EXPRESSION AND CHARACTERISATION

OF THE N-TERMINAL HALF

OF HUMAN LACTOFERRIN

his thesis is submitted to Massey University as partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Biochemistry.

Catherine Louise Day

January, 1993

ACKNOWLEDGEMENTS

CONTENTS

Chapter One

INTRODUCTION

Part A   CLONING, CHARACTERISATION AND MUTAGENESIS OF THE N-TERMINAL HALF OF HUMAN LACTOFERRIN

Part A : Chapter One

MATERIALS AND METHODS

Part A : Chapter Two

CLONING AND EXPRESSION OF THE N-TERMINAL HALF OF HUMAN LACTOFERRIN

Part A : Chapter Three

CHARACTERISATION OF Lf$_N$

Part A : Chapter Four

MUTAGENESIS STUDIES ON HUMAN LACTOFERRIN

Part B   **CRYSTALLOGRAPHIC ANALYSIS OF Lf$_N$**

Part B : Chapter One

METHODOLOGY AND PROGRAM DESCRIPTION

Part B : Chapter Two

CRYSTALLISATION OF Lf_N

Part B : Chapter Three

FeLf_N STRUCTURE SOLUTION AND ANALYSIS

Part B : Chapter Four

ApoLf$_N$ DATA COLLECTION AND STRUCTURE SOLUTION

Chapter Two

CONCLUSIONS AND FUTURE DIRECTIONS

ABSTRACT

Lactoferrin is an 80 kDa iron binding protein which is found in human milk and other exocrine solutions. Each molecule contains two metal binding sites which each bind a single iron atom with high affinity. The properties of the two sites are slightly different. In an attempt to more fully understand the nature of these differences a construct for the high level expression of the N-terminal half of human lactoferrin (Lf_N) has been constructed and protein expressed from this construct has been purified and characterised. Characterisation of the recombinant protein has shown that the signal peptide is correctly removed from Lf_N and that an N-linked carbohydrate moiety is added to Lf_N. Lf_N has been shown to bind one iron atom and the spectral properties are very similar to those of Fe_2Lf. The most significant difference between hLf and Lf_N is in the pH stability of iron binding. Iron is released from Lf_N 2 pH units higher than from hLf.

In an attempt to understand the bases for this difference a structural analysis of Lf_N was initiated. Using deglycosylated protein high quality crystals of both iron free and iron saturated Lf_N have been grown. The structures of both $FeLf_N$ and $ApoLf_N$ have both been solved by molecular replacement using the coordinates from the N-lobe of Fe_2Lf as the starting model.

The structure of $FeLf_N$ has been refined using data between 8.0 and 2.0 Å. The current model has good geometry and is believed to accurately represent the structure of $FeLf_N$. The structure of $FeLf_N$ provides the highest resolution and most accurate structure of a member of the transferrin family. Analysis of the structure has shown that the folding pattern and the environment of the iron atom in $FeLf_N$ are very similar to the N-lobe of Fe_2Lf although several differences exist. Most of the differences seen are due to the absence of the C-lobe and the rearrangement of residues 315 - 327. The altered conformation of residues 315 - 327 and the changes in the solvent accessibility to other residues are believed to be responsible for the different iron binding and release properties of Lf_N.

Although the structure of $ApoLf_N$ is not complete analysis of this structure has shown that unlike the N-lobe of intact apo hLf the domains are closed in $ApoLf_N$. The structure of $ApoLf_N$ is very similar to that of $FeLf_N$ even though the crystal packing is quite different. In addition although the protein was believed to be iron free there is some density in the iron site which is unaccounted for at present. This study continues.

Several mutants of Lf_N have also been created. These mutants have shown that the carbohydrate groups attached to lactoferrin probably have a role in folding and secretion of lactoferrin by BHK cells. Several mutants involving changes to residues involved in metal and anion binding have also been created. These mutants have helped us begin to define the changes responsible for preventing iron binding in the C-lobe of melanotransferrin. In addition the role of arginine 121 has been investigated however further analysis of these mutants is required before the structural changes responsible for the different properties can be defined.

ABBREVIATIONS

A_{280}	Absorbance at 280 nm
A_{406}	Absorbance at 406 nm
A_{454}	Absorbance at 454 nm
ApoLf	Iron free native lactoferrin
ApoLf$_N$	Iron free Lf$_N$
ATP	adenosine triphosphate
BG	background
BHK	baby hamster kidney
bLf	bovine lactoferrin
bp	base pair
BRL	Bethesda research laboratories
BSA	bovine serum albumin
cDNA	complementary DNA
cfu	colony forming units
CIF	colony inhibitory factor
CML	chronic myeloid leukaemia
CM-sephadex	carboxymethyl sephadex
cpm	counts per minute
C-terminal	carboxy terminal
Cu$_2$Lf	copper saturated native lactoferrin
CuLf$_N$	copper saturated Lf$_N$
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEPC	diethylpyrocarbonate
DG	deglycosylated
DGLf$_N$	deglycosylated Lf$_N$
dGTP	deoxyguanosine triphosphate
DHFR	dihydrofolate reductase
DMEM	Dulbeco's modification of Eagle's medium
DMF	dimethylformamide
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
EDTA	ethylene diamine tetra-acetate
EEO	electroendosmosis
ELISA	enzyme linked immunoabsorbant assay

ESR	electron spin resonance
F's	structure factors
F_{calc}	calculated structure factors
F_{obs}	observed structure factors
F12	Hams-F12 medium
FCS	foetal calf serum
$FeLf_N$	iron saturated Lf_N
Fe_2Lf	iron saturated native lactoferrin
gDNA	genomic DNA
GF/A	glass fibre/A
GM	granulocyte macrophage
GM-CSF	granulocyte macrophage colony stimulating factor
HBS	HEPES buffered saline
HBV	hepatitis b virus
HEPES	N-2-hyroxethyl piperazine-N'-2-ethane sulfonic acid
hGH	human growth hormone
hLf	human lactoferrin
hTf	human transferrin
I	intensities
IPA	isopropanol
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
LB	Luria-Bertani
Lf	lactoferrin
Lf_N	the amino terminal half of human lactoferrin
Lf_{C50}	50 kDa carboxy terminal fragment from lactoferrin
Lf_{N30}	30 kDa amino terminal fragment from lactoferrin
LMP	low melting point
MEM	minimal essential medium
MES	2(N-morpholino) ethane sulfonic acid
MIR	multiple isomorphous replacement
MPD	2-methyl-2,3-pentanediol
MR	molecular replacement
MT-1	metallothionein-1
mtx	methotrexate
N-terminal	amino terminal
NTA	nitrilotriacetate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline

PCR	polymerase chain reaction
pdb	protein data bank
PEG	polyethylene glycol
pfu	plaque forming units
PNGase	peptide:N-glycosidase
RIA	radio immunoassay
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid]
SDS	sodium dodecylsulphate
SDS-PAGE	sodium dodecylsulphate polyacrylamide gel electrophoresis
SSC	standard saline citrate
sTf	serum transferrin
sT_N	the amino terminal half of serum transferrin
sTf_{N35}	35 kDa fragment from the amino terminal half of human lactoferrin
SV-40	simian virus 40
TA	Tris acetate
TAE	Tris acetate EDTA
TCA	trichloroacetic acid
TE	10 mM Tris, 1 mM EDTA
Tris	Tris-(hydroxymethyl) aminomethane
U-DNA	uracil containing DNA
UV	ultraviolet
X-gal	5-bromo-4-chlor-3-indoyl β-D-galactopyranoside

LIST OF TABLES

LIST OF FIGURES

Chapter One

INTRODUCTION

Iron (Fe) is an essential element for the growth and survival of all living cells. Iron in aqueous solution can be in either the ferrous, Fe (II) state, or the ferric, Fe (III) state. Under the conditions which prevail in the cell ferric iron is the stable form; however in this form iron is essentially insoluble and if the concentration of uncomplexed iron exceeds 10^{-17} M insoluble ferric hydroxides form (Aisen and Listowsky, 1980). In order to overcome the problem of maintaining iron in an available non-toxic form, most organisms, from bacteria to mammals, contain specialised proteins for the transport and storage of iron.

The iron-binding proteins of vertebrates can be considered in three classes. Most of the iron (~70%) in a normal human is found complexed in the iron containing proteins which have either catalytic or carrier functions. This class of proteins includes the haem containing cytochromes, oxygen carriers such as haemoglobin and myoglobin, and the non-haem iron-containing electron carriers eg.cytochromes b, c and c_1. The second class of iron-binding proteins includes the proteins of iron storage. In normal humans ~27% of the total body iron is complexed with ferritin which is the main iron storage protein (Smith *et al.*, 1983). Iron in ferritin exists as an iron-oxide-phosphate complex containing up to 4500 iron atoms surrounded by 24 ferritin monomers. The third class of iron containing proteins are those of iron transport and it is to this class that the transferrin family of proteins belongs. The transferrins are not as widely distributed as the other iron binding proteins and only ~0.1% of the total body iron is found complexed with these proteins (Smith *et al.*, 1983). Despite the low proportion of total iron in transferrins, these proteins are believed to have an important role in the transfer of iron between sites of storage and all other tissues. Serum transferrin (sTf) is the major iron transport protein, however the family includes ovotransferrin from egg whites, melanotransferrin from melanocytes, and lactoferrin (Lf) from mammalian milks and other secretions. A role for lactoferrin in iron transport has not been conclusively demonstrated and the term 'iron-transport protein' may be somewhat of a misnomer in the case of this protein. However, lactoferrin does share many properties with serum transferrin and binds iron in the same manner as other members of the family. A more appropriate name for the transferrins may be 'iron regulators' as they share the common function of controlling the levels of free iron in biological fluids through their ability to solubilise and sequester iron. There is a high level of homology between the four classes of transferrins and all of the secreted transferrins are glycoproteins with a molecular weight of approximately

80 kDa. The single exception is melanotransferrin which, while homologous to the other transferrins, is slightly larger due to the presence of a 25 amino acid C-terminal hydrophobic extension. This extension is believed to be involved in anchoring melanotransferrin to the plasma membrane of melanocytes (Rose *et al.*, 1986). In addition melanotransferrin binds only one iron atom while all the other members bind two (Baker *et al.*, 1992).

The work described in this thesis has been carried out in an attempt to advance the understanding of both the structure and function of lactoferrin and how the two are related. As a consequence this introduction will concentrate on the structure and putative functions of lactoferrin and it is not intended to be a complete review of the transferrins. However, where appropriate the general properties of the transferrins will be referred to in order to place lactoferrin in perspective.

1.1   Distribution and regulation of lactoferrin gene expression

Lactoferrin is a major protein in mammalian vertebrate milk and in a variety of other secretory fluids including tears, saliva, seminal fluid, gastric and cervical mucus; it is also found in the secondary granules of neutrophils (Williams, 1985). Lactoferrin from these various sources in humans appears to be identical based on immunological evidence (Moguilevsky *et al.*, 1985) and on protein and cDNA sequence data. Lactoferrin cDNA sequences from both mammary tissue (Rey *et al.*, 1990) and neutrophils (Rado *et al.,* 1987; Rado *et al.*, unpublished; Stowell *et al.*, 1991) have been determined and the few minor differences which have been discovered appear not to be significant (Stowell *et al.*, 1991). Spik *et al.* (1988) have sequenced the carbohydrate groups of lactoferrin from a variety of sources and noted some differences although the significance of these differences remains elusive.

The level of lactoferrin expression and the timing of expression in the mammary gland varies considerably between species. Human colostrum contains up to 6 mg/ml but later in lactation the level falls to 1-2 mg/ml (Lonnerdal *et al.*, 1976). Lactoferrin expression in cows is somewhat different with colostrum levels of about 1-5 mg/ml which fall later in lactation to only about 0.1 mg/ml (Smith & Schanbacher, 1977). In cattle the highest level of lactoferrin expression occurs during involution of the mammary gland when the level of lactoferrin in the secretory fluid rises to around 50-100 mg/ml (Welty *et al.*, 1976). This variation in the level of expression throughout lactation suggests that the genes are hormonally regulated.

Hormonal regulation of lactoferrin expression is also suggested by the finding that the major oestrogen-stimulated protein synthesised in the uterine tissue of the mouse is lactoferrin (Pentecost and Teng, 1987). Evidence for the tissue specific expression of lactoferrin has also been provided by this group (Teng *et al.*, 1989). They reported that in the uterus the synthesis and secretion of lactoferrin was dependent on the amount and timing of oestrogen administration, but as yet no correlation between the level of lactoferrin in the mammary tissue and the amount of circulating oestrogen has been observed. This group has subsequently cloned and characterised the 5' regions of both mouse and human lactoferrin (Liu and Teng, 1991; Teng *et al.*, 1992). Analysis of the sequence in the 5' regions has identified elements, within the promoter, that are probably involved in both constitutive and inducible gene expression. One of the elements, common to the 5' regions in both mice and humans, has been identified as an oestrogen responsive element. At this stage analysis of these regions is incomplete but it appears that the oestrogen stimulated response differs between mice and humans. Taken together these results present a complex pattern of lactoferrin expression that is both tissue and animal specific.

Variation in the level of lactoferrin expressed in normal and cancerous tissues has also been reported. Lactoferrin is expressed in both normal breast duct epithelium cells and neutrophils but their malignant counterparts have been found to have no lactoferrin or if present the level is considerably reduced (Charpin *et al.*, 1985; Rado *et al.*, 1987). Panella *et al.* (1991) have shown that the methylation pattern differs between normal and leukaemic cell lines. On this basis they suggest that methylation may play a role in controlling the level of lactoferrin expressed. The absence of lactoferrin from HL60 cells has also been examined (Johnston *et al.*, 1992). They found that the promoter region of lactoferrin was intact and suggested that the absence of lactoferrin was due to the absence of a critical transacting factor. These studies continue and it is hoped that eventually the factors controlling lactoferrin expression will be identified and that this will help define the function of lactoferrin.

1.2 Biological roles of lactoferrin

For many years a considerable amount of effort has been directed at determining the functions of lactoferrin. As a consequence of these investigations many different functions have been postulated for lactoferrin. In the following sections the putative functions will be reviewed. Unfortunately much of the evidence is indirect and in many cases contradictory.

1.2.1   **Bacteriostatic and bactericidal properties of lactoferrin**

In 1948 Alexander reported that babies fed breast milk suffered fewer and less severe episodes of gastric enteritis than did bottle fed infants. The implications of this study were that there was some factor(s) in breast milk that reduced the frequency and severity of enteric infections. In babies enteritis is associated with specific types of *E.coli*, but the serotypes identified in babies suffering from enteritis can also be found in healthy babies (Bullen *et al.*, 1972). This suggested that enteritis is associated with an increased population of *E. coli* rather than a new infection. Several studies have been directed at defining the components of human milk which confer resistance to enteritis. A role for lactoferrin in resistance to enteritis was suggested at an early stage following the identification of this protein in many secretory fluids and in 1966 Masson *et al.* reported that "iron-free protein (lactoferrin) plays a part in the defence of epithelial surfaces against microbial agression". One of the first studies to actually demonstrate the inhibitory affect of human milk and of lactoferrin in particular was published by Bullen *et al.* in 1972. These workers reported that, *in vitro*, the growth of *E. coli* 0111 was inhibited by milk and that the inhibition could be reversed by the addition of iron to the milk. Purified lactoferrin was shown to have a similar affect, although not quite as dramatic as that of whole milk. The affect of purified lactoferrin could also be abolished by saturation of the protein with iron. Addition of another component termed '*E. coli* specific antibody' to the purified lactoferrin was capable of restoring the level of antibacterial activity obtained with whole milk. These results suggested that lactoferrin, in combination with specific antibodies to *E. coli*, is responsible for the bacteriostatic effect of human milk. In the same study the suppressive effect of guinea pig milk in preventing growth of *E. coli* was monitored. Guinea pigs produce a milk which has quite high concentrations of lactoferrin and it was observed that when iron-free the milk prevented bacterial growth but when iron-saturated the affect was abolished (Bullen *et al.*, 1972).

In subsequent investigations lactoferrin has been reported to have bactericidal activities. Using concentrations of protein which were within the physiological range Arnold *et al.* (1977) reported that lactoferrin was capable of killing *V. cholerae* and *S. mutans*. This bactericidal activity was also reported to be reversible by iron saturation of the protein. In 1980 the range of organisms which could be killed by lactoferrin was extended (Arnold *et al.*, 1980) although considerable variation between strains and within some strains, depending on the growth conditions, was noted. Finkelstein *et al.* (1983) have questioned this work as their results, using the same bacterial strains, only demonstrated a bacteriostatic affect for lactoferrin.

By the 1980's the anti-bacterial properties of lactoferrin *in vitro* were quite well established; however if these affects were to be relevant in the infant it was necessary to demonstrate that lactoferrin can escape digestion in the gastrointestinal tract. Samson *et al.* (1980) reported that pepsin digestion *in vitro* (or low pH alone) released iron from milk and abolished its bacteriostatic properties. However, trypsin digestion of milk did not affect its iron binding or bacteriostatic properties. On the basis of these studies they postulated that milk, and lactoferrin in particular, did have a bacteriostatic role in the infant. These results are supported by those of Britton and Koldovsky (1980) who analysed the ability of gastric juices from preterm infants to degrade lactoferrin. In this study lactoferrin was found to be completely degraded at low pH but at pH 5.8 the breakdown was minimal. In the infant it has been reported (Mason, 1962) that feeding increases the pH of the gut and that the pH is probably between 5 and 6.5 for up to two hours after feeding. On the basis of these results it is reasonable to suggest that at least a fraction of the lactoferrin from human milk would survive digestion. This suggestion is supported by the results of Prentice *et al.* (1988) who claimed that the amount of lactoferrin in the stool of breast-fed infants was greater than that of formulae-fed controls. Additional evidence for the survival of at least some lactoferrin is provided by the finding that lactoferrin excretion declines in breast-fed infants with increasing post-natal age paralleling the decline of lactoferrin in milk (Davidson and Lonnerdal, 1988). Taken together these results all suggest that some lactoferrin survives digestion in the gut and that it could have an antibacterial role. This is also supported by the finding that almost all of the lactoferrin in milk is iron free (Brock, 1985; Lonnerdal, 1985) and thus capable of binding any free iron encountered.

A role for lactoferrin in protection of the mammary gland from infection has also been suggested. In particular the high levels of lactoferrin in the involuting bovine mammary gland suggest that lactoferrin has a protective role for the cow (Welty *et al.*, 1976). These results have not been confirmed for humans.

More recently controversy has centred on the actual mechanism by which lactoferrin prevents bacterial growth. In 1972 Bullen *et al.* reported that the presence of specific antibodies greatly enhanced the antibacterial properties of lactoferrin. Other workers have suggested that the interactions between these other proteins and lactoferrin are either hydrophobic or ionic in nature (Hekman, 1971; Mason, 1976) but more recently Akin *et al.* (1992) have suggested that covalent interactions may be involved. They suggest that disulphide bridges may be involved in linking other proteins to lactoferrin. Covalent interactions of this type

seem somewhat unlikely but the involvement of other proteins is quite plausible. *In vivo* evidence from trials using infant formula supplemented with bovine lactoferrin support this (Wharton *et al.*, 1992). If, as reported previously (Rainard, 1986), bovine lactoferrin can inhibit bacterial growth in the absence of any associated proteins addition of bovine lactoferrin to milk formula would be expected to alter the feacal flora of infants fed this formula. No difference in the feacal flora of infants fed a lactoferrin supplemented formula was noted (Wharton *et al.*, 1992). Although this is not conclusive as other factors may be involved it does suggest that the addition of bovine lactoferrin to milk formulae on its own does not alter bacterial growth in the infant.

While the bacteriostatic properties of lactoferrin are now accepted by most authors to be a consequence of the iron binding ability of apolactoferrin controversy still surrounds the putative bactericidal capacity of lactoferrin. Some researchers believe that this capacity is also due to iron deprivation (Arnold *et al*, 1977), although it is believed by most that if lactoferrin does have a bactericidal role it is due to a different mechanism. Of particular interest are the reports from a research group at the Morinaga Milk Company. This group have analysed the peptides produced by enzymatic hydrolysis of bovine lactoferrin and have identified a peptide which is 18 times more potent in preventing bacterial growth than intact lactoferrin (Tomita *et al.*, 1991). The peptide was shown to retain its activity in the presence of iron demonstrating that iron deprivation was not involved in the bactericidal properties of this fragment. Subsequently a similar peptide has been identified in the hydrolytic peptides of human lactoferrin (Bellamy *et al.*, 1992). The active peptide has been localised to the N-terminal region of the protein (residues 18 - 40 in hLf and 17 - 41 in bLf). Both peptides are rich in basic amino acids and it has been postulated that these peptides act by increasing cell permeability leading to disruption of salt gradients and loss of the cell contents. This theory is supported by the finding that the peptide must interact with the cell walls to exert its bactericidal properties (Tomita *et al.*, 1992). This remains an area of interest as the possible commercial applications are extensive.

1.2.2 Lactoferrin as an iron carrier

A role for lactoferrin as an iron carrier has also been proposed. It has been observed that exclusively breast-fed infants maintain adequate levels of iron longer than infants fed bovine milk based formula (Saarinen, 1978). This suggests that the iron in human milk is present in a more available form. The occurrence of a putative intestinal receptor specific for lactoferrin was first reported in 1979 by Cox *et al.* They showed that, in contrast to transferrin and ovotransferrin,

lactoferrin was able to donate iron to intestinal tissue but unable to release iron to reticulocytes. This suggested that lactoferrin may interact with specific binding sites on the brush border membrane and release iron to these cells. A 100 kDa protein capable of binding lactoferrin was subsequently identified from rabbit brush border cells by ligand blotting (Mazurier *et al.*, 1989). The ability of lactoferrin to bind to receptors in the brush border cells of rhesus monkeys has also been analysed and this is probably a better model for the human system (Davidson & Lonnerdal,1988). These experiments demonstrated the specific saturable binding of lactoferrin to rhesus monkey brush border cells and suggested that removal of the fucose groups resulted in a significant decrease in binding of lactoferrin to the cells. The role of fucose had been implicated earlier in receptor binding when it was found that human lactoferrin was rapidly removed from mouse blood by a receptor specific for α 1-3 linked fucose (Prieels *et al.*, 1978). In this study periodate oxidation or treatment of lactoferrin with glycosidases considerably prolonged clearance rates suggesting that the carbohydrate moiety is required for recognition by the receptor. However it is equally likely that these treatments may have altered the protein in some other way which prevented uptake. Davidson & Lonnerdal (1989) have also shown that brush-border membrane vesicles accumulated iron efficiently from lactoferrin half-molecules. Fragments of lactoferrin are likely to occur in the gastrointestinal tract and therefore may contribute to the absorption of iron from human milk. Another location where specific lactoferrin receptors have been found is on mouse small intestinal brush border membrane sections. This receptor was found to bind several lactoferrins (human, bovine and mouse) although it had a higher affinity for mouse lactoferrin (Hu *et al.*, 1988). If lactoferrin does act as a source of iron for the infant the problem remains as to how the iron is released from the protein. The conventional view of iron release from transferrin is that the iron is released spontaneously at the lower intracellular pH (5-6) following receptor mediated internalisation of transferrin. Such a mechanism could presumably not apply to lactoferrin as reduction of the pH alone does not remove iron from lactoferrin until pH 3.0. Recently, however, studies on transferrin have shown that when transferrin is complexed with the receptor release of iron from the protein is promoted in the endosome (Bali *et al.*, 1991; Bali and Aisen, 1992). This type of iron removal could presumably apply equally well to lactoferrin irrespective of any pH affect. Unfortunately this aspect of lactoferrin has not been studied yet and until a suitable *in vitro* system is developed it is likely to remain poorly understood.

Lactoferrin receptors have also been reported to occur on the human pathogen *Neisseria meningitides*. In contrast to most bacteria Neisseria do not produce

siderophores but instead they appear to have specific receptors for both transferrin and lactoferrin. Schryvers & Morris (1988) have successfully isolated a protein approximately 105 kDa in size which specifically binds lactoferrin. This putative receptor was able to discriminate between human lactoferrin and human transferrin and between lactoferrin from different species.

Lactoferrin as a growth promoter

A role for lactoferrin as a growth promoter has been proposed. It is well known that transferrin is essential for the growth of many cell lines and that cell surface receptors for transferrin are expressed on proliferating cells *in vitro* and *in vivo* (Huebers and Finch, 1987; Casey *et al.*, 1989). A similar role for lactoferrin has been suggested based on the report that lactoferrin had a greater ability to stimulate BriT cell (a human B lymphocytic cell line) growth than did transferrin (Hashizume *et al.*, 1983). Evidence has also been reported supporting the presence of lactoferrin-specific receptors on both a cultured myeloid leukaemic cell line (Hashizume *et al.*, 1983) and a human adenocarcinoma cell line HT29 (Roiron *et al.*, 1989). Possibly more convincing is the report by Mazurier *et al.* (1989) which demonstrated that phytohaemaglutinin-stimulated human peripheral blood lymphocytes express surface receptors which bind human lactoferrin. Once the receptors had been synthesised lactoferrin binding was found to stimulate growth. The stimulatory affect of lactoferrin was concentration and iron status dependent; maximal stimulation occurred at a concentration of 0.014 mg/ml using 30% saturated lactoferrin. These results clearly demonstrated that lactoferrin possesses a growth stimulatory activity for phytohaemaglutinin-stimulated lymphocytes and that the activity is similar to that previously described for human serotransferrin (Brock, 1981).

Evidence that lactoferrin may have a role in cell growth, and that the role is similar to that of serum transferrin, is also provided by sequence homology studies. At the N-terminus of both serum transferrin and hLf there is a region, which corresponds to exon 2, that shows a similar pattern of conserved and divergent changes to that seen in the Blym-1 transforming genes (Diamond *et al.*, 1984). Diamond *et al.* postulated that the Blym genes may promote cell proliferation via a pathway similar to transferrin. This theory was supported by the crystal structure of rabbit sTf which showed that the conserved region is on the surface of the protein where it could interact with a receptor (Bailey *et al.*, 1988). As the configuration of these residues is similar in hLf (Anderson *et al.*, 1989) the analogy between Blym and sTf can be extended to include lactoferrin. Although there is no direct evidence that this region is involved in promoting cell

proliferation the analogy is quite suggestive. Once a receptor for lactoferrin is isolated mutagenesis studies should allow the role of the conserved residues, and the role of lactoferrin as a growth promoter in general, to be evaluated.

1.2.4 Lactoferrin and inflammation

Besides secretory fluids the other main location of lactoferrin is in the secondary granules of neutrophils (Baggiolini *et al.*, 1970). Neutrophils are phagocytic cells and their primary function is to engulf invading microorganisms, particularly bacteria. They are capable of responding to stimuli and migrating to the site of damage where they are an important part of the acute inflammatory response. The role of lactoferrin in the neutrophils is not clear although it is known that during inflammatory states plasma lactoferrin levels increase and the lactoferrin content of the secretory granules declines suggesting that it may have a role in the plasma at this time (Malmquist *et al.*, 1978). On this basis a role for neutrophil derived lactoferrin in inflammation, immune responses, myelopoiesis and as a regulator of oxidation has been proposed. The role of lactoferrin in inflammation will be discussed first.

Polymorphonuclear leucocytes (PMN), including neutrophils, degranulate, aggregate and are rendered immobile following exposure to high concentrations of chemotactic stimuli. Experiments by Lash *et al.* (1983) have suggested that this response may in part be related to the release of lactoferrin from the specific granules of neutrophils. Once released from the neutrophils lactoferrin is believed to bind to the the surface of PMN altering the surface properties of these cells (Boxer *et al.*, 1982). In particular their experiments suggest that binding of lactoferrin to the surface of PMN's increases cell 'stickiness' resulting in a variety of cell-cell interactions. This correlates with the observation that during inflammation there is an increased adherence of PMN cells to the endothelial cells which serves to retain them at sites of inflammation and thus amplify the inflammatory response (Oseas *et al.*, 1981).

Besides regulating neutrophil adherence and migration the lactoferrin released from neutrophils has also been suggested to have an antibacterial role (van Snick *et al.*, 1974). It is postulated that lactoferrin binds free iron, which may have been derived from transferrin, and that once iron saturated the lactoferrin binds to macrophages. The macrophage bound lactoferrin is then removed by the reticuloendothelial system (RES). This can result in localised anaemia. This function of lactoferrin was questioned by Baynes and Bezwoda (1992) after they studied bone marrow transplant patients exhibiting a hyperferritinemic response

but they could not detect significant levels of lactoferrin in the plasma of these patients.

Although this area is rather contradictory and confusing a general role for lactoferrin in inflammation is supported although the exact mechanism is poorly defined. Most likely the causes and affects are complex and as long as *in vitro* models are inadequate the role of Lf in this response is likely to remain elusive.

1.2.5 Lactoferrin: oxidant or antioxidant ?

A role of lactoferrin as a catalyst of OH· production was suggested by Ambruso and Johnston (1981). Hydroxy radicals can act to destroy most known biomolecules and it was consequently proposed that lactoferrin may act as an anti-microbial agent by catalysing hydroxyl radical formation. Bannister *et al.* (1982) also found evidence to support this hypothesis. Following this initial suggestion several studies have shown that lactoferrin is in fact a very poor catalyst of hydroxyl radical formation (Winterbourn, 1983; Gutteridge, 1983; Britigan *et al.*, 1989; Baldwin *et al.*, 1984). In contrast these authors propose that lactoferrin actually complexes free iron, preventing hydroxyl radical formation and lipid peroxidation. It is suggested that the earlier reports used lactoferrin which was contaminated by complexing agents such as EDTA or NTA which are known to affect the properties of lactoferrin (Winterbourn, 1983). This work has been supported by Monteiro and Winterbourn (1988) who have shown that lactoferrin binds the iron released from ferritin. Transferrin was found to have a similar role although lactoferrin retains iron in conditions under which iron is released from sTf suggesting that lactoferrin may be a more potent antioxidation agent. This work has been extended and a role for neutrophil lactoferrin in trapping iron released from ingested microorganisms and removing it from sites of inflammation has been suggested (Molloy and Winterbourn, 1990). This would prevent the free iron from catalysing undesirable oxidation reactions and also prevent growth of any microorganisms that survived the killing process. Another role for lactoferrin in the protection of monocytes has also been suggested (Britigan *et al.*, 1991). It is proposed that lactoferrin binds to a lactoferrin receptor on monocytes, as reported by Birgens *et al.* (1983), preventing hydroxyl damage to these and neighbouring cells.

1.2.6 Lactoferrin and myelopoiesis

The proliferation of granulocyte-macrophage progenitor cells is controlled by both positive and negative feedback mechanisms. One of the factors with an inhibitory

role, called colony inhibitory factor (CIF), has been identified as lactoferrin (Broxmeyer *et al.*, 1978). Previously CIF had been shown to have a role in inhibiting granulopoiesis *in vivo* (Broxmeyer, 1978) and limiting colony formation *in vitro* (Broxmeyer and Ralph, 1977). As CIF is now known to be lactoferrin these functions are attributed to lactoferrin. The actual mechanism involved is poorly understood although lactoferrin appears to decrease the production and release of granulocyte-macrophage colony stimulating factors (GM-CSF) from normal and leukaemic monocytes (Broxmeyer *et al.*, 1978). It has been reported that apolactoferrin is more effective (Broxmeyer *et al.*, 1978) and that lactoferrin exerted its inhibitory affect after it had bound to cells (Broxmeyer *et al.*, 1980). Although this study appears quite conclusive Winton *et al.* (1981) have been unable to repeat these experiments and have thus raised questions about the inhibitory affect of lactoferrin. Several features of Broxmeyer's study are quite surprising. Firstly the requirement for iron saturation of lactoferrin was not expected given that transferrin in the blood is not iron saturated (Halliwell and Gutteridge, 1990) and that the amount of free iron available is limited. At physiological pH both transferrin and lactoferrin have a similar affinity for iron so under these conditions lactoferrin could not take advantage of its ability to bind iron more strongly at lower pH. A possible explanation for the contradictory results is that the apolactoferrin used by Broxmeyer and coworkers was prepared by dialysis against low pH buffers and this may have irreversibly damaged the protein. In addition Broxmeyer and coworkers have always used human lactoferrin and mouse cells. If, as they suggest, the mechanism depends upon receptor binding it is quite surprising that an effect is seen with human lactoferrin as most receptors are species specific (Davidson and Lonnerdal, 1989).

Despite these complications a role for lactoferrin in myelopoiesis is generally supported. The likely course of events involved in myelopoiesis is that a specific stimulus induces the release of lactoferrin from the granules of neutrophils. The free lactoferrin then binds to specific receptors on sub populations of monocytes and macrophages suppressing the production of granulocyte-macrophage colony stimulating factors (GM-CSF). The absence of GM-CSF inhibits the proliferation and differentiation of progenitor cells of the GM lineage. Lactoferrin has no direct affect on the progenitor cells themselves (Broxmeyer, 1983).

1.2   Structure of lactoferrin

Primary structure

A considerable amount of sequence data is now available for the transferrin family.

The genes sequenced and the origin of the sequence data, cDNA or protein, is summarised in Table 1.

Table 1 Transferrin genes sequenced to date.

Gene	Type of sequence	Reference
human serum transferrin	protein	MacGillivray et al., 1983
	cDNA	Park et al., 1985
rabbit serum transferrin	cDNA	Banfield et al., 1991
chicken ovotransferrin	cDNA	Jeltsch and Chambon, 1982
	protein	Williams et al., 1982
mouse lactoferrin	cDNA	Pentecost and Teng, 1987
bovine lactoferrin	cDNA	Mead and Tweedie, 1991
		Pierce et al., 1991
human lactoferrin	protein	Metz-Boutigue et al., 1984
	cDNA	Powell and Ogden, 1990; Rey et al., 1990
		Rado et al., 1987; Rado et al., unpublished
porcine lactoferrin	cDNA	Lydon et al., 1990
Xenopus laevis transferrin	cDNA	Moskaitis et al., 1991
melanotransferrin	cDNA	Rose et al., 1986
Horn worm transferrin	cDNA	Bartfield and Law, 1990
Horse transferrin	cDNA	Carpenter et al., unpublished

All members of the transferrin family sequenced so far demonstrate a high degree of similarity with approximately 65% amino acid identity between the lactoferrins and 55 - 60% identity between the lactoferrins and the transferrins (Mead, 1992). In addition there is a high level of sequence identity between the N- and C-terminal halves of all transferrin gene sequences determined to date. In human lactoferrin there exists ~40% amino acid identity between the two halves. This homology suggests that not only have the transferrins evolved from a common ancestor but that they are probably also the result of a gene duplication event which occurred at an early stage in the evolution of these proteins (Bowman et al., 1988).

In addition to the cDNA sequences listed above 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. These three genes differ in size but the position of the seventeen intron/exon boundaries are highly conserved. The major difference in the size of the genes (human transferrin is 33 kb while the ovotransferrin and mouse

lactoferrin genes are 10.6 and 12 kb respectively), is due to variation in the size of the introns. Not only is the position of the intron/exon boundaries conserved between genes but they also have homologous positions in the N- and C-terminal halves. This provides strong evidence for the theory that 'gene duplication' was the primary event involved in generating this bilobal protein (Bowman *et al.*, 1988).

Tertiary structure

The first transferrin structure solved was that of iron saturated human lactoferrin (Fe$_2$Lf) (Anderson *et al.*, 1987). A cartoon diagram of this structure is shown in Fig. 1.

Fig. 1 Ribbon structure of iron saturated lactoferrin.
The diagram was prepared using the program Ribbon (Richardson, 1985; Priestle, 1988). α-helices are shown as green spirals and β-sheets are shown as pink arrows.

Although the initial structure was not fully refined and data to only 3.2 Å were used it provided a considerable amount of information about the overall configuration of lactoferrin and the transferrins in general. The main points to note are that the N- and C- terminal halves form two separate globular lobes linked by a short helix and that each lobe contains a single iron binding site. The iron atom was found to be coordinated to four protein ligands (2 Tyr, 1 Asp and 1 His) and the remaining two ligands appeared to be provided by the adjacent anion. Each lobe could be further divided into two domains, which are based on 5 or 6 central

β-strands against which several α-helices are packed, and between which the iron atom is bound.

With the aid of the Fe₂Lf structure the structure of diferric rabbit sTf was completed, and reported (Bailey *et al.*, 1988). The overall folding pattern of transferrin was found to be very similar to that of Fe₂Lf. The main difference between the two structures was found to be in the orientation of the two lobes. If the N-terminal halves of the two proteins are superimposed a rotation of 15° is required to obtain the best match of the C-lobes (Baker and Lindley, 1992). This difference, together with other more subtle differences between the two structures may account for the differences in the metal binding properties of the two proteins (Section 1.4). After these initial moderate resolution structures had been published a refined structure of human lactoferrin at 2.8 Å was reported (Anderson *et al.*, 1989). In this structure the details of protein folding and the environment of the iron site were shown with greater accuracy. A stylised representation of the metal binding site in the two lobes of hLf is shown in Fig. 2.

Fig. 2. Stylised representation of the iron binding site in lactoferrin.

The ligands involved in iron binding in the N-lobe together with arginine 121 and helix 5, which interact with the anion, are shown.
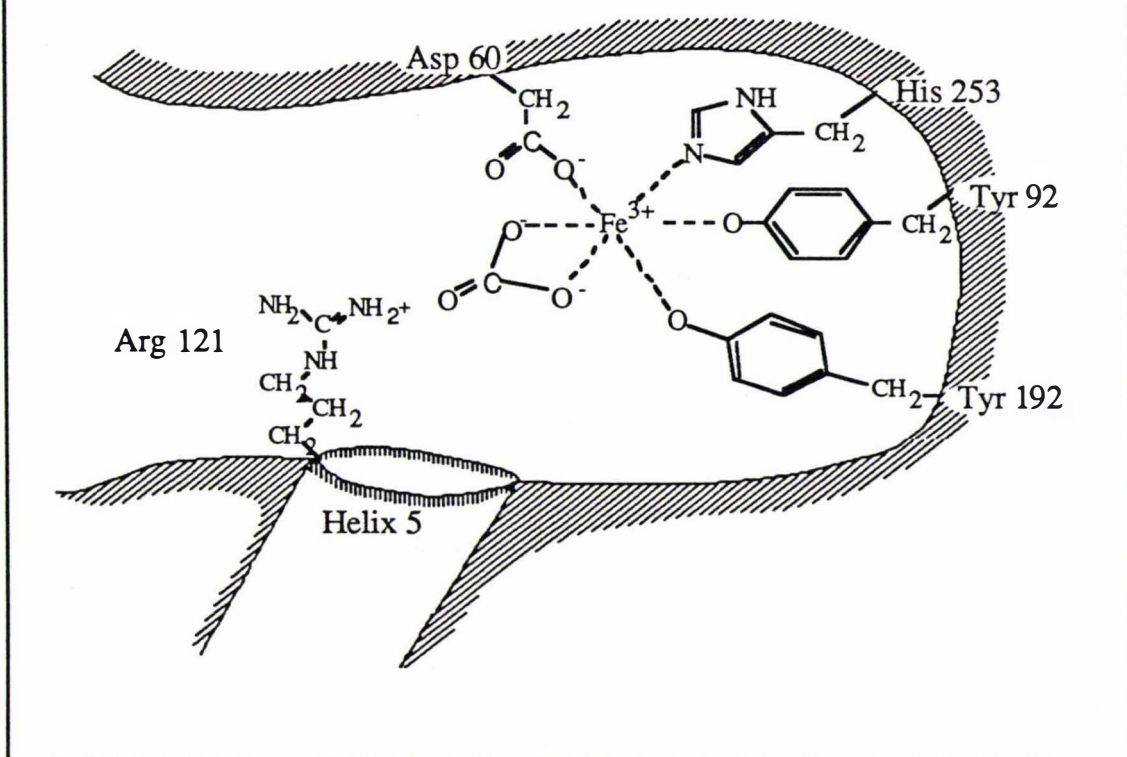
The local environment of the two sites is very similar and in both cases four of the ligands are provided by the protein. The protein ligands are provided by very different parts of the structure; one ligand, Asp 60 (395) from domain 1; one

ligand, Tyr 192 (528) from domain 2; and two ligands, Tyr 92 (435) and His 253 (597) from the β-strands that run across the back of the iron site connecting the two domains. The two remaining ligands are provided by the anion, believed to be carbonate, which is bound in a bidentate manner. The refined structure suggested that arginine 121 and the N-terminus of helix 5 interact with the anion and stabilise its position (Anderson *et al.*, 1989). The corresponding features are also present in the C-lobe. These interactions were confirmed when the structure of the N-terminal half of rabbit sTf using data to 2.3 Å was published (Sarra *et al.*, 1990). In sTf hydrogen bonds are formed between the anion and the sidechains of residues corresponding to Arg 121 and Thr 117, and to the mainchain nitrogen atoms of Ala 123 and Gly 124.

The structure of iron free human lactoferrin (ApoLf) was the next transferrin structure to be solved (Anderson *et al.*, 1990; 1991). This structure was particularly interesting as domain 2 of the N-lobe (N2) was found to be rotated 53° relative to domain 1 of the N-lobe (N1). This rotation causes the N-lobe binding cleft to be wide open. Domain opening appears to involve a simple hinge movement at the back of the iron site preserving the individual domain structures. No corresponding structural change occurs in the C-lobe of ApoLf. The C-lobe has essentially the same structure as in Fe_2Lf although there is no metal or carbonate present in the binding site. The domain opening in the N-lobe of ApoLf resembles the 'hinge bending' conformational changes observed in the bacterial periplasmic binding proteins. The structure of both liganded and unliganded forms of these proteins have been determined (Quiocho, 1990, Sack *et al.*, 1989a; Sack *et al.*, 1989b) although in no case have both open and closed states been determined for the same protein. Comparison of the bacterial binding protein structures with those of human lactoferrin has allowed a model for the mechanism of iron (and ligand) binding to be proposed. A schematic representation of this model is shown in Fig. 3 (taken from Baker *et al.*, 1990). Using this model as a basis it is proposed that an equilibrium exists between the open and closed unliganded forms and that the open form binds the ligands (anion and metal in the case of lactoferrin). Ligand binding causes the domains to close and once bound a specific trigger is required to allow opening to occur again (Section 1.4). The only form not seen to date in the structures of lactoferrin is the open ligand bound form although this is believed to occur. The model presented here is believed to be essentially correct although it provides little information about the order and kinetics of iron binding. These features are discussed in the next section.

Fig. 3 Model of the mechanism of iron binding and release.

Open ligand free

Open ligand bound

Closed ligand free

Closed ligand bound
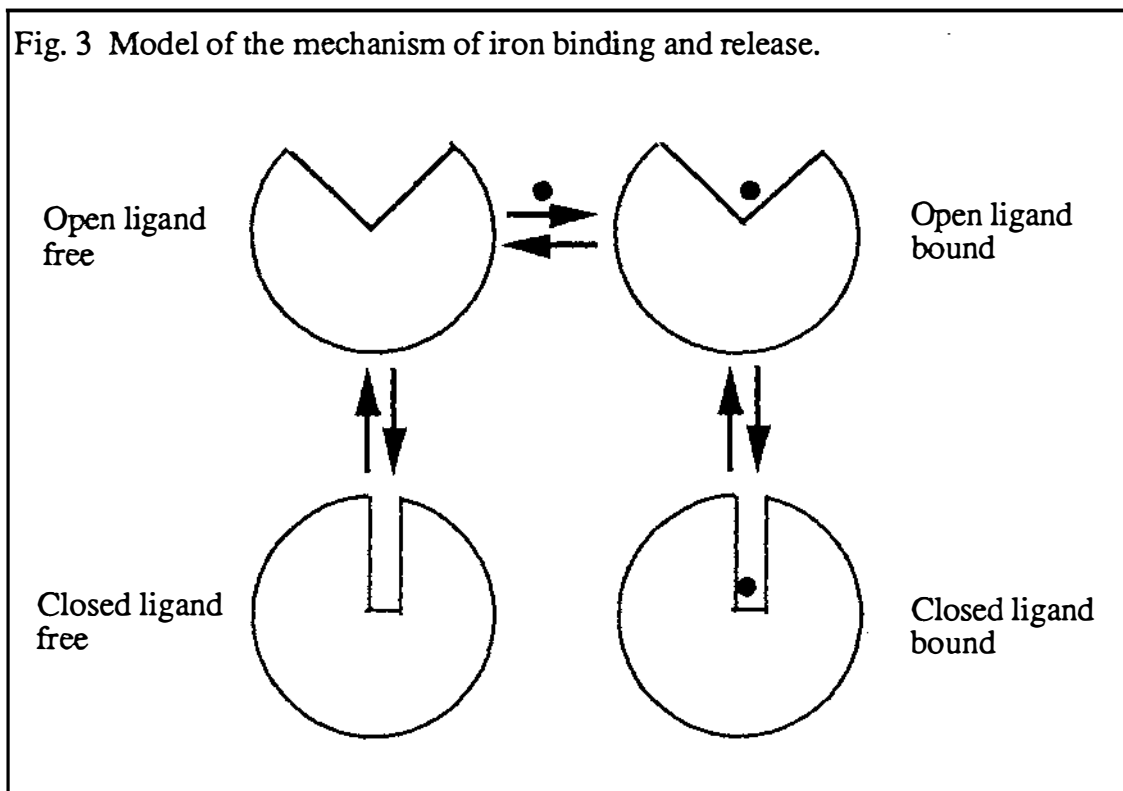
1.4 Binding properties of lactoferrin

Early studies on cation and anion binding by the transferrins were mainly carried out with serum transferrin (Aisen *et al.*, 1967). One of the first studies describing the binding properties of lactoferrin was that of Aisen and Leibman (1972). These authors analysed the spectroscopic properties of lactoferrin and demonstrated that the properties of lactoferrin were very similar to those of serum transferrin (Aisen *et al.*, 1967). Their studies suggested that both lactoferrin and transferrin contained two equivalent non-interacting metal binding sites and suggested that Lf and Tf had probably evolved from a common ancestor and hence belonged to the same protein family. These studies also suggested that the iron binding ligands included 2 tyrosines and 1 nitrogen containing amino acid, possibly histidine. The involvement of these residues in iron binding was confirmed when the structure of Fe_2Lf was determined (Anderson *et al.*, 1987). Although lactoferrin and transferrin were found to have many properties in common the strength of iron binding was found to be significantly different. Human lactoferrin bound iron much more tightly than serum transferrin and in fact when the K_D for the first Fe(III) ion bound to hLf was measured at pH 6.4 it was reported to be 300 times greater than that obtained for sTf at pH 6.7 (Aisen and Leibman, 1972).

