R.W., Townes-Andrews, E., Lindley, P.F, Appel, H., Thies, W.-G. and Hasnain, S.S. (1992) *J. Mol. Biol.* **255,** 811-819.

Harris, D.C. and Aisen, P. (1989) chapter 3 (pgs 239-351) "*Physical Biochemistry of the Transferrins*" in "Iron Carriers and Iron Proteins" edited by Thomas M. Loehr.

Hashizume, S., Kuroda, K. and Murakami, H. (1983) *Biochim.Biophys.Acta* **763,** 377-382

Heery, D.M., Gannon, F. and Powell, R. (1990) *Trends in Genetics* **6,** 173.

Hill, C.P., Yee, J., Selsted, M.E. and Eisenburg, D. (1991) *Science* **251,** 1481-1484.

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

Huettinger, M., Retzek, H., Eder, M. and Goldenberg, H. (1988) *Clin. Biochem.* **21,** 87-92.

Ismail, M. and Brock, J.H. (1993) *J.Biol.Chem.* **268,** 21618-21625.

Iyer, S. and Lonnerdal, B. (1993) *Eur.Jou.Clin.Nutr.* **47,** 232-241.

Iyer, S., Yuen, M. and Lonnerdal, B. (1993) *FASEB abstracts,* 370.

Iyer, S., Hara, P.S., Hutchens, T.W. and Lonnerdal, B. (1994) *FASEB abstracts,* 5412.

Jacobson, B.L., He, J.J., Lemon, D.D. and Quiocho, F.A. (1992) *J. Mol. Biol.* **223,** 27-30.

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.

Jeltsch, J-M., Hen, R., Maroteaux, L., Garnier, J-M. and Chambon, P. (1987) *Nucleic Acid Research* **15,** 7643-7645

Johansson, B.G. (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.

Karlsson, R., Zheng, J.H., Xuong, N.H., Taylor, S.S. and Sowadski, J.M. (1993) *Acta. Crystallogr.* D49, 381

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

Kraulis, P.J.J. (1991) *J. Appl. Cryst.* **24,** 94 6-950.

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

Kretchmar, S.A. and Raymond, K.N. (1988) *Inorg. Chem.* **27,** 1436-1441.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods in enzymology* **154,** 367-382.

Kvingedal, A.M., Rorvik, K.A. and Alestrom, P. (1994) *Mol. Marine Biol. Biotech.* **2,** 233-238.

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

Legrand, D., Mazurier, J., Elass, A., Rochard, E. and Vergoten, G. (1992) *Biochemistry* **31,** 9243-9251.

Le Provost, F., Nocart, M., Guerin, G. and Martin, P. (1994) *Biochim. Biophys. Acta* **203,** 1324-1332.

Lestas, A.N. (1976) *Br.J.Haematol.* **32,** 341-350.

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

Lindley, P.F., Bajaj, M., Evans, R.W., Garratt, R.C., Hasnain, S., Jhoti, H., Kuser, P., Neu, M., Patel, K., Sarra, R., Strange, R. and Walton, A. (1993) *Acta. Crystallogr.* D49, 292

Lonnerdal, B. (1985) *Am.J.Clin.Nutr.* **42,** 1299-1317.

Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C. and Jordan, P.M. (1992) *Nature* **359,** 33-39.

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

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

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.

Magdoff-Fairchild, B. and Low, B.W. (1970) *Arch. Biochem. Biophys.* **138,** 703

Makey, D.G. and Seal, U.S. (1976) *Biochimica et Biophysica Acta* **453,** 250-256.

Malmquist, J., Hansen, N.E. and Karlf, H. (1978) *Scand.J.Haematol.* **21,** 5-8.

Marquardt, D.W. (1963) J. *Soc. Ind. Appl. Math.* **11,** 431-441.

Masson, P.L., Heremans, J.F. and Schonne, E. (1969) *J. Exp. Med.* **130,** 643-656.

Masson, P.L. and Heremans, J.F. (1971) *Comp.Biochem.Physiol.* **39B,** 119-129.

Mazurier, J. and Spik, G. (1980) *Biochimica et Biophysica Acta* 6 **29,** 399-408.