One of the most intriguing features of iron binding by the transferrins is the obligatory requirement for the concurrent binding of a suitable synergistic anion. In 1975 Bates and Schlabach reported that "in the absence of synergistic anions the affinity of the specific metal binding sites of transferrin for Fe^{3+} is so low as to not

compete favourably with hydrolytic polymerisation and non-specific binding". As a consequence of this study the essential role of the anion was established. Bates and Schlabach extended their investigation by examining the capacity of a range of anions to act in a synergistic fashion allowing iron-binding to transferrin (Schlabach and Bates, 1975). Following studies of over 30 anions they concluded that if the anion is to promote iron binding it must possess a carboxylate group and a proximal electron donor group, and obey certain steric limitations. Based on these results they proposed the 'interlocking sites' hypothesis in which the anion binds to the metal through the electron donor group and also to a cationic group on the protein through the carboxylate group.

The next step in understanding the mechanism of iron binding was to determine the order of metal and anion binding. Using ^{13}C NMR it was shown that weak anion binding occurred in the absence of the metal and that specific anion binding may be a preliminary step in iron binding (Zweier et al., 1981). These studies also demonstrated that the anion bound directly to both the iron atom and an arginine residue. The involvement of an arginine residue in iron binding had previously been suggested on the basis of chemical modification studies (Rogers et al., 1978) and homology studies with other proteins which bind anionic substrates or cofactors (Riordan et al., 1977). The involvement of an arginine in anion binding was confirmed by Harris (1985). He also suggested that the metal and anion binding sites should be considered as a single site and that anion binding led to the formation of a reactive intermediate to which the metal ion could bind. In this study it was postulated that binding of the metal led to the loss of several hydrogen bonds within the protein but the complex is stabilised by the formation of several, more stable, ligand-metal bonds. Although the actual mechanism of iron binding has not been elucidated, the role of an arginine in stabilising the position of the anion, and the overlapping nature of the metal and anion sites have been confirmed by crystallographic studies of Fe_2Lf and Fe_2Tf (Anderson et al., 1987; 1989; Bailey et al., 1988; Sarra et al., 1990).

While iron, as Fe(III), is bound most strongly to the transferrins, other metals ions, including Cr(III), Mn(III), Co(III) and Cu(II), can bind to lactoferrin and transferrin in the place of iron (Ainscough et al., 1980). Analysis of the various metal transferrin complexes has demonstrated that while lactoferrin and transferrin are very similar, differences do exist. For example it is possible to prepare monoferric serum transferrins, in which either the N- or the C-terminal site is selectively saturated, and to distinguish the two sites by the ESR spectra obtained (Aisen et al., 1978). Many other studies on serum transferrin have also allowed the two sites to be distinguished. In contrast, it has not proved possible to prepare

monoferric lactoferrins, and many of the other metal-binding studies have been difficult to repeat with lactoferrin. Most metals examined, including some large lanthanides and actinides (Th^{4+}), bind to both lobes although there are differences in the rate and strength of binding (C.Smith, *pers. comm.*). In particular when Cr(III) or vanadyl are bound to lactoferrin the sites can be distinguished (Ainscough *et al.*, 1980; Mazurier *et al.*, 1983). More recently differences between the two lobes of lactoferrin have been highlighted when different metals have been bound with anions other than carbonate (Shongwe *et al.*, 1992).

Most of the results discussed in this section were obtained before publication of the crystallographically determined structure. The crystal structure has allowed many points to be clarified, some of which have been noted above. Other questions remain unanswered. In particular the mechanism of iron release remains somewhat of an enigma although several different factors have been implicated. One of the first factors considered to be important was reduction. It was suggested that loss of iron from transferrin was associated with reduction of Fe^{3+} to Fe^{2+} and that as Fe^{2+} is only bound weakly to transferrin it would not reassociate with the protein without prior oxidation (Kojima and Bates, 1979; Chasteen *et al.*, 1983; Harris, 1983). In another study Kretchmar and Raymond (1986) showed that iron was lost more readily from the N-terminal site and suggested that this was due to the greater flexibility of this lobe allowing the required conformational changes to occur. These authors also showed that the rate of iron release was dependent on the ionic strength of the environment (Kretchmar and Raymond, 1988). Extrapolation of their results actually suggested that no iron would be lost from transferrin in the absence of salt. These results suggest that electrostatic interactions and/or conformational changes, induced by the ions in solution, may play a role in promoting iron release; however they provide little information about the actual mechanism involved. Several groups have speculated that one of the first steps in iron release might be protonation of the carbonate anion (Williams and Woodworth, 1973; Sarra *et al.*, 1990). They suggest that protonation would disrupt the anion site and that once protonated the anion may bind to the metal in a monodentate manner, similar to that seen in the structure of dicupric lactoferrin (Smith *et al.*, 1992). This mechanism for the first steps in iron release is quite plausible but there is little direct evidence to support it.

The bases for the subtle differences in binding and release between the two metal binding sites in lactoferrin (and other transferrins) are not completely clear, although some clues are provided by the crystal structure. Firstly the individual domains of the lobes superimpose significantly better than the complete lobes indicating that the extent of closure of the lobes differs. These differences have the

effect of making the C-lobe more closed over the iron site than the N-lobe (Anderson *et al.*, 1989) and thus may partly explain the greater thermodynamic stability of, and slower release of iron from this lobe compared to the N-lobe of lactoferrin (Brock, 1985; Kretchmar & Raymond, 1986). The C-lobe has several additional disulphide bridges and these are also likely to contribute towards the greater thermodynamic stability of the C-lobe.

In serum transferrin the differences in binding of the two lobes have been extended to suggestions that the two halves of the protein have separate functional roles (Fletcher and Huehns, 1968). The hypothesis proposed is that one binding site preferentially delivers iron to non-erythroid cells while the other site delivers iron to reticulocytes and is concerned with iron storage. The proposal is controversial, however, and other researchers (Heubers *et al.*, 1978 and 1984) claim that the rate of iron donation from the two lobes is identical. The idea has been raised again by the structure of apolactoferrin (Anderson *et al.*, 1990) in which one lobe is closed and the other open. This could be consistent with different functional roles for the two lobes. In addition the observation that release of iron from sTf is modulated by interaction with the Tf receptor (Bali *et al.*, 1991), and that release from the C-lobe is primarily affected (Bali and Aisen, 1991), supports this proposal.

1.5 Fragment studies of lactoferrin

In light of the previous sections there are a number of questions which could be addressed through a study of lactoferrin fragments in which the two metal binding sites have been physically separated.

- ❑ Do the binding and release properties of the two individual lobes differ and are these properties affected by joining them together in a bilobal fashion ?
- ❑ Can the spectroscopic properties of the two lobes be distinguished?
- ❑ Do the two lobes of lactoferrin have separate functional roles?
- ❑ Can a receptor binding fragment be isolated?

Several different proteolytic enzymes have been used in attempts to produce functional fragments of lactoferrin. Bluard-Deconinck *et al.* (1978) produced the C-terminal half by pepsin digestion of human lactoferrin at pH 3.0. They were also able to produce, by digestion with both trypsin and chymotrypsin, fragments representing the N- and C-terminal halves of human lactoferrin. Unfortunately these fragments were produced at very low yield, with only between 1 and 3% of the protein recovered, precluding a complete analysis. Legrand *et al.* (1984) altered the conditions and obtained a higher yield. The two fragments produced

were a 30 kDa N-tryptic fragment (Lf$_{N30}$) and a 50 kDa C-tryptic fragment (Lf$_{C50}$). Characterisation of the fragments showed them to comprise the following residues; Lf$_{N30}$, residues 4 - 281; Lf$_{C50}$ residues 282 - 703.

Iron release studies using these fragments showed that Lf$_{C50}$ released iron at pH 4.5, while Lf$_{N30}$ released iron at pH 6.2 compared to pH 3.0 for hLf. When Lf$_{N30}$ and Lf$_{C50}$ were allowed to reassociate the iron release curve of the mixed proteins was almost identical to native diferric human lactoferrin (Legrand *et al.*, 1990). These results suggest that the separate N- and C-terminal halves exhibit a lower stability towards iron binding than reassociated or native human lactoferrin and they imply that interactions between the two lobes help stabilise the two iron-binding sites (Legrand *et al.*, 1990). The lesser stability of the N-terminal fragment compared to the C-terminal fragment suggests that residues 282 - 333, which are missing from Lf$_{N30}$, have a role in stabilising iron binding as Lf$_{C50}$ contains the equivalent of these residues and it binds iron more strongly. Residues 282 - 333 form part of one wall of the cleft and residues 296, 301 and 303 are involved interdomain contacts. The residues missing from Lf$_{N30}$ compared to a true half molecule (residues 1 - 333) are highlighted in Fig. 4.

Fig. 4 Comparison of Lf$_{N30}$ and a fragment representing the entire N-lobe. Residues 1 - 281 which correspond to Lf$_{N30}$ are shown in blue and the rest of the N-lobe, residues 282 - 333, is shown in pink.
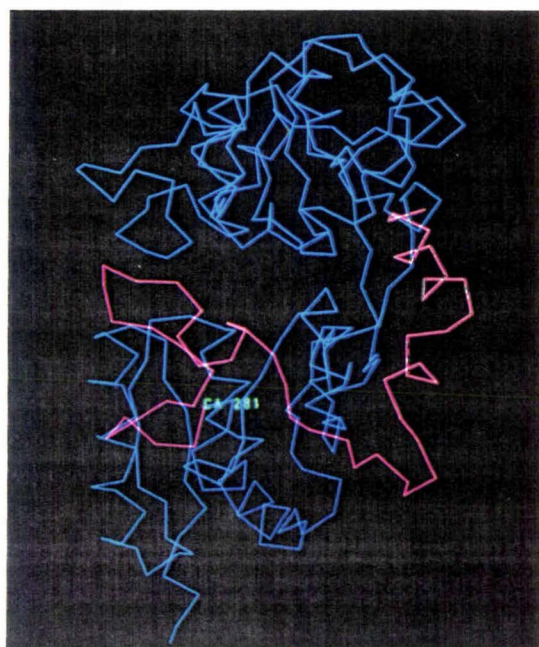
Further digestion of Lf$_{N30}$ gave an 18 kDa glycopeptide which contained domain 2 (residues 91 - 253). The 18 kDa fragment was coloured indicating that it still bound iron even though, Asp 60, one of the ligands involved in iron binding was

absent. Iron release studies using these fragments showed that the pH of iron release from this fragment was similar to that obtained for Lf_{N30}. This was quite surprising as it implies that the presence of domain 1 does not significantly stabilise iron binding in the N-lobe. Subsequent studies on a similar 18 kDa fragment from ovotransferrin suggest that this also binds iron and that the ligand role of Asp 60 has been replaced by a water molecule (Jhoti *et al.*, 1988; Baker and Lindley, 1992).

Rochard *et al.* (1989) have used these same fragments in an attempt to determine the site on lactoferrin which is recognised by receptors on mitogen-stimulated human lymphocytes. They found that Lf_{N30} and the reassociated NC tryptic complex bound to the lactoferrin lymphocyte receptor with kinetics similar to that of the native lactoferrin. However, neither the 18 kDa domain 2 fragment nor Lf_{C50} were capable of binding to the receptor. These results suggested that the receptor binding site on lactoferrin is located in the N-lobe and most probably in residues 4-90 or residues 258-281 as these are absent in the 18 kDa glycopeptide which can no longer bind but are present in Lf_{N30} which can bind to the receptor.

Clearly fragments of lactoferrin are potentially very useful in elucidating the function of lactoferrin and the differences between the two lobes, but to date these studies have been limited by the small amounts of material available and the fortuitous nature of the cleavage sites. If a system could be developed for producing large amounts of selected fragments of known amino acid composition the investigations which could be undertaken would be significantly increased. Such a system would also be amenable to mutagenesis and allow the role of specific amino acids to be investigated.

1.6 Perspectives and aims of this project

Perspectives

The previous sections have described, in general terms, the putative functions and structure of lactoferrin. The iron binding and release properties have also been reviewed, together with the properties of the proteolytically derived fragments of human lactoferrin. At the time this study was commenced even less was known about lactoferrin and it was thought that production of true half molecules might allow the properties of lactoferrin to be defined more precisely. Also it was thought that mutagenesis would prove to be a powerful tool in understanding how the structure was related to the various functions.

Aims

The aims of the work outlined in this thesis were threefold.

❑   To express the amino terminal half of human lactoferrin (Lf_N) *in vitro*.

❑   To characterise the recombinant protein and determine the tertiary structure of this fragment.

❑   To initiate mutagenesis studies and begin characterising any mutants produced.

As this thesis involved diverse procedures and methods the results are presented in two parts. Part A describes the cloning, expression and spectroscopic characterisation of the amino terminal half of human lactoferrin. The mutagenesis experiments are also discussed, in the last chapter of part A. Part B details the solution and analysis of the Lf_N structure. The majority of this part describes the $FeLf_N$ structure although a short section is devoted to the solution and initial analysis of the $ApoLf_N$ structure.

In the final chapter the results from the different sections are summarised and the prospects for future studies are briefly discussed.

CLONING, CHARACTERISATION AND MUTAGENESIS OF THE N-TERMINAL HALF OF HUMAN LACTOFERRIN

As discussed in section 1.5 fragments of human lactoferrin produced by proteolytic cleavage are not true half molecules making it difficult to investigate the properties of the individual lobes of lactoferrin. In order to produce a fragment which represents the N-terminal half of hLf a cDNA coding for the N-terminal lobe has been produced. This cDNA has been cloned into an expression vector and protein has been produced *in vitro* . The spectroscopic and iron binding properties of this fragment have been investigated. In addition several mutants of lactoferrin have also been constructed, expressed and a preliminary characterisation of these proteins has been carried out. This study will hopefully allow the role of specific residues to be investigated.

Part A : Chapter One

MATERIALS AND METHODS

Materials and Reagents

All restriction endonucleases, T_4 polynucleotide kinase, T_4 DNA polynucleotide ligase, *E. coli* DNA polymerase I (Klenow), proteinase K, low melting point agarose, ammonium persulphate, guanidine thiocyanate, 1 kb DNA ladder, random primers labelling system and the Photogene™ Nucleic Acid Detection System were obtained from Bethesda Research Laboratories (Maryland, USA).

Calf alkaline phosphatase was from Boehringer Mannheim (West Germany). pGEM®-3Zf(+) was supplid by Promega (Madison, Wisconsin, USA). The strain of *E. coli* used for the maintenance of all plasmids was XL-1 (Stratagene, La Jolla, California, USA).

Pure SDS, tetra-sodium pyrophosphate, formamide, formaldehyde, chloroform, dimethyldichlorosilane, isoamylalcohol, absolute ethanol, HPLC grade methanol and acetonitrile were obtained from BDH Ltd. (Poole, England).

X-ray film was from Amersham (Sweden); Eastman Kodak (New York, USA); or Fuji Photo Film Company Ltd. (Japan). Polaroid film was from Polaroid Corporation (Cambridge, Massachusetts, USA). Nitrocellulose was supplied by Schleicher & Schull (West Germany).

Gibco (Grand Island, New York, USA) supplied culture media and foetal calf serum. All tissue culture plasticware, 50 ml conical screw cap centrifuge tubes and cryotubes were from Nunc (Denmark). Acrocap filters were obtained from Gelman Sciences (USA).

Sigma Chemical Company (Missouri, USA) provided all antibiotics, all dNTP's, ethidium bromide, glycerol, DNAse, RNAse, X-gal, IPTG, tRNA, DTT, Coomassie brilliant blue R-250, tissue culture grade sodium bicarbonate, DMSO, goat anti-rabbit IgG-alkaline phosphatase conjugate, protein molecular weight markers, Low EEO agarose, N-lauryl sarkosine, 8-hydroyquinoline, PEG 6000 and PEG 8000.

CM-sephadex C-50 and the Cell-Phect kit used for the transfection of BHK cells was from Pharmacia LKB Biotechnology (Uppsala, Sweden). GF/A filters, 3 MM paper and SEP-PAK C_{18}-cartridge tubes were from Whatman (England). Cellulose acetate membrane filters (0.2 μm) were obtained from Micro Filtration Systems (California, USA). All ultrafiltration membranes and Centricon devices were supplied by Amicon (Massachusetts, USA).

The mutagenesis kit, acrylamide, bis-acrylamide, TEMED, mixed bed deionising resin and the Chelex® resin were from Biorad (California, USA). All radiolabelled nucleotides were supplied by New England Research Products, (Boston, Massachusetts, USA).

All high grade buffers and Sequenase version 2.0 were from United States Biochemical Corporation (Cleveland, Ohio, USA). Geneclean was from BIO 101 Inc. (La Jolla, California, USA). Diethylpyrocarbonate was obtained from Aldrich Chemical Company Inc. (Wisconsin, USA).

Human lactoferrin was prepared according to the methods of Norris *et al.* (1989) and PNGase was prepared according to the method of Elder and Alexander (1982). Both human lactoferrin and PNGase were prepared by Dr G. Norris and Mrs H. Baker.

PCR-primer oligonucleotides and sequencing primers were obtained from the Department of Immunobiology, University of Auckland Medical School and from the Department of Biochemistry, University of British Columbia, Vancouver, Canada. All mutagenic oligonucleotides were obtained from Oligos Etc. Inc. (Guilford, Connecticut, USA). The eucaryotic expression vector pNUT (Palmiter *et al.*, 1987) and baby hamster kidney (BHK) cells were generously provided by

Dr R.T.A. MacGillivray (Department of Biochemistry, University of British Columbia, Vancouver, Canada). Methotrexate was obtained from the Palmerston North Hospital.

A1.2   General methods for the manipulation of DNA and RNA

A1.2.1   General precautions for handling DNA and RNA

All solutions, glassware and plasticware in direct contact with DNA were autoclaved at 15 psi for 20 min unless specified. Volatile or temperature sensitive compounds (eg. antibiotics) were sterilised by filtration through 0.2 µm membrane filters. Dialysis tubing was sterilised as described in Sambrook *et al.* (1989). Gloves were worn at all times to minimise contamination by nucleases. Solutions for use with RNA were made 0.1% with diethylpyrocarbonate (DEPC), an inhibitor of ribonucleases. After overnight treatment excess DEPC was removed by autoclaving. Plasticware for use with RNA was also treated with DEPC to inactivate residual ribonuclease activity.

A1.2.2   Quantitation of DNA and RNA

The concentration of purified DNA and RNA in solution was determined by spectrophotometric analysis. Pure preparations of DNA have an A_{260}/A_{280} ratio of ~1.8 while pure RNA has an A_{260}/A_{280} ratio of 2.0 (Sambrook *et al.*, 1989). Concentrated solutions of DNA were diluted with distilled water before determining the absorbance values. RNA was diluted in 10 mM HEPES, 1 mM EDTA, pH 7.6 before spectrophotometric analysis. The concentration of nucleic acid solutions was calculated using the following relationships (Sambrook *et al.*, 1989)

$$1 \ A_{260} \text{ unit of dsDNA} = 50 \ \mu g/ml$$
$$1 \ A_{260} \text{ unit of ssRNA} = 40 \ \mu g/ml$$

Quantitation of small amounts of plasmid DNA was achieved by comparison of aliquots of DNA with standards of known concentration on agarose gels (Section A1.2.8). The standards were prepared by dilution of a stock plasmid of known concentration.

A1.2.3   Phenol / chloroform extraction of DNA

Protein was removed from DNA by phenol/chloroform extraction using the method described by Sambrook *et al* . (1989). Phenol was obtained in a

crystalline form which was made 0.1% with 8-hydroxyquinoline and distilled at 180°C then equilibrated with 50 mM Tris, pH 8.0 prior to use.

A1.2.4 Ethanol precipitation of DNA

DNA was precipitated by making solutions 0.3 M with Na acetate and then adding 2.5 volumes of 95% ethanol. After 15 minutes at -20°C the samples were centrifuged at 12 000 rpm for 10 minutes to pellet the DNA. DNA pellets were washed with 70% ethanol to remove excess salt and then dried in a vacuum concentrator.

A1.2.5 Maintenance and storage of bacterial strains

The *E. coli* strain XL-1 was used in most manipulations. Stock cultures were maintained on Luria broth (LB) plates (1% tryptone, 0.5% yeast extract, 0.5% NaCl containing 1.5% agar) containing 10 μg/ml of tetracycline. Transformed cells were also maintained on LB plates supplemented with antibiotics, where necessary, by the addition of an appropriate volume of a concentrated stock solution, before the plates were poured. Fresh plates were streaked every 2-3 weeks and allowed to grow at 37°C overnight. Clones of interest were stored frozen at -20°C in 50% glycerol and at -70°C in 20% glycerol. Details of the protocols followed for culturing, maintenance and storage of bacteria can be found in Miller *et al.* (1987).

A1.2.6 Preparation and transformation of competent cells

E. coli (strain XL-1), which were used for the maintenance of plasmid DNA in most instances, were made competent by treatment with $CaCl_2$ and transformed according to the methods of Sambrook *et al.* (1989). When α-complementation was to be used for the selection of recombinant plasmids cells were plated out in the presence of IPTG and X-gal. Both were prepared as concentrated solutions [IPTG (200 mg/ml); X-gal (20 mg/ml in DMF)] and then added to the agar plates with the cells. Normally 20 μl of IPTG and 50 μl X-gal was used per petri dish.

A1.2.7 Digestion of DNA with restriction endonucleases

Restriction endonuclease digestions were typically carried out in a total volume of 30 μl for one hour using React® buffers (Bethesda Research Laboratories, MD, USA) at 37°C. For digests using multiple restriction endonucleases with incompatible buffer requirements, TA buffer (33 mM Tris-acetate, pH 7.8, 6.6

mM potassium acetate, 1.0 mM magnesium acetate, 0.05 mM β-mercaptoethanol and 5 μg/ml BSA) was used. When digests were carried out using minipreparation DNA, 1 μl of RNAse (10 mg/ml, boiled for 15 min to inactivate DNAses) was added after restriction endonuclease digestion to degrade any contaminating RNA. Incubation for 5 minutes at room temperature was generally sufficient for complete degradation of the RNA.

A1.2.8   Agarose gel electrophoresis of DNA

For general use 1% agarose in 1x TAE (0.04 M Tris, 0.02 M acetate, 1.0 mM EDTA, pH 8.0) was used and the gels were electrophoresed at 100 V for 1-2 hrs. When it was necessary to purify DNA bands after electrophoresis 1% low melting point agarose gels were used. These gels were electrophoresed at 50 V to prevent overheating and stop the gel from melting. Ethidium bromide (5 μl of 10 mg/ml solution per 500 ml of agarose) was added to all agarose solutions used for DNA gels before the gels were poured to allow visualisation of DNA bands using a UV transilluminator.

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

A1.2.9   Autoradiography

Autoradiography was carried out by placing either the dried sequencing gels or the nitrocellulose filters in a light-tight cassette adjacent to X-ray film (Kodak or Fuji). After an appropriate exposure time the X-ray film was developed using a Kodak X-Omat automatic film processor.

A1.2.10   Preparation of plasmid DNA

Small scale preparations

For the rapid purification of small amounts of plasmid DNA the rapid-boil method of Sambrook *et al.* (1989) was used. Typically 5 μg of DNA was obtained from 1.5 ml of an overnight culture and was resuspended in 50 μl of TE.

Large scale preparations

For the large scale preparation of plasmid DNA 500 ml broth cultures were grown to an optical density of ~ 0.4 at 600 nm. Plasmids were amplified by adding 2.5 ml chloramphenicol (34 mg/ml in ethanol) and incubation of the cultures overnight with continuous shaking. The cells were harvested by centrifugation at 5000 rpm for 15 min in a GSA rotor, after which the DNA was isolated by the alkaline lysis method (Sambrook *et al.*, 1989). The DNA was purified by equilibrium centrifugation in CsCl-ethidium bromide gradients (Sambrook *et al.*, 1989). The band corresponding to the plasmid was recovered and then the ethidium bromide was removed by extraction with water saturated butanol. Caesium chloride was removed by dialysis against several changes of 1xTE (10 mM Tris, 1 mM EDTA, pH 8.0) buffer.

Isolation of phagemid ssDNA

Small amounts of ssDNA were purified from 2 ml of 2xYT (18% tryptone, 10% yeast extract, 10% NaCl, pH 7.5) broths which had been grown as described by the suppliers Promega®. The purified DNA was quantitated by electrophoresing a sample on a gel (Section A1.2.8). Where a single template was to be used with several sequencing primers 25 ml broth cultures were grown and the ssDNA purified in the same way.

A1.2.11   DNA Ligations

All ligations were carried out using T_4 DNA ligase (Bethesda Research Laboratories, MD, USA) and the 5x stock buffer supplied with the enzyme. For ligations involving staggered ends a vector : insert molar ratio of 1:2 or 3 was used and ligations were performed at 12°C overnight. Ligations involving blunt ended temini were incubated at 18°C overnight with a vector : insert molar ratio of 1:5 or 6.

A1.2.12   Isolation of DNA fragments from agarose gels

Fragments of DNA between 500 bp and 2 kb were isolated using Geneclean (Bio-101 La Jolla, CA, USA) and the protocols recommended by the suppliers. Larger fragments were purified using the 'freeze squeeze' method. In this method the band is excised from the gel, wrapped in parafilm and then frozen. Once frozen the gel slice is squashed and the liquid, containing the DNA, is recovered. This proved a very efficient method for recovering large fragments of DNA.

A1.2.13   **End filling of cohesive ends**

DNA fragments with staggered ends were rendered blunt ended using the large (Klenow) fragment of *E. coli* DNA polymerase I. This fragment has both 5'→3' DNA polymerase activity and 3'→5' exonuclease activity, resulting in the filling in of 5' overhangs and the nibbling back of 3' overhangs. As Klenow is active in most buffers used for restriction enzyme digestion it is only necessary to add enzyme and dNTP's to a final concentration of 0.1 mM to obtain activity. After incubation at 30°C for 15 min the enzyme was inactivated by heating at 75°C for 10 minutes.

A1.2.14   **Phosphorylation of blunt ended DNA fragments**

Terminal phosphate groups were added to DNA fragments using T_4 polynucleotide kinase which catalyses the transfer of the γ-phosphate from ATP to the 5' terminus of DNA. The reaction was carried out as described in Sambrook *et al.* (1989).

A1.2.15   **Dephosphorylation of linearised plasmid DNA**

The 5' terminal phosphate groups of DNA fragments were removed with calf alkaline phosphatase using the protocols described in Sambrook *et al.* (1989).

A1.2.16   **DNA sequencing**

Sequencing was carried out using both single stranded and double stranded DNA as templates. Appropriate primers were annealed to the template DNA and extended using Sequenase version 2.0 according to the manufacturer's instructions (USB). ssDNA templates were prepared from phagemids as described in section A1.2.10. dsDNA was denatured with alkali prior to use as a template. Typically 4 μg of dsDNA from a large scale preparation (Section 2.2) in 18 μl of H_2O was denatured by the addition of 2 μl of 2 M NaOH, 2 mM EDTA. The mixture was neutralized by the addition of 2 μl of 2 M ammonium acetate pH 4.6. The denatured DNA was then precipitated with ethanol and resuspended in 7 μl of H_2O before annealing of the primer and labelling with Sequenase version 2.0.

Sequencing gels were prepared and electrophoresed as described in Sambrook *et al.* (1989). Dried gels were autoradiographed overnight at room temperature.

A1.2.17   Agarose gel electrophoresis of RNA

Denaturing gels containing formaldehyde were prepared as described in Sambrook *et al.* (1989). Typically 10 µg of RNA was resuspended in 15 µl of sample buffer (50% glycerol, 0.4% Bromophenol Blue, 0.4% Xylene Cyanol, 1 mM EDTA) and then loaded onto gels and electrophoresed until the dye front had moved three quarters of the way down the gel. One µl of ethidium bromide (10 mg/ml) was added to samples prior to loading on to gels decreasing the high background often present when ethidium bromide is added to the agarose prior to pouring the gel.

A1.2.18   Transfer of RNA to nitrocellulose (RNA blotting)

Following separation of RNA on a denaturing agarose gel and photography to record the position of the RNA bands, the gel was soaked in 2x SSC (0.3 M NaCl, 0.03 M citrate) for approximately 30 minutes, during which time the transfer stand was assembled as described in Sambrook *et al.* (1989). In all instances RNA was transferred to nitrocellulose overnight using 20x SSC (3.0 M NaCl, 0.3 M citrate) as the carrier buffer. The following morning the gel and the nitrocellulose were photographed to ensure that complete transfer had been achieved. The RNA was then fixed to the nitrocellulose by heating the blots in a vacuum oven at 80°C for 2 hours.

A1.2.19   Labelling of cDNA probes with ^{32}P

cDNA fragments which were to be used as probes were purified from agarose gels and then labelled with [α-^{32}P]-dCTP using the Random Primers DNA labelling system (Bethesda Research Laboratories, MD, USA) as described by the manufacturers. This procedure normally produced probes which had a specific activity greater than 5×10^8 cpm/µg DNA.

A1.2.20   Hybridisation using cDNA probes

Prehybridisation

Blots were first wet with 2x SSC (0.3 M NaCl, 0.03 M sodium citrate) then prehybridised for 3 hours in 50 ml of prehybridisation solution [6x SSC (0.9 M NaCl, 0.09 M sodium citrate) containing 1x Denhardts reagent]. This was carried out in roller bottles in an air bath rotary oven at the same temperature which was to be used for hybridisation.

Hybridisation

The radioactive probe was denatured by boiling and then added to the roller bottle to which 10 ml of hybridisation solution (1 M NaCl, 50 mM phosphate pH 6.5, 2 mM EDTA, 0.5% SDS, 1x Denhardts) had been added. The hybridisation reaction was carried out overnight at the selected temperature.

Washing

Blots were washed in the roller bottles at the temperature of hybridisation. Firstly two washes in 6x SSC (0.9 M NaCl, 0.09 M sodium citrate) containing 1% SDS for 1 hour were carried out, followed by two 30 minute washes in 6x SSC containing 0.05% SDS. The final stringent wash was always in 1x SSC for 30 minutes.

A1.2.21   Preparation and purification of oligonucleotides

All oligonucleotides were obtained from commercial suppliers. Oligonucleotides to be used as sequencing primers were dissolved in water at a concentration of 10 $\mu g/\mu l$ and then a working solution of 1 ng/μl was prepared. Oligonucleotides to be used as PCR primers or in mutagenesis reactions were purified prior to use by electrophoresis through 15% denaturing polyacrylamide gels according to the method of Sambrook *et al.* (1989). The oligonucleotides were visualised by UV shadowing using a thin layer chromatography plate that fluoresced at 254 nm. The area containing the least mobile band (highest molecular weight) was excised and then eluted from the gel slices by incubation at 37°C overnight in 1 ml of gel elution buffer (0.5 M ammonium acetate, 10 mM magnesium acetate). After incubation overnight in gel elution buffer the sample was vortexed and then centrifuged at 12 000 rpm for five minutes before the supernatant was recovered. To ensure that most of the oligonucleotide was recovered the gel slices were washed with a further 0.5 ml of elution buffer which was recovered in a similar manner. The pooled supernatants were further purified by reverse phase chromatography using a SEP-PAK C_{18} column as described in Sambrook *et al.* (1989). The oligonucleotide containing fraction from the column was dried in a vacuum desiccator and the concentration determined spectrophometrically.

A1.2.23   Amplification of DNA sequences by PCR

Specific DNA sequences were amplified using a Gene Amp® kit which contains the thermostable enzyme Amplitaq® Taq polymerase. A Perkin-Elmer-Cetus DNA

thermal cycler was used to carry out the reaction. 100 μl reactions were routinely used with the concentration of reagents as recommended by the suppliers.

A1.3   Mutagenesis of cloned cDNA

In all cases the method of Kunkel *et al.* (1987) was used. A kit was obtained from Biorad® that contained all the essential reagents and strains of cells along with the recommended protocols. Although for most steps the method recommended by the suppliers was followed the preparation of good quality ss uracil-containing DNA was a critical step; therefore the main steps are outlined below.

A1.3.1   Preparation of uracil-containing phagemids

A single CJ236 colony containing the plasmid to be mutated was grown overnight at 37°C in 5 ml of LB containing 30 μg/ml of chloramphenicol. One ml of this culture was used to inoculate 50 ml of 2x YT broth containing 100 μg/ml of ampicillin. This culture was grown until the OD_{600} was 0.3, approximately 5 hours, when it would have a concentration of approximately 1×10^7 cfu/ml. Helper phage (M13K07) was then added to obtain a multiplicity of infection of around 20, using a stock solution of phage which had a titre of 4.8×10^{11} pfu/ml. Once the phage had been added the culture was incubated at 37°C for one hour and then 70 μl of kanamycin (50 mg/ml) was added. After a further 5 hours of incubation, 30 ml of the culture was centrifuged for 15 minutes at 12 000 rpm in a Sorvall SS-34 rotor. The supernatant was recovered and recentrifuged to ensure that all of the cellular debris had been removed. Contaminating RNA was then degraded by addition of 150 μg of RNAse A and incubation for 30 minutes at room temperature. The phagemids were precipitated by adding 0.25 volumes of a 3.5 M ammonium acetate/20% PEG 6000 solution and recovered by centrifuging for 15 minutes at 12 000 rpm. Traces of PEG were removed by thoroughly draining the tube and wiping the sides with tissue before the pellet was resuspended in 200 μl of high salt buffer (300 mM NaCl, 100 mM Tris, pH 8.0, 1 mM EDTA).

A1.3.2   Extraction of phagemid DNA

Uracil-containing DNA (U-DNA) was extracted in essentially the same way as for all DNA samples, except that all phenol/chloroform layers were back extracted with TE and several additional chloroform extractions were carried out at the end to ensure that all traces of contaminants had been removed. Once all the aqueous phases had been combined and extracted the U-DNA was precipitated for 30

minutes at -70°C after 0.1 volumes of 7.7 M ammonium acetate and 2.5 volumes of ethanol had been added. The pellet was washed with 70% ethanol and then resuspended in 20 μl of TE.

A1.3.3   Phosphorylation of the oligonucleotides

Terminal phosphates were added to purified oligonucleotides (Section A1.2.21) using 4.5 units of T4 polynucleotide kinase in a reaction mix containing 200 pmoles of oligonucleotide, 100 mM Tris pH 8.0, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP. After incubation at 37°C for 45 minutes the reaction was terminated by heating at 65°C for ten minutes. Typically 30 μl reactions were used giving a final oligonucleotide concentration of 6 pmoles/μl.

A1.3.4   Annealing and extension of the primer

Six pmoles of phosphorylated oligonucleotide were annealed to 200 ng of ss U-DNA in a total volume of 10 μl with a final buffer concentration of 20 mM Tris pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. Once the oligonucleotide had been added the reactions were placed in one litre of water at 70°C which was then allowed to cool slowly (~ 45 minutes) to 30°C. After the reactions had cooled they were placed on ice and 1 μl of 10x synthesis buffer (4 mM of each dNTP, 76 mM ATP, 175 mM Tris pH 7.4, 37.5 mM $MgCl_2$, 5 mM DTT) along with 1 μl of T4 DNA ligase (3 units) and 1 μl of T4 DNA polymerase (1 unit) were added. The reactions were then left on ice for 5 minutes before being placed at 25°C for a further 5 minutes before finally being incubated at 37°C for 90 minutes. The gradual increase in temperature helped stabilise the template/primer complex allowing extension to occur. After extension of the primer and ligation of the new strand the reaction was stopped by adding 90 μl of 10 mM Tris pH 8.0 containing 10 mM EDTA before freezing the samples at -20°C.

A1.4   Expression of cloned Lf$_N$ cDNA

A1.4.1   Maintenance of cells in tissue culture media

All mammalian cells were grown at 37°C in a humidified CO_2 incubator which maintained an atmosphere of 5% CO_2 in air. All manipulations were carried out in a laminar flow hood. The working surface was routinely swabbed with 70% ethanol and the interior of the cabinet sterilised with UV light. Cells were grown in petri-type dishes which allowed the rapid exchange of air. For routine work, BHK cells were grown in a 1:1 mixture of Dulbecco's modified Eagle's medium

(DMEM) and Hams-F12 (F12), supplemented with 10% foetal calf serum (FCS) and the antibiotics penicillin (100U/ml) and streptomycin (100 μg/ml). All plasticware was of tissue culture grade and was supplied in a sterile form. Glass pipettes were plugged with cotton wool and baked at 160°C prior to use.

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 0.2 μm Acrocap™ filters into sterile bottles. A masterfile pump with a flow rate of ~50 ml/minute was used to pump the medium. Sterile medium was stored at 4°C.

100x stocks of antibiotics were prepared in water, filter sterilised through a 0.2 μm filter and stored in 1 ml aliquots at -70°C.

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

Complete medium was made up from the constituents as required and used within a week. Typically 200 ml or 500 ml were prepared at a time.

Freezing and thawing of cells.

Cells were grown on 100 mm^2 plates (see below) until ~90% confluence then 'passed' (see below) in the normal way except that the detached cells were resuspended in 2 ml of freezing medium (10% DMSO, 90% FCS) which was aliquoted into 2 cryotubes. To allow freezing to occur slowly the tubes were wrapped in paper towels and placed at -70°C overnight. The next day the tubes were transferred to liquid nitrogen.

Cells were thawed rapidly by placing the tubes at 37°C for 5 minutes and then they were transferred to 160 mm^2 plates containing 20 ml of medium. After the cells had adhered to the plates (4-6 hours) the medium containing DMSO was removed and replaced by complete medium.

Passage of cells.

Cells were 'passed' every 2-3 days. To 'pass' cells the medium was removed from the plate and then sufficient virsene (0.2 g Na$_2$EDTA, 8.0 g NaCl, 0.2 g

KCl, 1.15 g Na_2HPO_4, 0.2 g KH_2PO_4 per litre, pH 7.2) was added to cover all the cells. The virsene was immediately removed using a plastic pasteur pipette and then the plates were left for 5 minutes during which time the cells became detached from the plate surface. Medium (2 ml) was then added to each plate and the cells resuspended using a plastic pasteur pipetter. (Lumps were dispersed by vigorously squirting the cells against the dish). Once an even suspension had been obtained 10% of the cells were inoculated into a fresh petri dish containing medium. Dishes were returned to the incubator and left undisturbed for 1 day allowing adhesion of the cells to the dish.

1.4.2 Transfection and selection of BHK cells

Plasmids for transfection into BHK cells were purified by CsCl purification (Section A1.2.10). Pure DNA was precipitated and introduced into cells using the standard calcium phosphate coprecipitation method and the protocols outlined in the Cell Phect kit (Pharmacia LKB Biotechnology, Uppsala, Sweden). Typically 6 µg of DNA was used to transfect 2 x 60 mm^2 plates that were approximately 75% confluent with BHK cells. The coprecipitate was left on the cells for 4 hours, then the plates were rinsed several times with media that did not contain FCS to remove all of the precipitated material. The efficiency of transformation was increased by treating the cells with 15% glycerol in HEPES buffered saline [(HBS) 0.05 M HEPES, 0.15 M NaCl, 0.005 M KCl, pH 7.5] for 1.5 min. Again it was important to thoroughly rinse the cells as glycerol will kill the cells if left on the cells too long. After incubation overnight in fresh medium the cells were passed into 100 mm^2 plates and the medium was supplemented with 500 µM methotrexate, resulting in the selection of transformed cells after 1-2 weeks. Once colonies of transformed cells had reached confluence on the plates they were 'passed' in the normal manner.

1.4.3 Large scale growth of cells in roller bottles

The production of large amounts of recombinant protein was achieved by growing the cells in roller bottles. Cells were grown until 150 mm^2 plates were confluent; they were then 'passed' and the cells from one plate were used to inoculate a single roller bottle. Initially ~50 ml of medium (1:1 DMEM/F12 containing 5% FCS and no methotrexate) was added to each bottle. The bottles containing the cells were gassed with 5% CO_2 for 2-3 min, sealed tightly and then incubated at 37°C on a rotating machine. Cells generally reached confluence in 5-7 days, after which time the medium was changed daily until extensive cell death started to occur (typically 3 - 4 weeks).

A1.4.4   Preparation of DNA from BHK Cells

DNA was isolated from approximately 10^6 BHK cells according to the method of Karlinsey *et al.* (1989). This preparation produced several micrograms of DNA that was suitable for analysis by PCR.

A1.4.5   Preparation of total cellular RNA from BHK cells

BHK cells in 160 mm^2 plates were grown to confluence. The cells were released from the plates using 4 ml of guanidine thiocyanate buffer (4.0 M guanidine thiocyanate, pH 7.0; 0.025 M sodium citrate, pH 7.0; 0.7% β-mercaptoethanol; 0.5% N-lauroyl sarkosine; filter sterilised and stored in a foil covered bottle, for no longer than 1 month) per plate. The lysed cells from two plates were combined and the RNA isolated according to the methanol of Chirgwin *et al.* (1979). Typically 100-200 μg of RNA was obtained from 2 x 160 mm^2 plates of BHK cells.

A1.5   Analysis of recombinant proteins

A1.5.1   SDS-polyacrylamide gel electrophoresis of proteins

In general the protocols in the Hoefer Mighty Small II instruction manual were used. [A slight modification of the Laemmli (1970) method]. The resolving gels were 12% acrylamide, 2.7% bisacrylamide, 0.375 M Tris pH 8.8, 0.1% SDS, 0.1% w/v ammonium persulfate and 0.07% v/v TEMED. Stacking gels were always used and had the following composition: 4% acrylamide, 2.7% bisacrylamide, 0.125 M Tris pH 6.8, 0.1% SDS, 0.05% w/v ammonium persulfate and 0.05% v/v TEMED. The tank buffer was 0.02 M Tris base, 0.192 M glycine, pH 8.3 and 0.1% SDS. Prior to loading, samples were diluted 1:1 with 2x treatment buffer (0.125 M Tris/HCl, pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol containing a drop of Bromophenol Blue) and placed in a boiling water bath for 3 minutes.

A1.5.2   Staining and destaining polyacrylamide gels

Gels were stained using Coomassie brilliant Blue R-250 (0.125% Coomassie blue R-250, 50% methanol, 10% acetic acid) for 1 hour at room temperature. The stain solution was filtered before use and stored in a dark bottle. Destaining was accomplished by first using two changes of 50% methanol, 10% acetic acid and

then several washes in 7% acetic acid, 5% methanol until there was no background colour. When completely destained, gels were dried on to blotting paper using a Biorad® gel dryer at 80°C for 45 minutes.

A1.5.3 Preparation of antibodies against human lactoferrin

Antibodies against human lactoferrin had been raised in New Zealand white rabbits and purified (Stowell, 1990). Antibodies typically had a specific activity of 150 μg/mg (i.e. 150 μg of antigen can be precipitated by 1 mg of antibody) and a concentration of 2 mg/ml. The antibodies had been demonstrated to be be specific for human lactoferrin, showing no cross reactivity with bovine lactoferrin in either immunotitration or protein blotting experiments (Stowell, 1990).

A1.5.4 Immunoprecipitation of human lactoferrin

Immunoprecipitates were formed over a step-gradient of 0.5 M sucrose and 1 M sucrose in phosphate buffered saline [(PBS) 0.15 M NaCl, 0.01 M NaH_2PO_4, pH 7.2] containing 1% sodium deoxycholate and 1% Triton X-100. Typically 250 μl of medium or cell extract was layered on the gradient, followed by an excess of anti-human lactoferrin γ-globulin. (This was usually 25 μl of a γ-globulin solution with a specific activity of ~150 μg/mg and a protein concentration of ~2 mg/ml). Immunoprecipitates were allowed to form during incubation for 1 hour at 37°C and were then collected by certrifugation at 12000 rpm for 10 minutes at room temperature in a microcentrifuge. The supernatant was discarded and the pellet washed 3 times with PBS containing 1% deoxycholate and 1% Triton X-100, standing for 15 minutes on ice between washes.

A1.5.5 ^{35}S-Labelling of newly synthesised proteins

Newly synthesised proteins produced by BHK cells were labelled with ^{35}S-methionine as described by Stowell (1990), and summarised as follows. Cells were grown in small 30 mm^2 petri dishes until confluence was reached, after which the medium was removed and replaced by supplemented MEM containing 10% FCS and ~50 μCi of ^{35}S-methionine. The cells were incubated overnight, the medium was then recovered and centrifuged to remove cell debris. The amount of radioactivity incorporated into lactoferrin was then determined by forming immunoprecipitates, using 200 μl of media, over a sucrose gradient as described previously (Section A1.4.6). The total amount of newly synthesised protein was precipitated by the addition of 1 ml of 25% TCA to 250 μl of decolourised cell medium (solutions were decolourised to prevent quenching

during scintillation counting). After transfer of the precipitate to a GF/A filter it was extensively washed with several volumes of ice-cold 10% TCA and ice-cold 95% ethanol. The filters were dried at 60°C, transferred to a scintillation vial and 3 ml of scintillation fluid was added to each filter. The immunoprecipitates were also transferred to scintillation vials after being resuspended in 1 ml of 0.1 M NaOH. Six ml of scintillation fluid was added to each of these samples. The amount of radioactivity in the vials was determined by scintillation counting using a 14C window (programme 4, channel 2 of a Beckman LS 8000 scintillation counter).

In some instances it was necessary to determine the amount of ^{35}S-methionine incorporated into protein within the cells. This was done in exactly the same manner as described above except the cells were lysed first. Lysis was achieved by incubation of the cell pellet in radioimmunoassay extraction buffer (0.15 M NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, in 30 mM HEPES, pH 7.3) for 2 hours or overnight. Cell debri was then removed from the extracted samples by centrifugation for 30 minutes at 12 000 rpm in a microfuge at 4°C. The recovered supernatant was processed as for cell media (see above).

A1.5.6   Protein blotting

After electrophoresis on a polyacrylamide gel, proteins were transferred electrophoretically to Photogene™ Nylon membranes in 25 mM sodium carbonate buffer pH 10.5 for 45 min at 0.5 A. After transfer, the membranes were blocked in phosphate buffered saline solution [(PBS) 0.01 M phosphate pH 7.2, 0.15 M NaCl] containing 10% non-fat milk for one hour. The membrane was then washed with PBS before being agitated for 30 minutes at room temperature with rabbit anti-human lactoferrin antiserum diluted 1:100 in PBS containing 1% non-fat milk. After extensive washing with PBS the membrane was incubated for 30 minutes with the secondary antibody (goat anti-rabbit γG conjugated with alkaline phosphatase), which was diluted 1:1000 in PBS containing 1% non-fat milk powder. Finally the blot was washed with final wash buffer (Photogene™ diluted 1/20 with water) and blotted dry. A chemiluminescent signal was generated by placing the membrane between clear plastic sheets and applying 300 μl of detection reagent (Photogene™). After 6 minutes X-ray film (Fuji, Japan) was placed over the membrane. The film was exposed for 5 minutes then developed using Kodak developer.

1.5.7   Purification of recombinant lactoferrins

Once the cell media had been centrifuged at 5000 rpm for 10 minutes to remove
cell debri the recombinant protein was purified from tissue culture media using C-
50 CM-Sephadex® resin. The resin (~5 g/litre) was stirred with the media for 1-2
hours then filtered using a Buchner funnel. Tris buffer (0.025 M pH 7.8
containing 0.2 M NaCl) was used to wash the resin until the A_{280} of the eluate
was < 0.1. Lf_N was then eluted using 0.8 M NaCl in the same buffer. At this
stage Lf_N was still contaminated by several minor impurities which were removed
by ion exchange chromatography on a C-50 CM-Sephadex® column equilibrated
in 0.01 M HEPES pH 7.8, 0.2 M NaCl. Before loading Lf_N onto the column it
was first dialysed against several changes of 0.01 M HEPES pH 7.8, 0.2 M NaCl
to remove the excess salt. Lf_N bound as a tight band which was washed with 0.35
M NaCl in 0.01 M HEPES pH 7.8 and then eluted using 0.4 M NaCl in the same
buffer. Fractions were collected, analysed by SDS-PAGE and the fractions
containing pure protein were combined and concentrated using a YM30 or YM10
membrane (Amicon).

1.5.8   Determination of protein concentration

The approximate concentration of both hLf and Lf_N was estimated from the
absorbance at 280 nm given the relationship

$$E_{1cm}^{1\ mg/ml}\ \text{at 280 nm} = 1.2$$

for iron saturated protein and

$$E_{1cm}^{1\ mg/ml}\ \text{at 280 nm} = 1.09$$

for apo protein (Ainscough *et al.*, 1979).

1.5.9   Concentration of protein samples

Dilute solutions of lactoferrin were concentrated using Amicon concentrators and
YM-30 membranes. For smaller volumes (up to 2 ml) Centricon-30®
concentrators with a molecular weight cut off of 30 000 Da were used. At all
times buffers were supplemented with 0.2 M NaCl to avoid precipitation of the
protein. The absorbance at 280 nm of eluates was always recorded to ensure that
no protein had leaked through the ultrafiltration membrane.

A1.5.10   **Deglycosylation of lactoferrin**

Samples of Lf$_N$ were deglycosylated essentially as described by Norris *et al.*
(1989). In brief protein samples were dialysed against several changes of Nonidet
buffer (0.1 M sodium phosphate, pH 6.1, 0.5% Nonidet P-40, 50 mM EDTA)
and then 5 - 10 µl of active PNGase, purified following the method of Elder and
Alexander (1982), was added. The sample was then incubated at either room
temperature or at 37°C until deglycosylation, as judged by SDS - PAGE, was
complete.

A1.5.11   **Spectroscopy**

All UV-visible spectra of lactoferrin were recorded using a Hewlett Packard diode
array spectrophotometer (Model HP 8452A) and the data were recorded on disk.
The ESR spectra were recorded at ~ 100K, using a Varian E-104A
spectrophotometer according to the protocols of Ainscough *et al.* (1980; 1983).
Protein concentrations for ESR spectroscopy were 20 mg/ml for FeLf$_N$ and 10
mg/ml for CuLf$_N$ samples.

A1.5.12   **Iron release from lactoferrin**

The affect of pH on iron release from lactoferrin was determined by dialysis of
protein samples against buffers at different pH values. The buffers used for the
various pH ranges were: pH 8-7, 0.05 M Tris HCl; pH 6.5-6.0, 0.05 M MES;
pH 5.5-3.5, sodium acetate (I=0.05), pH 3.0-2.0, 0.1 M glycine HCl. All
buffers contained 0.2 M NaCl. Samples were always dialysed for 48 hours to
ensure that equilibrium was attained. In most instances separate samples were
used for each pH value measured. However, when the amount of protein
available was limited samples were shifted from high pH values to low pH values
and equilibrated a second time. One sample was never used more than twice.
Absorbances were recorded and then the % of iron bound was determined after the
background absorbance had been subtracted.

Part A : Chapter Two

CLONING AND EXPRESSION OF THE N-TERMINAL HALF OF HUMAN LACTOFERRIN

2.1   INTRODUCTION

At the outset of this project the cDNA encoding intact human lactoferrin was available (Stowell *et al.*, 1991) and was used as the template for the construction of a cDNA encoding the N-lobe of lactoferrin. Lysine 333 was chosen as the last residue in the N-lobe as it is after helix 11, the last helix in the N-lobe, and before the short linking helix which connects the two lobes in lactoferrin.

Several possible approaches were considered for the construction of a cDNA encoding the N-terminal half of human lactoferrin (Lf_N). One approach would be to introduce a stop codon and unique restriction site at the end of the N-lobe using site specific mutagenesis, in a manner similar to that used for the construction of a cDNA encoding the amino terminal half of human serum transferrin (sTf) (Funk *et al.*, 1990). Alternatively the cDNA could be constructed using PCR and specific primers complementary to the 5' end of the hLf cDNA and to the proposed termination region at the end of the N-lobe. In this approach a stop codon and unique restriction site would be introduced on the 3' PCR primer. This second PCR-based approach was chosen because PCR-technology was available to us and had been successfully used on several other occasions. PCR is now a routinely used technique but care is still required as the enzyme used does not have a proof reading 3'-5' exonuclease function and can thus introduce errors (Tindall & Kunkel, 1988; Denney & Weissman, 1990). PCR-induced errors are reported to occur at low frequencies but they remain a real concern when PCR-generated sequences are cloned and used for expression studies. Because of these problems it is important that all PCR generated sequences are sequenced to ensure that no errors have been introduced.

When attempting to express a new protein, especially complex eucaryotic proteins such as lactoferrin, it is often necessary to try several expression systems before a successful one is found. In relation to the present case, however, the expression of the amino terminal half of human serotransferrin (sTf_N) had been accomplished using the expression vector pNUT (Palmiter *et al.*, 1987) and transformed BHK cells. Because of the high level of similarity between sTf_N and Lf_N it was decided to attempt to express Lf_N in the same system.

RESULTS

A2.2.1   PCR synthesis and cloning of the 1.1 kb cDNA encoding Lf$_N$

Preparation of templates

The full length cDNA clone available was called pUC:hLf (Stowell, 1990). This clone had been produced by joining together three fragments isolated from a λgt10 library. The clone had been sequenced by Dr K.M.Stowell and was reported to contain the complete coding sequence for human lactoferrin (Appendix 1). A partial restriction map of this clone is shown in Fig. 5.

Fig. 5 Partial restriction map of pUC:hLf.
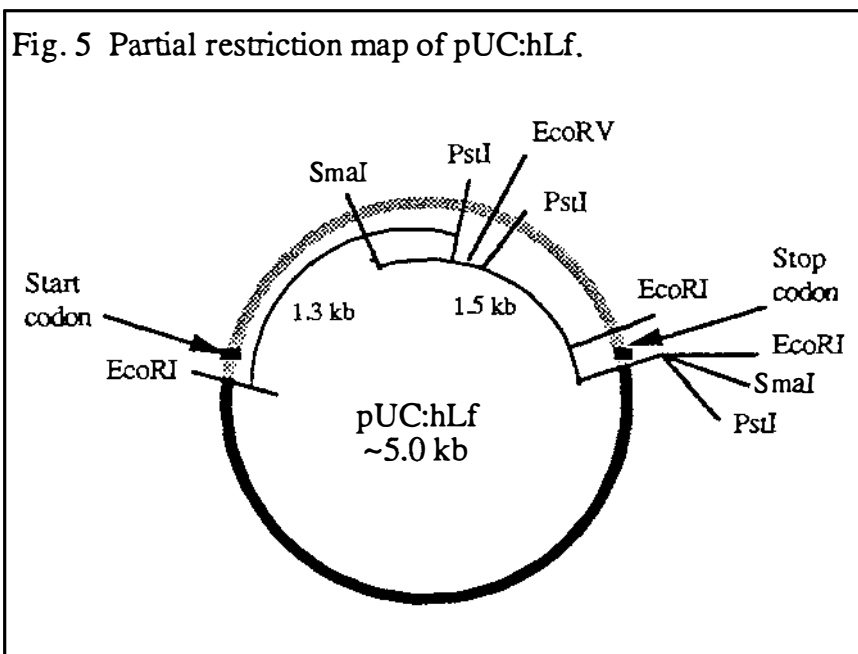
The sequence identity between the two lobes of lactoferrin is high (~40%) introducing the possibility of priming from homologous sequences in both the N- and C- terminal lobes. In order to reduce the possibility of spurious priming from homologous sequences in the C-lobe a template was prepared that only included bases 1 to 1300. The insert was produced by digesting pUC:hLf with EcoRI and PstI, releasing a 1.3 kb band including bases 1 to 1300. This fragment encoded all of the amino terminal lobe and the first 50 amino acids of the C-terminal lobe. The 1.3 kb fragment was then ligated into EcoRI/PstI cut pUC9 and transformed into *E. coli* (strain XL-1) cells. Recombinant clones were selected as white colonies when transformed cells were grown in the presence of X-gal and IPTG. The presence of the expected insert was confirmed by restriction digestion with the appropriate restriction enzymes. This clone was designated pUC:hLf1.3 (contains the 1.3 kb EcoRI/PstI fragment) and is depicted in Fig. 6.

Fig. 6 Restriction map of pUC:hLf1.3.
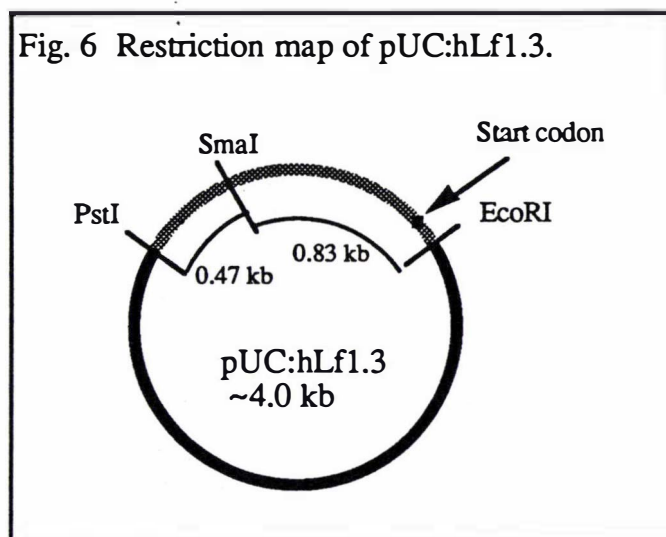
Synthesis of the cDNA by PCR

Two oligonucleotide primers were required to produce a PCR product encoding the upstream untranslated region, the signal peptide and residues 1 to 333. An oligonucleotide primer complementary to the 5' upstream region of human lactoferrin called JT30 was already available to us (Fig. 7). A second oligonucleotide (shown in Fig. 7) complementary to the (+) strand of the cDNA encoding amino acids 329 to 333 was obtained and called JT11. In addition to the region of homology with the cDNA, JT11 also encoded a single stop codon (bold) and had three additional G's at the 5' end (underlined).

Fig. 7 Sequence of JT30 and JT11.

The position of the primers relative to the DNA sequence and the protein sequence is indicated. In both cases the coding strand of the DNA is shown.

Primer to 5' end of the coding sequence

protein	non coding region of the DNA	Met 1
DNA	1 GGCCGGTCGCCTCCAGACCGCAGAC AT 27	
JT30	5' GGCCGGTCGCCTCCAGACCGCAGAC AT 3'	

Primer to 3' end of the coding sequence

protein	327	Ala	Ile	Gln	Asn	Leu	Arg	Lys	333
DNA	1040	CC	ATC	CAG	AAC	TTG	AGG	AAA	1059
JT11	3'	GG	TAG	GTC	TTG	AAC	TCC	TTT	**ATT** GGG 5'

Before use the oligonucleotides were purified as described in section A1.2.21. PCR was carried out as described in section A1.2.22 using 10 ng of pUC:hLf1.3

as the template and 5 ng of both JT11 and JT30. The reaction products were analysed by agarose gel electrophoresis (Fig. 8).

Fig. 8 Analysis of the PCR products by agarose gel electrophoresis.

Lane (1) BRL 1 kb ladder; (2 and 3) 10 µl of a PCR reaction that contained 10 ng of pUC:hlf1.3 template; (4) 10 µl of a reaction which did not contain any template DNA.
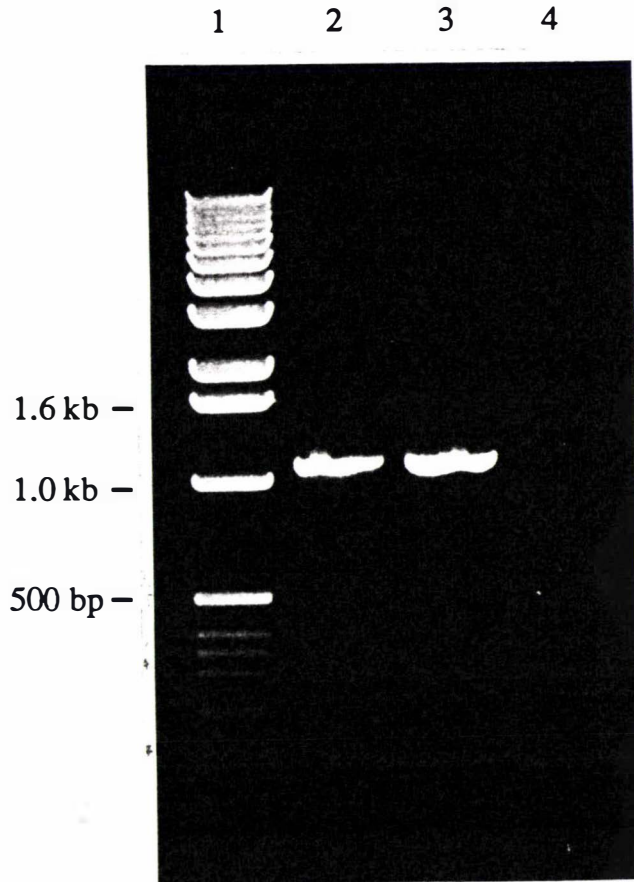
The absence of a band in lane 4 demonstrates that the reagents were not contaminated with any DNA which could act as a template. A faint lower molecular weight band is present in lane 3 suggesting that a small amount of nonspecific priming may have occurred. However the main band is clearly 1.1 kb in size indicating that amplification from pUC:hLf1.3 was successful. The DNA from six identical reactions was combined and concentrated ready for use in subsequent steps.

Restriction enzyme analysis of the 1.1 kb PCR fragment

Prior to cloning, the 1.1 kb PCR product was analysed by digestion with restriction enzymes diagnostic for Lf_N (Fig. 9). Fragments of the expected size were obtained indicating that the 1.1 kb PCR product contained no major

insertions or deletions. The absence of any extraneous bands also demonstrated that PCR had produced only one product.

Fig. 9 Restriction analysis of the 1.1 kb PCR product.

The purified product was analysed by restriction endonuclease digestion and gel electrophoresis. Lane (1) uncut PCR product; (2) SmaI cut PCR product; (3) BglII cut PCR product; (4) RsaI cut PCR product. No molecular weight markers are shown on this gel but the positions of the restriction endonuclease recognition sites and the expected size of each fragment are shown in the adjacent panel. The 130 bp fragment and the 60 bp fragment are not visible in this photo but were visible on the original.
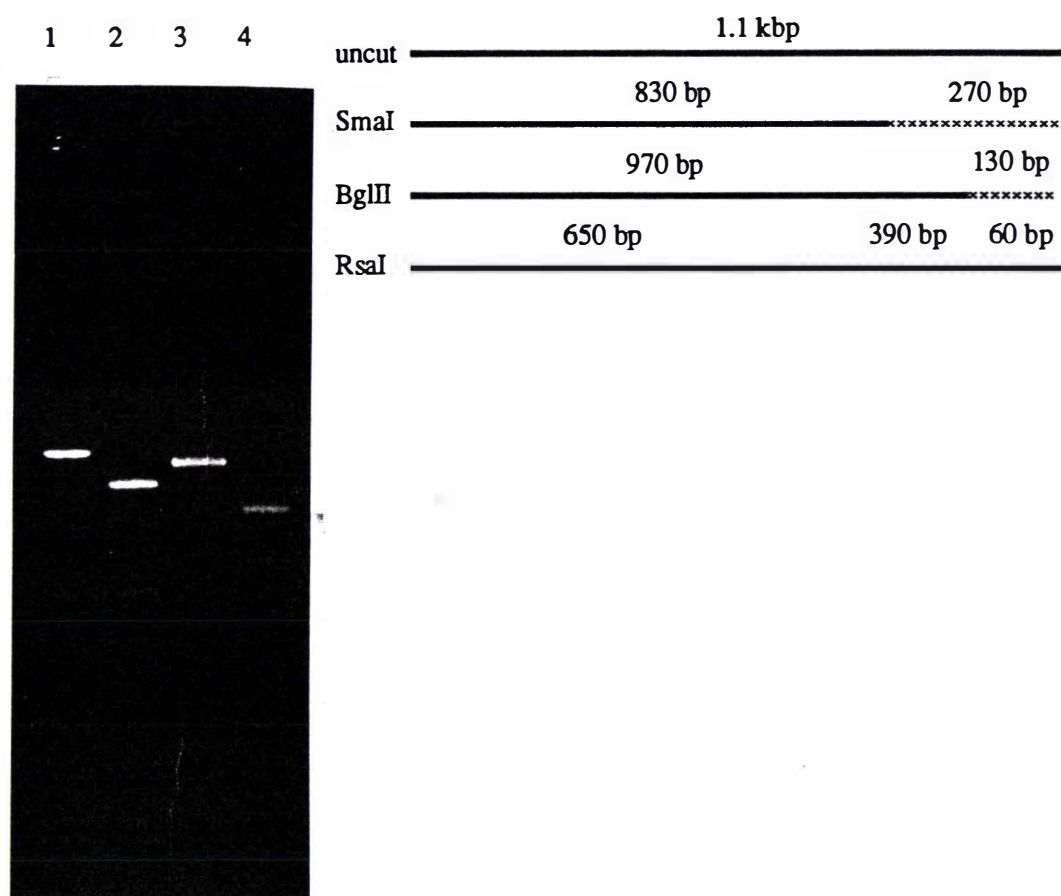
Cloning of the 1.1 kb PCR product

The PCR product was cloned into pUC18 in preparation for subsequent subcloning and sequencing using the following strategy. Since restriction endonuclease sites had not been incorporated into the PCR primers it was necessary to carry out a blunt end ligation. Approximately 10 µg of the PCR product was precipitated and terminal phosphate groups were added using T4 polynucleotide kinase (Section A1.2.14). The 1.1 kb fragment was then purified from a 1% LMP gel using Geneclean (Section A1.2.12) and quantitated on an agarose gel (Section A1.2.8). pUC18 was prepared by digesting with SmaI

followed by treatment with calf alkaline phosphatase to remove the 5' phosphate groups (Section A1.2.15). The vector was treated in this way to decrease the background of non-recombinant clones which might arise from religation of the vector without incorporation of an insert. A ligation reaction was set up using a 2:1 molar ratio of insert to vector in a total volume of 10 μl. Half of this ligation reaction was used to transform *E. coli* (strain XL-1). The cells were plated out in the presence of X-gal and IPTG, allowing selection of recombinant clones which appeared as white colonies in a background of non-recombinant blue colonies. Minipreparation DNA (Section A1.2.10) isolated from several colonies was screened by restriction analysis with BamHI and EcoRI which should cleave on either side of the insert releasing a 1.1 kb fragment. Three clones were found to contain the desired insert, one of which was selected and designated pUC:Lf$_N$1. DNA was prepared on a large scale from this clone and digested with diagnostic restriction enzymes to confirm its authenticity (Fig. 10).

Fig. 10 Restriction enzyme analysis of pUC:Lf$_N$1.

Aliquots of pUC:Lf$_N$1 plasmid DNA were analysed by digestion and gel electrophoresis. Lane (1) BRL 1 kb ladder; (2) BamHI/EcoRI digest; (3) SmaI digest; (4) EcoRI/SmaI digest; (5) BamHI/BglII digest; (6) EcoRI/BglII digest.
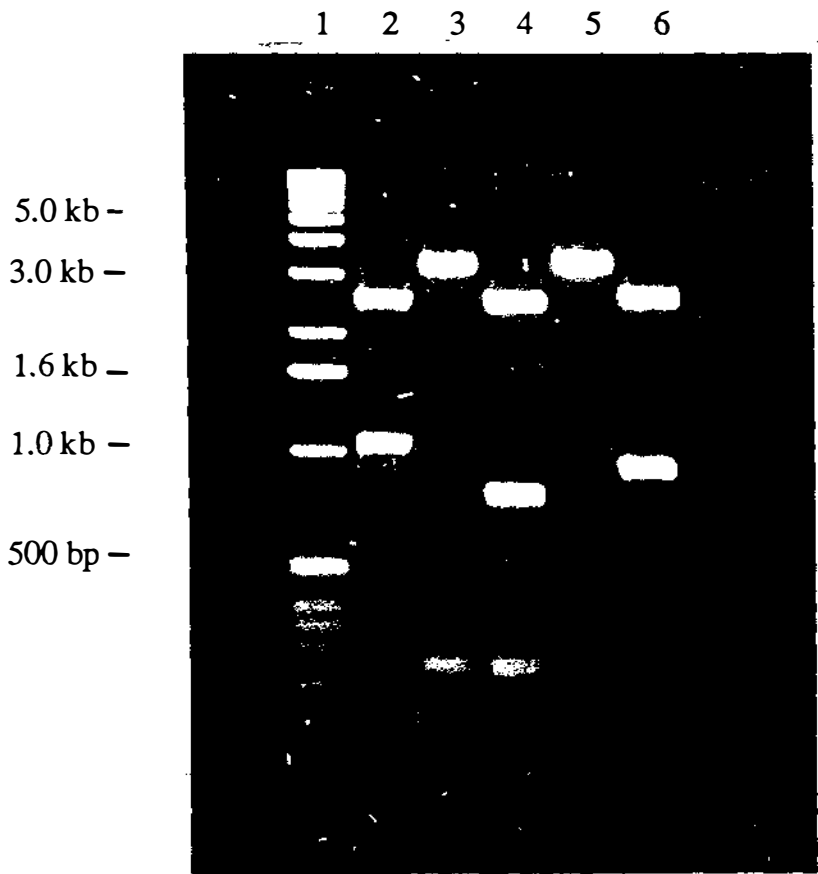
The pattern of restriction fragments predicted from the sequence of the hLf cDNA was obtained demonstrating that the cDNA for Lf$_N$ had been successfully cloned into pUC18.

Preliminary sequence analysis of the 1.1 kb insert

The Lf$_N$ cDNA insert was sequenced in order to ascertain that there had been no errors introduced by PCR. The standard method for determining a DNA sequence is to subclone small fragments and then combine the data from overlapping clones. Although the cDNA for lactoferrin contains very few convenient restriction sites the sequence of the entire cDNA was obtained from four fragments as outlined in Fig. 11. Approximately 300 bp of sequence data was obtained from either end of the intact cDNA after it had been cloned into M13mp18 and M13mp19. The universal -40 sequencing primer was used as the primer in all cases. The internal sequence was obtained by subcloning the EcoRI/SmaI and EcoRI/RsaI fragments into M13mp18 and again preparing single strand templates for use with the universal -40 sequencing primer. The sequence obtained was compared to pUC:hLf and thought to be correct (see section A2.2.4 for a further discussion).

Fig. 11 Sequence analysis of the 1.1 kb fragment.

The regions of DNA sequenced from each primer/template combination and the direction of sequencing are indicated.
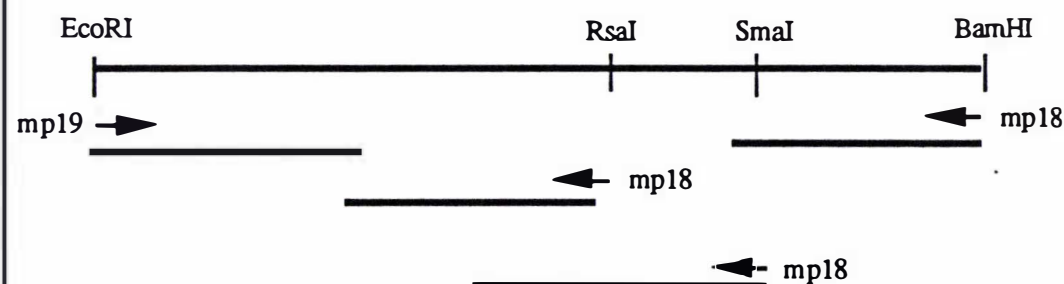
2.2.2   Cloning of Lf$_N$ cDNA into the expression vector pNUT

Preparation of pNUT for use as a vector

A restriction map of pNUT (Palmiter *et al.*, 1987), the vector used in this study, is shown in Fig. 12. pNUT is not a particularly convenient vector to work with, as it does not contain a polylinker and recognition sites for most of the commonly used restriction enzymes are scattered throughout the plasmid. The vector is also relatively unstable and deletion mutants are often obtained making routine manipulations of the plasmid difficult. Despite these problems, pNUT remains a

useful vector because when digested with SmaI the human growth hormone (hGH) genomic sequence is removed but the polyadenylation and transcription termination signals, along with the metallothionein (MT-1) promotor are left intact. This allows fragments of DNA to be inserted into pNUT between the MT-1 promotor and the hGH polyadenylation sequences, obviating the need for inserts to contain their own processing signals.

Fig. 12 Restriction map of pNUT.
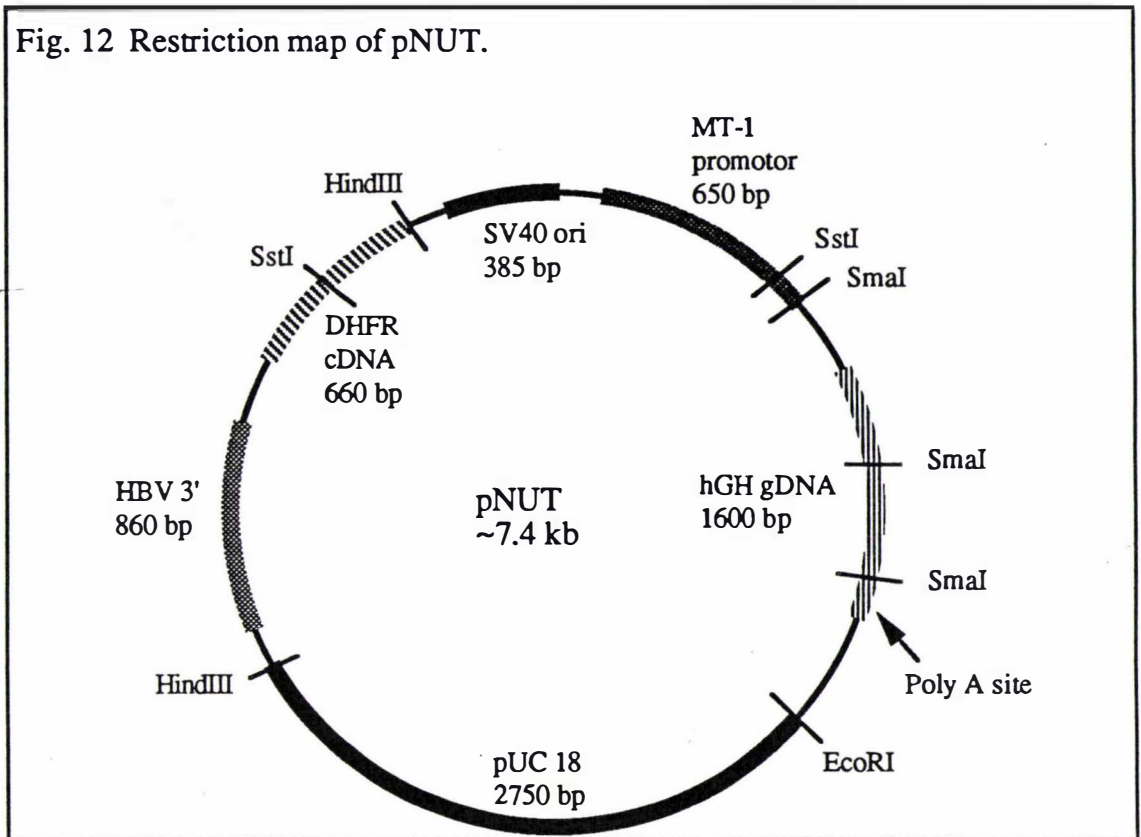
The vector was prepared by complete digestion with SmaI, as verified by agarose gel electrophoresis, followed by treatment with calf alkaline phosphatase (Section A1.2.15) which removes the terminal phosphate groups from the vector preventing religation of the vector without incorporation of an insert. After dephosphorylation the reaction mixture was electrophoresed on a 1% LMP gel and the 5.8 kb vector band was purified using Geneclean (Section A1.2.11). The concentration of pNUT was determined on an agarose gel (Section A1.2.2). Dephosphorylation of pNUT was shown to be essentially complete because when ~50 ng of the vector was religated and transformed into *E. coli* (strain XL-1) only 3-4 colonies were obtained. This low level of self religation implied that virtually all colonies obtained using this vector should contain an insert eliminating the need to screen by colony hybridisation. It was important to demonstrate this as there is no blue/white selection in pNUT.

Cloning of the Lf$_N$ cDNA into pNUT

The LF$_N$ cDNA was excised from pUC:Lf$_N$1 by digestion with EcoRI and BamHI. The insert was made blunt ended (Section A1.2.13) and purified from a 1% LMP agarose gel using Geneclean (Section A1.2.11). The 1.1 kb Lf$_N$ cDNA fragment was quantitated (Section A1.2.2) and used in a ligation reaction containing pNUT prepared as described in the previous section. A vector to insert ratio molar ratio of 1:5 was used. After ligation for 24 hours the reaction was transformed into *E. coli* (strain XL-1) cells and plated out on LB plates containing ampicillin. After overnight incubation 50 colonies were obtained. Plasmid DNA was prepared from 12 of these colonies and analysed by digestion with EcoRI and Bgl II. These enzymes are diagnostic for both the presence and the orientation of the insert because pNUT contains a single EcoRI site and Lf$_N$ contains a single Bgl II site. It is important to be able to determine the orientation of the insert since ligation of Lf$_N$ into pNUT involves blunt ends, thus allowing the insert to be present in either of two possible orientations, only one of which is likely to be expressed. The possible orientations and the fragments that would be expected when digested with EcoRI and Bgl II are shown in Fig. 13.

Fig. 13 Possible orientations of the Lf$_N$ cDNA in pNUT.
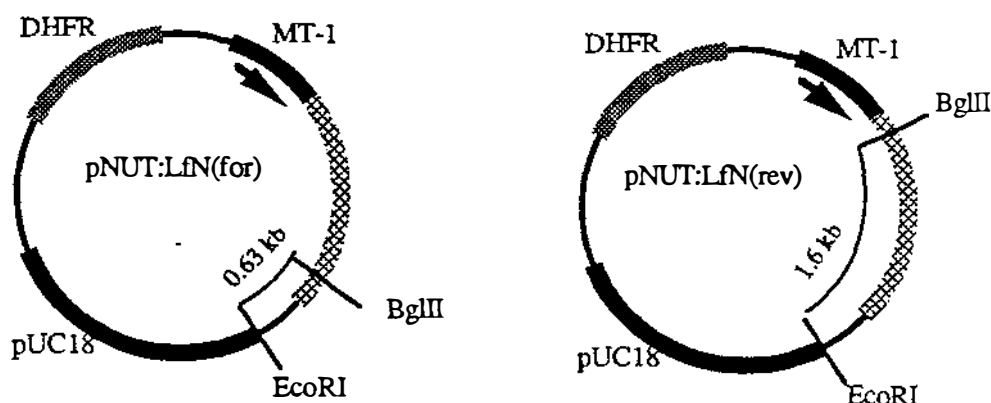
Analysis of the 12 clones showed that one contained Lf$_N$ in the correct, forward orientation and three contained it in the reverse orientation. Two clones were chosen, one forward [pNUT:Lf$_N$(for)1] and one reverse [pNUT:Lf$_N$(rev)1]. DNA was prepared on a large scale from both clones, as described in section A1.2.10, and analysed by restriction enzyme digestion to determine its authenticity (Fig. 14). The restriction fragments obtained are the same as those predicted, demonstrating that no major insertions or deletions had occurred.

Fig. 14 Restriction analysis of pNUT:Lf$_N$(for)1 and pNUT:Lf$_N$(rev)1.

Aliquots of restriction digests from both plasmids were analysed by gel electrophoresis. Lane (1) BRL 1 kb ladder; Lanes 2-7 contain pNUT:Lf$_N$(for)1 digested with (2) HindIII; (3) EcoRI/BglII; (4) EcoRI/SmaI; (5) SstI; (6) SstI/SmaI; (7) SstI/BglII; and lanes 8-13 contain pNUT:Lf$_N$(rev)1 digested with (8) HindIII; (9) EcoRI/BglII; (10) EcoRI/SmaI; (11) SstI; (12) SstI/SmaI; (13) SstI/BglII.
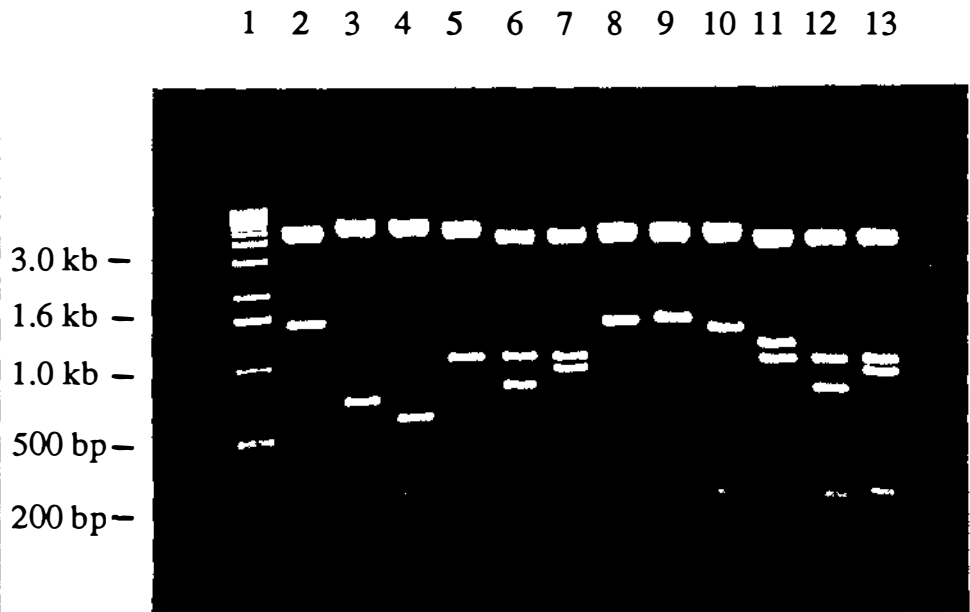
An oligonucleotide complementary to the MT-1 sequence was available and was used to sequence across the junction between pNUT and Lf$_N$ in both clones using double strand sequencing techniques (Section A1.2.16). Approximately 100 bp of sequence data was obtained that was identical to the sequnce of Lf$_N$ obtained when Lf$_N$ was sequenced in M13, indicating that no errors had been introduced at the cloning junctions.

A2.2.3   Expression of Lf$_N$ in transfected BHK cells

Transfection and selection of BHK cells

BHK cells were transfected with pNUT:Lf$_N$(for)1 and pNUT:Lf$_N$(rev)1 and then selected using 500 µM methotrexate (mtx) (Section A1.4.2). After 2 weeks the transformed cells had reached confluence and the cells were passed (Section A1.4.1) and used to seed several more plates. Small 30 mm^2 plates were seeded and grown in the presence of ^{35}S-methionine (Section A1.4.5) allowing all newly synthesised protein to be radiolabelled. After growth for 24 hours the amount of radioactivity incorporated into total protein and the amount incorporated into protein that could be immunoprecipitated by anti-human lactoferrin antibodies was

determined (Section A1.5.5). Analysis of the counts obtained showed that there was no difference between cells transfected with the forward clone or the reverse clone. Both results were similar to the counts obtained from media from untransformed cells. This demonstrated that there was no detectable secretion of lactoferrin from cells transfected with pNUT:Lf$_N$(for)1. The possibility remained that the cells synthesised Lf$_N$ but that it was not secreted. To assay for the presence of Lf$_N$ within the cells small plates of cells were grown in an identical fashion as before except that after removal of the media the cells were thoroughly washed and then lysed (Section A1.5.5). The lysate was assayed for the presence of lactoferrin by determining the amount of radioactivity precipitated by anti-human lactoferrin antibodies. As before the counts precipitable by anti-human lactoferrin antibodies was no different to the controls. These results demonstrated that BHK cells transfected with pNUT:Lf$_N$(for)1 did not synthesise any protein that was recognised by antibodies against human lactoferrin. It had previously been shown (data not shown) that these antibodies could recognise proteolytically derived fragments of human lactoferrin implying that the failure to form an immunoprecipitate was due to the absence of protein rather than the inability of the antibodies to recognise the half molecule.

Analysis of BHK cell genomic DNA

Although BHK cells transfected with pNUT:Lf$_N$(for)1 were resistant to methotrexate the possibility existed that the cDNA for Lf$_N$ had not been incorporated. To investigate this possibility genomic DNA was isolated from untransformed BHK cells and BHK cells transfected with pNUT:Lf$_N$(for)1 and pNUT:Lf$_N$(rev)1. The DNA was used in PCR reactions with the primers JT11 and JT30. The products of each reaction were analysed by agarose gel electrophoresis (Fig.15). The presence of a 1.1 kb band in lanes 2 and 3, that is identical in size to the product obtained from the positive control (lane 5) showed that the cells did contain the cDNA. The negative control containing no DNA (lane 6) did not give rise to a product, demonstrating that contamination had not occurred. This result showed that the lack of protein synthesis was not due to the absence of the cDNA.

Analysis of BHK cell RNA

While the cDNA had been shown to be present in the BHK cells it was possible that it was not being transcribed in these cells. To analyse for this possibility RNA was prepared (Section A1.4.5) from untransformed BHK cells and from BHK cells transformed with pNUT:Lf$_N$(for)1 and pNUT:Lf$_N$(rev)1. The total RNA

Fig. 15 PCR analysis of genomic DNA isolated from BHK cells.

DNA was isolated from BHK cells as described in section A1.4.4 and then used in PCR with the primers JT11 and JT31. An aliquot of each reaction was analysed by electrophoresis. Lane (1) BRL 1 kb ladder; (2) Analysis of DNA from untransfected cells; (3) Analysis of DNA from cells transfected with pNUT:Lf$_N$(for)1; (4) Analysis of DNA from cells transfected with pNUT:Lf$_N$(rev)1; (5) Positive control using pNUT:Lf$_N$(for)1 DNA as the template; (6) Negative control (No template DNA).
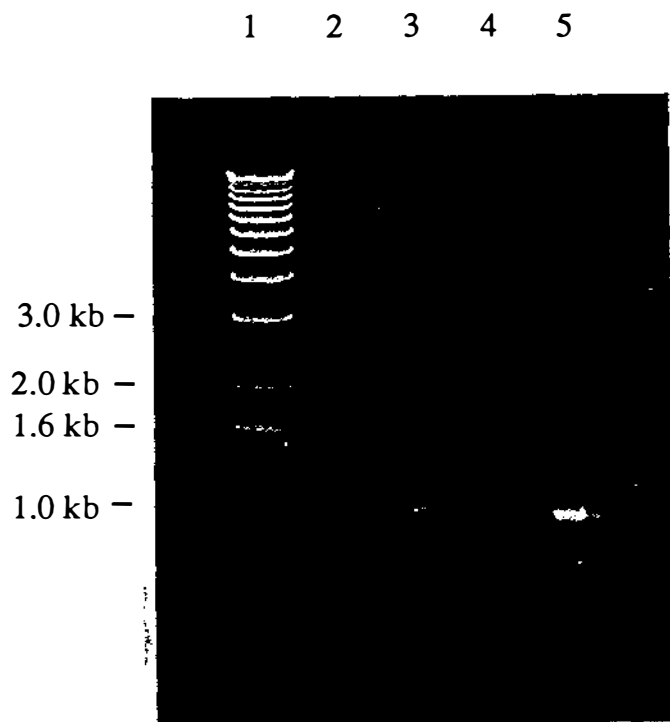
was separated by formaldehyde-agarose gel electrophoresis (Fig.16a) and shown to be intact because the 28S and 18S rRNA bands are clearly visible along with several other minor bands. The RNA from duplicate gels (identical to the one shown in Fig. 16a) was transferred to nitrocellulose and hybridised to radioactively labelled probes (Section A1.2.20). One blot was hybridised with Lf$_N$ cDNA from pUC:Lf$_N$ (Fig. 16b) and a second was hybridised with the 200 bp HindIII/SstI fragment from pNUT (Fig. 16c). The 200 bp HindIII/SstI fragment (Fig. 12) represented part of the cDNA encoding dihydrofolate reductase (DHFR) and was included as a positive control demonstrating that this cDNA was being transcribed. The 1.1 kb Lf$_N$ cDNA hybridised strongly to an RNA species of ~1.1 kb in size (determined by reference to the position of the 28S and 18S RNA species) in RNA obtained from pNUT:Lf$_N$(for)1 cells (Fig. 16b, lane 3). Only a weak signal was obtained with RNA from the reverse clone indicating that there was either a low level of lactoferrin mRNA present or that the probe had a low affinity for the RNA from the reverse clone. As expected a strong signal was obtained in all transfected cells hybridised with the DHFR cDNA. The control

untransformed BHK cells did not hybridise to either probe demonstrating that these cells did not contain DHFR or lactoferrin mRNA sequences.

Fig. 16 Analysis of the RNA isolated from BHK cells.

10 μg of of total cellular RNA was separated on 1.5% formaldehyde-agarose gels (Fig. 14a) and transferred to nitrocellulose (Section A1.2.18). Duplicate blots were prepared, hybridised to the appropriate probe at 68°C and then washed at high stringency (1x SSC, 68°C, 30 min). Fig. 16b was probed with Lf_N cDNA and Fig. 16c was probed with DHFR cDNA. The samples on the gel and on both blots were, Lane (1) untransfected BHK cell RNA; (2 and 4) RNA isolated from BHK cells transfected with pNUT:Lf_N(rev)1; (3 and 5) RNA isolated from BHK cells transfected with pNUT:Lf_N(for)1.

Fig. 16a   1 2 3 4 5
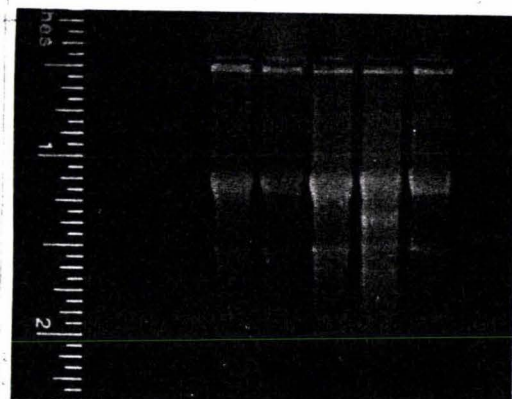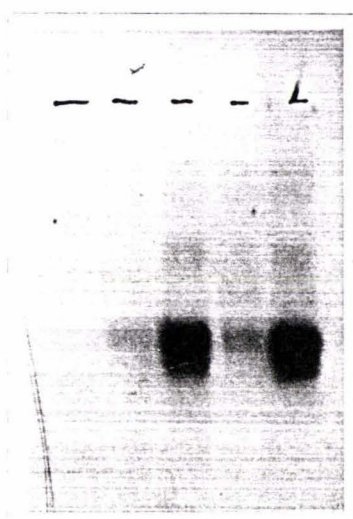
Fig. 16b 1 2 3 4 5   Fig. 16c 1 2 3 4 5
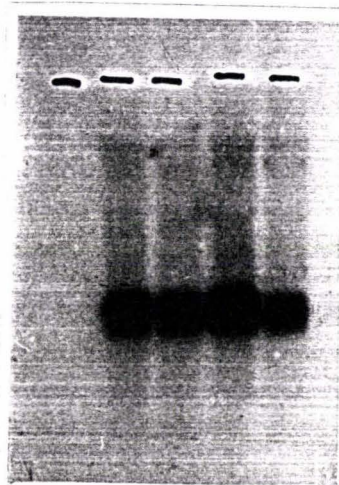
These results demonstrated that the inability of the BHK cells transfected with pNUT:Lf_N(for)1 to synthesise Lf_N protein was not due to inefficient transcription of the cDNA. The reason for the weak Lf_N signal in pNUT:Lf_N(rev)1 cells is not clear although it is quite possible that the antisense strand of Lf_N is less stable. As the DNA and mRNA were both present in the BHK cells there must exist another

reason for the absence of protein. Possibilities considered were that the protein was unstable or the mRNA was not translated or that perhaps the cDNA contained an error preventing expression of a lactoferrin-like species.