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

McAbee, D.D. and Esbensen, K. (1991) *J. Biol. Chem.* **266,** 23624-23631.

McPherson, A. (1982) *The preperation and analysis of protein crystals*. John Wiley and sons. New York.

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

Mead, P.E. and Tweedie, J.W. (1990) *Nucleic Acids Research* **18,** 7167-7168.

Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) *Science* **257,** 1251-1255.

Meador, W.E., Means, A.R. and Quiocho, F.A. (1993) *Science* **262,** 1718-1721.

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

Moore, S.A., Haridas, M., Anderson, B.F. and Baker, E.N. (1996) manuscipt in preparation.

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

Mowbray, S.L. (1992) *J. Mol. Biol.* **227,** 418-440.

Murray, M.J., Murray, A.B., Murray, M.B. and Murray, C.J. (1978) *Brit.Med.J.* **2,** 1113-1115.

Naidu, S.S., Svensson, U., Kishore, A.R. and Naidu, A.S. (1993) *Antimicrob.Agents Chemother.* **37,** 240-245.

Nicholson, H.H., Anderson, B.F., Tweedie, J.W. and Baker, E.N., manuscript in preparation.

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

Oh, B-H., Pandit, J., Kang, C-H., Nikaido, K., Gokcen, S., Ames, G.F-L. and Kim, S-H. (1993) *J. Biol. Chem.* **268,** 11348-11355.

Olson, A.J., Bricogne, G. and Harrison, S.C. (1983) *J. Mol. Biol.* **171,** 61-93.

Osbourne, T.B. and Campbell, G.F. (1900) *J.Amer.Chem.Soc.* **22,** 422-450.

Oseas, R., Yang, H-H., Baehner, R.L. and Boxer, L.A. (1981) *Blood* **57,** 939-945.

Palmiter, R.D., Behringer, R.R., Quaife, C.J., Maxwell, F., Maxwell, I.H. and Brinster, R.L. (1987) *Cell* **50,** 435-443.

Park, I., Schaeffer, E., Sidoli, A., Baralle, F.E., Cohen, G.N. and Zakin, M.M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3149-3153.

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

Pflugrath, J.W. and Quiocho, F.A. (1988) *Nature* **314,** 257-260.

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

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

Quiocho, F.A. (1990) *Phil.Trans.Royal.Soc.* **326,** 341-351.

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

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

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

Rochard, E., Legrand, D., Mazurier, J., Montreuil, J. and Spik, G. (1989) *FEBS letters* **255,** 201-204.

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

Sack, J.S., Saper, M.A. and Quiocho, F.A. (1989) *J. Mol. Biol.* **206,** 171-191.

Sack, J.S., Trakhanov, S.D., Tsigannik, I.H. and Quiocho, F.A. (1989) *J. Mol. Biol.* **206,** 193-207.

Sambrook, J., Fritsch, E.F. and Maniatis (1989) in *Molecular Cloning: A Laboritory Manual Second edn.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Sanchez, L., Calvo, M. and Brock, J.H. (1992) *Arch.Dis.Child.* **67,** 657-661.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *PNAS. (USA)* **74,** 5463-5467.

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

Schade, A.L., Reinhart, R.W. and Levy, H. (1949) *Arch.Biochem.Biophys.* 20, 170

Schade, A.L. and Caroline, L. (1944) *Science* **100,** 14-15.

Schaeffer, E., Lucero, M.A., Jeltsch, J-M., Pu, M-C., Levin, M.J., Chambon, P., Cohen, G.N. and Zakin, M.M. (1987) *Gene* **56,** 109-116.

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

Schreiber, G., Dryburgh, H., Millership, A., Matsuday, Y., Inglis, A., Phillips, J., Edwards, K. and Maggs, J. (1979) *J. Biol. Chem.* **254,** 12013-12019.

Schulz, G.E., Muller, C.W. and Diederichs, K. (1990) *J. Mol. Biol.* **213,** 627-630.

Sharff, A.J., Rodseth, L.E., Spurlino, J.C. and Quiocho, F.A. (1992) *Biochemistry* **31,** 10657-10663.