Further analysis of the pNUT clones

At this stage I travelled to Vancouver to work in the laboratory of Dr Ross MacGillivray where expression of the amino terminal half of human serotransferrin (sTf$_N$) had been obtained. Here I had access to an oligonucleotide sythesiser and decided to make sequencing primers ~220 bp apart along the Lf$_N$ cDNA. These primers would allow the complete sequence analysis of the Lf$_N$ cDNA in pNUT to be carried out. Using these primers (see Appendix 2 for the primer sequences) sequence was obtained that extended from inside the the MT-1 promotor to inside the hGH termination sequence. Analysis of this sequence data demonstrated that the cDNA was correct up to base 964 but at this position a single G nucleotide was absent when compared to the reported sequence of Rado *et al.* (1987) and Rado *et al.* (unpublished). The absence of base 964 would cause a frameshift mutation scrambling the end of the lactoferrin sequence. Even given this error it was quite surprising that no lactoferrin species had been detected as the protein produced would still contain ~300 correct amino acids which should be recognised by the antibodies. Clearly there was something else preventing recognition by the antibodies, but this was not pursued here. Analysis of my original sequence data showed that base 964 was also missing in the PCR product but the error had not been detected. At this stage it was assumed that the missing base was a PCR induced error although PCR is reported (Tindall and Kunkel, 1988) to mainly introduce transitions or transversions rather than deletions or insertions.

A2.2.4   Construction of a new cDNA

By this time Dr K.M. Stowell had successfully obtained expression of intact human lactoferrin using cDNA derived from human bone marrow tissue (Stowell *et al.*, 1991). These results demonstrated that the system chosen was amenable to synthesis of recombinant lactoferrins. It was hoped that by repeating the steps undertaken previously a clone capable of expressing lactoferrin could be constructed.

Synthesis and cloning of cDNA number 2!

For the second attempt a slightly different strategy was used. A new 3'
oligonucleotide primer was synthesised (CLD7, shown in Fig. 17) that contained
two translational stop codons (underlined) and an EcoRI site (bold).

Fig. 17 Sequence of the 3' primer used to construct the second Lf$_N$ cDNA.
The position of this primer with respect to the DNA and protein sequences is also indicated.


```
protein 327 Ala Ile Gln Asn Leu Arg Lys  333
DNA  1040  CC ATC CAG AAC TTG AGG AAA  1059
CLD7   3'  GG TAG GTC TTG AAC TCC TTT ATTACTCTTAAGG 5'
```

The two stop codons were included in this oligonucleotide because the construct
used to express sTf$_N$ had two stop codons in the corresponding position (Funk *et
al.*, 1990). An EcoRI site was incorporated to aid cloning of the recombinant
cDNA. A 1.1 kb PCR product was generated from pUC:hLf1.3 in the same
manner as described in section A2.2.1 except that JT11 was replaced by CLD7.
In an attempt to reduce the possibility of incorporating PCR-induced errors into the
cDNA encoding Lf$_N$ it was decided to include only a small piece of the PCR
product by utilising the BglII site at position 950 and the EcoRI site in the 3'
primer. To allow the 3' EcoRI site to be utilised it was necessary to have a clone
containing the full length human lactoferrin cDNA with the EcoRI site from the
polylinker at the 3' end of the construct. The steps involved in producing this
construct are given in Appendix 3, essentially the human lactoferrin cDNA in
pGEM:hLf (Stowell, 1990) was inverted. The new construct was called
pGEM:hLf3 and a partial restriction map of this clone is shown in Fig. 18. For
future reference it should be noted that pGEM:hLf3 was derived from pGEM:hLf
which contained the lactoferrin cDNA, produced from bone marrow by PCR, and
not the cDNA derived from the λgt10 library used up to this stage. The vector
used in this case was pGEM:3Zf(+), a phagemid, which contains the F' origin of
replication allowing generation of single stranded DNA without having to subclone
into M13.

After concentration of the PCR products by ethanol precipitation (Section A1.2.4)
the material was digested with BglII and EcoRI releasing a 130 bp fragment which
was isolated from a 1% LMP agarose gel (Section A1.2.12). A complete cDNA
for Lf$_N$ was then constructed by ligating the ~ 130 bp fragment into EcoRI/BglII

cut pGEM:hLf3 which had also been gel purified by 'freeze squeeze' (Section A1.2.12). An outline of the steps involved is shown in Fig. 19.

Fig. 18 Partial restriction map of pGEM:hLf3.
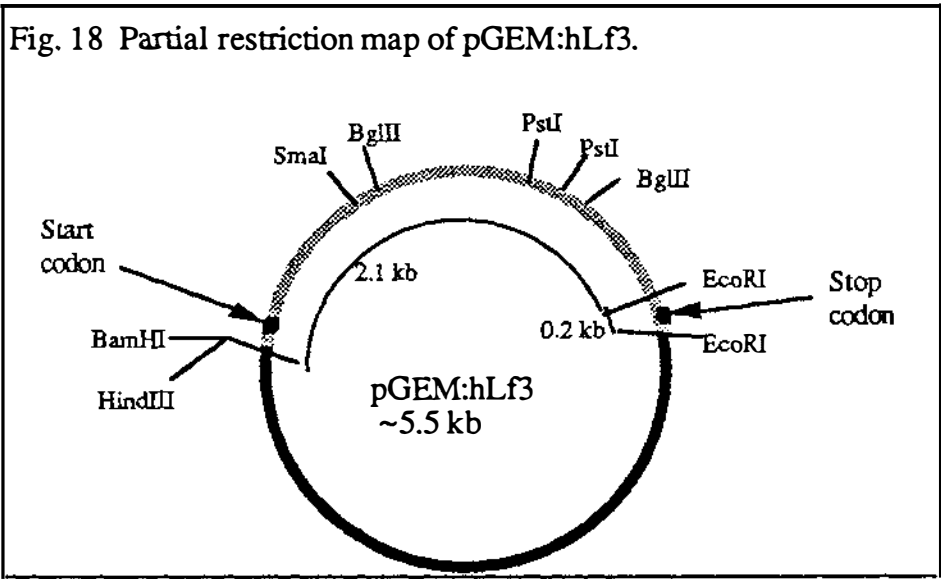
Fig. 19 Strategy used for the construction of the Lf$_N$ cDNA (version 2!).
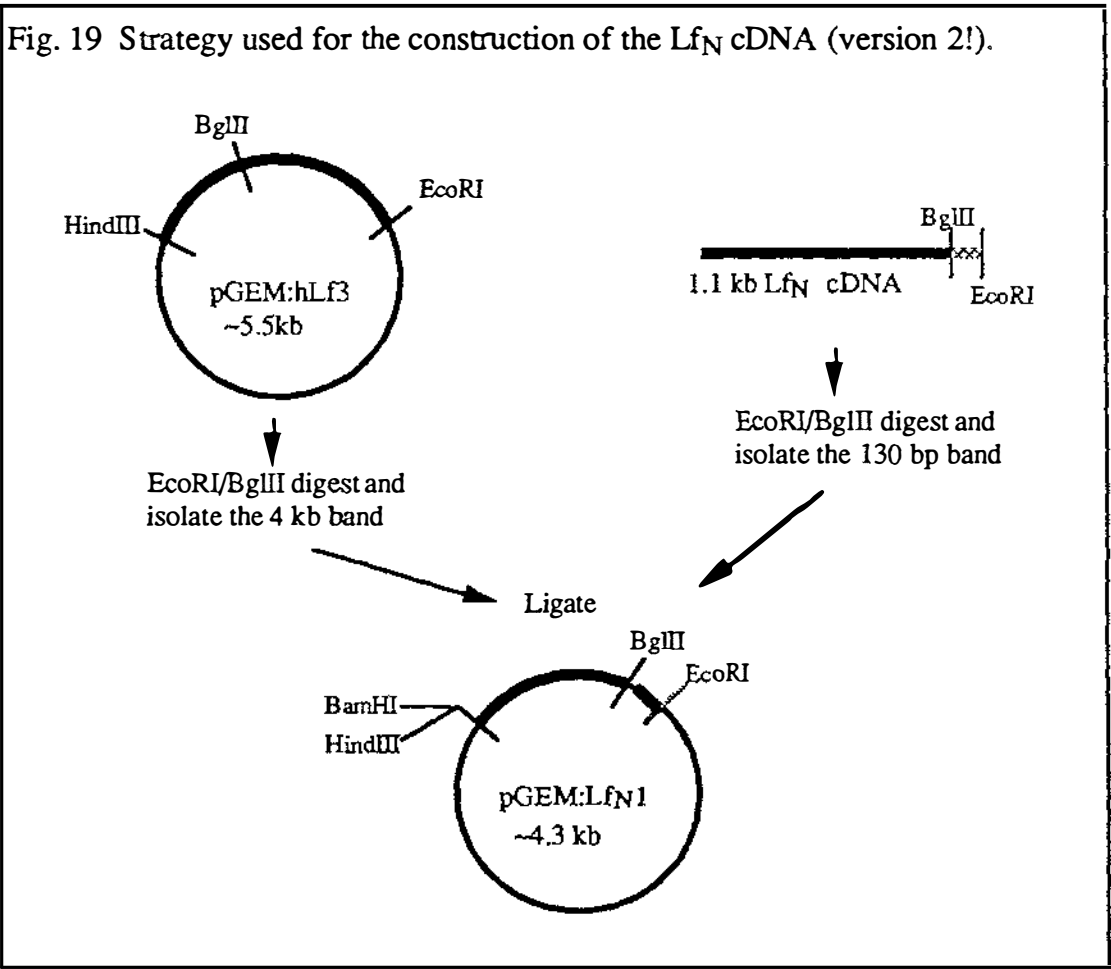
The products of ligation were transformed into _E. coli_ (XL-1). Eight colonies were chosen and small scale DNA preparations were made. The minipreparation DNA was digested with EcoRI and BamHI which were expected to release a 1.1 kb fragment. Two clones were found to contain the expected fragment, one of which was chosen and called pGEM:Lf$_N$. Single strand DNA was prepared

directly from this clone and sequenced using the -40 universal sequencing primer. Approximately 200 bp of sequence data was obtained and compared with the reported sequence for human lactoferrin (Rado *et al.*, 1987; Rado *et al.*, unpublished). To my dismay I found that once again base 964 was absent.

The repeated absence of base 964 was a source of great concern as it was improbable that a PCR error would occur in exactly the same place in two independent reactions. The error observed was consistent with an error in the template used despite the fact that it had been reported to be correct. Careful analysis of the sequence gels used to sequence pUC:hLf revealed that pUC:hLf was also missing base 964 but that this error had not been detected in the original sequencing.

Reassessment of the results obtained so far

At this stage a brief review of the different clones available is required in an attempt to clarify the results obtained so far and to explain the steps that were subsequently taken. The original pUC:hLf clone isolated by Dr K.M. Stowell was constructed from three cDNA's isolated from a λgt10 library generated from a patient with chronic myeloid leukaemia. It was this clone which was used as the template for generating the 1.1 kb fragments discussed so far. Another cDNA clone, pGEM:hLf, encoding human lactoferrin was subsequently made in our laboratory by PCR, from first strand cDNA isolated from human bone marrow (Stowell *et al.*, 1991). The cDNA from both pUC:hLf and pGEM:hLf was cloned into pNUT and transformed into BHK cells but only the cDNA derived from pGEM:hLf was shown to direct the synthesis of lactoferrin (Stowell, 1990). It had been postulated that the cDNA in pNUT:hLf3 (which was derived from pUC:hLf) was not expressed because of differences in the 3' region compared to the PCR derived clone (Stowell, 1990). In light of the data presented here it is unlikely that differences in the 3' sequence account for the inability to obtain expression. A more likely reason is the absence of base 964 which would cause a frame shift mutation resulting in premature termination of the protein, and quite possibly incorrect folding. At the time I began this project expression of pNUT:hLf3 had not been attempted and the error had not been recognised therefore there had been no reason to suspect that the cDNA was faulty. By this stage expression of lactoferrin from pNUT:hLf5 had been obtained and it was assumed that the lactoferrin cDNA present in pGEM:hLf had the correct sequence. Consequently lactoferrin cDNA derived from pGEM:hLf was used in all subsequent steps.

Subcloning to remove the error

A new cDNA could be produced by either making a new PCR product, using pGEM:hLf as the template, or by subcloning to remove the error. Examination of the cDNA sequence showed that the restriction enzyme TaqI only cleaves the Lf_N cDNA at a single site (position 983). A strategy, summarised in Fig. 20, was designed utilising this site.

Fig. 20 Strategy used for the construction of Lf_N cDNA (version 3!).
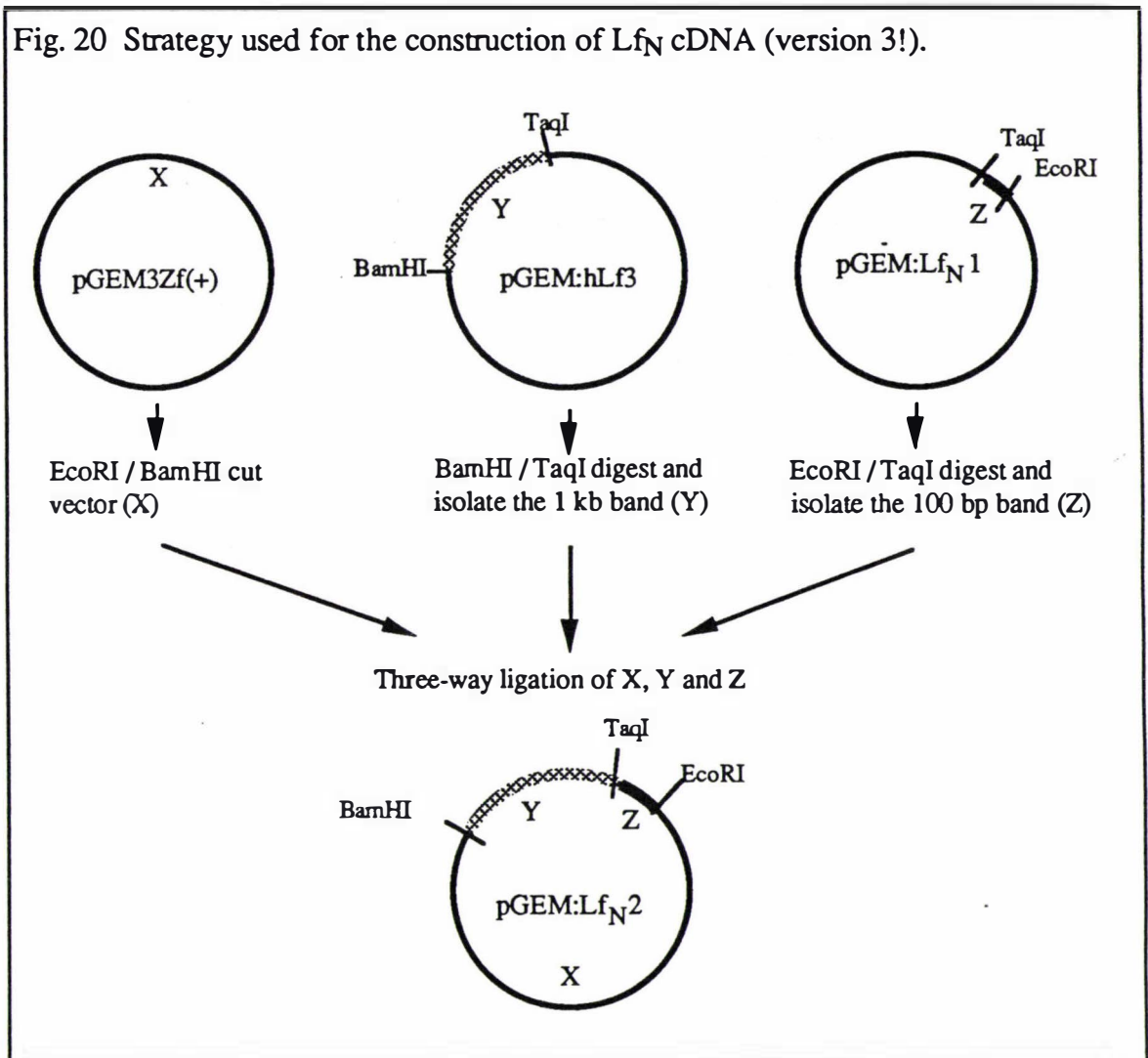
Because pGEM:3Zf(+) contains several TaqI sites it was necessary to prepare the three fragments (labelled **X**, **Y** and **Z** in Fig. 20) separately and carry out a three-way ligation to produce pGEM:Lf_N2. A ligation was carried out (Section A1.2.11) containing equimolar amounts of each fragment and then the products were transformed into *E. coli* (strain XL-1) and grown in the presence of X-gal and IPTG. Twelve white colonies were chosen and DNA from these was analysed by restriction digestion using BamHI and EcoRI. Ten clones appeared to contain the correct fragment, one of which was chosen and called pGEM:Lf_N2. Before proceeding further, single strand DNA was prepared (Section A1.2.10) from this clone and sequenced. The cDNA insert was shown to have the same

sequence as the cDNA in pGEM:hLf suggesting that it was correct. DNA was prepared on a large scale from pGEM:Lf$_N$2 and then analysed by restriction digestion (Fig. 21). The fragments obtained were the same as those which had been predicted indicating that no major insertions or deletions had occurred.

Fig. 21 Restriction analysis of pGEM:Lf$_N$2.

Aliquots of pGEM:Lf$_N$2 were digested with the following restriction enzymes and analysed by gel electrophoresis. Lane (1) BRL 1 kb ladder; (2) EcoRI/BamHI digest; (3) BamHI/SmaI digest; (4) EcoRI/SmaI digest; (5) BamHI/BglII digest; (6) EcoRI/BglII digest.
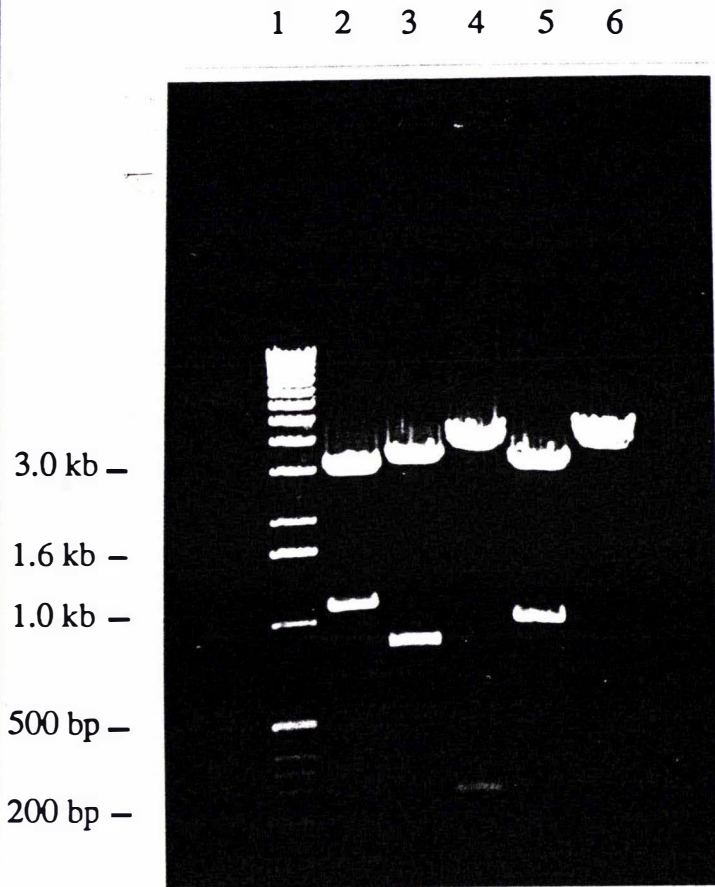
Cloning into pNUT

The 1.1 kb insert from pGEM:Lf$_N$2 was released using EcoRI and BamHI and then treated in the same manner as described previously (Section A2.2.2). pNUT clones containing Lf$_N$ in both the forward and reverse orientations were chosen and grown on a large scale (Section A1.2.10). Fragments of the expected size were obtained when pNUT:Lf$_N$ was analysed by restriction (data not shown). To further confirm that the forward clone was correct the insert was sequenced by double strand sequencing (Section A1.2.16) using MT30, the primer complementary to the metallothionein promotor, and the primers complementary to lactoferrin which have been described previously (Section A2.2.3). This time the forward clone was shown to have the correct sequence. The clones were called

pNUT:Lf$_N$(for)2 and pNUT:Lf$_N$(rev)2. The sequence of the cDNA in the reverse clone was not determined.

Transfection of BHK cells and analysis of the RNA

DNA from the new forward and reverse clones was used to transfect BHK cells (Section A1.4.2). The transformed cells were analysed for the expression of an appropriate RNA species. RNA was prepared and analysed as described

Fig. 22 Analysis of the RNA isolated from BHK cells.

10 μg of total cellular RNA was separated on 1.5% formaldehyde-agarose gels (Fig. 22a) and transferred to nitrocellulose (Section A1.2.18). Duplicate blots were prepared, hybridised to the appropriate probe at 68°C and then washed at high stringency (1x SSC, 68°C, 30 min). Fig. 22b was probed with Lf$_N$ cDNA and Fig. 22c was probed with DHFR cDNA. The samples were isolated from the following cell lines (1) untransfected BHK cells; (2) BHK cells transfected with pNUT:hLf-5; (3) BHK cells transfected with pNUT:Lf$_N$(for)2; (4) BHK cells transfected with pNUT:Lf$_N$(rev)2.
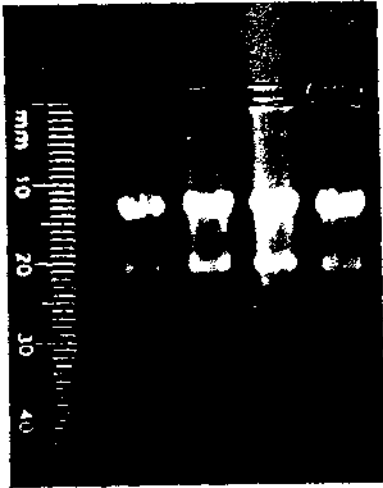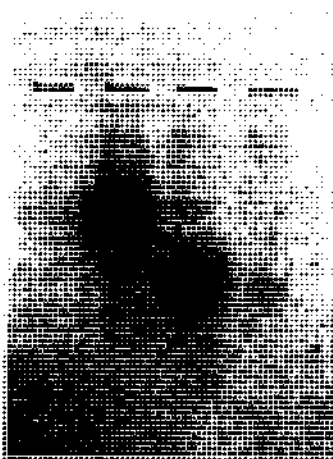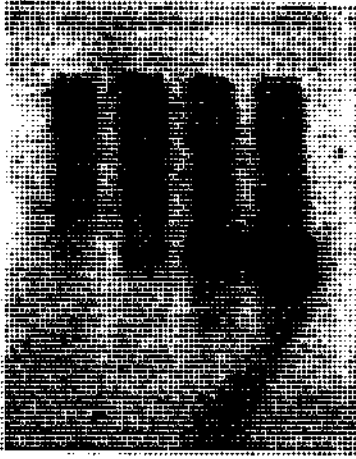
previously (Section A2.2.3). This time RNA was also isolated from pNUT:hLf5 cells and used as a positive control. An RNA species ~ 1.1 kb in size was found to be synthesised in pNUT:Lf$_N$(for)2 cells (Fig. 22b, lane 3) and again a low level of expression was obtained from the reverse clone (Fig. 22b, lane 4). An interesting difference in the level of DHFR RNA was detected with the cells containing hLf expressing considerably less than the Lf$_N$(for)2 cells. The reason for this difference is not clear as both cell lines appear to grow equally well in the presence of methotrexate.

Analysis of the cell media

Transformed cells were grown to confluence and the media collected. Instead of analysis by ^{35}S labelling of newly synthesised proteins the cell media was analysed by immunoprecipitation and subsequent electrophoresis of the immunoprecipitates. This method had been successfully used to detect expression of intact human lactoferrin by BHK cells. Initially a quick analysis was carried out by adding 20 μl of anti-human lactoferrin antibodies to 1 ml of media. After incubation at 37°C for one hour the samples were centrifuged and a small light pellet was clearly visible suggesting that lactoferrin was being produced. Following this analysis immunoprecipitates were formed over a gradient (Section A1.5.4) and the washed pellets were separated on a 12% polyacrylamide gel (Section A1.5.1) (Fig. 23a). A band of ~ 40 kDa in size is visible in the precipitate obtained using the media from cells which contain Lf$_N$(for) cDNA (lane 7). To confirm that this band was Lf$_N$ a gel equivalent to that in Fig. 23a was transferred to nitrocellulose and reacted with anti-human lactoferrin antibodies and then with a fluorescent detection reagent (Section A1.5.6). A band is clearly visible in lane (Fig. 23b, lane 7) corresponding to the 40 kDa protein.

3 DISCUSSION

BHK cells containing pNUT:Lf$_N$(for)2 have been shown to synthesise and secrete a protein ~40 kDa in size. The protein can be precipitated from culture media with anti-human lactoferrin antibodies and reacts with the same antibodies in a protein blot. These results demonstrate that the 40 kDa protein produced is a fragment of human lactoferrin and is probably the N-terminal half. Lf$_N$ was easily detected in the medium suggesting that this system would be suitable for the expression of large quantities of recombinant protein. The level of expression was estimated to be between 20 and 30 mg/ml although it could not be accurately determined as there was no RIA or ELISA system available to us at this time.

Fig. 23 Analysis of cell culture media.

Duplicate samples were separated on 12% SDS-polyacrylamide gels and stained with Coomassie blue (Fig. 23a) or transferred to nitrocellulose and reacted with antibodies to human lactoferrin (Section A1.5.6) (Fig. 23b). Samples for both gels were (1) Molecular mass markers (kilodaltons); (2) native human lactoferrin; (3) anti-human lactoferrin antibodies; (4) 5 µl of medium from untransformed BHK cells; (5) immunoprecipitate from 250 µl of untransformed BHK cell medium; (6) 5 µl of medium from cells transfected with pNUT:Lf$_N$(for)2; (7) immunoprecipitate from 250 µl of pNUT:Lf$_N$(for)2 cell medium.

Fig. 23a
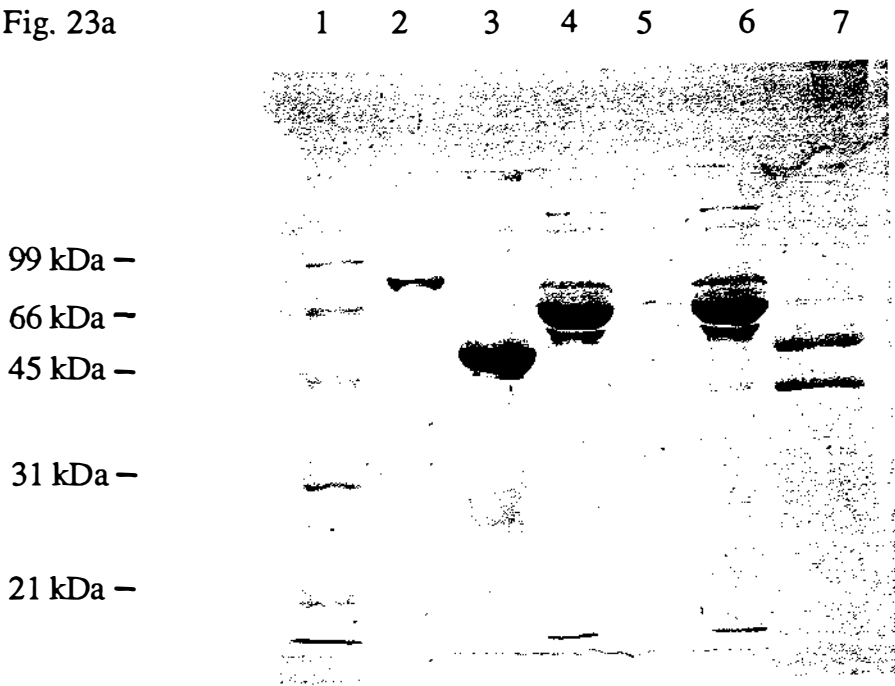
Fig. 23b
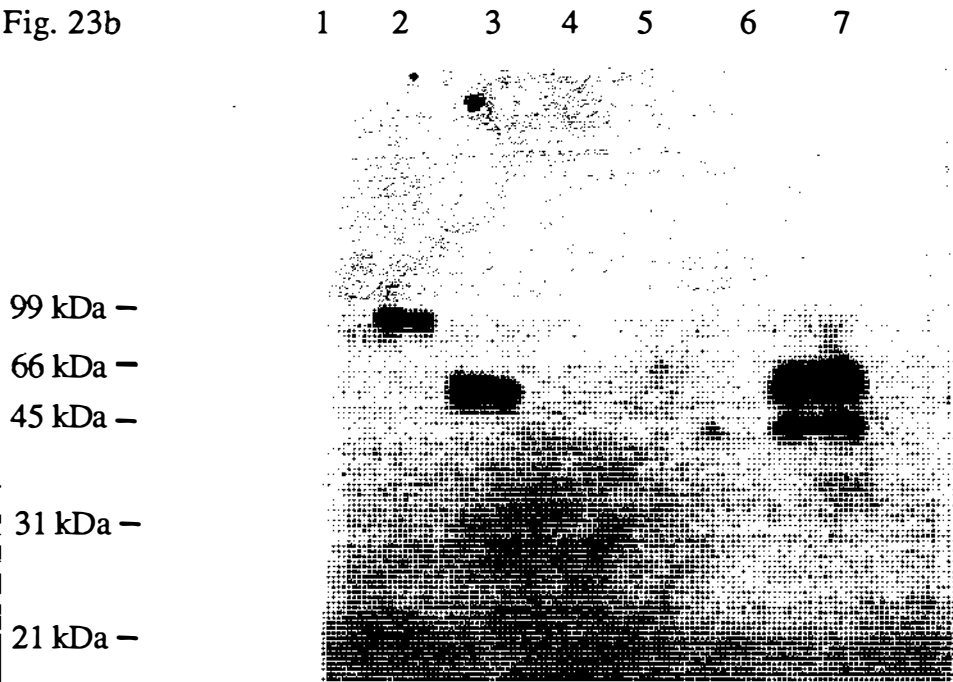
The problems encountered with the initial constructs emphasise the necessity to take great care when sequencing DNA and the need to have sequence data, especially that for new clones, subjected to independent analysis. Even though the initial constructs were missing a nucleotide it is quite surprising that the cells didn't produce a protien that could be recognised by the antibodies to lactoferrin as the cDNA still encoded ~300 amino acids from lactoferrin. This may suggest that in the absence of the last residues the protein does not fold correctly and is thus unstable.

Part A : Chapter Three

CHARACTERISATION OF Lf$_N$

3.1   INTRODUCTION

Following the demonstration that transformed BHK cells were synthesising a fragment of lactoferrin it was necessary to establish that the recombinant protein was correctly processed and that it was synthesised in sufficient quantities to allow complete characterisation. In addition, before the role of individual amino acids could be investigated through site specific mutagenesis it was necessary to fully characterise the native N-terminal fragment.

3.2   RESULTS

3.2.1   Purification of recombinant Lf$_N$

BHK cells that had been stably transfected with pNUT:Lf$_N$(for) were grown in roller bottles (Section A1.4.3). The medium was collected daily once the cells had reached confluence and stored at 4°C prior to purification of the recombinant protein (Section A1.5.7). The recombinant protein was eluted from the CM-Sephadex column with 0.4 M NaCl in HEPES as a dilute broad peak. All fractions which had A$_{280}$ readings greater than 0.05 were analysed by SDS-PAGE to determine their purity. A typical elution profile and SDS-PAGE analysis of the corresponding fractions is shown in Fig. 24. Analyses of this type demonstrated that the predominant protein species purified had a molecular weight of ~ 40 000 Da although fractions 40 to 44 appeared to contain a minor contaminant with a molecular weight of ~37 000 Da. The nature of the lower molecular weight contaminant was not known at this stage. Fractions 28 to 39 were pooled and concentrated in an ultrafiltration cell ready for characterisation while fractions 40 to 44 were combined and subjected to a further chromatographic step. Because Lf$_N$

was not eluted from the column until at least three column volumes of 0.4 M NaCl buffer had passed through the column it was not considered necessary to dialyse the protein against a standard buffer and unless otherwise stated Lf_N was used in 0.4 M NaCl, 0.01 M HEPES pH 7.8. Up to 30 mg of purified protein was recovered from a litre of medium when the cells in the roller bottles were confluent. This level of expression was considered sufficient to allow the complete characterisation of Lf_N including X-ray diffraction analysis.

Fig. 24 Purification of Lf_N by ion exchange chromatography.

Lf_N was eluted from the column using 0.4 M NaCl. The absorbance of each fraction was monitored at 280 nm (top panel) and a sample of each fraction was analysed by SDS-PAGE to determine the purity (bottom panel).
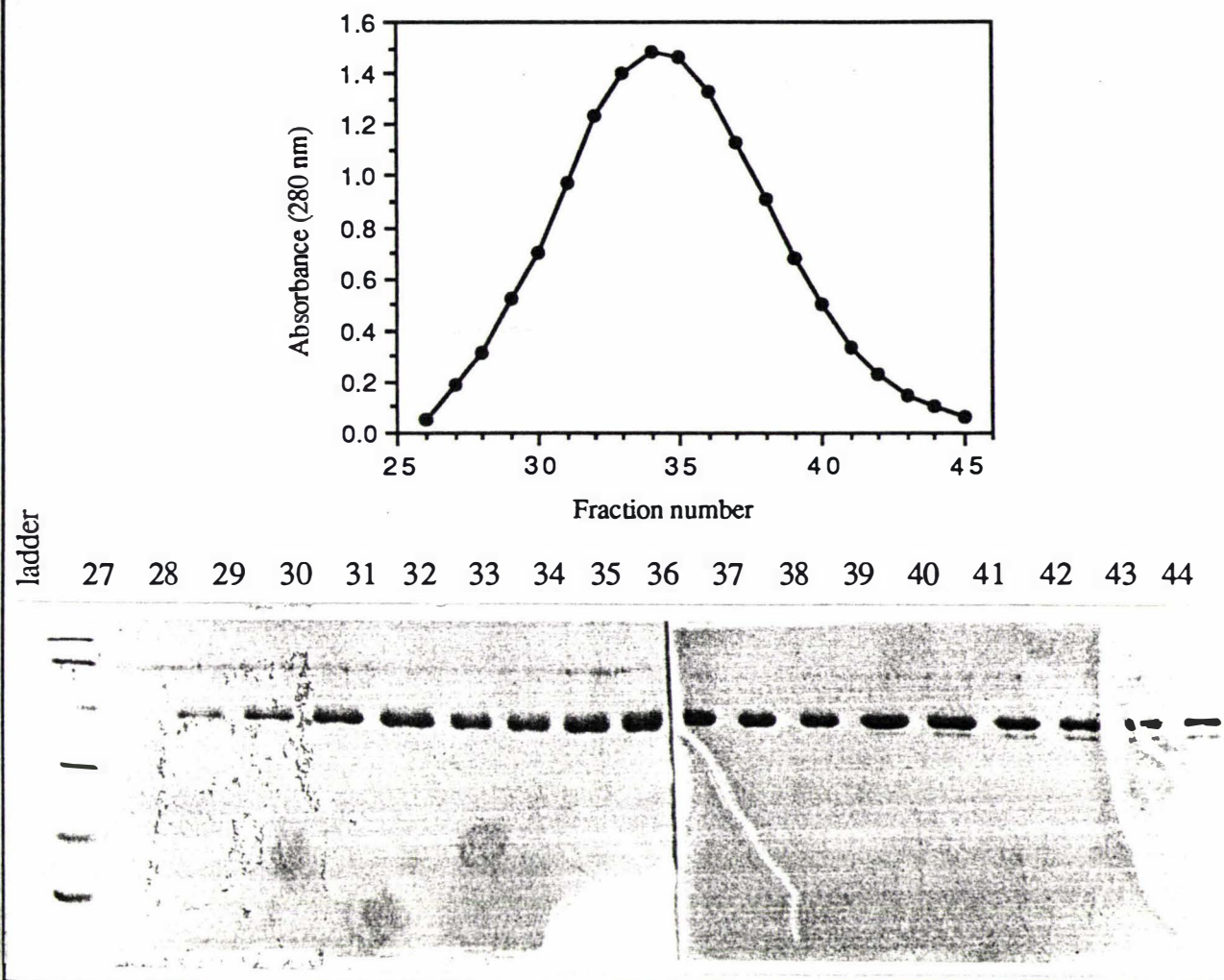
A3.2.2   Characterisation of recombinant Lf_N

A3.2.2.1   Absorption spectrum

The absorption spectrum of recombinant Lf_N was measured between 260 nm and 700 nm (Fig. 25).

Fig. 25 Comparison of the UV-visible spectra for iron saturated Lf_N and hLf.
Spectra were recorded between 250 nm and 700 nm for both Lf_N (solid line) and hLf (dashed line).
The inset contains an enlargement of the region where the shift in the maximum has occurred.
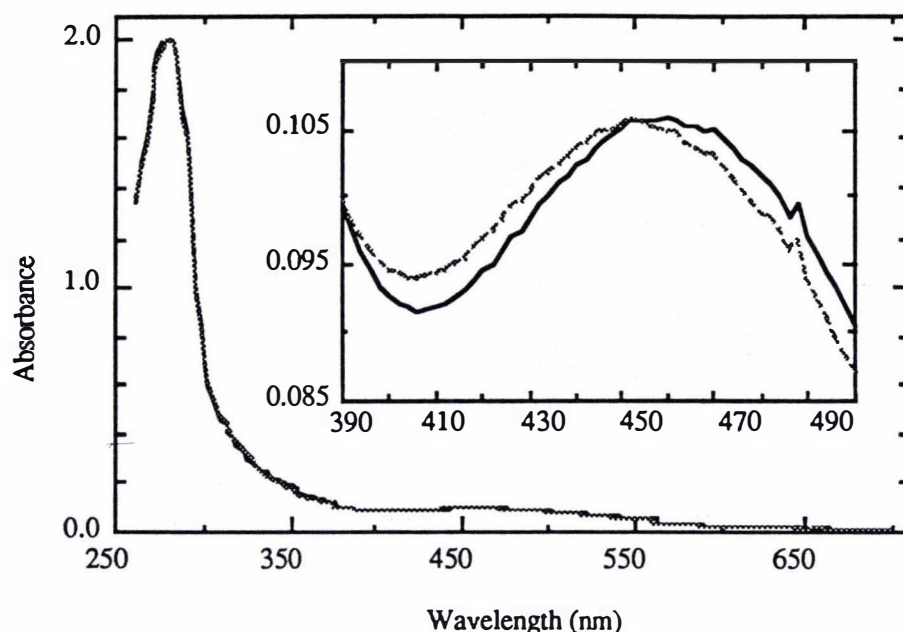
The concentration of the protein was 1.8 mg/ml as determined from the absorbance at 280 nm (Section A1.5.8). The inset contains an enlargement of the region between 400 nm and 500 nm which includes the iron-tyrosine charge transfer band characteristic of lactoferrin (Ainscough *et al.*, 1979). The native lactoferrin spectrum is overlayed on the Lf_N spectrum clearly demonstrating that the 466 nm maximum found for human lactoferrin has shifted to 454 nm for Lf_N. The maximum at 454 nm was repeatedly obtained and was used in all subsequent work. The purity and the % iron saturation of lactoferrin can be estimated by calculating the A_{280}/A_{454} and the A_{406}/A_{454} ratio (Parry and Brown, 1974). The values obtained for pure iron saturated Lf_N were 19.0 and 0.8, respectively, which are in the range expected for pure native lactoferrin. This of course assumes that the values for the pure half molecule are comparable with intact native human lactoferrin.

.2.2.2   N-terminal sequence analysis

Correct processing of the signal peptide from Lf_N was demonstrated by subjecting 60 pmoles of Lf_N to gas phase peptide sequence analysis. The sequence of the first 16 amino acids was determined and shown to be GRRRRSVQWCAVSQPE. This sequence corresponds to the first 16 amino acids of the mature protein (Metz-Boutigue *et al.*, 1984), indicating that the signal peptide had been correctly removed during expression and secretion by the BHK cells. As expected, this is

the same as the sequence obtained by Stowell *et al.* (1991) but it is different to the sequence derived from the cDNA reported by Rado *et al.* (unpublished) which was subsequently reported by Anderson *et al.* (1989). In these instances only three contiguous arginines were found in the first five positions. The extra arginine in Lf_N and in recombinant hLf was not expected to pose any problems but for future reference it is important to clarify the amino acid numbering system used. As the numbering system proposed by Anderson *et al.* (1989) has been extensively used in previous work this system was retained and the additional arginine was referred to as 3a. The sequence and the numbering system used is shown in Table 2.

Table 2 Sequence alignment of the N-terminus of human lactoferrin

Numbering chosen	1	2	3	3a	4	5	6	7	8	etc
Anderson *et al* sequence for hLf	Gly-	Arg-	Arg-		-Arg-	Ser-	Val-	Gln-	Trp-	etc
Sequence found for Lf_N	Gly-	Arg-	Arg-	Arg-	Arg-	Ser-	Val-	Gln-	Trp-	etc

A3.2.2.3 Glycosylation

The amino terminal half of human lactoferrin contains a single recognition site for the addition of an N-linked carbohydrate chain. The molecular weight of Lf_N estimated from the mobility on SDS gels was ~40 000 Da suggesting that the fragment is indeed glycosylated since the amino acid composition of Lf_N would give a molecular weight of 37 084 Da. The presence of an N-linked carbohydrate moiety was investigated by treating Lf_N with the endoglycosidases Endo F and PNGase F (Section A1.5.10) according to the procedures previously used for the deglycosylation of native lactoferrin (Norris *et al.*, 1989). Treatment with these enzymes decreased the molecular weight of Lf_N by ~3 kDa (Fig. 26, lane 4) whereas no change in the molecular weight was detected when Lf_N was treated with N-octyl β-D-glycopyranosidase which is specific for O-linked carbohydrate groups (Fig. 26, lane 3). This result clearly demonstrated that Lf_N did indeed have at least one N-linked carbohydrate chain.

When Lf_N was purified it should be recalled that a minor contaminant ~37 kDa in size was found to copurify with Lf_N. The contaminating band appeared to be very similar in molecular weight to deglycosylated Lf_N and to that expected for unglycosylated Lf_N. The possibility that this band was in fact unglycosylated Lf_N was investigated by treating a small amount of Lf_N from fraction 2 (Fig. 24) with PNGase F. This treatment converted the doublet into a single band with a molecular weight of 37 kDa suggesting that the impurity seen was unglycosylated Lf_N. The nature of the lower molecular weight band was unequivocally identified

Fig. 26 Deglycosylation of Lf$_N$.

Samples of pure Lf$_N$ were treated with deglycosylation enzymes (Section A1.5.10) and then the samples were subjected to gel electrophoresis on a 12% polyacrylamide gel. Lane (1) molecular weight markers (kilodaltons); (2) pure Lf$_N$; (3) Lf$_N$ which has been treated with N-octyl β-D-glycopyranosidase; (4) Lf$_N$ treated with PNGase and Endo F.
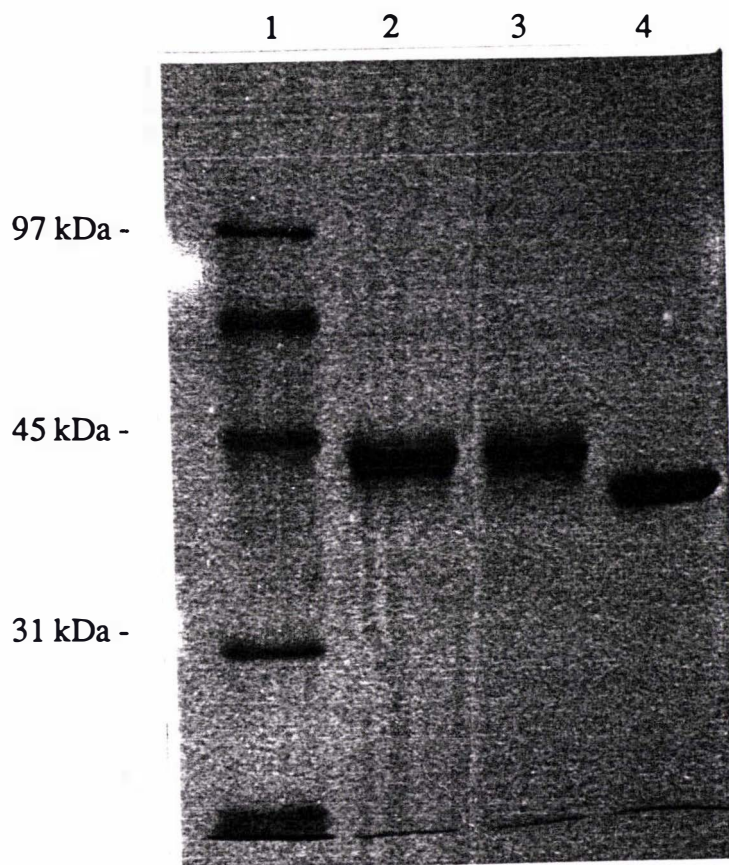
by western blotting. A gel equivalent to that shown in Fig. 27a was transferred to nitrocellulose and reacted with antibodies to human lactoferrin (Section A1.5.6). It is clearly evident that in lane 3 (Fig. 27b) both the high molecular weight band and the less intense lower molecular weight band have reacted with the antibodies, confirming that they are both lactoferrin species. The difference in molecular weight is thus attributed to the complete absence of carbohydrate groups as the mobility of the lower band does not change when it is treated with PNGase F.

The presence of the lower molecular weight band in preparations of Lf$_N$ raises several interesting possibilities: (1) Is a small amount of Lf$_N$ synthesised and secreted by the BHK cells without the addition of carbohydrate groups? (2) Is the lower molecular weight band due to the gradual loss of carbohydrate with time?

Although no studies have been specifically directed at this problem it seems most

Fig. 27 Identification of the lower 37 kDa band found in preparations of Lf$_N$.
In an attempt to positively identify the the lower molecular weight which copurified with Lf$_N$
samples, equivalent to fraction 1 and fraction 2 (Fig. 24), were separated on a 12%
polyacrylamide gel (Fig. 27a), transferred to nitrocellulose (Fig. 27b) and then reacted with anti
human lactoferrin antibodies (Section A1.5.6). Samples for both gels were (1) Molecular weight
markers (kilodaltons); (2) fraction 1; (3) fraction 2.

Fig. 27a   1   2   3   Fig. 27b   1   2   3

97 kDa -

45 kDa -

31 kDa -
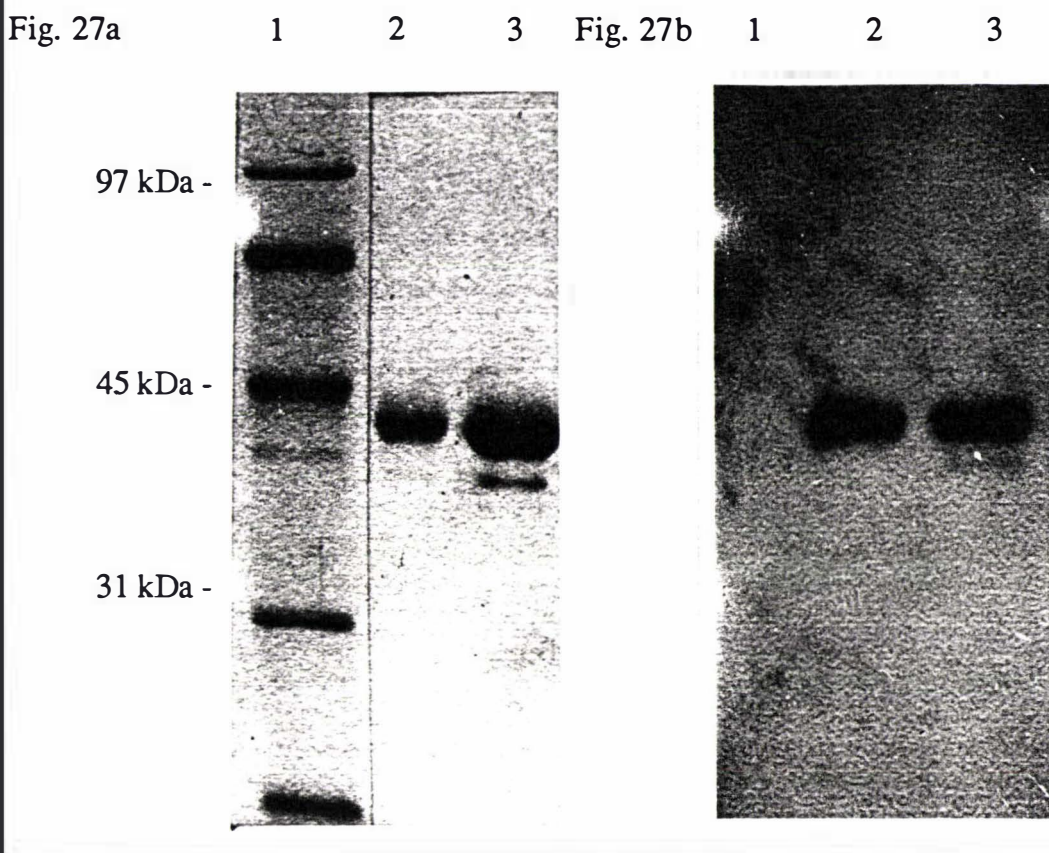
likely that a small amount of Lf$_N$ escapes glycosylation as it passes through the
endoplasmic reticulum. This is supported by the observation that the lower band
is present in freshly prepared Lf$_N$ and that the intensity of this band in samples of
Lf$_N$ which were up to six months old did not appear to increase. Initially the small
amount of unglycosylated protein was not a cause for concern as it could be
separated from the higher molecular weight band and it did not represent more than
~2% of the protein synthesised. In later batches, however, the multiplicity of Lf$_N$
species became more of a problem because often up to four bands were present at
quite high levels. The variation in the extent of glycosylation between batches is
shown in Fig. 28. In the later batches it was not possible to purify one form of
Lf$_N$ away from the others. This made crystallisation of this protein more difficult
(Section B3.2.1). Reasons for the difference in the extent of glycosylation
between batches are unclear as the cells used to produce the protein have always
been derived from the same original batch of frozen cells. The problems

encountered here raise several issues regarding the capacity of BHK cells to produce recombinant glycoproteins at high levels in a consistent manner. The time available did not allow an investigation into the conditions affecting this processing but it would make an interesting topic of study for the future.

Fig. 28 Analysis of the variation in protein purity between different batches. A small amount of protein was saved from several different batches and compared on a 12% polyacrylamide gel. Lane (1) Molecular weight markers (kilodaltons); (2, 3 and 4) Protein samples from several of the earlier preparations.; (5, 6 and 7) Protein samples from several of the later batches of protein.
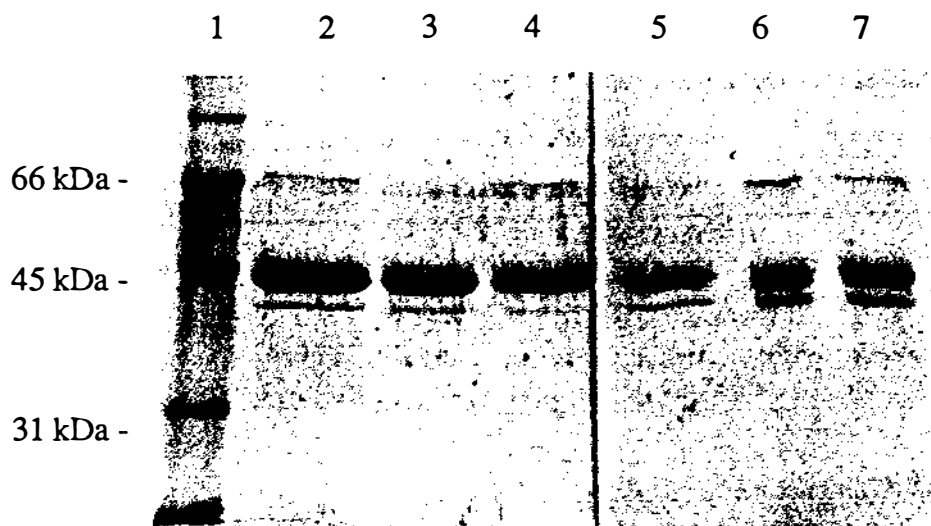
2.2.4 Spectroscopy

The shift in the visible absorption maximum from 466 nm in native human lactoferrin to 454 nm in Lf_N suggested that there existed small differences in the iron coordination between the two proteins. In an attempt to analyse the differences an ESR spectrum was recorded for iron saturated Lf_N (Fig. 29). The spectrum obtained was essentially indistinguishable from that obtained for native lactoferrin with both spectra characterised by a split peak at 1500 Gauss. The similarity of these ESR spectra suggested that differences between the sites were likely to be minor.

In an attempt to understand the nature of the differences in the metal sites Lf_N was saturated with Cu^{2+} instead of Fe^{3+}. Copper is a more informative metal as it gives rise to a more interpretable ESR spectrum (Aisen and Leibman, 1972). When Cu^{2+} was added to $ApoLf_N$ (prepared by lowering the pH to 4.0, section A3.2.2.5), as $CuCl_2$, the protein became yellow in colour. The yellow colour is due to copper-tyrosine charge transfer absorption and clearly demonstrated that

Fig. 29 ESR spectra of iron saturated Lf_N and hLf.

The ESR spectra were recorded as described in section A1.5.11.

$FeLf_N$

700   1500   2300

magnetic field (Gauss)

Fe_2Lf

500   1500   2500

magnetic field (Gauss)
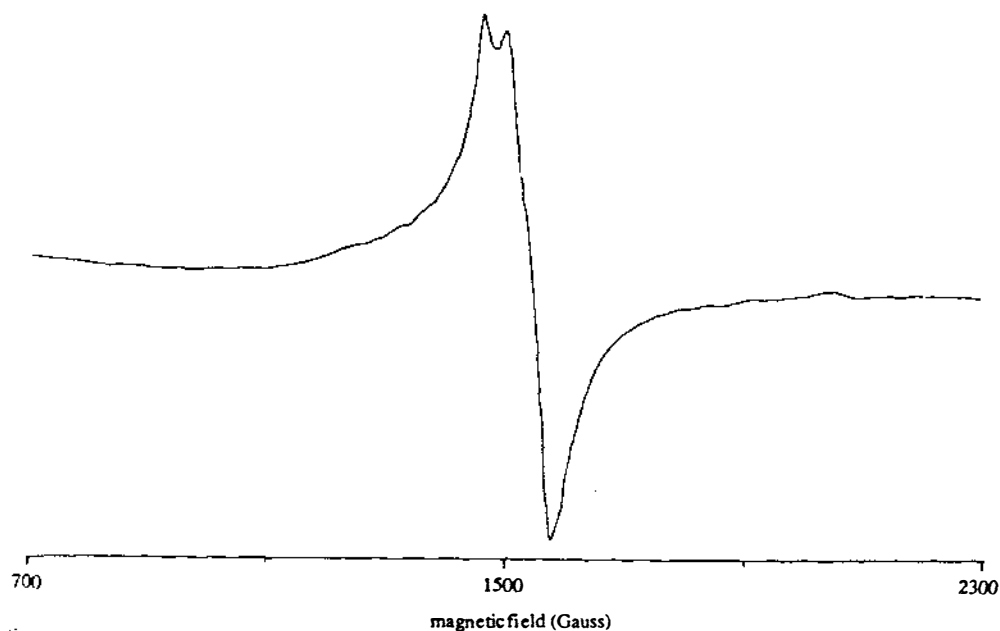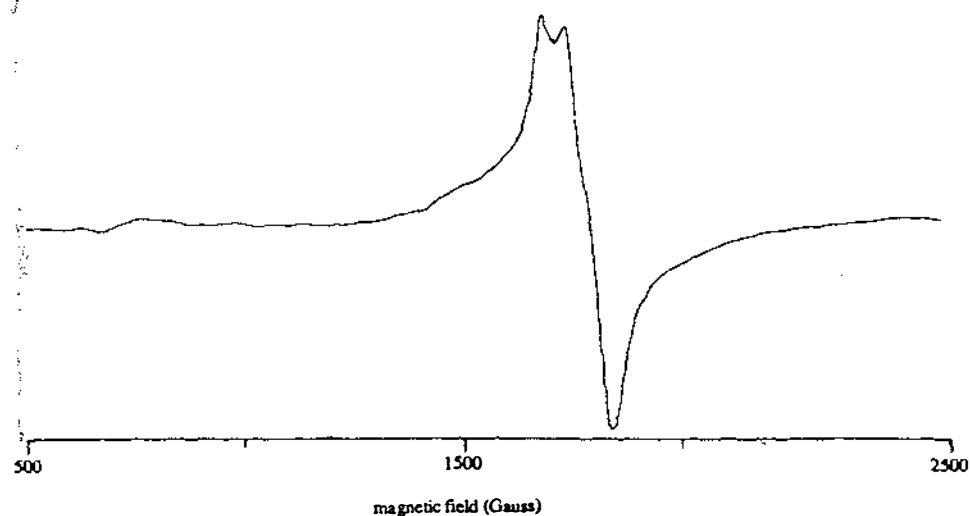
binding had occurred. Copper saturated native lactoferrin (Cu_2Lf) was prepared as described by Smith *et al.* (1991). Samples prepared in this way were used to record UV-visible and ESR spectra. Analysis of the UV-visible spectra (Fig. 30) shows that, as for $FeLf_N$, the absorption band has shifted to a lower wavelength in $CuLf_N$, compared with Cu_2Lf (absorption maxima at 424 nm instead of 434 nm). The same protein sample used for the UV-visible analysis were concentrated to 10

mg/ml and used to record ESR spectra (Section A1.4.10). The two copper ESR spectra (Fig. 31) are similar, although not identical, and the calculated splitting constants (A_{II} and g_{II} and $g_\perp$) agree very well (data not shown). The values obtained are characteristic of Cu(II) in a rhombic environment providing further evidence that no major changes have occurred at the metal binding site of Lf_N. The more complex pattern of peaks seen at ~3250 Gauss in Cu_2Lf compared to $CuLf_N$ may be due to the fact that Cu_2Lf contains two metal binding sites, which may be slightly different, while $CuLf_N$ contains only a single metal binding site.

Fig. 30 Comparison of the copper UV-visible spectra obtained from Lf_N and hLf. Spectra were recorded between 250 nm and 700 nm for both Lf_N (solid line) and hLf (dashed line). The inset contains an enlargement of the region where the shift in the maximum has occurred.
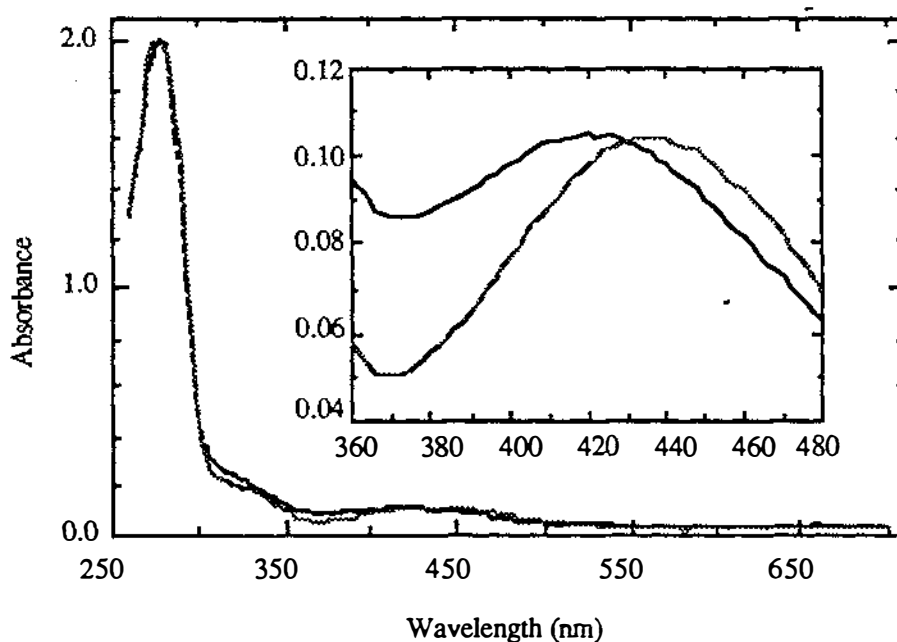
2.2.5 Metal binding and release

The pH dependence of iron release from Lf_N was measured and compared with the value obtained for native human lactoferrin. The extent of iron saturation, estimated from the absorbance at 466 nm (454 nm for native Lf_N), was monitored over the pH range 8.0 to 2.0 (Section A1.5.12). For Lf_N iron removal began at pH ~6.0 and was essentially complete at pH ~4.0. When native lactoferrin is treated in the same way iron removal does not begin until pH ~ 4.0 and is not complete until pH ~ 2.5 (Fig. 32).

Fig. 31 ESR spectra of copper saturated Lf$_N$ and hLf.

The ESR spectra were recorded as described in section A1.5.11.

CuLf$_N$

CuLf$_N$ ESR spectrum plotted against magnetic field (Gauss), x-axis range 1900 to 3900 with midpoint 2900.

Cu$_2$Lf

Cu$_2$Lf ESR spectrum plotted against magnetic field (Gauss), x-axis range 2400 to 3400 with midpoint 2900.
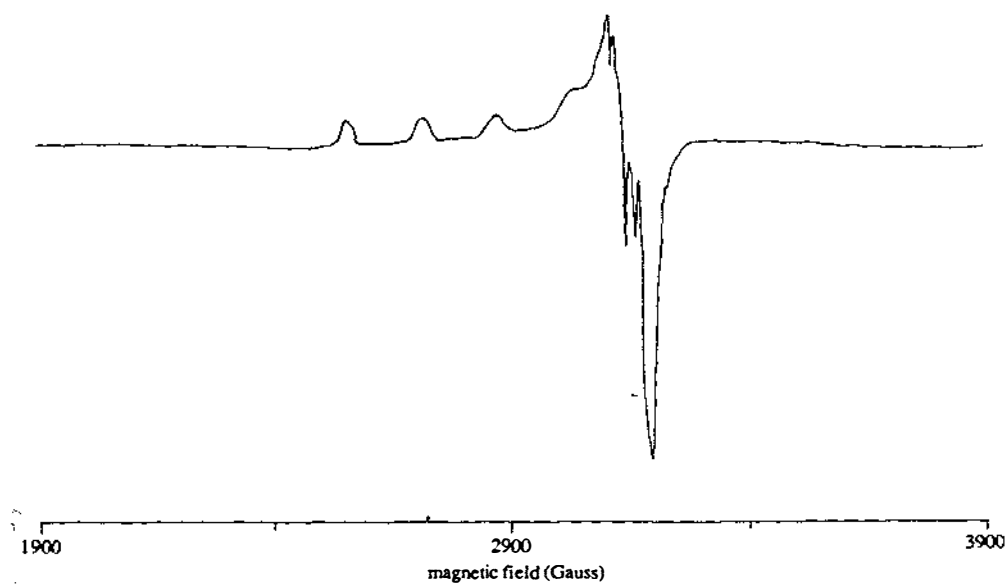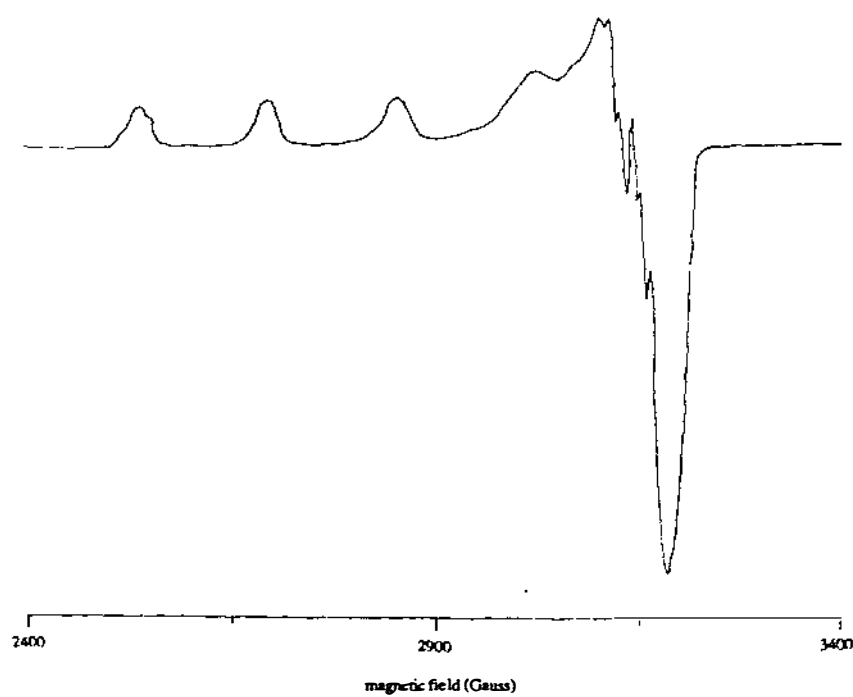
Fig. 32 Analysis of the pH dependent release of iron from Lf$_N$.

Samples of human lactoferrin (●) and Lf$_N$ (O) were dialysed for 48 hours against buffers of different pH (Section A.5.12). The spectra of the dialysed samples were recorded and the percentage of iron bound determined.
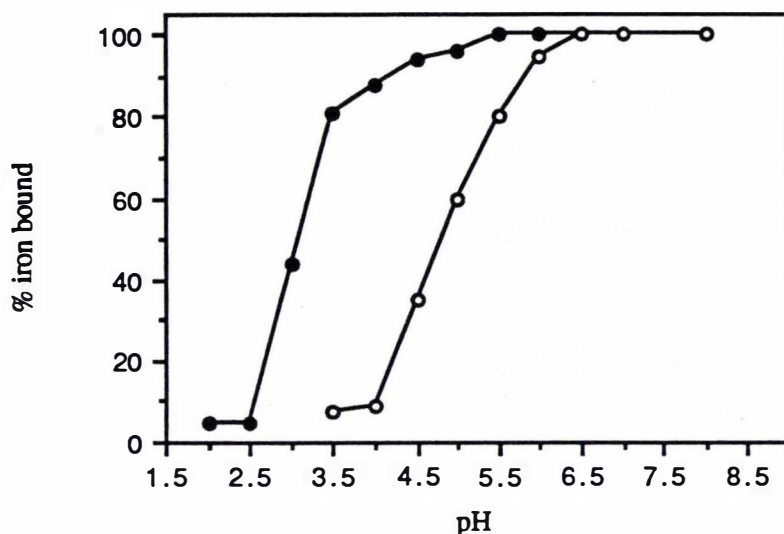
The reversibility of iron binding to Lf$_N$ was shown by binding studies carried out after complete removal of iron at pH 4·0. Titration of Lf$_N$ with iron (added as ferric-nitrilotriacetate complex) at pH 7·8 showed that the iron free Lf$_N$ binds one molar equivalent of iron when either fluoresence quenching or absorbance at 454 nm were measured (Fig. 33). The curve obtained for fluoresence quenching does not flatten out as much as for other lactoferrins. The reason for this difference is unknown although several tyrosines at the back of the molecule are exposed to solvent in the half molecule (see section B3.3) and may be free to interact with the iron in solution in a non-specific manner.

.3   DISCUSSION

Analysis of Lf$_N$ has shown that the signal peptide has been correctly processed during transfer of the recombinant protein across the endoplasmic reticulum. The protein has also been shown to be glycosylated although the precise nature of the carbohydrate groups added has not been determined. In some preparations of Lf$_N$ several bands of different mobility were obtained when analysed by SDS-PAGE. Treatment of this protein with glycosidases specific for N-linked groups converted all species to a single discrete band indicating that the difference in molecular weight was due to variation in the nature of N-linked carbohydrate groups. These results are similar to those obtained using mutant forms of human lysozyme containing artifical glycosylation sites and a BHK cell based expression system

Fig. 33 Titration of Lf$_N$ with iron at pH 7.8.

Iron was removed from Lf$_N$ by dialysis against pH 4.0 buffer (Section A1.5.12) and then iron was added back to the protein after it had been equilibrated in 0.01 M HEPES, pH 7.7, 0.2 M NaCl. In Fig. 33a iron binding was measured by monitoring the absorbance at 454 nm while in Fig. 33b the extent of fluorescence quenching was measured.

Fig. 33a
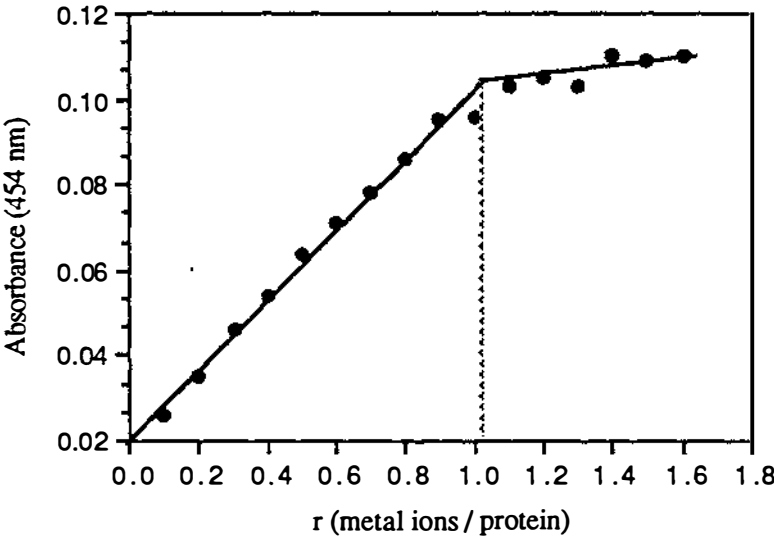
Fig. 33b
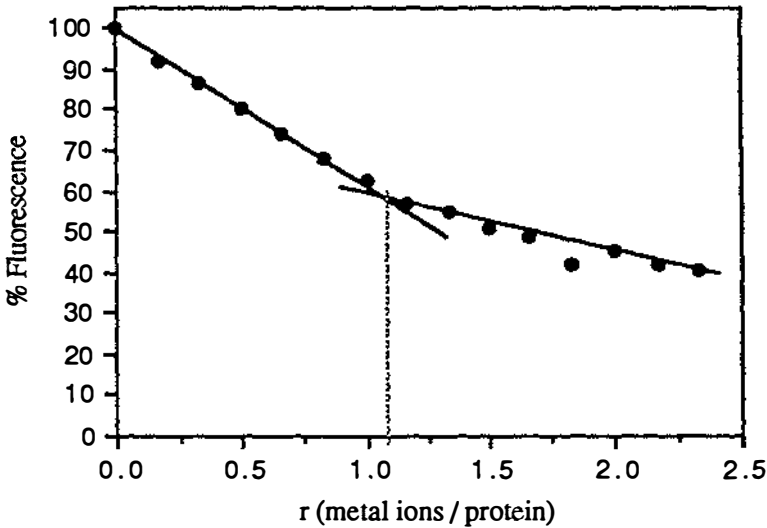
(Horst *et al.*, 1991). In this study they showed that the amount of carbohydrate added to the recombinant lysozyme produced from a single batch of cells varied and that the level of glycosylation was dependent on the position of the glycosylation site in the polypeptide chain. This apparent variation in the extent of protein glycosylation raises questions about the fidelity of the glycosylation machinery in BHK cells. This has not been pursued in this study.

Spectral analysis of Lf$_N$ has shown that the metal binding characteristics are very similar to those of the native protein. The similarity of the UV-visible and ESR spectra obtained from Lf$_N$ and hLf indicates that the ligands involved in metal

binding have essentially the same geometry in both proteins. The small reduction in λ_{max} to 454 nm implies a slight change in the ligand field around the iron atom. A similar shift has been reported to occur in Lf_{N30}, the proteolytically derived fragment of human lactoferrin, where a maxima at 440 nm was found (Legrand *et al.*, 1990). In the case of transferrin no such consistent affect is seen with the maxima of sTf_N, the recombinant N-terminal half of human transferrin, reported to be higher at ~473 nm (Woodworth *et al.*, 1991). No shift has been reported for sTf_{N35}, the proteolytically derived fragment of human transferrin (Lineback-Zins and Brew, 1980). Presumably there are very small changes in the precise nature of the metal site, but the peaks are broad and the significance of these results is not clear.

The most striking difference between the properties of Lf_N and native hLf is the lesser pH stability of the half molecule with respect to iron release. The iron release curve for Lf_N is similar in form to that of native lactoferrin but is displaced 2 pH units higher and is in fact very similar to that of human serum transferrin which is 50% desaturated at pH 5.2. A similar effect has been noted for the proteolytically derived N-terminal fragment of lactoferrin, Lf_{N30} (Legrand *et al.*, 1990), although the later is even less stable, with iron release beginning at pH ~6.6 and being complete at pH 6.0. If the pH stability is described in terms of pH_{50}, the pH at which 50% of the iron has been released, the values are Lf 3.0, Lf_N 4.8, Lf_{N30} 6.3, compared with 5.2 for sTf and 5.4 for sTf_{N35}, a proteolytically-derived transferrin half-molecule (Lineback-Zins & Brew, 1980). These results suggest that the interaction of the two lobes of lactoferrin may contribute to the greater stability of iron binding at low pH in lactoferrin compared to transferrin. The possible reasons for these differences will be discussed more fully later (Section 2.1).

Part A : Chapter Four

MUTAGENESIS STUDIES ON HUMAN LACTOFERRIN

A4.1   INTRODUCTION

Theoretically 20^{691} mutants of lactoferrin could be constructed! From this range
of possible targets, mutants directed at understanding : (a) the role of specific
residues in the metal binding site of lactoferrin and (b) the role of the N-linked
carbohydrate groups attached to lactoferrin, were designed. Other features which
were considered for investigation were the molecular basis for receptor binding
and the regions of lactoferrin involved in the bacteriostatic and bactericidal
properties of human lactoferrin. Although both of these aspects are likely to be of
fundamental importance to the function of lactoferrin, at the time of this work
suitable assays were not readily available.

The metal and anion binding sites of lactoferrin have been described earlier
(Section 1.4). Two residues with distinctly different roles in the iron binding site,
Asp 60 and Arg 121, were selected for initial analysis by mutagenesis. The
structural and functional roles of these residues are described below, together with
the details of lactoferrin glycosylation.

Aspartic acid 60

Of the four protein ligands bound to iron in lactoferrin, the aspartate (Asp 60 in the
N lobe and Asp 395 in the C-lobe) was predicted to play the single most important
role. This residue has three obvious functions in iron binding:
(i) As an anionic oxygen ligand it is highly suitable for binding to Fe^{3+} ions, and
the negative charge of the sidechain helps to balance the positive charge of the
cation,
(ii) In the closed diferric structure the carboxylate group lies between the N-termini
of two helices, helix 3 from the N1 domain and helix 5 from the N2 domain (Fig.
34). The charge on the aspartate will assist in neutralising the two partially
positively charged helix N-termini.
(iii) It forms, through its non-coordinated carboxylate group, a strong H-bond
(Asp 60 OD1 to Arg 122 N) between the two domains which probably helps to
'lock' the two domains together after they close over the bound metal.

Legrand *et al.* (1990) have questioned the importance of aspartic acid 60 in iron
binding following the observation that iron could bind to the N2 domain

Fig. 34 Environment of aspartic acid 60 in Fe₂Lf.

The two hydrogen bonds from the carboxylate group of aspartic acid 60 to the end of helix 3 (62 N) and helix 5 (122 N) are shown. The interactions between the metal ion and the ligands are also indicated.
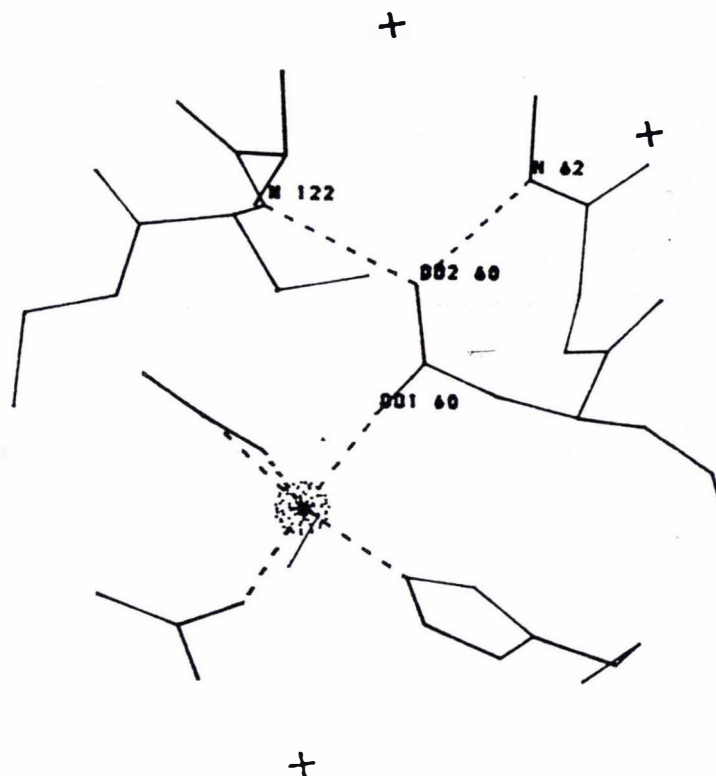
proteolytic fragment of lactoferrin despite the fact that it lacked this residue. However, no detailed analysis of the iron binding properties of this fragment were reported. The importance of interactions involving this residue are further suggested by the fact that in the C-lobe of melanotransferrin the equivalent residue is changed to a serine. This is likely to be one of the factors which contribute to the lack of iron binding to this site (Baker *et al.*, 1992).

Arginine 121

On the basis of chemical modification studies early workers (Rogers *et al.*, 1978) had suggested that at least one arginine was directly involved in iron binding. More recent structural studies (Anderson *et al*, 1989) have shown that the critical arginine is likely to be 121 in the N-lobe (465 in the C-lobe) as this residue interacts directly with the anion through two hydrogen bonds (Fig. 35) and provides a positive charge which helps neutralise the CO_3^{2-} ion. The importance of Arg 121 is further suggested by the observation that, apart from the C-lobe of melanotransferrin, it is invariant in the sequences of all transferrins examined so

far. In melanotransferrin the residue corresponding to arginine 121 is changed to serine. This change and the change from aspartate to serine are the only two mutations, to residues which interact directly with the metal or the anion, in the non-functional C-lobe metal binding site. It is likely that iron binding is prevented by the combination of these two mutations and also by changes to several other residues which make important hydrogen bonds in the binding site. However, the exact role of these residues can only be determined if they are individually changed and the effect of the change is studied in isolation and in combination with other mutations.

Fig. 35 Environment of arginine 121 in Fe₂Lf.

The two hydrogen bonds from arginine 121 to the carbonate group are shown. The position of the other metal binding ligands and the interactions they make with the iron atom are also shown.
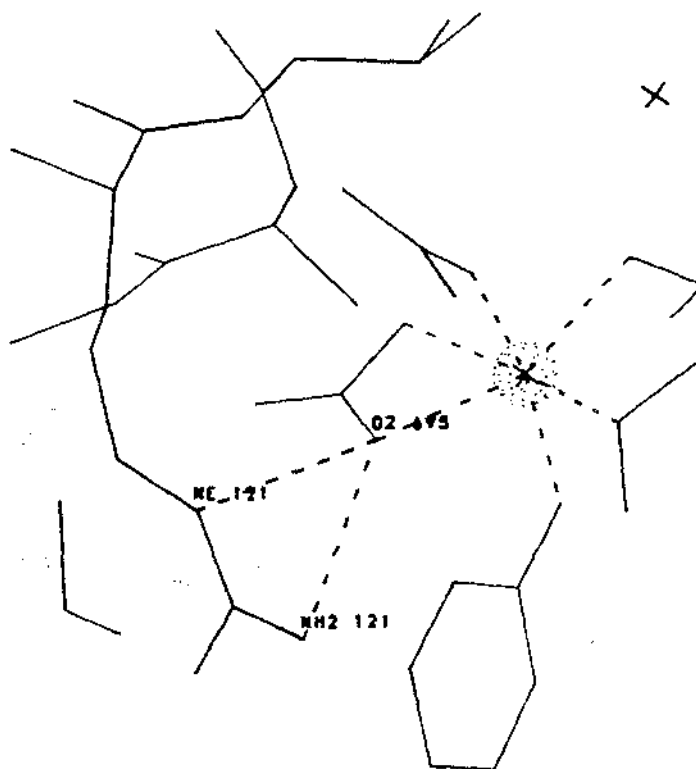
It would also be interesting to replace arginine 121 by a variety of other residues to allow an investigation of the role played by this residue in iron release. It has been proposed that one of the key steps in iron release is the protonation of CO_3^{2-} leading to the repulsion of arginine 121 and consequent destabilisation of the metal binding site (Williams and Woodworth, 1973; Sarra et al., 1990). Replacement of Arg 121 by Ser would remove this positive charge. This would be expected to weaken anion binding and indirectly reduce the strength of iron binding. More significantly it would leave no possibility of charge repulsion between this residue

and the protonated anion, and would therefore test this hypothesis of iron release. A change from Arg to Asp or Glu would introduce further possibilities. The negative charge is likely to prevent anion binding, but the carboxylate group could possibly substitute for the anion and bind directly to the metal ion. If this were the case there would be no possibility of protonation of the anion, and in fact no need for anion binding at all since the positive charge of Fe^{3+} could be partially neutralised directly by the carboxylate group.

Glycosylation of lactoferrin

Human lactoferrin contains three potential recognition sites (Asn -X-Ser/Thr) for the addition of N-linked carbohydrate groups. Of the three sites at Asn 137, Asn 478 and Asn 623 only the first two have been reported to be glycosylated in human tissues (Metz-Boutigue *et al.*, 1984). The N-linked carbohydrate groups have been implicated in the indirect stabilisation of iron binding (Legrand *et al.*, 1990) and in receptor interactions. However, conclusions to the contrary have been drawn by other investigators (H.M. Baker *pers. comm.*; Padda and Schryvers, 1990). Synthesis of non-glycosylated lactoferrin would allow the role of the carbohydrate groups to be examined under conditions where there was no possibility of other changes to the protein, such as can occur during enzymatic deglycosylation.

A further advantage of producing deglycosylated protein is that it may crystallise better than the native glycoprotein. This has already been shown in the case of intact human apolactoferrin (Norris *et al.*, 1989). In this instance, the carbohydrate groups were removed using the enzymes Endoglycosidase F (Endo F) and Peptide *N*:glycosidase *F* (PNGase) both isolated from *Flavobacterium meningosepticum*. The production of non-glycosylated protein in cells would simplify purification as it would remove the necessity for enzymatic deglycosylation. This involves several steps which always results in the loss of some protein and has the potential to affect the protein in unknown ways.

Mutagenesis protocols

A wide variety of techniques are available for introducing site-specific mutations into cloned cDNA. Most of the earlier methods are reviewed by Leatherbarrow and Fersht (1980). Although variations of many of these methods are still used they have largely been superseded by the method described by Kunkel *et al.* (1987) which uses a *dut⁻, ung⁻* strain of bacteria to produce the template. The advantage of this approach is that *dut⁻ ung⁻* strains produce DNA in which a high

proportion of the thymine residues have been replaced by uracil. The lack of dUTPase (dut^-) activity results in high intracellular levels of dUTP which increases the frequency of uracil incorporation into the DNA. Once uracil has been incorporated into the DNA in these cells it is not efficiently removed because uracil N-glycosylase (ung^-) is also inactivated. This enzyme is responsible for initiating the DNA repair mechanism which allows replacement of uracil by thymine in wild type cells. Single strand DNA which has been prepared from phage or phagemids grown in dut^- ung^- bacterial hosts can be used as the template for mutagenesis by elongation from the annealed mutagenic oligonucleotide. Transformation of the reaction products, following elongation and ligation, into dut^+ ung^+ bacteria will result in degradation of the uracil-containing template strand, ensuring that only the mutated strand survives.

A4.2   RESULTS

A4.2.1   Production of a mutant

Cloning into pTZ and preparation of the template phage

Although pGEM:Lf$_N$2 (Section A2.2.4) would have been suitable for the preparation of single strand template phage, it was decided to clone the cDNA encoding Lf$_N$ into pTZ18U, the vector supplied with the mutagenesis kit available. The main feature of pTZ18U is that, like pGEM-3Zf(+), it contains the f1 origin of replication, allowing production of phage-like particles containing ssDNA when cells have been superinfected with helper phage. The cDNA encoding Lf$_N$ was excised from pGEM:Lf$_N$ using EcoR1 and BamH1 and ligated into pTZ18U which had been cut with the same enzymes. After transformation (Section A1.2.6) into XL-1 cells a single colony was chosen, analysed by restriction enzyme digestion and then used to transform competent CJ236 cells which were selected on plates containing ampicillin. This clone was called pTZ:Lf$_N$ and a partial restriction map is shown in Fig. 36. Once a clone had been obtained in CJ236, uracil-containing phagemids were prepared as described in section A1.3.1.

Titration and extraction of the phage

Before the ssDNA was extracted from the phage they were titred in an attempt to estimate the level of uracil incorporation. Two bacterial strains, one dut^+, ung^+ and the other dut^-, ung^-, were used to titre the phagemids. Approximately 10^5-fold fewer colonies were obtained from MV1190 cells (dut^+, ung^+) than from the CJ236 cells (dut^-, ung^-). This difference was greater than the value suggested by

Fig. 36 Partial restriction map of pTZ:Lf$_N$.
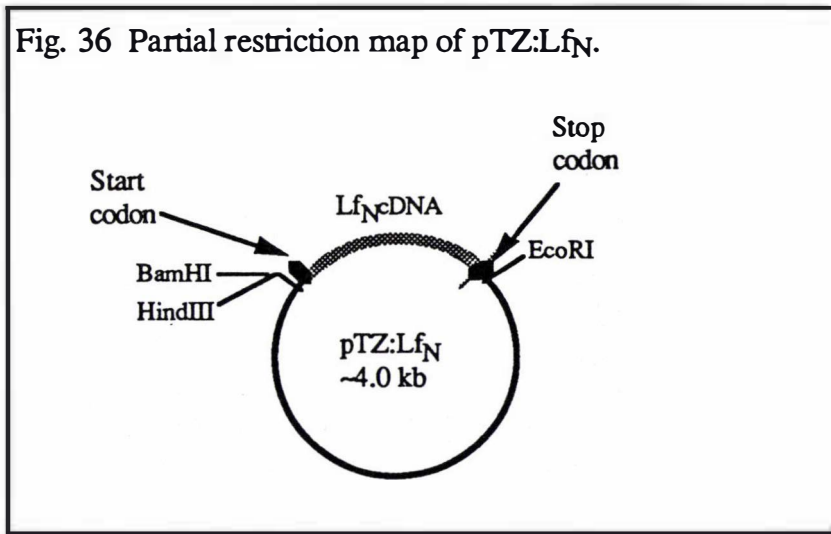
the suppliers indicating that almost all of the phage contained uracil. If a large number of colonies are obtained in the *dut+*, *ung+* strain the template is not suitable for use in mutagenesis experiments as many of the parental strands will survive when transformed into the *dut+*, *ung+* strain, decreasing the efficiency of mutagenesis. The results obtained here suggested that ssDNA extracted (Section A1.3.2) from these phage would serve as a good template for mutagenesis because almost all of the parental strands would be degraded in a *dut+*, *ung+* host.

Synthesis of the complementary strand

Oligonucleotides were obtained from commercial suppliers and if they had not already been purified they were purified as described in section A1.2.21. The oligonucleotides were then phosphorylated (Section A1.3.3) prior to addition to the purified template at a ratio of between 20:1 and 10:1. The reaction mixture was then slowly cooled from 70°C to room temperature (~1 hour). During this time the oligonucleotides annealed to the template. Once the primer had annealed the second strand was formed using the enzyme T4 DNA polymerase and the complete strand was ligated with T4 DNA ligase. T4 DNA polymerase was chosen because it has no strand displacement activity preventing it from displacing the annealed oligonucleotide and cycling around on the same template strand (Masamume & Richardson, 1971). Other enzymes (Klenow and T7 DNA polymerase) have higher processivity activities but they have the ability to displace the oligonucleotide resulting in a lower proportion of mutagenic clones (Biorad mutagenesis manual). Details of the annealing, extension and termination protocols are given in section A1.3.4. The sequences and position of all the mutagenic oligonucleotides used in this study are given in Appendix 4.

Analysis of the reaction products

For every oligonucleotide used two reactions were carried out in parallel. One reaction contained oligonucleotide and the other was performed without oligonucleotide, acting as a control. These reactions allowed the extent of conversion from ss template to covalently closed circular DNA (cccDNA) to be monitored. The controls recommended and supplied by Biorad were also carried out in the initial mutagenesis experiment. Once extension was complete and the reaction had been terminated one tenth of each mutagenic reaction was analysed by gel electrophoresis; a typical gel is shown in Fig. 37. The aim of the mutagenesis reaction is to produce cccDNA (RF-IV DNA) which becomes positively supercoiled in the presence of ethidium bromide causing it to migrate at a similar speed to negatively supercoiled RF-1 DNA (lower band lane 8). In lanes 5, 6 and 7 it is clear that much of the template DNA has been converted to cccDNA although very little conversion has occurred in the control reaction (lane 4). This indicates that the template was not contaminated by any small fragments of DNA capable of priming strand extension. Lanes 2 and 3 contain controls provided by the suppliers of the kit. Although the original double stranded template has not been run in this instance it is again clear that most of the template (lane 3) has been converted to a higher molecular weight product suggesting that the reaction has succeeded.

Following analysis on an agarose gel another portion of each reaction was used to transform (Section A1.2.6) competent *E. coli* cells (strain XL-1). Successful mutagenesis of the control template provided by Biorad removed a reading frame error in the β-galactosidase gene, restoring enzyme activity. This allowed the efficiency of mutagenesis to be monitored by comparing the ratio of blue colonies to white colonies when the cells were grown in the presence of X-gal and IPTG. In the example shown here 100x as many colonies were obtained from the reaction shown in lane 3 as were obtained from the reaction shown in lane 2. All the colonies from the reaction shown in lane 3 were blue, indicating that β-galactosidase activity had been restored. The reactions shown in lanes 5, 6 and 7 also appeared to have worked relatively well because 10x as many colonies were obtained from these reactions as from the reaction shown in lane 4. It should be noted that there was no colour selection in this case therefore the number of mutated colonies could not be accurately quantitated.