Shirsat, N.V., Bittenbender, S., Kreider, B.L. and Rovera, G. (1992) *Gene* **1120,** 229-234

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

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

Spik, G., Brunet, B., Mazurier-Dehaine, C., Fontaine, G. and Montreuil, J. (1982) *Acta Paediatr.Scand.* **71,** 979-985.

Spurlino, J.C., Lu, G-Y. and Quiocho, F.A. (1991) *J. Biol. Chem.* **266,** 5202-5219.

Stehle, T. and Schulz, G.E. (1990) *J. Mol. Biol.* **211,** 249-254.

Stillman, T.J., Baker, B.J., Britton, K.L. and Rice, D.W. (1993) *J. Mol. Biol.* **234,** 1131-1139.

Stowell, K.M. Thesis, PhD, Massey University, 1990.

Stratil, P., Bobak, P. and Valenta, M. (1983) *Comp.Biochem.Physiol.* B74, 603

Stryer (1988)

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

Thibodeau, S.N., Lee, D.C. and Palmiter, R.D. (1978) *J. Biol. Chem.* **253,** 3771-3774.

Tomita, M., Bellamy, W., Takase, M., Yamauchi, K., Wakabayashi, H. and Kawase, K. (1991) *J.Dairy Science* **74,** 4137-4142.

Tomita, M., Matsue, M ., Matsuyama, J. and Kiyosawa, I. (1994) *Biosci.Biotech.Biochem.* **58,** 722-726.

Tronrud, D.E., Ten Eyck, L.F. and Mathews, B.W. (1987) *Acta Cryst.* **A43,** 489-501.

Valenti, P., Visca, P., Nicoletti, M., Antonini, G. and Orsi, N. (1985) in *Proteins of Iron Storage and Transport.* (Spik, G., Montreuil, J., Crighton, R.R. and Mazurier, J. eds) pgs 245-249. Elsevier, Amsterdam.

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

Van Dijk, M.C.M., Ziere, G.J., Boers, W., Linthorst, C., Bijsterbosch, M.K. and Van Berkel, T.J.C. (1991) *Biochem. J.* **279,** 863-870.

Van Snick, J.L., Masson, P.L. and Heremans, J.F. (1974) *J.Exp.Medicine* **140,** 1068-1084.

Vyas, N.K., Vyas, M.N. and Quiocho, F.A. (1988) *Science* **242,** 1290-1295.

Vyas, N.K., Vyas, M.N. and Quiocho, F.A. (1991) *J. Biol. Chem.* **266,** 5226-5237.

Ward, P.P., Lo, J-Y., Duke, M., May, G.S., Headon, D.R. and Conneely, O.M. (1992) *Bio/technology* **10,** 784-789.

Weber, I.T. and Steitz, T.A. (1987) *J. Mol. Biol.* **198,** 311-326.

Wenn, R.V. and Williams, J. (1968) *Biochem.J.* **198,** 69-74.

Westerhoff, H.V., Juretic, D., Hendler, R.W. and Zasloft, M. (1989) *PNAS .(USA)* **86,** 6597-6601.

Wienburg, E. (1984) *Physiol.rev.* **64,** 65-102.

Williams, J. and Moreton, K. (1980) *Biochem.J.* **185,** 483-488.

Williams, J., Grace, S.A. and Williams, J.M. (1982) *J.Biochem.* 201, 417

Yu, R. and Schryvers, A.B. (1993) *Microbial Pathogenesis* **14,** 343-353.

Yuen, M., Iyer, S. and Lonnerdal, B. (1993) *FASEB abstracts,* 931.

Zak, O., Aisen, P., Crawley, J.B., Joannou, C.L., Patel, K.J., Rafiq, M. and Evans, R.W. (1996) *Biochemistry* (submitted).

Ziere, G.J., van Dijk, M.C.M.. Bijsterbosch, M.K. and van Berkel, T.J.C. (1992) *J.Biol.Chem.* **267,** 11229-11235

Ziere, G.J., Bijsterbosch, M.K. and van Berkel, T.J.C. (1993) *J.Biol.Chem.* **268,** 27069-27075.

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