Selection and complete sequence analysis of the mutants

Following mutagenesis, carried out as described above, four colonies were

Fig. 37 Analysis of the reaction products from a mutagenesis experiment.
Mutagenesis experiments were carried out as described in section A1.3. 10 µl of each reaction
was analysed on a 1% agarose gel. Lane (1) BRL size markers; (2) Reaction containing the
control template with no primer added.; (3) Reaction containing control template and control
primer; (4) Reaction containing pTZ:Lf$_N$ U-DNA template without primer; (5, 6, 7) Three
different reactions containing pTZ:Lf$_N$ U-DNA template and mutagenic primers; (8) pTZ:Lf$_N$
plasmid DNA.
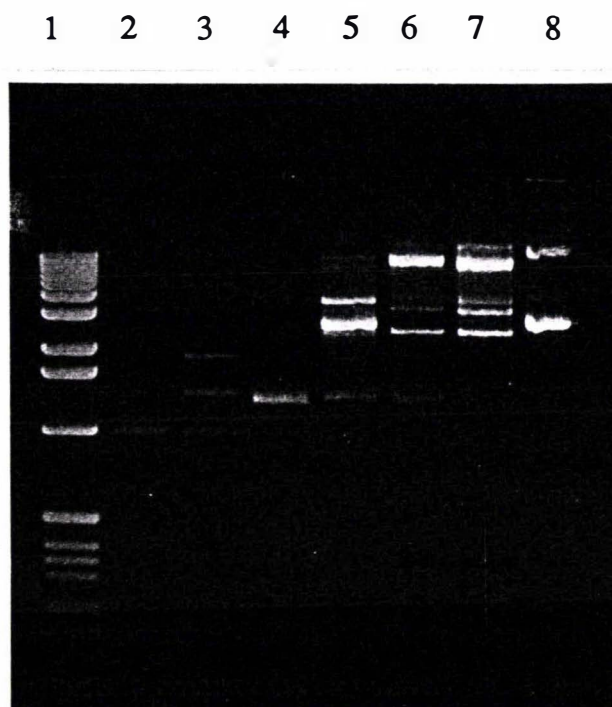
selected from each plate and single stranded DNA was prepared from each of these
(Section A1.2.10). Using one of the internal sequencing primers (Appendix 2) the
region around the proposed site of change was sequenced (Section A1.2.16),
allowing any changes to be detected. Typically more than 90% of the clones
sequenced contained the desired mutation. The high rate of mutagenesis achieved
meant that for most of the later mutants it was only necessary to sequence two
clones before one with the desired change was identified. Once a clone for each
mutant had been selected at least one complete strand was sequenced, and in most
cases both strands, ensuring that no spurious mutations had been introduced. The
sequencing strategy used is summarised in Fig. 38.

As this generally involved 2.2 kb of sequence data per mutant a considerable
amount of time was involved in reading and processing the data. In an attempt to
decrease the amount of time required to process the data many of the mutants were

Fig. 38 Summary of the sequencing strategy used to sequence the mutants.

The position of the internal sequencing primers and the direction of sequencing is summarised below.
The sequence and exact position of the oligonucleotides used is given in Appendix 2.
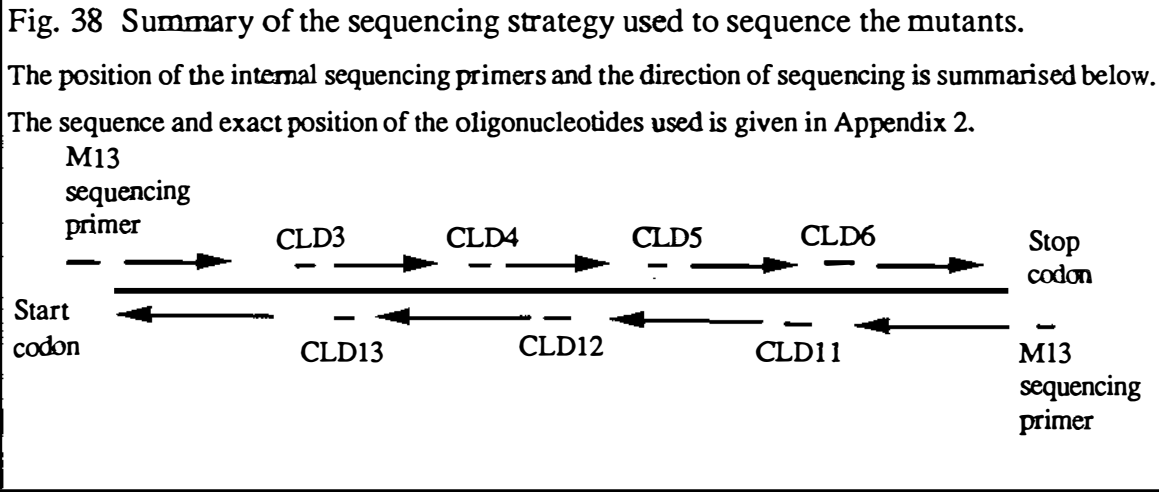
sequenced by running them in tracks adjacent to the template molecule which had been carefully sequenced. Typically four or five A tracks (or G or C or T) were run in adjacent lanes (Fig. 39). The region of sequence shown here surrounds a region where several of the introduced mutations were located. Careful analysis reveals that the loss of a band from one track is associated with the appearance of a band in another track. Sequencing of the clones in this way allowed a lot of sequence data to be processed quickly.

;. 39 Sequence analysis of mutated clones.

e clones were sequenced using standard protocols (Section A1.2.16) and then the samples were analysed in allel. Track one of each type contains pTZ:Lf$_N$ and the other four tracks in each set contain mutants.
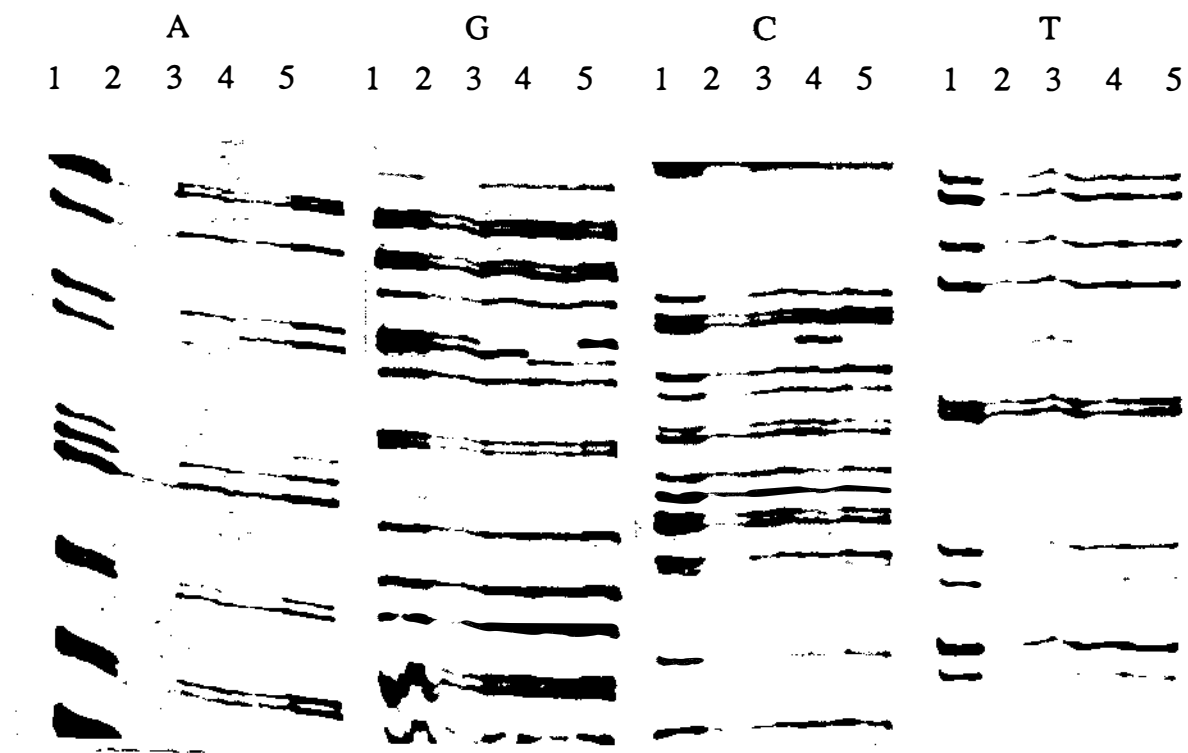
4.2.2   **Cloning and expression of the mutants**

Once the sequence of all the mutants had been verified all of the mutant cDNA's were cloned into pNUT according to the steps outlined previously (Section A2.2.2). Before the pNUT plasmids were used to transfect BHK cells the site of mutation was resequenced by double strand sequencing (Section A1.2.16) allowing the identity of each clone to be confirmed. Transformed BHK cells were selected with methotrexate as before (Section A1.4.2) and lactoferrin expression was confirmed by carrying out immunoprecipitation reactions prior to growth of the cells on a large scale (Section A1.4.3). In all cases the protein was purified from the media (Section A1.5.7) and analysed by SDS-PAGE (Section A1.5.1) before it was used in subsequent experiments.

In order to clarify the discussion in the following sections an example of the nomenclature used is given here. For the mutation of asparagine 137 into alanine the pNUT DNA containing the change is called pNUT:$_{N137A}$ and the mutant protein is called N137A. All of the other mutants are named in a similar way and the sequences of the primers used to create these mutants are given in Appendix 4.

4.2.3   **Glycosylation site mutants**

Asparagine 137

The first mutant made involved the conversion of asparagine 137 to alanine, removing the only site where N-linked carbohydrate groups are reported to be attached in the amino terminal half of human lactoferrin. The protein obtained from cells transfected with pNUT:$_{N137A}$ was purified as in section A1.5.7, except that 0.45 M NaCl was used to elute the protein from the resin. The molecular weight of this protein matched that of protein prepared by enzymatic deglycosylation of Lf$_N$ when analysed by SDS-PAGE (data not shown).

It had been the intention to prepare all other mutants of Lf$_N$ in this non-glycosylated form, but cells transfected with pNUT:$_{N137A}$ were found to produce a significantly lower level of protein than cells producing Lf$_N$. The expression studies were repeated several times but levels of no greater than 5 mg/l were obtained compared with up to 30 mg/l for Lf$_N$. Such a decrease in the level of expression was somewhat surprising although there have been several reports of studies in which removal of glycosylation sites has either affected the level of expression obtained or affected the functional properties of the protein (Grigera *et al.*, 1991; Matzuk and Boime, 1988; Dube *et al.*, 1988). In an attempt to

determine if the low level of expression was related to the site of cDNA incorporation in the BHK cell genome a second batch of BHK cells were transfected with pNUT:$_{N137A}$. Once again a low level of expression was obtained, suggesting that the level of expression was influenced by some other factor. One possible explanation is that the protein was synthesised by the cells but it was not secreted properly. To investigate this possibility cell lysates were prepared (Section A1.5.5) and immunoprecipitation reactions were carried out (Section A1.5.4). The lack of a sensitive ELISA or RIA test made analysis of expression levels quite difficult, but there did not appear to be significant amounts of N137A trapped in the cells.

Several further possibilities that could account for the difference in expression levels include mRNA instability or increased protein turnover rates. Another alternative is that conversion of asparagine 137 to an alanine may have affected the folding of the protein. This is suggested by the UV-visible spectrum of the protein which is clearly different to that obtained for Lf$_N$ (Fig. 40), even though it was difficult to measure accurately.

Fig. 40 Comparison of the spectrum of N137A and Lf$_N$ (400 nm to 500 nm). Only the region between 400 nm and 500 nm is shown as the remainder of the spectrum was essentially identical. Lf$_N$ is shown as a solid line and N137A as a broken line. Both protein samples had a concentration of 1.8 mg/ml.
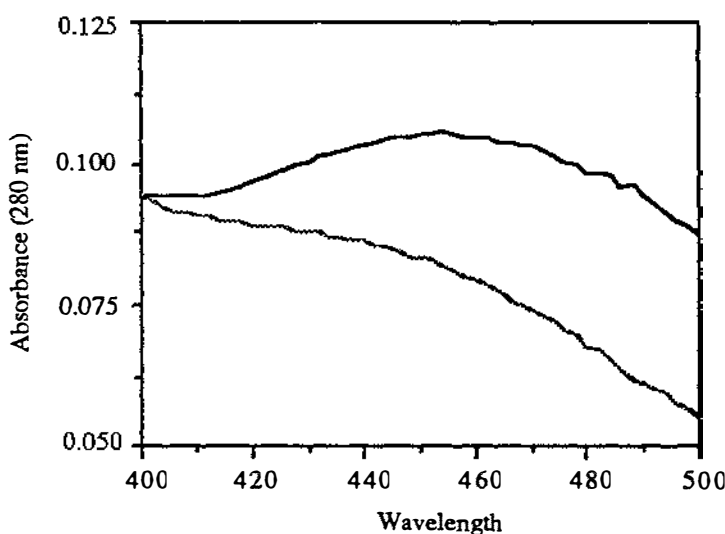
Asparagine 137 is just past the C-terminus of an α-helix and alanine is a helix stabilising amino acid (O'Neil and DeGrado, 1990) therefore it is possible that the change from asparagine to alanine may have caused helix 5 to be extended. Arginine 121 is at the other end of helix 5 and elongation of this helix may have slightly altered the orientation of Arg 121 altering the environment of the iron atom. In retrospect it may have been more appropriate to change asparagine 137 to

serine or glutamine. In bovine lactoferrin the residue corresponding to Asn 137 is actually serine. The recognition site for the addition of N-linked carbohydrate groups at position 137 could have also been abolished by changing the threonine at position 139 to any amino acid except serine. Although it would have been interesting to pursue these possibilities the aim of producing N137A was to simplify the project, not extend it in other directions. Because of the problems encountered here it was decided to abandon expression of the other mutants in a carbohydrate free protein background and to continue using enzymatic removal of the carbohydrate where necessary.

Asparagine 137 / asparagine 478 double mutant

As a prelude to the preparation of mutants of full length lactoferrin both of the sites reported to be glycosylated in lactoferrin were removed from the intact human lactoferrin cDNA. This mutant was prepared in the same way as the N-terminal mutants described in section A4.2.1 except that the template used was derived from pGEM:hLf4 (Appendix 5) and the two changes were introduced in a sequential manner. Firstly asparagine 137 was converted to alanine, and the sequence determined around the site of the mutation. New uracil containing single strand DNA template was then prepared from this mutant and the second change was introduced, converting asparagine 478 into an alanine. Before the clone was sequenced it was used to transfect BHK cells and the expression of lactoferrin was then detected by immunoprecipitation (Section A1.5.4). To our surprise BHK cells containing the double glycosylation mutant cDNA were found to produce lactoferrin of two different mobilities (Fig. 41, lane 2). Neither of the protein bands had a mobility corresponding to that of the fully glycosylated molecule. The lower band had a mobility similar to deglycosylated lactoferrin while the upper band was intermediate in mobility between fully glycosylated and unglycosylated lactoferrin. Both bands were demonstrated to be lactoferrin by protein blotting and immunodetection (Section A1.5.6) (Fig. 41, lane 6). It appeared that the upper band in this mixture may represent lactoferrin which has one attached carbohydrate group. Treatment of the mixture with PNGase converted the doublet into a single band equal in size to that of the lower doublet band (data not shown), confirming that the band of intermediate size was indeed due to the presence of N-linked carbohydrate. Treatment of the doublet with enzymes specific for O-linked carbohydrate groups did not result in any change in molecular weight.

The appearance of the doublet was completely unexpected and the obvious conclusion was that somehow the cells had been contaminated. This possibility was discounted by transfecting another batch of BHK cells with

Fig. 41 Analysis of the N137A/N478A mutant by PAGE and immunodetection. Protein samples were purified and analysed SDS-PAGE. Deglycosylated samples were prepared using standard procedures (Section A1.5.10). Lane (1) Molecular mass markers (kilodaltons); (2) N137A/N478A protein as purified from a column; (3) N137A/N478A which has been purified and then deglycosylated; (4) Pure hLf; (5) Deglycosylated hLf; (6) N137A/N478A as in lane 2 except that it has been transferred to nitrocellulose and reacted with antibodies to human lactoferrin (Section A1.5.6).

97 kDa —
66 kDa —
45 kDa —
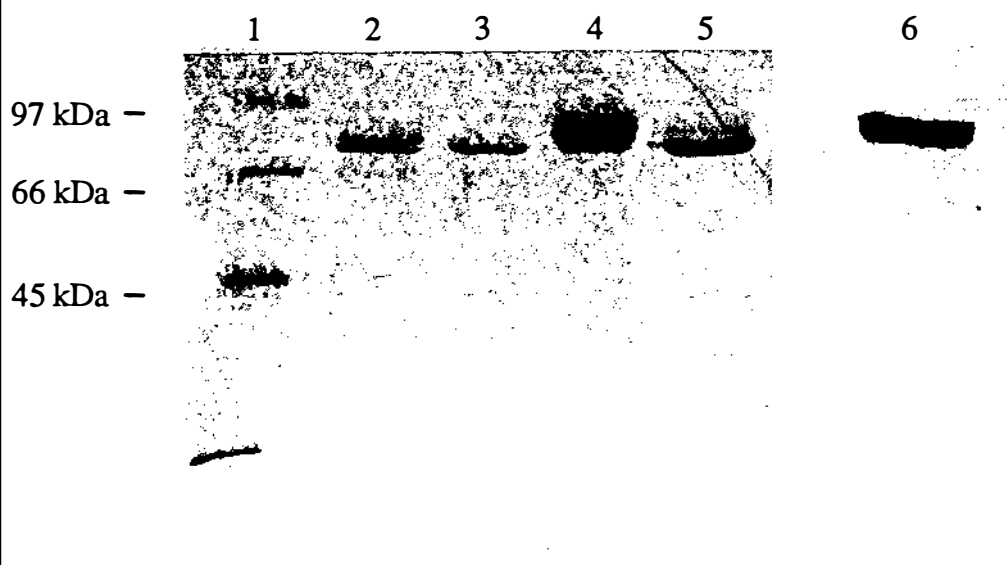
pNUT:N137A/N478A and monitoring protein expression. Once again both bands were present in what appeared to be equal levels. Contamination of the cells also seemed unlikely on the grounds that neither band obtained was equal in size to any lactoferrin species previously produced in BHK cells. A more likely possibility was contamination of the pNUT:N137A/N478A DNA at some stage. In an attempt to test this possibility the sites where the mutations had been introduced were resequenced in the pNUT clone by double strand sequencing (Section A1.2.16). At both positions it was clear that the desired changes had been introduced and no shadow bands were apparent which would have been indicative of a mixed clone. There still remains a slight possibility that contamination may have occurred or that a mutation has been introduced elsewhere in the protein. However, the fact that the two bands of protein occurred in equal amounts and that they were consistently obtained from different batches of cells suggested that there may be some other reason for the changes.

In lactoferrin there are two sites which are reported to be glycosylated (Spik et al., 1988) and these are the two which have been changed here. There does, however, exist another potential site for the addition of N-linked carbohydrates, at position 623. This site is located in a loop on the surface of the protein, a site which would be accessible to glycosylation and which is in fact glycosylated in the corresponding position in human serum transferrin (Metz-Boutigue et al., 1984).

In native lactoferrin this site is apparently not glycosylated, but it is possible that in BHK cells it is recognised in some instances and glycosylated efficiently. This observation is consistent with the results of Horst *et al.* (1991) who claim that the extent of glycosylation decreases with increasing distance of the signal sequences from the N-terminus. If this is indeed correct a likely scenario is that in the native protein positions 137 and 478 are glycosylated and satisfy the requirement for protein glycosylation. However, in the mutant protein missing both of these sites the less favourable site at position 623 is utilised. Again this work has not been pursued to completion but is interesting and warrants further investigation

2.4 Binding site mutants

Aspartic acid 60

As the first step towards understanding the role of the aspartate ligand in the metal binding site, Asp 60 was changed to serine, mimicking one of the changes that occurs in the C-lobe of melanotransferrin. Protein was purified and characterised as described in section A3.2.1. In the initial purification steps it was clear that D60S was different to native lactoferrin because it was yellow in colour rather than the characteristic pink colour typical of iron saturated lactoferrin. This difference was clearly shown by UV-visible absorption spectroscopy which revealed that the λ_{max} for the visible charge transfer band had decreased to ~ 434 nm (Fig. 42a), a shift of 20 nm relative to native Lf$_N$. Note that the actual position of the maximum was difficult to determine accurately because the spectrum did not have a well defined peak in this region. A difference was also obvious in the ESR spectrum where the doublet at 1500 Gauss, which is typical of all members of the transferrin family, was absent. Instead a single peak in the same position was obtained (Fig. 42b).

The pH dependence of iron release (Section A1.5.12) was also shown to be different but it was difficult to measure accurately because the bound iron appeared to be lost gradually from D60S even at high pH. Although a complete analysis has not been carried out these results clearly demonstrate that iron is bound by D60S but that the properties of iron binding and release are quite different. In addition to this mutant a second mutation was also prepared in which aspartic acid 60 was converted to threonine forming D60T. This mutant appeared to have iron binding properties which were identical to those obtained with the serine mutant. Although a complete structural analysis is required it is likely that the role of Asp 60, as a ligand, has been replaced by a water molecule as is reported to occur in the

Fig. 42 Comparison of the spectral properties of D60S and Lf$_N$.

Iron saturated samples of Lf$_N$ and D60S at 1.8 mg/ml were prepared and used to obtain
absorption spectra (Fig. 42a). The inset shows the region of difference. In Fig. 42b the ESR
spectrum for D60T, recorded as described in section A1.5.11, is shown. Compare this spectrum
with those shown in Fig. 29.

Fig. 42a
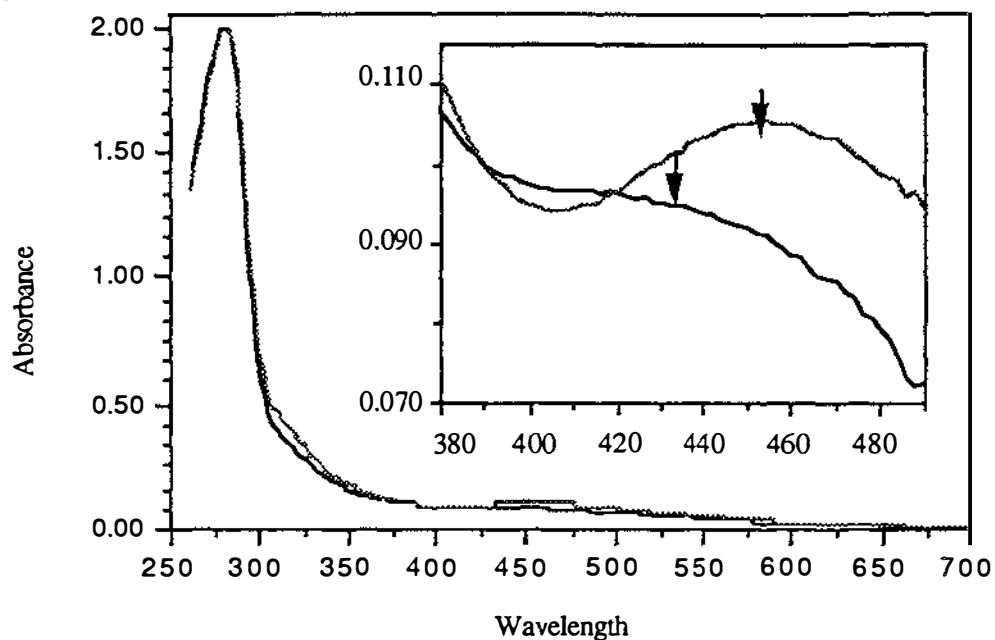
Fig. 42b
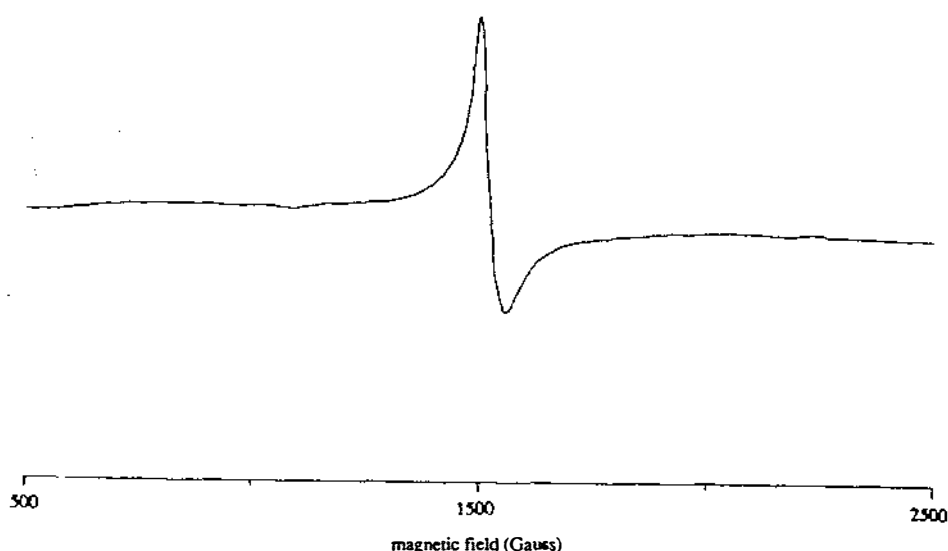
iron binding 18 kDa fragment of ovotransferrin which is missing the equivalent of
Asp 60 (Baker and Lindley, 1992). Determination of the tertiary structure of one
of these mutants should allow further interpretation of the changes in iron binding.

Arginine 121

The role of arginine 121 in iron binding and release was investigated by construction of several mutants. In the first instance arginine 121 was replaced by serine, forming R121S. Pure R121S was characterised as for D60S. Several subtle but potentially interesting differences in both the λ_{max} and iron release profiles were noted. Firstly the position of λ_{max} was found to depend on the pH of the solution. At pH 7.0 λ_{max} was at 454 nm, identical to that obtained for native Lf$_N$, but dialysis of R121S against lower pH buffers shifted the λ_{max} to between 470 and 480 nm (Fig. 43). The increase in λ_{max} is clearly visible in Fig. 43 even though some iron has been lost at pH 5.5, decreasing the intensity of the charge transfer band. The same shift was consistently obtained with all samples dialysed against buffers lower than pH 6.0 suggesting that a group at or near the iron site had been protonated at about pH 6.0. A likely explanation for the phenomenon is that in the wild type protein the presence of the positive charge associated with arginine 121 maintains the anion as carbonate. However, in R121S the absence of the positive charge means that the anion is more easily protonated forming HCO_3^- which has different coordination properties.

Fig. 43 Change in R121S λ_{max} with decreasing pH.

A sample of R121S was dialysed at pH 8.0, the UV-visible spectrum was recorded (dashed lines) and then the sample was dialysed for 48 hours against a buffer of pH 5.5 (Section A1.5.12) before the spectrum was recorded again (solid lines). The inset shows the region of the spectrum where the shift in λ_{max} occurs.
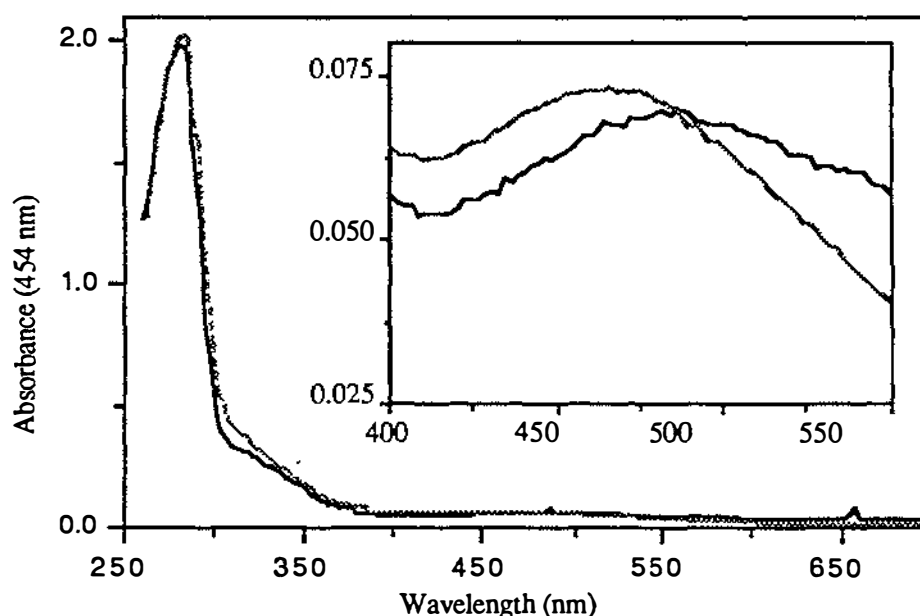
The second change of interest is the shape of the pH mediated iron release profile (Fig. 44). Although the values obtained are subject to considerable error, over the range where the iron is lost, most of the points have been recorded in triplicate for both samples. These results clearly show that replacement of arginine 121 by serine still allows iron to be bound by the protein, but they suggest that the mechanism of iron release has somehow been altered because iron is lost from R121S over a more extended range (7.0 to 3.0) than from Lf_N (6.0 to 4.0). In this study the shape of the R121S curve is only slightly different from that of Lf_N however, it is quite possible that in native lactoferrin a greater differences would be seen. Although at present it is difficult to propose a model which can account for the difference in the iron release profiles, it is likely to be associated with the changes in the iron site which give rise to the changes in the absorption spectrum over this pH range. The determination of the tertiary structure should help clarify this issue.

Fig. 44 Comparison of the iron binding properties of Lf_N and R121S.

Samples of Lf_N (●) and R121S (○) were dialysed for 48 hours against buffers of different pH (Section A1.5.12). The spectra of the dialysed samples were recorded and the percentage of iron bound determined. The point of maximum absorbance between 454 nm and 474 nm was used to calculate the percent of iron bound.
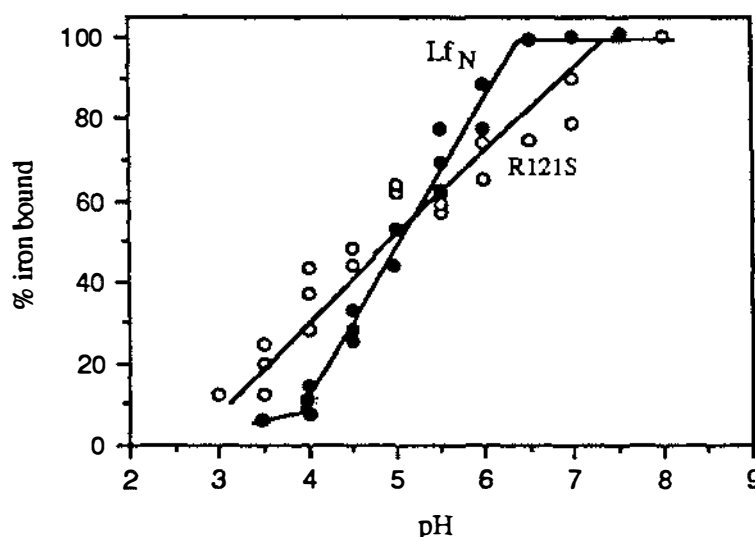
In the second series of mutations at position 121 the arginine was changed to the two acidic amino acids, aspartic acid and glutamic acid. Both of these mutants had a similar effect on iron binding, increasing λ_{max} to 472 nm, changing the shape of the ESR spectra (Fig. 45) and increasing the pH at which iron was released to between ~7.0 and ~5.5 (Data not shown). The pH at which iron was released was difficult to measure and the value obtained is not very accurate. The main point to note is that iron was bound, but not very tightly, and the ESR spectrum was quite different suggesting that the environment of the iron atom was altered. It is

difficult to interpret the effects of this mutation in terms of structural changes at the iron site. If the new carboxylate group associates with the N-terminus of helix 5 as suggested by modelling it is difficult to imagine that carbonate would still be found in the metal binding site as binding would be sterically restricted and it would be electrostatically unfavourable. Perhaps aspartic acid or glutamic acid can act as one of the ligands and the other position is filled by a water molecule? Alternatively the mutated residue at position 121 may point away from the metal site, allowing the anion to remain bound but destabilising the closed structure in some other way. A structure of at least one of these mutants is eagerly awaited.

Fig. 45 ESR spectra of iron saturated R121D.

The ESR spectra was recorded as described in section A1.5.11 and should be compared with those shown in Fig. 29.
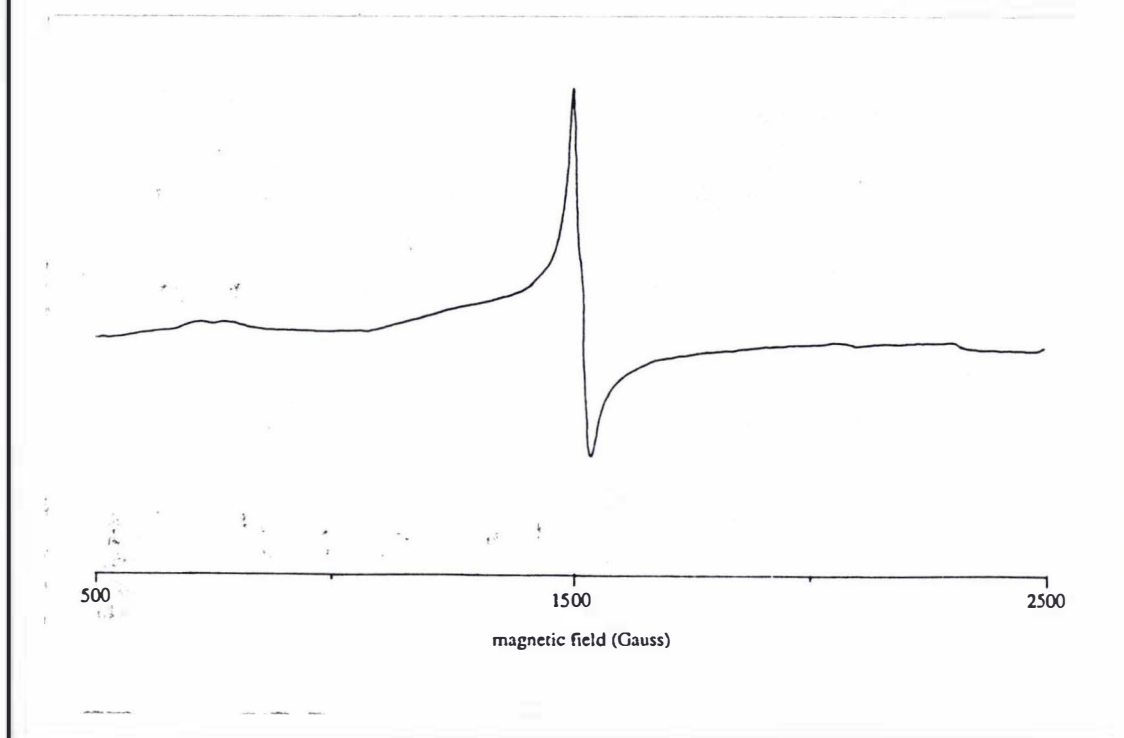
Arginine 121 / Aspartic acid 60

In an attempt to further define the factors which prevent iron binding to the C-lobe of melanotransferrin a double mutant where both Asp 60 and Arg 121 were converted to serine was constructed. The changes were made sequentially, by introducing the change at Asp 60 into the pTZ:R121S mutant which had already been sequenced and expressed. Although only a small amount of protein has been purified this mutant clearly binds iron and initial analysis suggests that it has properties similar to those of R121E and R121D. These results were quite surprising given that the λ_{max} at pH 7.5 for the two individual mutants was 454 nm and 434 nm yet the λ_{max} for the double mutant is 474 nm. Once again it is

difficult to understand what has happened at the iron site and a structural analysis is required before the nature of specific interactions can be determined.

4.3   DISCUSSION

The mutagenesis experiments presented here are preliminary and incomplete, but do serve to demonstrate that a mutagenesis protocol for lactoferrin has been established and that these mutants can be expressed. Characterisation of the mutants must be developed further if the effects of each mutation are to be properly evaluated. In particular, it is clear that high resolution structures of each mutant need to be determined if the spectral changes are to be explained in terms of structural and eventually functional differences. Although analysis is incomplete the mutations made to date have proved interesting. The properties of the iron binding mutants are summarised in Table 3 and discussed below.

Table 3 Summary of the iron binding properties of the mutants.

Mutant	pH range of iron release	pH of ~50% iron saturation	λ_{max}
Lf$_N$	6.0 - 4.0	4.8	454
R121S	7.0 - 3.0	4.8	454/474
R121D	7.0 - 5.0	~6.0	474
R121E	7.0 - 5.0	~6.0	474
D60S/R121S	7.0 - 5.0	~6.0	474
D60S	7.0 - 5.0	~6.0	434

Firstly the N-lobe mutant in which both arginine 121 and aspartic acid 60 have been changed to serine still binds iron. This suggests that changes other than, or in addition to, the mutations of Asp and Arg in the C-lobe of melanotransferrin prevent iron binding. The other mutations likely to be involved are the conversion of the equivalents of Thr 117 and Thr 122 in the C-lobe of melanotransferrin to alanine and proline, respectively. These residues are conserved in all members of the transferrin family which are known to bind iron suggesting that they are important for stable iron binding. A quadruple Lf$_N$ mutant, currently being constructed by Ms S. Shaw, will allow the importance of Thr 117 and Thr 122 to be investigated. Although these four mutations are most likely to be responsible for preventing iron binding it is possible that more subtle changes at positions distant from the metal binding site may also have a role. This is because the strength of iron binding is likely to depend not only on the direct interaction between metal, anion and protein, but also on the interdomain interactions which

define the stable closed configuration of the binding cleft seen in wild type lactoferrin. If changes other than those at the metal binding site are involved in preventing iron binding it is likely that these will only be revealed through a structural analysis of intact melanotransferrin and not by mutagenesis studies based on Lf_N. Garrett and Jhoti (1992) have modelled the structure of melanotransferrin using the iron saturated rabbit serum transferrin structure. However, they do not give details about the iron site, and it is doubtful whether modelling can give the reliability and precision required to allow interpretation of these changes.

Mutations of Arg 121 showed that this residue does influence the metal site (shown by the change in λ_{max}) even though its affect can only be indirect through interaction with the anion. The change to serine showed a pH dependence for λ_{max} which suggested that protonation of a group at the iron site may occur as the pH is lowered and that a change in metal coordination results. HCO_3^- is reported to be more likely to bind in a monodentate fashion than CO_3^{2-} and it is possible that reducing the pH may cause a change towards HCO_3^- and a five coordinate iron atom which could result in an altered iron release profile. It has previously been suggested that a change to a monodentate anion may be the first step in iron release (Sarra et al., 1990). On the basis of these observations and previous suggestions it is proposed that the role of arginine 121 may be to maintain the anion in the dinegative (CO_3^{2-}) form.

Conversion of arginine 121 to aspartic acid or glutamic acid considerably destabilised iron binding and changed the λ_{max}. At this stage it is difficult to be certain as to what changes have occurred and the actual interactions involved will only be revealed by structural studies of the mutant proteins. Likewise conversion of Asp 60 to serine destabilised iron binding although in this instance the λ_{max} was at 434 nm suggesting that removal of Asp 60 had a different affect on the iron site than did changing Arg 121 to serine.

Removal of the N-linked glycosylation site from Lf_N resulted in the production of the expected non-glycosylated half molecule. However, the level of expression of this protein was much lower than that of Lf_N and the strategy of using this mutant as the basis for subsequent mutations was abandoned. The sample available was used to record a UV-visible spectrum which appeared to be different to the spectrum obtained for Lf_N suggesting that the protein may not be folded correctly. Small crystals of this mutant have been prepared and it is hoped that a crystal structure of this protein may reveal what the changes are (J.Tweedie, pers. comm.). In the second instance removal of both glycosylation sites reported to

carry carbohydrate groups in hLf resulted in the production of two forms of lactoferrin which could be separated by SDS-PAGE. The upper band is believed to correspond to a protein which contains one N-linked carbohydrate group and the lower band appeared to be completely carbohydrate free. A likely explanation for this result is that the third, normally cryptic, glycosylation site at position 623 is glycosylated in this double mutant. It would be interesting to express the cDNA in which in which all three of the glycosylation sites have been removed. This would hopefully produce a protein that was carbohydrate free.

There exist many more lactoferrin mutants which could be made; however if the biological significance of such changes is to be understood it will be necessary to develop an *in vitro* assay at least for receptor binding, and ultimately for bacteriostasis and myelopoiesis. A functional analysis carried out in parallel with a structural analysis should help to achieve the ultimate goal of understanding the biological significance of the tight and specific binding of iron by lactoferrin.

In order to complete the characterisation of Lf$_N$ the tertiary structure of the protein was determined. A high resolution structure would allow a detailed comparison of the metal binding site in Lf$_N$ with the metal binding sites in native human lactoferrin. Such an analysis may help explain the differences between Lf$_N$ and native hLf observed in spectroscopic and iron release studies. Although major changes at the iron site are unlikely it is quite possible that small changes in the geometry of the iron site may have occurred. A high resolution structure of Lf$_N$ was also required before a detailed comparative analysis of mutant structures based on Lf$_N$ could be undertaken.

Part B : Chapter One

METHODOLOGY AND PROGRAM DESCRIPTION

.1   Crystal growth

Crystal growth

Preparation of high quality crystals is often the limiting step in analysis of protein structures. Many factors affect the growth of crystals including pH, temperature, protein purity, solubility and ionic strength. In an attempt to obtain good crystals all these factors can be varied along with the method of crystal growth. Many of the approaches available are discussed in detail in several papers by McPherson (1985; 1990). In this study crystals were mainly grown by microdialysis although hanging drop and sitting drop methods were used to screen conditions when only a small amount of protein was available.

2   Data collection

Collection of a high resolution data set from a protein crystal involves the measurement of many thousands (normally greater than 10 000) of reflections which are often quite weak. Previously the collection of high resolution data sets was a time consuming process as individual reflections were collected in a sequential manner. However, in recent years several techniques have been developed which allow the rapid and efficient collection of data, with often many hundreds of reflections collected on each image. The rapidity of these methods has been of critical importance as X-ray induced damage of crystals is a very real

problem. The availability of intense synchrotron radiation sources has also decreased X-ray exposure times and many small crystals can now be used for data collection.

CAD4 data collection

In this study low resolution data to 3.2 Å was collected using an Enraf-Nonius CAD4 diffractometer fitted with a 600 mm helium filled long arm and using Cu-K_α filtered radiation ($\lambda = 1.5418$). Once aligned, the hkl values of a series of strong reflections, used to monitor crystal decay, were input together with the theta limits, collection mode and scan speed. This information was then used by the control program to collect the data as described in the CAD4 operating manual. In this study the limited step scan method of data collection (Hansen *et al.*, 1968) was used and background measurements were not made. The output from the diffractometer was a series of measurements representing the profile of each reflection as the crystal was rotated. Backgrounds were extracted from weak reflections and the integrated intensity of each reflection was obtained using a profile fitting method during data processing (Section B1.3.2). This approach significantly increases the amount of data which can be collected in the life of each crystal.

Image plate data collection

Data were collected using imaging plates and Weissenberg (moving crystal/moving film) photography at the Photon factory, Tskuba, Japan (Sakabe, 1991). Crystals were mounted and then aligned using small angle oscillation photographs and a polaroid cassette. Reflections were recorded on Fuji imaging plates and measured as laser stimulated luminescence with a BA100 image plate densitometer. All of these processes are described in the Synchrotron users guide (Rehse and Zubak, 1992).

Rigaku - R axis axis data collection

ApoLf$_N$ data was collected at the Molecular Structure Corporation (USA) using a Rigaku R-Axis II image plate detector, a rotating anode generator and oscillation photography (moving crystal/stationary film). The data were recorded using an automated image plate collection system.

1.3   **Data processing**

1.3.1   **CAD4 data processing**

Once the CAD4 data had been collected a series of programs, summarised here and described in detail elsewhere (Shepard, 1991; Smith 1992), were used to reduce the data. In essence the three programs converted the raw step profile CAD4 data into a file containing hkl and scaled intensity values. Two important features of the method used were the manner in which backgrounds were determined and the use of a limited step scan mode of data measurement. These are therefore described in more detail.

PREPRO

Background measurements were extracted from the weak reflections as described by Baker and Dodson (1980). In this method reflections are defined as weak if Peak < Ratio x BG; where Peak is the average of the two highest adjacent steps and BG is the average of the two outermost steps, or the two smallest steps (the smaller value was used). The value of Ratio input depended on the number of weak reflections in a shell but was typically in the range of 1.5 to 2.0. Once the weak reflections had been identified the steps to be used as backgrounds were extracted. Background variation as a function of θ was then calculated and a table of backgrounds, divided into $10°$ blocks of ϕ and χ, was created. This table was then used as a look up table to obtain background corrections for all the reflections. Besides creating this table PREPRO also computed and smoothed the absorption curve which had been obtained from an azimuthal scan, with steps of $10°$ in ϕ, according to the method of North *et al.* (1968). A radiation decay correction, calculated from the falloff in intensity of the standard reflections, was also determined.

FITC3

In FITC3 the peak widths of strong reflections, above a given threshold, are used to determine a seven parameter Gaussian function which can be used to describe the half width of all reflections. This function accounts for variation in the peak width over reciprocal space due to the non uniform shape of the crystal. Typically profiles were used if Peak/BG > 2.0, where Peak is the average of the two highest adjacent step movements and BG is the background determined in PREPRO. Once this profile function has been determined it is used in REDUCER to determine the half width and intensity, which is proportional to the product of the

peak height and the half width, of each reflection. This method assumes that the reflections are symmetrical and that a Gaussian function can be fitted to the peak profile. However, this is not always the case, especially for poor quality crystals where the reflections may be split or distorted, and at high resolution where there are often very few intense reflections and the Gaussian function may not be accurately calculated. Although of concern, these problems are largely offset by the greatly increased number of reflections which can be measureed in the life of the crystal when a limited step scan method is used.

REDUCER

REDUCER converts the step scan data into integrated intensities. These are calculated using the calculated peak widths (from FITC3), the position of the centroid for each scan, and the intensity distribution over each scan. Once the intensities have been calculated corrections are applied for crystal absorption, intensity falloff, and Lorentz and polarisation effects. The final output file contains h, k, l values, corrected structure factor amplitudes and sigma values.

B1.3.2   Image plate data processing

Convmt

Images which had been produced using the BA100 image plate scanner in Japan were reformatted for processing on the Silicon Graphics IRIS 4D/30 computer (SGI) using this program. An empirical mask supplied by Dr Sakabe, of the Photon Factory, was also applied at this stage. This mask corrects for the observed variation in response of the scanner over the area of the image.

Image

Once the images had been converted each image was displayed on the SGI, using the program *Image*. This program allows the full diffraction image to be displayed and individual spots to be examined in detail. The coordinates of the fiducial spots can be recorded allowing the position of the film coordinates relative to the instrument coordinates to be determined. In a similar manner the absolute start and end of rotation for data collection relative to the crystal orientation can be measured directly. This was determined by recognising an axial row of spots bisecting the first two fiducial spots. From the absolute rotation range for this image the rotation ranges for all the images can be calculated. It is also possible to overlay

files using this program allowing comparison of the positions of calculated and observed reflections.

Weis

This is a multifaceted program which carries out several different calculations once the cell dimensions, space group and setting parameters used for data collection have been input. Firstly the fiducials are accurately located and used to determine the orientation of the image. Once orientated, a set of well defined reflections is chosen and a least squares refinement of the setting parameters is carried out. Initially only low resolution data was used but once DR and DS (the rms deviations between the calculated and observed spot positions in the x and y directions) had been minimised refinement was carried out using higher resolution data. Typically data to 2.5 Å was used in refinement. Once refinement was complete two files were output, one containing the positions of all the expected spots and the other containing the refined setting parameters (cell dimensions, missetting angles etc). The first file could be displayed on the SGI and compared with the image obtained from the densitometer. This was particularly useful if the orientation of the crystal was uncertain as the arcs of predicted reflections agreed very badly with the position of the observed spots if the setting parameters had been incorrectly refined. If the position of the observed and predicted reflections agreed well the intensities of the predicted reflections, within the sin θ limits specified, were determined using a profile fitting method. Since the spot shape may vary over a film image each image was divided into 9 different regions and a mean spot profile was determined for each region using a series of well defined spots of medium intensity. Once the average peak profile for the 9 regions had been determined and corrected for the background level of absorbance, the intensity of all the reflections was calculated. Lorentz and polarisation corrections were also applied at this time and then a file containing hkl (true and unique), F^2, sigma (F^2), image number and a full/partial reflection flag was output. The process of alignment and integration was repeated for each image.

WeisdataJ

Lastly before the integrated intensities obtained from *Weis* could be merged the data from each image were combined and converted into a form which could be input into INTAV3 and INTAV4 (Section B1.4) using the program *WeisdataJ*.

B1.4   Scaling and merging of data

INTAV3

This program uses structure factors (F's) to calculate scale factors by which different sets of data can be scaled to the same level. The scale factors are calculated in a two step process. In the the first pass through the data all common reflections are identified and a scale factor is calculated. Tighter criteria for the selection of the common reflections are applied in the second pass. Finally the scale factors are output after the reflections that agree poorly have been rejected. Reflections were given unit weights in these calculations.

INTAV4

Scale factors calculated by INTAV3 were applied in INTAV4 and the data were then merged and a unique data set output. There are a variety of acceptance criteria which can be varied allowing reflections that agree poorly to be rejected. Unless stated otherwise the acceptance criteria were
- accept as observed if $F_{obs}/sig > 2.0$
- accept all values if $|F_{max} - F_{min}|/sig < 4.0$
- reject lowest if $|F_{ave} - F_{min}|/sig > 2.0$
- reject highest if $|F_{max} - F_{ave}|/sig > 2.0$
- for two measurements reject the smaller if $|F_{max} - F_{min}|/sig > 4.0$.

When the data is merged in INTAV4 the reflections can either be given equal weight (unit weights) or they can be weighted according to their standard deviations.

B1.5   Structure solution

When diffraction data are collected from a crystal the position in reciprocal space (specified by the indices h, k and l) of each reflection is known (or can be determined) and the amplitude of each reflection is measured, however, the phase relationship between reflections is unknown. Two methods, known as multiple isomorphous replacement (MIR) and molecular replacement (MR) are commonly used to solve protein structures. Essentially these two methods differ in the way in which the phase problem is overcome.

Multiple isomorphous replacement (MIR)

The key feature of this method is that it requires the preparation of heavy atom derivatives where a heavy metal eg. Au, Pt, or Hg is attached to the protein without causing changes in protein conformation or in the crystal packing. Once suitable heavy metal derivatives have been found both native and derivative data sets are collected and the positions of the heavy atoms are approximately determined using Patterson methods (Stout and Jensen, 1989). The information from the derivative and native data sets is then combined allowing the phases to be determined using the vector equation

$$F_H = F_{PH} - F_P$$

where F_P, F_{PH} and F_H are the structure factors of the protein, derivative and heavy atom respectively. Following calculation of the phases an electron density map can be calculated and used to trace the protein backbone and eventually to position all atoms in the structure.

Molecular replacement (MR)

Molecular replacement as a method of protein structure solution was developed between 1962 and 1967 at the MRC laboratory in Cambridge (UK) (Lattman, 1985). In this method a model structure is used to obtain information about the phases instead of relying on heavy atom derivatives. This method is dependent on the availability of a similar or homologous structure which can be used as the model. Once a suitable model has been identified the structure is solved in two steps. Firstly the rotation matrix required to correctly orientate the model structure is determined and then the position of the model in the cell is calculated. The two functions used are termed the rotation function and the translation function. Despite the obvious advantages of this method it is limited by the availability of the coordinates for a closely related structure.

Molecular replacement was the method chosen in this study because the structure of several forms of lactoferrin, both metal saturated and metal free were available for use as models.

ALMN

The rotation function can be calculated in various ways. One way is to use Patterson space, in which a Patterson map is calculated from the crystal intensity data and compared with the Patterson map which is obtained from the model. Once calculated one map is rotated onto the other and the overlap function is

calculated for each discrete rotation. Since the Patterson map is the Fourier transform of the intensity data an equivalent process can be carried out in reciprocal space by comparing intensities of the crystal with those derived from the model. These comparisons are carried out in ALMN (Crowther, 1972) using Fast Fourier methods. The output, in Eulerian angles α, β and γ, shows those rotations where maximum overlap occurs. The peaks are listed in order of peak height allowing the top peaks to be easily recognised. The position of all symmetry related peaks is also output.

Tsearch

The relationship between molecules in a unit cell is defined by a translation vector which can be determined using the program Tsearch. In this program the model in its correct orientation (determined by ALMN) is translated through the cell on a 3 dimensional grid. Structure factors are determined and the crystallographic R value is calculated at each grid point. If the model is appropriate the correct location for the model will be recognised by a lower R value. The shifts on x, y and z required to obtain the lowest R value represent the vector required to give the best agreement between the observed $F_{(hkl)}$ and the calculated values. Once the position of one molecule is defined the position of the other molecules in the unit cell can be determined by applying the symmetry operators.

B1.6   Protein structure refinement

Although much information about the overall configuration of a protein is provided by the unrefined structure the precise details of interactions can only be obtained when a protein structure has been refined. For structures which have been solved by molecular replacement it is especially important to refine the structure as it provides the only verification that the structure is correct. The aim of refinement is to minimise the difference between observed and calculated structure amplitudes ie. minimise $|F_{obs}| - |F_{calc}|$, by varying the atom parameters. In macromolecular structures refinement is often difficult because the number of parameters approaches the number of observations. For a well determined structure the ratio of observations to variables should be greater than ten. This is usually not possible with protein structures but the ratio can be improved by including additional observations in the form of constant B values and occupancies, and constraining the atom positions to known geometries.

Several packages incorporating different approaches are available for refining protein structures. In this study the restrained least squares refinement package TNT was used (Tronrud *et al.*, 1987).

TNT - the restrained least squares refinement package.

The main unique feature of TNT is the modular nature of the program which allows components of the program to be exchanged and altered without changing the main body of the program. Details of the package are provided in the paper by Tronrud *et al.* (1987) and the manual which accompanies the package. For the purposes of this thesis three types of refinement termed

- rigid body refinement
- XYZ refinement
- B value refinement

will be described. The first type of refinement used was rigid body refinement. In this instance shift vectors are calculated for each atom and then summed to give a single shift vector which is used to shift the model as a single rigid body. This type of refinement was used to refine the position of the model provided by the initial rotation and translation function solution. Once the orientation and position were more accurately known the position of each atom was refined using XYZ refinement. As before shift vectors are calculated for each individual atom but in this case the shift vector is applied to each separate atom. Restraints are, however, applied to bond lengths, bond angles, planar groups and non-bonded contact distances. The relative weight of each restraint can be varied, as can the relative weighting of the geometrical terms and the diffraction terms. The program also allows a correction for the solvent continuum in the crystals to be applied, but this was not done in XYZ refinement. B values and occupancies were also kept constant. Occupancies were never refined but in the later stages of refinement the B values were refined while the XYZ shifts were kept constant. For B value refinement no overall B was used, allowing shifts to be taken up as changes in the B values of the individual atoms. Also in B value refinement a correction for the solvent continuum was applied. After the final cycle of each round of refinement a list of the worst atoms with regard to bad bond angles, lengths, non-bonded contacts etc was obtained, highlighting areas which required manual rebuilding.

B1.7   Map calculations and model building

$2F_{obs} - F_{calc}$ / $F_{obs} - Fc_{alc}$

Electron density maps were calculated using a series of standard programs from the CCP4 suite of programs. Firstly the calculated structure factors from the model were scaled to the observed data and then the Fourier synthesis for the unique area of space was calculated. The data was then sorted and used to calculate a map. However, in the case of the monoclinic space group used in this study it was necessary to expand the data into a hemisphere of reciprocal space before it was sorted. Once the map had been calculated in the unique area of space it was extended to cover the model and "bricked" so that it could be displayed on the Evans and Sutherland. Two different types of maps were calculated, one with the coefficients $2F_{obs} - F_{calc}$, and the other with coefficients $F_{obs} - F_{calc}$. The first type of map shows the electron density of the molecule including that of the omitted residues. Typically this map was contoured at 1σ for well defined parts of the structure but at 0.5σ in poorly defined regions. The second map is called a difference map and was used to highlight incorrect parts of the structure. This type of map was especially useful for positioning omitted or incorrectly placed sidechains and for locating solvent molecules. Incorrectly placed atoms included in the model appeared as negative peaks while if atoms were missing a peak of positive density was often visible.

FRODO - the model building package

In between rounds of refinement models were manually rebuilt using an interactive graphics system (Evans and Sutherland PS-300) and the program FRODO (Jones, 1978).

Part B : Chapter Two

CRYSTALLISATION OF Lf$_N$

̦.4.1 INTRODUCTION

Before a structural analysis could be undertaken high quality crystals of Lf$_N$ were
required. Crystals of several different forms of lactoferrin have previously been
grown and the structures of these solved. The conditions used to obtain high
quality crystals of the various types of lactoferrin are summarised in table 4.

Table 4 Crystallisation conditions for different lactoferrin species.
Isopropanol has been abbreviated to IPA.

Protein	Concentration	Crystallisation conditions
Native human Fe$_2$Lf (Baker and Rumball, 1977)	100 mg/ml	0.01 M phosphate pH, 7.8, 10% v/v ethanol, 4°C
Native human apoLf (Norris *et al.*, 1989)	20 - 30 mg/ml	0.01 M phosphate, pH 7.8, 10% v/v methanol, 4°C
DG human apoLf (Norris *et al.*, 1989)	12 mg/ml	0.05 M Tris, pH 7.8, 4°C 5% v/v MPD, 5% v/v ethanol
Native bovine Lf (G.Norris, *pers. comm.*)	100 mg/ml	0.01 Tris, pH 8.25, 4°C 10% IPA or 10% ethanol

Although the details of the conditions used to obtain good crystals vary depending
on the type and origin of the protein all of the crystals were grown by
microdialysis against low ionic strength buffers. It was predicted that Lf$_N$ would
crystallise using similar conditions.

B2.2 RESULTS

B2.2.1 Crystallisation of native Lf$_N$

Lf$_N$ was purified by ion exchange chromatography as described previously
(Section A1.5.7). Pure Lf$_N$, as judged by SDS-PAGE, was prepared for
crystallisation by filtration through a 0.2 μm membrane and then concentrated to
~70 mg/ml using a microconcentrator (Section A1.5.9) which had a molecular
weight cut off of 10 kDa. A preliminary screen of various crystallisation
conditions was carried out using 20 μl microdialysis buttons (Table 5). Although
this table does not detail every condition used the range of conditions screened is

summarised and the typical result is indicated. In the majority of samples oils or small needles were obtained. The best crystals were obtained in 0.01 M Tris, pH 7.8 containing 10% isopropanol. Phosphate buffers were abandoned after the initial screen as oils were obtained in every instance. Subsequent batches of protein were set to crystallise over a range of conditions similar to those where small crystals had been obtained (Table 5). After 6-8 weeks complete crystallisation was obtained in buttons dialysed against 0.01 M Tris, pH 8.0 containing either 10% or 12% isopropanol. Most of the crystals formed were long thin needles or very thin plates although some had cell dimensions of up to 0.8 x 0.8 x 0.2 mm and were able to be used for initial diffraction studies. Unfortunately, these crystals were highly unstable at room temperature reverting to an oil in approximately one hour. In an attempt to prevent rapid deterioration of the crystals they were mounted in capillaries at 4°C and then transferred to 12°C but they still appeared to be very unstable and dissolved after several hours.

Table 5 Summary of the conditions used to obtain crystals of FeLf$_N$.

* Conditions scereened in the initial attempt to find crystallisation conditions. All samples were kept at 4°C for crystallisation.

Buffer	pH range	Precipitant	Result
*0.01 M phosphate	7.0 - 7.8	5% ethanol/ 5% MPD	Oils formed in all instances
*0.01 M phosphate	7.0 - 7.8	10% IPA	Oils formed in all instances
*0.01 M Tris	7.0 - 7.8	5% ethanol/ 5% MPD	Oils and a few small needle like crystals
*0.01 M Tris	7.0 - 7.8	10% IPA	Many small needle crystals
0.01 M Tris	8.0	10% - 14% IPA	Lots of small crystals
0.01 M Tris	8.1 - 8.2	10% - 12% IPA	Lots of crystals some thick rods

Throughout this work the production of better crystals was hampered by an apparent variation between batches of protein with several batches of protein precipitating under conditions that had previously been used to obtain crystals. Also when the protein concentration was decreased, in an attempt to reduce the number of nucleation sites, the protein precipitated. These results were rather strange and were difficult to explain although subsequent experiments suggest that

the differences observed may be due to variation in the type and extent of protein glycosylation (see section A3.2.2.3). Despite these problems it was found that the presence of between 1 and 5 mM NaCl assisted the formation of better crystals although again the effect was somewhat random. After finally obtaining thicker crystals the problem of crystal stability remained. A slight improvement in crystal stability was achieved by slowly equilibrating the crystals into ~25% isopropanol although they were still not very stable and were difficult to photograph.

B2.2.2   Crystallisation of deglycosylated Lf$_N$

FeLf$_N$

In an attempt to improve crystal quality and the reproducibility of the crystallisation conditions attempts were made to crystallise deglycosylated iron saturated Lf$_N$ (DG FeLf$_N$). Approximately 40 mg of Lf$_N$ was deglycosylated using PNGase (Section A1.5.10). It was observed that deglycosylation of Lf$_N$ caused a total loss of protein colour, suggesting that removal of the carbohydrate group is somehow associated with iron release. The reason for the loss of iron is not very clear but is believed to be related, in part, to the actual removal of the carbohydrate group as the protein is still coloured after dialysis against the nonidet buffer and only becomes colourless after addition of PNGase. The loss of iron was not believed to be associated with permanent changes to the protein as it was possible to bind iron back to the protein in the normal way once it was purified and had been equilibrated in standard HEPES buffer.

Following removal of the carbohydrate groups the protein was equilibrated into 0.01 M HEPES, pH 7.8, 0.2 M NaCl and then purified by ion exchange chromatography on a CM-Sephadex column (Section A1.5.7). Essentially the purification was as described before (Section A3.2.1) except that 0.45 M NaCl was used to elute the protein because DGLf$_N$ binds to the resin slightly more tightly than native Lf$_N$. The eluted protein was analysed by SDS-PAGE and demonstrated to be a single band, the pure fractions were combined and then extensively dialysised against 0.01 M HEPES, pH 7.8, 0.2 M NaCl, 0.05 M NaHCO$_3$ to remove traces of EDTA. A molar excess of iron, as FeNTA, was added to the protein, allowed to bind and then excess iron was removed by dialysis against several changes of 0.01 M HEPES, pH 7.8 containing 0.2 M NaCl. The protein was then filtered through a 0.2 μm membrane and concentrated.

As the effect of deglycosylation on the crystallisation conditions could not be predicted a range of conditions, similar to those used to crystallise Lf$_N$, were screened. Surprisingly beautiful rod shaped crystals (Fig. 46) were obtained in

two buttons after 10 days of dialysis. The successful conditions used were 0.01 M Tris, pH 8.0 containing either 12% isopropanol or 12% ethanol. The stability of these crystals was not known but they appeared very similar in form to the native Lf$_N$ crystals suggesting that these crystals might also be unstable. In an attempt to find a better method of stabilising the crystals an extensive range of conditions were screened. After a considerable amount of trial and error it was found that the addition of 10% PEG 4000 to the mother liquor over a period of 5 days stabilised the crystals such that they could now be manipulated at room temperature. Crystals stabilised in this way were characterised and used for data collection.

Fig. 46 Crystals of DG FeLf$_N$.

These crystals were grown by microdialysis against 0.01 M Tris, pH 8.0 containing 12% ethanol.

At this stage better crystals of native Lf$_N$ were also obtained. They were grown in 0.01 M Tris, pH 8.1, 12% isopropanol, 4 mM NaCl, over 2 months. The protein used to obtain these crystals was actually the remains of several batches which had been combined and then passed down a second CM-Sephadex column. The appearance of these crystals was not anticipated as previously it had been very difficult to get discrete single crystals with nice sharp edges. These crystals were also stabilised by equilibration with 10% PEG 4000 and used for photographic analysis with the best crystals being saved for data collection.

ApoLf$_N$

As mentioned previously deglycosylation of Lf$_N$ causes bound iron to be released, effectively creating ApoLf$_N$. Although most of the deglycosylated protein was saturated with iron and used to grow FeLf$_N$ crystals a small amount of the protein was kept in the iron free form, dialysed against buffer to remove any remaining traces of EDTA and filtered. The purified protein was then used in hanging drop crystallisation trials at a concentration of 70 mg/ml. Small needles were obtained with well solutions of either 0.1 M Tris pH 8.0, 5% isopropanol/5% MPD or 0.1 M Tris pH 8.0, 10% isopropanol. On the basis of these preliminary results a small number of microdialysis buttons were set down to crystallise. After approximately 10 days colourless square plates (Fig. 47) up to 0.5 x 0.5 x 0.1 mm were obtained in buttons dialysed against 0.01 M Tris, pH 8.0, containing between 12% and 15% isopropanol as the precipitant. Slightly larger crystals were subsequently obtained when the concentration of isopropanol was reduced to 11%. The stability of these crystals was not known but in case they were also unstable 10% PEG 4000 was added to the mother liquor. This did not appear to have a detrimental effect on the crystals therefore all ApoLf$_N$ crystals were treated in this way.

Fig. 47 Crystals of DG ApoLf$_N$.

The crystals shown were grown by dialysis against 0.01 M Tris pH 8.0 containing 15% isopropanol. Subsequently bigger crystals have been grown using 12% isopropanol as the precipitant.
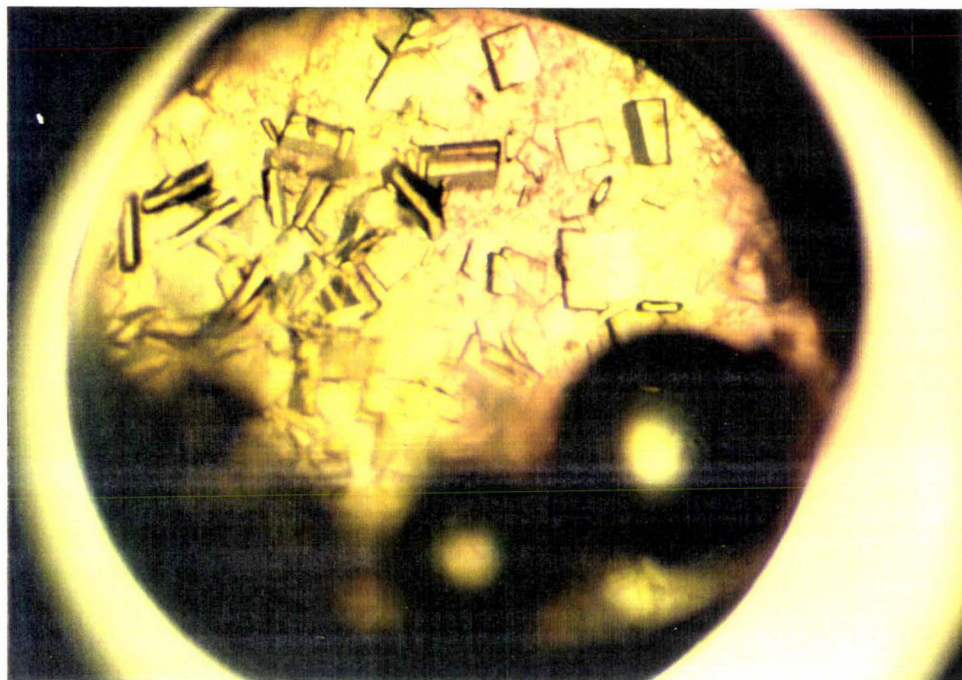
§2.2.3   Characterisation of the FeLf$_N$ and DG FeLf$_N$ crystals

Photography of FeLf$_N$ crystals

Initially photographic studies were carried out using the small unstable FeLf$_N$ crystals which had been equilibrated in 25% isopropanol. These crystals were very difficult to analyse due to the absence of strong crystal morphology and the weak nature of the reflections. From these initial studies the crystals were thought to be monoclinic, cell dimensions a = 57.9 Å, b = 58.4 Å, c = 120.0 Å , $\alpha = \gamma =$ 90° and β = 89.9°, but this space group was not assigned as the symmetry was not as expected for any of the standard monoclinic cells. Although the unit cell was not accurately known some of these crystals were taken to the Photon Factory in Japan where collection of a data set was attempted. The crystals diffracted moderately well (to ~ 2.5 Å) but it was not possible to collect a data set as the reflections were split. The split nature of the reflections (hence the crystals) was attributed to either crystal cracking during stabilisation in isopropanol or to deterioration of the crystals during transport to Japan.

Subsequent to this work larger crystals that diffracted more strongly were obtained and the stabilisation conditions were changed (Section B2.2.2). The more stable crystals were photographed in an attempt to find a conventional space group. Despite extensive precession photography no space group of conventional high symmetry could be found therefore it was decided that the crystals must be triclinic. It was difficult to determine which triclinic cell was the smallest, and hence the correct triclinic cell, therefore cell dimensions were not assigned at this stage.

Photography of DG FeLf$_N$ crystals

It had been hoped that DG FeLf$_N$ would crystallise in a different space group which could be more easily characterised. However this was not the case and crystals of DG FeLf$_N$ were found to be isomorphous to those of native Lf$_N$. They did appear to be of slightly better quality and diffracted to a higher resolution therefore they were used for data collection.

Photography of ApoLf$_N$ crystals

Visual analysis of the ApoLf$_N$ crystals suggested that they were tetragonal, this was confirmed by photographic methods. The crystals were easily aligned even though they did not diffract very strongly and were shown to belong to either

space group $P4_12_12$ or $P4_32_12$ which cannot be distinguished by precession photographs. The cell dimensions (a = b = 58.4 Å, c = 215 Å and $\alpha = \beta = \gamma = 90°$) were estimated using 12° precession photographs although it was difficult to accurately measure the c axis because the reflections were very close together. The Vm was calculated, assuming one molecule per asymmetric unit. A value of 2.3 was obtained which is comparable to that obtained for other protein crystals (Matthews, 1968).

3.4.5   DISCUSSION

After some initial problems crystals of iron saturated and iron free Lf_N have been prepared and stabilised using PEG 4000. The crystals of $FeLf_N$ were difficult to characterise and were thus not assigned a space group although at this stage it was assumed that the crystals were triclinic (see section B3.2.5 for elucidation of the correct space group). Quite a few good crystals of $FeLf_N$ and DG $FeLf_N$ had been grown by this time therefore it was anticipated that the structure could be solved in the absence of the correct space group. Only two buttons of $ApoLf_N$ crystals had been grown but as these were easy to characterise it was hoped that there would be sufficient crystals to allow a structural analysis. These crystals belonged to either space group $P4_12_12$ or $P4_32_12$. These two space groups cannot be distinguished by precession photography.

FeLf$_N$ STRUCTURE SOLUTION AND ANALYSIS

B3.1   INTRODUCTION

As discussed previously the structures of several different forms of lactoferrin have been determined (Section 1.3). The first structure solved was that of iron saturated human lactoferrin (Fe$_2$Lf) which was solved by multiple isomorphous replacement (Anderson *et al.*, 1987). Subsequently the structures of bovine lactoferrin (Baker *et al.*, manuscript in preparation) and human apolactoferrin (Anderson *et al.*, 1991) have been solved by molecular replacement using the structure of Fe$_2$Lf as the starting model. It was anticipated that the structure of FeLf$_N$ could also be solved by molecular replacement using fragments of the Fe$_2$Lf structure as the search models. It was hoped that once a high resolution structure was available some of the differences between Lf$_N$ and hLf may be accounted for.

B3.2   RESULTS

B3.2.1   CAD4 data collection and processing

Data collection

X-ray diffraction data to 3.2 Å resolution were collected from two crystals of deglycosylated FeLf$_N$ using an Enraf-Nonius CAD4 four circle diffractometer (Section A1.4.2). The crystals were mounted in thin walled capillaries and then optically centred on the diffractometer (Section B1.2). As the unit cell dimensions were unknown the initial unit cell was determined using a series of low angle reflections ($\theta \sim 2.0°$). The approximate cell dimensions were a = 58.2 Å, b = 58.3 Å, c = 72.4 Å, $\alpha = 112.5°$, $\beta = 113.6°$ and $\gamma = 90.2°$. This triclinic cell could be transformed into several cells of higher symmetry but given the uncertainty in the choice of cell and the desire to collect data quickly the triclinic cell was used for data collection. It was also reasoned that if a triclinic cell was used to collect the data, and subsequently a cell of higher symmetry was found, the data could be converted and a complete set of data would still be obtained. Before the data was collected the cell dimensions were refined using 21 higher angle reflections (θ - 6°/7°) to give a unit cell of a = 58.21 Å, b = 58.34 Å, c = 72.44 Å, $\alpha = 112.47°$, $\beta = 113.68°$ and $\gamma = 90.1°$. This unit cell had a volume of 204 668 Å^3 which suggested that there were two molecules in the asymmetric

unit assuming Lf$_N$ has a molecular weight of ~40 000. It was not possible to refine the cell further at this stage because a high angle set of data was not available.

Data were collected using the limited step scan method (Section B1.2) and the setting parameters listed in Table 6.

Table 6 Data collection parameters.

Instrument	CAD4 diffractometer
Radiation	CuK$_\alpha$
Wavelength	1.5418 Å
Power	40 kV, 32 mA
Refinement of the cell	21 reflections
Intensity standards	4 reflections measured every two hours
Scan	ω - scan, 8 steps of 0.04°
Background	No measurement

In this method no attempt is made to cover the whole peak, thus maximising the amount of data which can be collected during the life of a crystal. The scan chosen measures about 75% of the peak with the rest determined using profile fitting methods (Section B1.3.1) during data processing. Background intensity measurements were not made either, further decreasing data collection time. Instead these were extracted from the scan during data processing as described by Baker and Dodson (1980) (Section B1.3.1). Azimuthal absorption curves were collected for each crystal at the completion of data collection allowing the variation in intensity due to crystal morphology to be accounted for (North *et al.*, 1968). Data collection was terminated if the intensity of the control reflections decreased, on average by more than 25%, during the life of the crystal. Five shells of data (summarised in Table 7) were collected starting at low resolution data and progressing to higher resolution. As the resolution of the data was increased the scan speed was decreased in order to obtain large enough peaks, that had a profile, which could be integrated. At the end of data collection on the first crystal, the unit cell was refined using a series of higher angle reflections. This refinement gave a unit cell with dimensions of a = 58.37 Å, b = 58.27 Å, c = 72.44 Å, α = 112.49°, β = 113.82° and γ = 89.97°. This was the unit cell used for structure solution.

Table 7 CAD4 Data collection statistics.

Crystal	Shell	Resolution θ	limits (Å)	Number of reflections	cumulative % decay	Scan speed
1	1	1.0 - 7.4	44.0 - 6.0	2114	4%	1.268
	2	7.4 - 9.8	6.0 - 4.5	2730	6%	0.916
	3	9.8 - 11.6	4.5 - 3.8	3095	25%	0.916
2	4	11.3 - 13.0	3.9 - 3.5	3769	14%	0.8
	5	13.0 - 14.0	3.5 - 3.2	2740	28%	0.659

Data processing

Once the data had been collected the raw CAD4 data were processed, giving a file of hkl and scaled intensity values. The details of this process, which corrects for background radiation levels, absorption, intensity fall off during data collection and Lp (Lorentz and polarisation) factors, are given in section B1.3.1. The quality of each processed shell is summarised in Table 8. In total 91% of the reflections had $I > 2\sigma$ and very few reflections were rejected indicating that the data was generally of good quality.

Table 8 Summary of the CAD4 data processing statistics.

Shell	Number of measurements	Number with $I > 2\sigma$	Rejected reflections		
			High on edge	Wyckoff step on edge	Negative I
1	2022	1984	9	0	47
2	2581	2526	19	1	34
3	2925	2746	83	0	75
4	3502	3226	64	7	98
5	2416	1987	93	14	166

Finally the shells of data were scaled together. Although somewhat unreliable the scale factors within each crystal were calculated using the intensity control reflections as there was no overlap between the shells of data. The data from the two crystals were then scaled together using the common reflections between shell 3 and shell 4. A merging R of 6.0% was obtained after 1% of the common 1055 pairs of reflections had been discarded. The final scaled data file contained 12916 unique reflections which represented 80% of the expected reflections to 3.2 Å. Of the unique data 81% of the reflections had $I > 2\sigma$ which indicated that

the data was of good quality. This data file was then used to determine the structure by molecular replacement.

B3.2.2   Structure solution in P1

The models

Three molecular replacement search models were created from the refined iron lactoferrin (Fe_2Lf) coordinates, taken from the Brookhaven data base. The models (Table 9) represented the amino terminal half of human lactoferrin and the two domains from this lobe.

Table 9 Molecular Replacement search models for $FeLf_N$.

Model name	Residues included
Lfnmodel	intact N lobe of Fe_2Lf, residues 1 - 333
Domain 1	domain 1 of Fe_2Lf, residues 1 - 91 and 251 - 333
Domain 2	domain 2 of Fe_2Lf, residues 92 - 250

The iron atom and the bicarbonate ion were omitted from all of the models, as were all water molecules. Structure factors were calculated for the models in an orthogonal triclinic P1 cell large enough (80 Å x 80 Å x 80 Å) to eliminate the contribution of intermolecular Patterson vectors. An artificial temperature factor of 25.0 $Å^2$ was applied to all atoms.

The rotation function

The rotation angle required to transform the model structures was determined using Crowthers' (1972) fast rotation programme (Section B1.5.1). The rotation angles required to give the three highest peaks for each model, using data between 20 Å and 3.5 Å, are given in Table 10. As anticipated, for an asymmetric unit containing two molecules, two rotation vectors giving peaks of equal height were obtained for each of the model structures used. The rotation angles required to position the model for all of the test structures are very similar. This suggested that the values obtained were correct. Because Lfnmodel represents the complete structure in question, and because the difference in peak height/rms was greatest in this case, the rotation angles obtained from this solution were applied in the next stage.

Table 10 Rotation function solutions for FeLf$_N$.

α, β and γ are Eulerian angles which were incremented in 2.5° steps through 360° in α and γ and 5° steps between 0 and 180° in β. The value of rms refers to the deviation of the map from its mean.

Model	Peak	α°	β°	γ°	peak height	height/rms
Lfnmodel rms = 25.5	1	50.4	78.8	165.5	292.0	11.4
	2	309.3	101.3	345.8	289.9	11.3
	3	148.6	113.2	49.0	117.3	4.6
Domain 1 rms = 19.4	1	308.6	102.6	346.0	127.7	6.6
	2	50.5	78.0	166.0	119.0	6.1
	3	9.0	138.4	199.0	81.4	4.2
Domain 2 rms = 30.69	— 1	49.8	79.0	165.8	229.2	7.5
	2	309.7	100.7	345.8	223.1	7.27
	3	138.0	73.2	219.5	163.6	5.3

The translation function

The next step was to position the molecules in the unit cell. One molecule, molecule A, which had been rotated by 50.4°, 78.8° and 165.5° was placed at the origin while the position of the second molecule, molecule B, which had been rotated by 309°, 101.3° and 345.8° was determined by carrying out an R factor search over the entire unit cell (Section B1.5.2). When data between 20.0 Å and 3.5 Å were used and 1.0 Å steps were made along each axis a single minimum of 0.418 positioned at 53.16 Å, 18.73 Å and 60.02 Å on x, y, and z respectively was obtained. The background average R value was 0.476. The position of the minimum was more precisely defined using a stepsize of 0.1 Å between 50 Å - 52 Å on x, 17 Å - 19 Å on y, and 59 Å - 61 Å on z. This "finesearch" gave a minimum R value of 0.414, at x = 53.24 Å, y = 18.39 Å and z = 59.98 Å.

The coordinates for molecule B were then obtained by applying the translation vector to the previously rotated molecule. It is important to note that because Tsearch calculates the shifts parallel to the cell edges, whereas standard pdb (protein data bank) coordinates are orthogonal, it was necessary to apply the translation vector to the rotated coordinates as fractional shifts. At this stage models of the two molecules in the asymmetric unit were created and examined for overlaps using FRODO and an interactive graphics system (Section B1.7). From this inspection it appeared that the two molecules in the asymmetric unit occupied unique areas of space except for the last helix of each molecule which

appeared to overlap. This suggested that the position of this helix might differ in Lf$_N$.

B3.2.3 Refinement of the structure in P1

Rigid body refinement

A round of rigid body refinement (Section B1.6) was carried out once the rotation and translation functions had been solved. Firstly the two molecules were each treated as single rigid bodies and then each molecule was broken into three fragments representing the two domains (domain 1: residues 1 - 91 and 251 - 333; domain 2: residues 92 -250). This round of refinement caused the R factor to drop from 0.401 to 0.333 (Table 11).

Table 11 Summary of P1 rigidbody refinement.

The weights used for all cycles were R factor = 1 x 10^{-2}.

Cycles	Data used	Model	R factor
1 - 3	20 - 8 Å	Intact Lf$_N$	0.401 - 0.368
4 - 6	20 - 8 Å	Domain fragments	0.368 - 0.363
7 - 9	10 - 5 Å	Domain fragments	0.363 - 0.333

After nine cycles of rigid body refinement a 2F$_{obs}$ - F$_{calc}$ electron density map (Section B1.7) was calculated and the agreement of the model with the data was determined. Most of the main chain atoms and many of the sidechains appeared to fit the density well. There was also well defined density for the the iron and bicarbonate which had been left out of the model giving further confidence in the solution. However, the density was poorly defined for residues 313 to 333 and for the first four residues therefore these residues were removed from the structure at this stage. In addition all of the sidechains which did not have well defined density, as well as the iron binding ligands, were removed in an attempt to reduce bias in the final structure. Forty six sidechains were removed from molecule A and 52 sidechains from molecule B.

Following the removal of these parts of the structure the position of all the atoms was refined by nine cycles of XYZ refinement (Section B1.6). The progress of this round of refinement is summarised in Table 12. The weighting regime used was not very systematic in that initially the weights were too loose and then they were tightened too quickly, however the overall effect appeared to be satisfactory.

Table 12 Summary of P1 XYZ refinement (Round one).

8 - 3.2 Å data were used in all cases.

Cycles	Change in Rfactor	Geometry Weighting	rms Δ bonds
1 - 4	0.355 - 0.229	Loose weights 1 x 10^{-2}	0.109
4 - 6	0.229 - 0.263	Tight weights 1 x 10^{-3}	0.033
6 - 9	0.263 - 0.266	Tight weights 1 x 10^{-3} (increased bond angle weight)	0.023

Another $2F_{obs} - F_{calc}$ map was calculated at this stage and many of the sidechains were added back to the model, including all the ligands and the iron and bicarbonate moieties. There was still very little density for the last two helices (residues 313 - 333) or for the first four residues therefore they were not added back to the model. After this round of rebuilding 21 sidechains were still absent from molecule A and 24 sidechains from molecule B between residues 4 and 312. These two rebuilt models were used as the starting models for the next round of refinement which is summarised in Table 13.

Table 13 Summary of P1 XYZ refinement (Round two).

8 - 3.2 Å data were used in all cases.

Cycles	Change in Rfactor	Geometry weighting	rms Δ bonds
1 - 4	0.331 - 0.237	Loose weights 4 x 10^{-3}	0.069
4 - 6	0.237 - 0.247	Medium weights 2 x 10^{-3}	0.036
6 - 9	0.247 - 0.249	Tight weights 1 x 10^{-3}	0.023

At the end of P1 (round two) XYZ refinement the rms values for bond lengths and bond angles were 0.025 and 3.47 respectively which was reasonable for a structure in the preliminary stages of refinement. Further rounds of P1 refinement were not carried out because the relationship between the two molecules in the asymmetric unit was determined (Section B3.2.5) making it clear that a space group of higher symmetry existed.

B3.2.5 Determination of the correct space group

The existence of a space group of higher symmetry had always been suspected but its nature had been difficult to identify. However, at this time it was found that the two molecules in the asymmetric unit were related by a two fold rotation about the a axis of the triclinic cell. At different times during analysis of the crystals several other space groups had been suggested including one with

approximate dimensions of a = b = 58 Å, c = 120 Å and $\alpha = \beta = \gamma = 90°$. It was recognised at this time that this unit cell is an unconventional monoclinic cell, space group (I2), with reflections h + k + l = 2n+1 absent. Conversion from the triclinic, P1 (hkl) space group to monoclinic, I2 (HKL) space group was achieved given that H = h, K = k and L = 2l + h + k. The I2 space group could be further rearranged to the more conventional space group C2 (laue group 5) with the b axis unique and h + k = 2n+1 absent. The cell dimensions of space group C2 were a = 132.7 Å, b = 58.2 Å, c = 58.3 Å, $\alpha = 90.0°$, $\beta = 113.7°$ and $\gamma = 90.0°$. The relationship between the three space groups, P1, I2 and C2 is shown in Fig. 48.

Fig. 48 Representation of the relationship between the space groups P1, I2 and C2.

In Fig.48a the relationship of the P1 cell (blue) to the I2 cell (teal) is shown and in Fig.48b the relationship of the I2 cell (teal) to the C2 cell (pink) is shown. Using the same colour code the relationship of all three cells is shown in Fig. 48c.

Fig. 48a

Fig. 48b
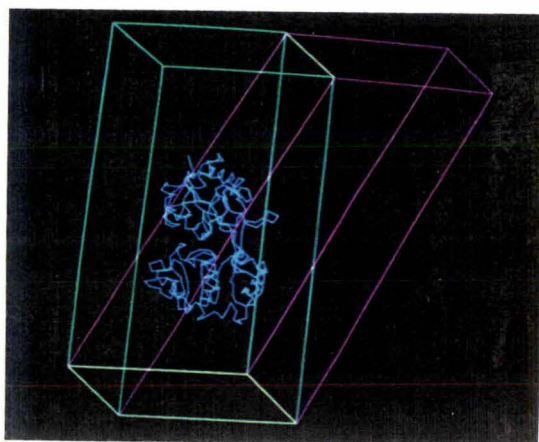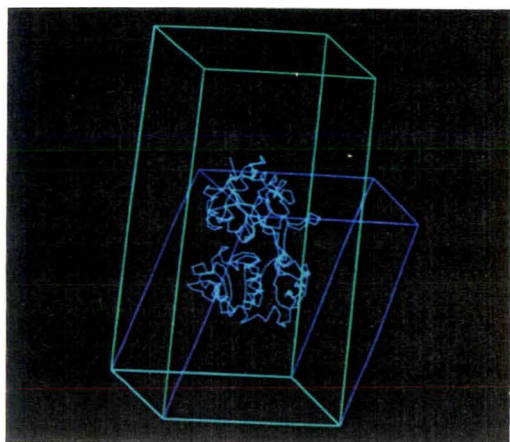
Fig. 48c
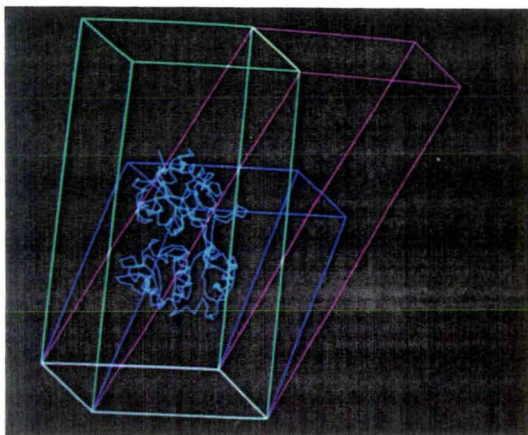
At this stage the CAD4 data was reindexed as monoclinic, space group C2, and merged (Section B1.4). A merging R of 4.5% for equivalent reflections was obtained confirming that C2 was indeed the correct space group. If the space group had not been correct a higher merging R value would have been obtained as the reindexing would have given nonequivalent reflections the same indices. Following elucidation of the correct space group it became possible to process the high resolution image plate data which had been collected at the Photon Factory (Tsukuba, Japan). This data had actually been collected with the same crystal as that used for the first three shells of CAD4 data prior to the CAD4 data collection but it was not possible to process the images until the correct unit cell was determined.

B3.2.5   Collection and processing of the high resolution data

Collection of the data

High resolution data was collected (Section B1.2) using synchrotron radiation and a Weissenberg camera equipped with imaging plates at the Photon Factory, Tsukuba (Sakabe, 1991). The camera settings and exposure times used to collect the data are summarised in Table 14.

Table 14 Image plate data collection statistics.

IP = image plate

Parameter	Value
Source	Beam-line 6A2, Photon Factory
Method	Screenless Weissenberg photography
Wavelength	1.0 Å
Crystal to IP distance	286 mm
Coupling constant	2.5
Oscillation range ($\omega°$/IP)	15.5°
Number of osc^n / IP	8
ω scan speed °/sec	2
Exposure per image (secs)	124
Number of IP's	13 split plates
Total exposure	27 minutes

When the crystals were taken to Japan the space group and cell dimensions of the crystals had not been assigned. Because of the problems encountered in determining the correct space group 180° of data were collected using an

orthogonal cell (a = 82.2 Å, b = 82.5 Å, c = 120.4 Å, $\alpha = \beta = \gamma = 90°$) of unknown space group. At the time of collection it was hoped that the correct space group could later be determined and the relationship of the true cell to the cell used for data collection would become clear. This was indeed the case and once the relationship of the triclinic (P1) cell to the unconventional (I2) cell and to the real (C2) cell was determined it was found that the synchrotron data had actually been collected about diagonals of the unconventional I2 cell.

Following elucidation of this relationship it became possible to index the image plate data given that initially the crystal had been aligned with the c axis parallel to the incident X-rays and the I2 a/b diagonal along the spindle axis. Pairs of horizontal plates were used to collect the data allowing all reflections at the edge of the sphere (radius = 1.8 Å) to be recorded.

Processing the data

The image plate data were processed using the program *Weis* after the position of the fiducial spots had been recorded from each image (Section B1.3.2). Essentially *Weis* orientates each image using the fiducial marks and then a number of medium-strong reflections are chosen which are used to refine the orientation matrix parameters. Initially data to 10.0 Å was used for refinement and then once the missetting parameters, DR and DS, had reached a minimum the resolution of the data used for refinement was progressively extended to 2.5 Å. The first image to be processed was the lower plate containing the zero order layer line of reflections. The setting file obtained from this image was then modified for all of the subsequent lower images. Upper images do not have fiducial spots therefore in these cases the setting file obtained from the corresponding lower image was modified and used for refinement. Although many of the images were processed without manual intervention the processing of some images required a considerable amount of manual manipulation before a setting file was obtained that could accurately predict the position of all the spots. Once the orientation matrix had been determined the intensity of each reflection was measured using a profile fitting method (Section B1.3.2). An integration sphere with a radius of 1.8 Å was used although only a low proportion of the spots between 1.8 Å and 2.0 Å were measurable. Details of the data processing statistics are given in Table 15. These statistics demonstrate that in general most of the images processed well and in most instances a large proportion of the expected reflections were observed. It was found that in order for the data from individual images to be successfully merged with the data from other images the

RMS/BGD value should be as low as possible and DR and DS should be similar, and also as low as possible.

Table 15 Data processing statistics for the FeLf$_N$ image plates.

ω range represents the oscillation range for which data were collected for each image. DR and DS give a measure of missetting of the centre of each reflection. DR is a measure of the horizontal displacement and DS is a measure of the vertical displacement. RMS/BGD is the ratio of the standard error to the background level of density.

Image plate	ω range	% expected reflections observed	number of whole reflections	DR	DS	RMS /BGD
1 top	-3.4° to 12.1°	90.7%	3426	0.53	0.60.	0.2
bottom		86.9%	3103	0.78	0.49	0.6
2 top	11.6° to 27.1°	87.2%	3334	0.48	0.46	0.2
bottom		93.5%	3590	0.49	0.5	0.3
3 top	26.6° to 42.1°	88.8%	4076	0.4	0.28	0.3
bottom		91.9%	3571	0.5	0.52	0.3
4 top	41.6° to 57.1°	90.0%	4003	0.4	0.33	0.5
bottom		85.5%	3908	0.39	0.30	0.1
5 top	56.6° to 72.1 °	86.0%	3904	0.34	0.37	0.2
bottom		84.8%	3119	0.43	0.31	0.4
6 top	71.6° to 87.1°	91.9%	4031	0.4	0.52	0.2
bottom		73.4%	3434	0.41	0.55	0.4
7 top	86.6° to 102.1°	88.1%	3585	0.57	0.58	0.3
bottom		86.6%	3916	0.26	0.56	0.2
8 top	101.8° to 117.1°	81.8%	3310	0.4	0.45	0.3
bottom		78.8%	3588	0.45	0.61	0.2
9 top	115.9° to 131.4°	79.0%	3073	0.57	0.34	0.2
9 top		90.0%	3827	0.61	0.61	0.4
bottom		85.2%	3836	0.57	0.37	0.7
10 top	131.7° to 147.2°	88.8%	3827	0.37	0.41	0.4
bottom		84.6%	3878	0.57	0.24	0.4
11 top	145.6° to 161.1°	84.1%	3541	0.53	0.56	0.4
bottom		84.0%	3849	0.39	0.34	0.2
12 top	164.5° to 179.5°	91.3%	4444	0.44	0.31	0.1
bottom		78.0%	3955	0.28	0.29	0.4
13 top	179.0° to 194.5°	91.9%	4427	0.34	0.35	0.2
bottom		Not processed due to alignment problems				

Merging the image plate data

After the first 12 images had been processed an attempt was made to scale and merge these in space group I2. Initially a high merging R was obtained even after over 20% of the reflections had been rejected. This was of some concern until it was realised that the h and k indices of these reflections needed to be inverted before the data was converted to the unique area of space. Once the h and k indices had been inverted and the scale factors recalculated a merging R of 6.3% was obtained after 11% of the 23692 reflections had been removed. Most of the interfilm R values were similar although one of the images agreed badly. This image was reprocessed along with the thirteen remaining images. Once all of the images had been processed they were converted using *WeisdataJ* (Section B1.3.2) and merged (Section B1.3.3) in space group I2. A merging R of 6.2% was obtained after 14% of the 97123 reflections had been rejected. The final output file contained 34 556 unique reflections. This time all of the images were found to agree quite well and all of the interfilm R values were less than 0.10.

Once the image plate data had been successfully merged in space group I2 the data was transformed to space group C2 and merged with the transformed but unscaled CAD4 data shells. Scale factors for the 31 batches of data were calculated using 20 503 pairs of common reflections out of a total of 111 279 reflections. The scale factors were then applied using the following acceptance criteria

a/ accept as observed if $F_{obs}/sig > 2.0$

b/ accept all if $|F_{max} - F_{min}|$ div sig < 6.0

c/ reject lowest if $|F_{ave} - F_{min}|$ div sig > 3.0

d/ reject highest if $|F_{ave} - F_{min}|$ div sig > 3.0

e/ For two reflections reject the smaller if $|F_{max} - F_{min}|$ div sig > 4.0.

Unit weights were used in this round of scaling which meant that all of the reflections were given equal weight.

Using these criteria a merging R of 7.5% was obtained after 11.8% of the data had been rejected. The final data file produced contained 35 000 unique reflections. Analysis of the interfilm R factors indicated that no batches disagreed more than any other (ie. most interfilm R factors were between 0.0 and 0.10) indicating that the CAD4 data and the image plate data had merged together well.

Analysis of the C2 data set

Once the complete data set had been obtained the quality and completeness of it was assessed. When the number of observed reflections was compared with the number of expected reflections [calculated as described in Blundell and Johnson, (1976)] it was found that, if the highest resolution was taken to be 2.0 Å, greater than 99% of the reflections had been measured. In Fig. 49 the percent of the observed reflections with I > 1σ, 2σ and 3σ is plotted against resolution. If the maximum resolution is taken to be 2.0 Å in total 93% of the reflections have I > 1σ and 80% have I > 2σ indicating that not only was the data set complete but also the data was of very good quality.

Fig. 49 Plot of resolution vs % of observed data with I greater than the specified sigma levels.

In this instance resolution is described in terms of sinθ/λ. The position equivalent to a resolution of 2.0 Å is indicated by an arrow.
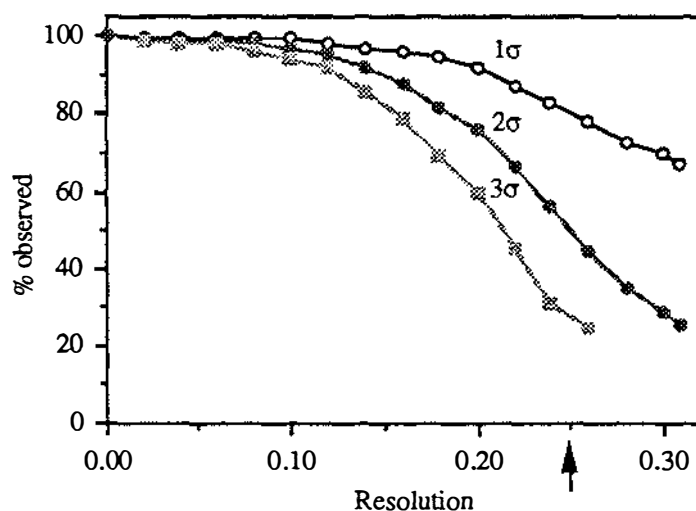
B3.2.6   Structure solution in space group C2

Following conversion of the CAD4 data to space group C2 and inclusion of the high resolution synchrotron data the position of the molecule in the new asymmetric unit was determined. Theoretically the position could be calculated and the coordinates of molecule A, in P1, transformed into the new cell. However, as the transformation required was not immediately obvious and to eliminate all possibility of error the position of the molecule in C2 was determined by repeating the rotation and translation searches (Section B1.5). ALMN (Section B1.5.1) gave a cluster of several closely positioned peaks all 17 standard deviations higher than the second cluster of peaks. The existence of

several peaks was accounted for by the fact that peaks are known to become spread when the rotation angles approach 0° or 90°. The solution obtained here involved a rotation of 0°, 65.3° and 90° about α, β and γ respectively. The model structure was rotated by these angles and then used to solve the translation function. In space group C2 the y coordinate of the origin is indeterminate and can have any value. In this instance y was fixed at 0.0 and then the required shifts on x and z were calculated by carrying out an R factor search in 1.0 Å steps along both axes (Section B1.5.2). As the entire unit cell was searched four equally sized minima (R factor = 0.318 compared to the average background value of 0.45) were obtained corresponding to the four different origins in the unit cell. The position of the four minima were refined (Table 16) using a step size of 0.1 Å.

Table 16 Translation function results for space group C2.

Molecule A from space group P1 was rotated (0°, 65.3°,90.0) and then used as the search model. Data between 20.0 Å and 3.5 Å were used.

Peak	Position in Å			R factor
	x	y	z	
1	53.03	0.0	33.83	0.30
2	53.03	0.0	4.68	0.30
3	119.38	0.0	4.68	0.30
4	119.38	0.0	33.83	0.30

B3.2.7   Refinement of the FeLf$_N$ structure in space group C2

Once the rotation and translation functions, using molecule A (from P1 refinement) as the model, had been solved for space group C2 six rounds of refinement (Section B1.6) and manual rebuilding (Section B1.7) were carried out. The course of refinement is summarised in Table 17.

Typically each round of refinement involved several cycles with tight weights followed by several cycles with loose weights and then the weights were tightened slowly until the geometry approached ideality. When tight weights were used the geometry was restrained whereas when loose weights were used the geometry was allowed to relax permitting the model to move to fit the X-ray data. The inclusion of a round of loose weights is important as models can become 'stuck' in local minima which may not be correct and only very loose weights will allow movement to occur. When temperature factors (B values) were refined (Section B1.6) the change in B values was restrained to, at most,

Table 17 Progress of FeLf$_N$ refinement in space group C2.

The number in parentheses refers to the number of cycles of temperature factor refinement.

* The Rfactor was computed by the program RFACTOR (TNT refinement package).

Round of refinement	Number of cycles #	Resolution limits (Å)	R factor*	rms bond lengths	rms bond angles
1 start	20	8.0 - 2.5	0.349	0.024	3.5
finish			0.283	0.032	3.72
2 start	29 (3)	8.0 - 1.8	0.347	0.033	3.94
finish			0.274	0.021	2.9
3 start	31 (4)	8.0 - 2.0	0.261	0.024	3.3
finish			0.22	0.022	2.83
4 start	40 (4)	8.0 - 2.0	0.231	0.028	3.23
finish			0.207	0.021	2.7
5 start	42 (4)	8.0 - 2.0	0.226	0.022	2.85
finish			0.194	0.017	2.856
6 start	26 (4)	8.0 - 2.0	0.203	0.019	2.93
finish			0.184	0.018	3.03

four times the average shift. Initially the B values were also restrained to lie between 10.0 Å^2 and 50.0 Å^2 but in round 5 the maximum allowable B value was increased to 80.0 Å^2 and then in round 6 the permitted range was extended to include B values from 5.0 Å^2 to 90.0 Å^2. The initial limits were included because the B values were otherwise found to increase dramatically, after which it was difficult to rescue them once the position of the atom had improved.

At the end of each round of refinement $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ maps were calculated and the position of all atoms inspected (Section B1.7). Side chains and water molecules were included as density became visible. The models used in each round of refinement are summarised in Table 18.

Criteria for the inclusion of additional sidechains

During refinement thirteen of the sidechains which were absent from the initial model were included. These were located using a $2F_{obs} - F_{calc}$ map contoured at 0.5σ and manual fitting on the interactive graphics system (Section B1.7). After their inclusion and the completion of another round of refinement the positions of the new sidechains were examined. At this stage maps contoured at 1σ were

Table 18 Summary of the model used for refinement in space group C2.

Round	Residue range refined	Number of sidechains absent or incomplete	Number of atoms in the model	Total number of water molecules in the model	First shell waters	Second shell waters
1	4 - 312	21	2298			
2	4 - 320	18	2396			
3	4 - 322	15	2429	83	78	5
4	4 - 325	8	2591	113	107	6
5	4 - 327	8	2653	158	144	14
6	4 - 327	8	2675	180	160	20

used and the sidechains were adjusted so that they fitted the density correctly. In several cases no 1σ density was visible and these sidechains were once again removed from the model structure. In addition the position of all the other residues was inspected at the end of each round of refinement and any sidechains that did not appear to fit the density properly were rebuilt. Particular attention was paid to those residues listed at the end of refinement as having bad bond angles or bond lengths. Atoms involved in bad contacts were also adjusted in an attempt to improve their position. The eight residues which are incomplete in the final structure are Arg 4, Lys 38, Lys 73, Glu 85, Arg 86, Asn 137, Thr 139, and Asp 315. With the exception of Asp 315 all of the incomplete residues are located on external loops which are also poorly defined in the structure of Fe_2Lf. It is most likely that these sidechains are quite mobile as they are on the surface of the molecule and consequently it is unlikely that their position can ever be accurately determined.

Positioning residues 313 - 333

In the initial stages of refinement residues 313 - 333 were excluded from the model because continuous density for these residues was not visible. However, at the end of round one continuous density could be seen beyond residue 312 when a $2F_{obs} - F_{calc}$ map contoured at 0.5σ map was used. The position of residues 312 - 320 were recognised in this map using, as a model, the coordinates for Fe_2Lf (N-lobe) which had been reorientated to match those of the $FeLf_N$ structure. Very little manual rebuilding was required to fit the Fe_2Lf structure for residues 313 - 320 to the $FeLf_N$ map, indicating that these residues had

essentially the same conformation as in the structure of Fe$_2$Lf. The sidechains for 315, 318 and 320 were left out at this stage as the density for them was not well defined. At the end of the next round of refinement density could be seen for the sidechain of Leu 320 and for the next two residues, Gly 321 and Ser 322. Residue 320 was repositioned and residues 321 and 322 were included at this stage. Although only two more amino acids had been included it was clear that the remaining residues must have a different orientation to that found in Fe$_2$Lf.

Interpretation of the map beyond residue 322 was ambiguous at this stage because of (i) the poor quality of the density and (ii) the fact that the density lay close to the crystallographic two fold axis, and thus the density for this same region of a symmetry related molecule was near by. In an attempt to determine the position of subsequent residues a minimap (Section B1.7), contoured at 0.5σ, 0.75σ and 1.0σ was calculated. Unfortunately this did not offer any clearer resolution and in fact the density appeared to branch beyond 322. Another round of refinement was then completed with residues 4 to 322 in the model. The presence of Gly 321 and Ser 322 in the model appeared to improve the map and it was now possible to include Gly 323, Tyr 324 and Phe 325 at the end of round 3. The inclusion of these residues was greatly aided by the fact that two of the amino acids included had aromatic sidechains, for which the density was clearer. In the absence of these two residues it would have been difficult to be confident about their exact orientation as the main chain density was poorly resolved. Two more residues, 326 and 327, were included at the end of the next round of refinement. Elucidation of the position of residues 320 to 327 has proved most intriguing as residues 321 - 323 are part of helix 10 while the remainder form an extended strand in this structure. This is in distinct contrast to the native structure where the last residue of helix 10 is Gly 321 and residues 323 to 333 form helix 11 The implications of these differences are discussed in section B3.2.7.4.

Solvent structure

After the first two rounds of refinement had been completed solvent molecules were included in the model. All of the solvent molecules included in the structure were treated as water molecules. Water molecules were located using a difference map with coefficients $F_{obs} - F_{calc}$ together with a $2F_{obs} - F_{calc}$ map. They were only included in the model if (i) the density was well defined; (ii) the peak height was greater than 3σ in a difference ($F_{obs} - F_{calc}$) map and (iii) there were potential hydrogen bonding partners within 3.5 Å. Using these criteria 83 water molecules were included at the end of round 2. Most of these water

molecules made good contacts with several different protein atoms and consequently had low B values when refined. Additional water molecules were included in each subsequent round of refinement as density became visible. After each round of B value refinement the thermal parameters of all the water molecules were checked and any that had B values greater than 50.0 Å^2 were omitted. Following the last round of refinement no waters were removed even though a few had high B values; therefore some of the water molecules in the final structure may not be accurately positioned. Although rigorous analysis of the water structure has not been carried out most of the waters included so far make at least one hydrogen bond to a protein atom and are not likely to be artefactual. Of the 20 second sphere waters included many are located in the cleft and have more than one hydrogen bonding partner and are also likely to be "real". See section B3.2.7.6 for a further discussion.

B3.2.8   Quality of the refined model

Accuracy of the structure

The structure of FeLf$_N$ has been determined and refined using data between 8.0 and 2.0 Å. The current model has an R factor of 0.184 and the *rms* deviations for the bond lengths and bond angles are 0.018 Å and 3.03° respectively. Apart from residues 1 to 4, 328 to 333 and the eight sidechains which have been omitted (Section B3.2.7) the structure is represented by well defined continuous electron density in the final $2F_{obs} - F_{calc}$ electron density map, when contoured at a level of 1σ. Two examples of the fit of the structure to the density are shown in Fig. 50.

In an attempt to quantitatively estimate the error in the structure the mean error of the coordinates was estimated from the variation in R factor with resolution as described by Luzzati (1952). The Luzzati plot obtained is shown in Fig. 51. Reference to the standard error lines allows the error in the structure to be determined. Except at low resolution, where the higher values can be attributed to a failure to fully account for the solvent structure, the error in the structure is ~0.25 Å. The value obtained from the graph actually represents the maximum average error and in fact the error is probably considerably less in well defined internal regions.

Fig. 50 Examples of the fit of the atoms to the density.

In both cases the $2F_{obs} - F_{calc}$ map has been contoured at 1.5σ. Fig. 50a shows a region of β-sheet and Fig. 50b shows the environment of Arg 210.

Fig. 50a
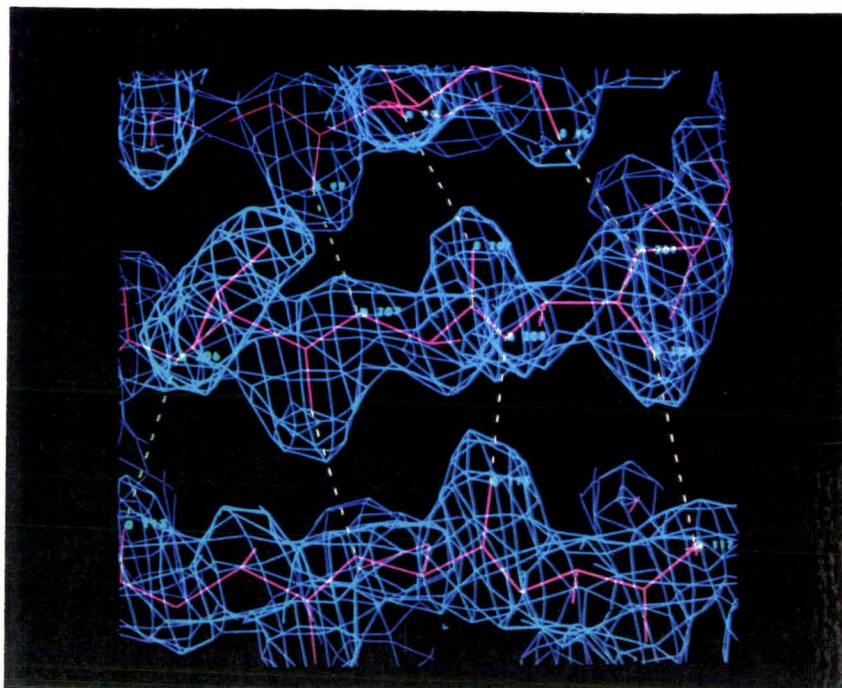
Fig. 50b
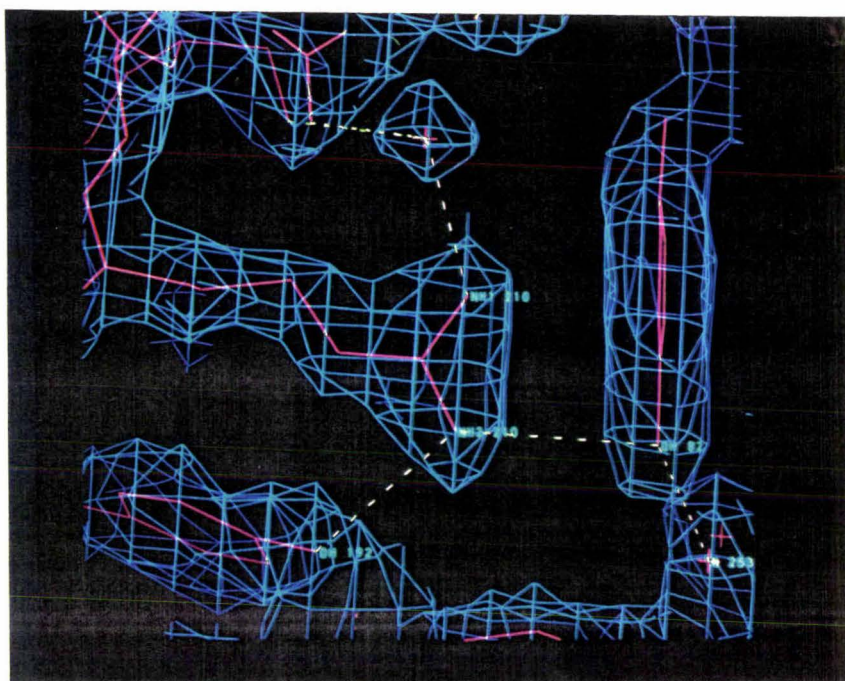
Fig. 51 Luzzati plot for the refined FeLf_N structure.

The position of the error lines equivalent to 0.1 Å, 0.2 Å, 0.3 Å and 0.4 Å are indicated. The resolution is defined in terms of sinθ/λ.
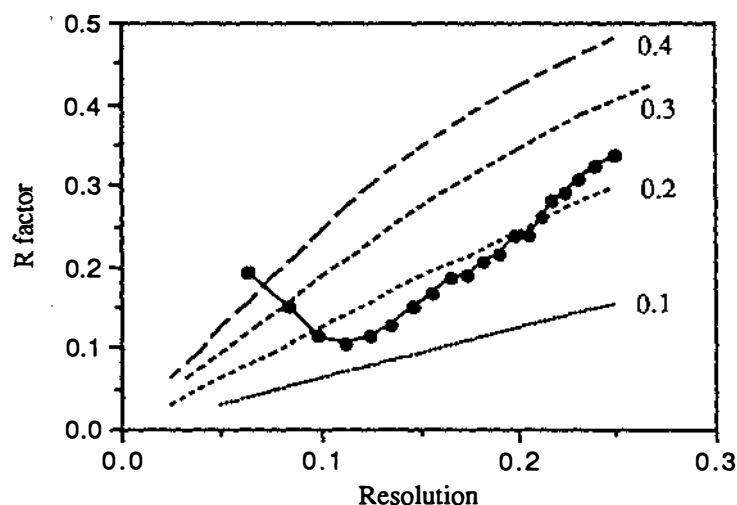
The fit of the atomic structure to the density was quantitated by calculating real space correlation coefficients using the program O (Jones *et al*, 1991). In general the correlation of the model with the data is very high (Fig. 52) with most of the mainchain atoms having correlation coefficients greater than 0.9. There are however some regions where the correlation of the structure to the density is reduced. In particular the mainchain correlation coefficients fall below 0.8 for residues 139 - 143, 260 - 261, 293, 313 - 314 and 326 - 327. The correlation coefficients for the sidechains follow a similar pattern to that seen for the mainchain except they are on average ~0.1 lower. Most of the residues that have low correlation coefficients are on the surface of the protein in more poorly defined, and probably more mobile regions, while the residues which have high correlation coefficients are found in internal regions. A similar trend is seen when the average mainchain temperature factors are plotted against residue number (see below).

Fig. 52 Plot of the mainchain real space correlation coefficients for FeLf$_N$.
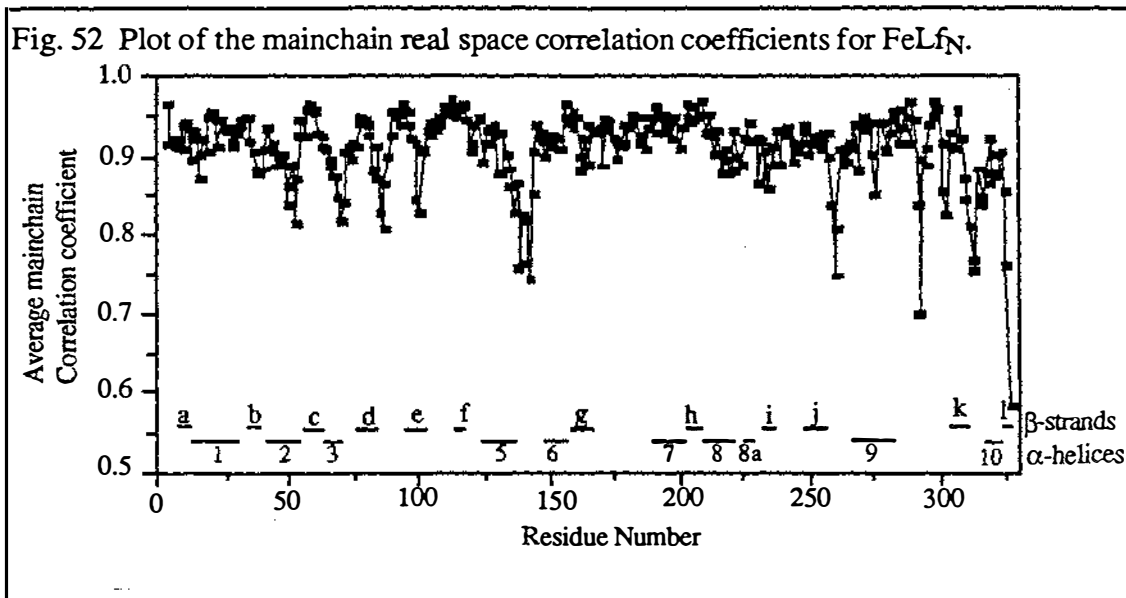
Thermal parameters

Analysis of the thermal parameters (B values) for the refined FeLf$_N$ structure has revealed that the distribution of high and low B values in FeLf$_N$ is similar to that found in the N lobe of Fe$_2$Lf although the average mainchain B value is 15 Å^2 higher for the N-lobe of Fe$_2$Lf (Table 19). Several factors probably contribute to the large difference in the average B values. Firstly the two structures have been refined using different programs and the number of cycles of B value refinem nt differs (19 cycles for FeLf$_N$ and >40 cycles for Fe$_2$Lf). In addition the B values of FeLf$_N$ have been restrained to lie between 5.0 Å^2 and 90.0 Å^2 and shifts of no greater than four times the average shift have been permitted. Although these factors may have contributed to the lower average B value in FeLf$_N$ it is most likely that the main reason is simply that the crystals of FeLf$_N$ are more ordered than those of Fe$_2$Lf. This assumption is supported by the fact that very few of the mainchain B values in FeLf$_N$ approach the limits mentioned above and even the average B value for the water molecules in FeLf$_N$ is lower than the average B value for the protein atoms in Fe$_2$Lf.

Table 19 Comparison of the average B values for FeLf$_N$ and Fe$_2$Lf.

Model structure	Average B value (Å^2)
Complete FeLf$_N$ structure	34.9
FeLf$_N$ protein only	34.1
FeLf$_N$ water only	46.4
N-lobe of Fe$_2$Lf; protein only	50.0

In Fig. 53 the average mainchain B value for each residue is plotted against residue

number and the regions of secondary structure are indicated.

Fig. 53 Analysis of the mainchain B values for the refined FeLf$_N$ structure.
The regions involved in α-helices and β-sheets are indicated by lines along the bottom of the graph.
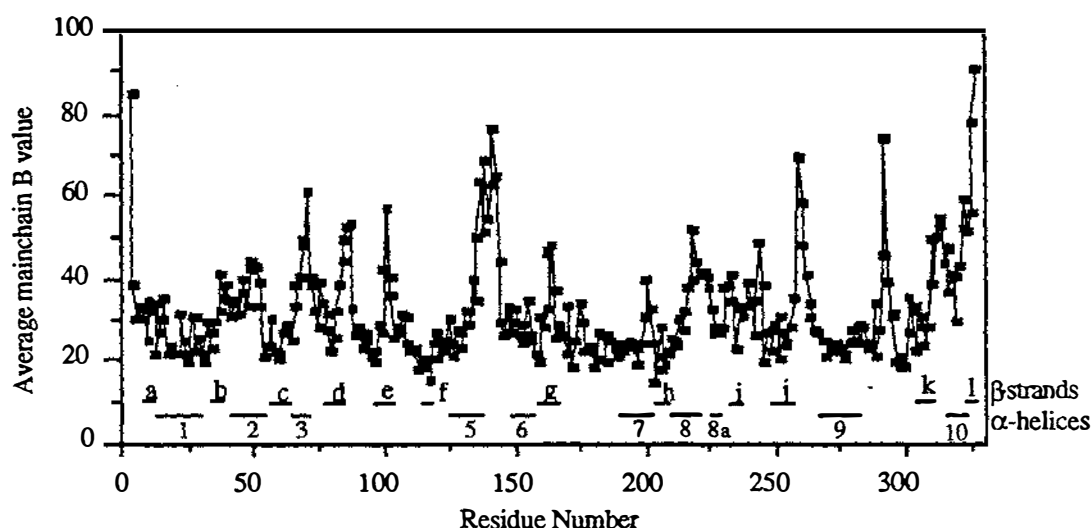
From this plot it is clear that most of the residues involved in α-helices and β-sheets have lower B values and that most of the residues with higher B values occur in connecting loops on the surface of the structure. While many of the surface residues do have higher B values indicating their greater mobility the course of the polypeptide chain is in no doubt. The difference in B values between a well defined internal residue and a more mobile external residue is clearly demonstrated by comparison of Lys 18 and Lys 100. Lysine 18 is internal, with an amino group which makes four potential hydrogen bonds, and has an average B value of 21 Å^2. In contrast Lys 100 is on the outside of the molecule, its sidechain only makes a single hydrogen bond with neighbouring atoms, and it has an average B value of ~52 Å^2. The position of both these residues and the contacts made by them are shown in Fig. 54.

Dihedral angles

The correctness of the model structure was also assessed by analysing the distribution of the dihedral angles. In Fig. 55 a plot of the mainchain torsion angles ϕ and ψ is shown, with the commonly occupied regions of conformational space indicated. Several features of the structure are highlighted by this Ramachandran plot (Ramakrishnan and Ramachandran, 1965). Firstly an estimate of the amount

Fig. 54 Comparison of the density and interactions for Lys 18 and Lys 100.
Lys 18
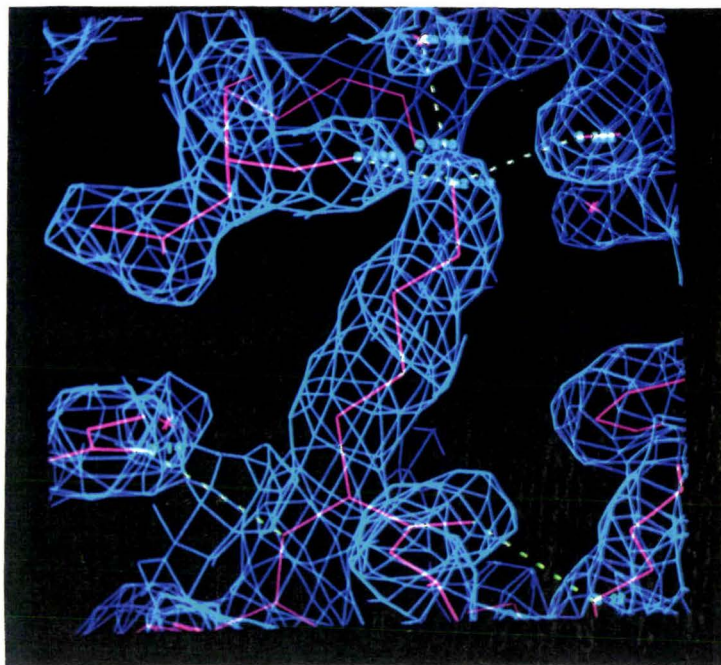
Lys 100
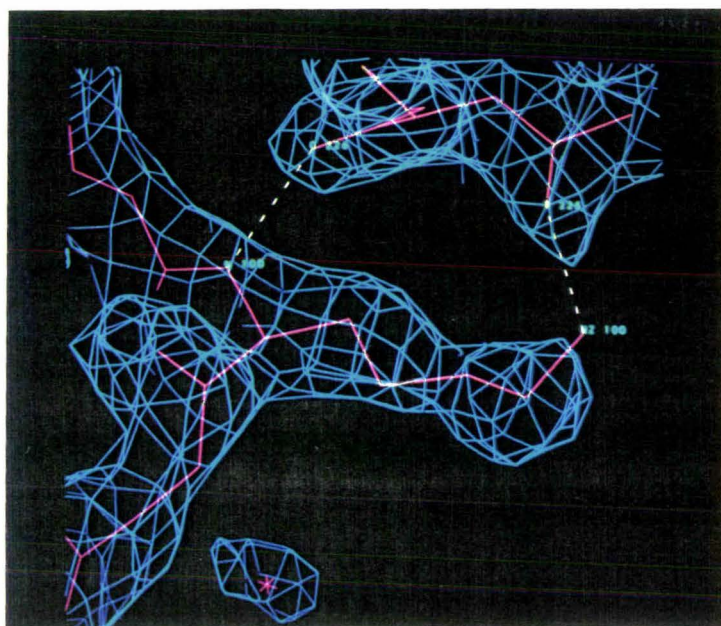
of α-helix and β-sheet in the structure can be obtained from the plot. The presence
of both secondary structural elements in FeLf$_N$ is reflected by the concentration of
residues around -60°, -40° (α-helix) and -120°, 135° (β-sheet). The helical torsion
angles are very tightly clustered. If the terminal residues in the helices are omitted
the mean φ and ψ angles are -64 ± 8° and -40 ± 7° respectively, as in other highly
refined structures (Baker, 1980). In addition the 22 proline residues have a mean φ

angle of -66 ± 11° which is close to the value of -65 reported by Morris *et al.*
(1992). The similarity of the geometry to that obtained for other highly refined
structures provides another means by which the structure can be assessed as the
mainchain torsion angles were not restrained during refinement.

Although most of the residues are located on the left hand side of the plot within the
acceptable regions there are several non-glycine residues on the right hand side of
the plot which must be accounted for. Those with angles around 60°, 30° represent
residues with a left handed α-helical configuration. The two apparent
conformational violations in the lower right quadrant correspond to Ser 191 and
Leu 299. Both residues, however, have similar φ and ψ angles in Fe_2Lf and have
well defined density in $FeLf_N$. Serine 191 lies at the N-terminus of helix 7 which
contains the ligand Tyr 192 and has a similar configuration (70, -175) to the active
site serine residues in lipase and hydrolase structures (Derewenda and Derewenda,
1991). The position of this residue is actually within an acceptable region if the
acceptable regions are defined in terms of observed conformations rather than
theoretically expected conformations (Morris *et al.*, 1992). Leucine 299 is the
central residue of a γ-turn (298 O-N 300). It has torsion angles of (74,-51) making
it a member of the class of γ-turns first observed in thermolysin (Matthews, 1972).

Fig. 55 Ramachandran plot for residues 4 to 327 in $FeLf_N$

The preferred regions, as reported by Ramakrishnan and Ramachandran (1965), are indicated by
broken lines and glycines are marked by a square rather than a cross.
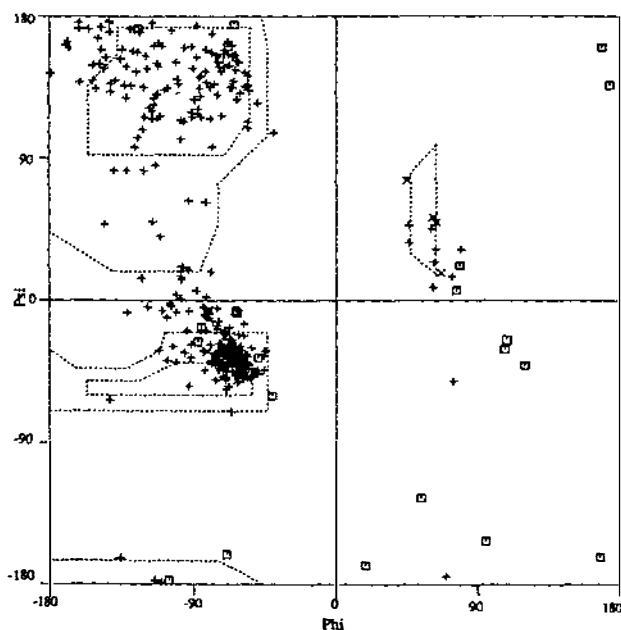
3.2.9   Analysis of the refined structure

3.2.9.1   General topology

The general organisation of FeLf$_N$ is very similar to that seen in the N lobe of Fe$_2$Lf. In both structures the polypeptide chain is folded into two similarly sized domains between which is bound an iron atom and a bicarbonate ion. The folding pattern and a ribbon diagram of FeLf$_N$ are shown in Fig. 56. A stereo view of the Cα trace is shown in Fig. 57. The current coordinates of FeLf$_N$ are provided on disc (Appendix 6) and the amino acid sequence of FeLf$_N$ is given in Appendix 7.

. 56 FeLf$_N$: ribbon and schematic diagrams.

56a This diagram was prepared using the program Ribbon (Richardson, 1985; Priestle, 1988). The α-helices depicted as spirals (pink), the β-strands as arrows (teal) and the iron atom is represented as a pink sphere. The rbonate ion and the water molecules are not shown. In Fig. 56b the helices are shown as cylinders and β-strands rrows. The α-helices and β-strands are numbered according to the scheme given in Table 20. The position of protein ligands, the iron atom and the anion, are indicated.

. 56a   Fig. 56b
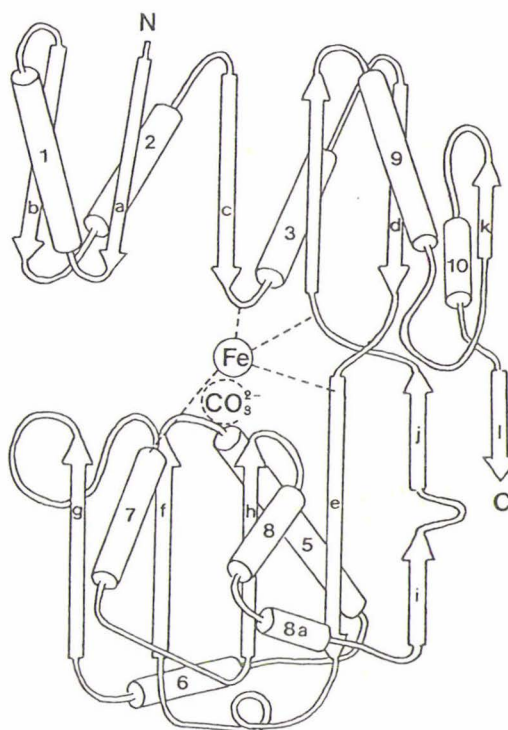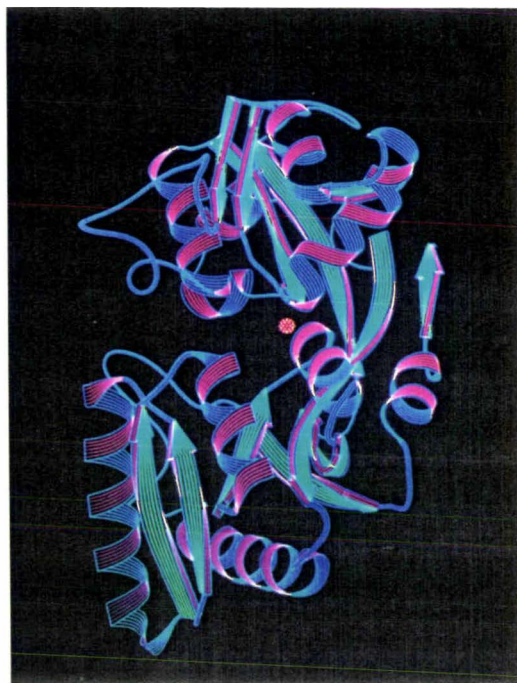
The residue ranges of the α-helices and the β-strands in FeLf$_N$ (Table 20) were determined by examining the hydrogen bonding pattern and the dihedral angles. In

Fig. 57 Stereoview of a Cα plot of FeLf$_N$.

The position of the iron atom is indicated and the disulphide bridges are shown as thickened lines. Approximately every twentieth residue is labelled.
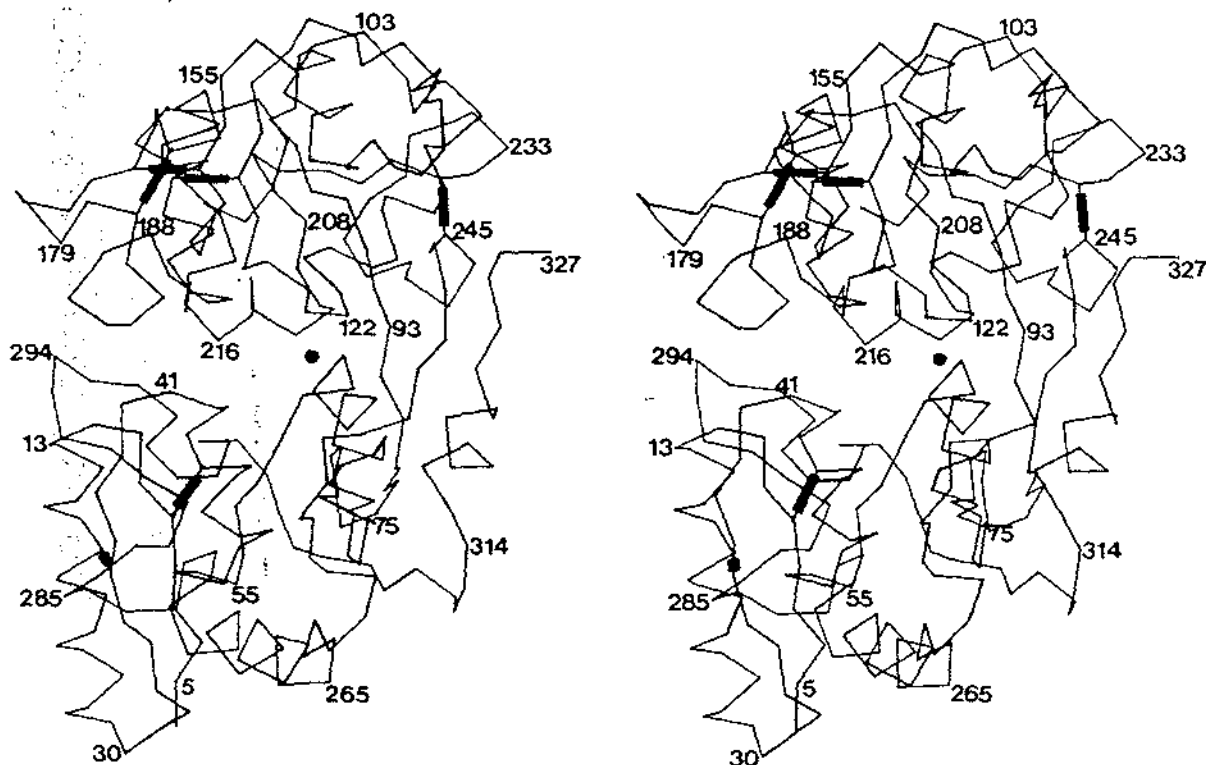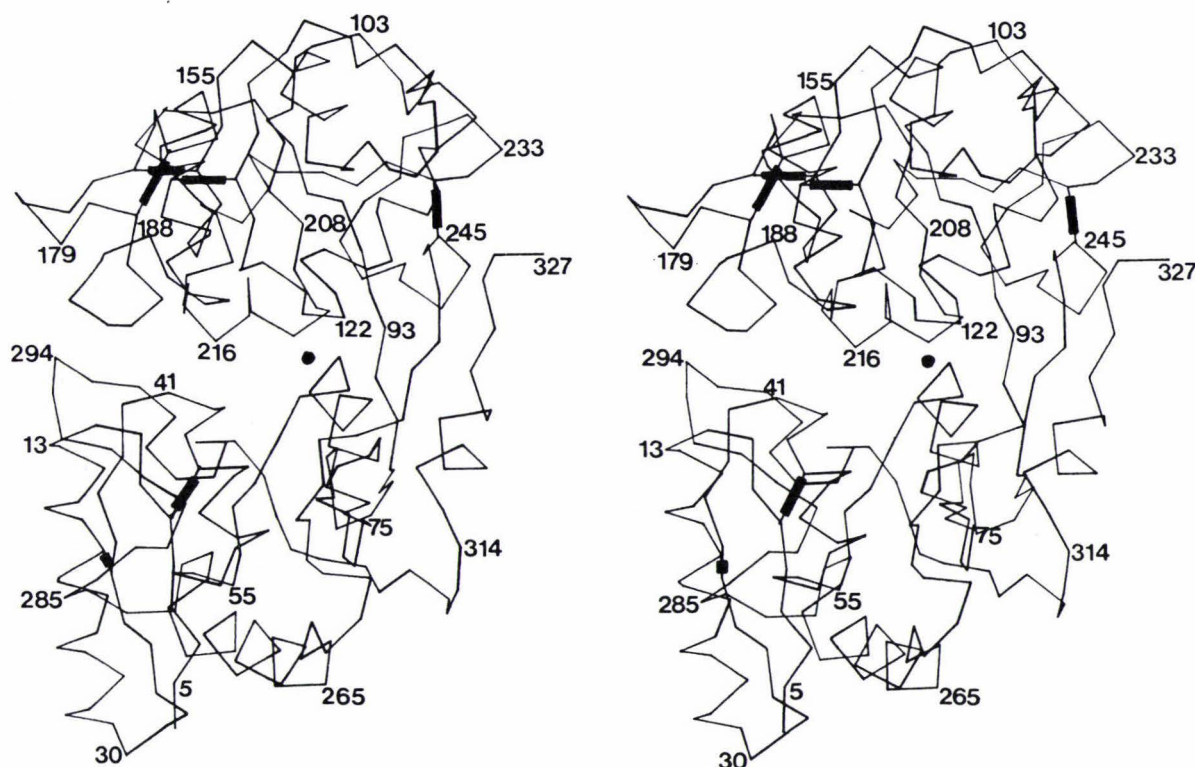
general the positions of the helices and the sheets were found to be very similar to those found in the N-lobe of Fe$_2$Lf (Anderson *et al.*, 1989) but several exceptions do exist. Firstly residues 105 - 110 form helix 4 in Fe$_2$Lf but, while roughly helical in FeLf$_N$, neither the hydrogen bonding pattern nor the dihedral angles allow them to be categorised as a helix in FeLf$_N$. The second change in the assignment of the α-helices and β-strands involves residues 315 to 333. In Fe$_2$Lf these residues form two α-helices (315 - 321 and 322 - 333) numbered 10 and 11. However, in FeLf$_N$ helix 10 is elongated and now includes residues 315 to 323 while the remaining residues, 323 - 327, form a piece of extended chain (Fig. 58). In Fe$_2$Lf the long α-helix formed by residues 322 - 333 runs along the back of the iron binding site making contacts with the N2 domain and the C-lobe. The extended strand in FeLf$_N$ has a different orientation and makes different contacts; in particular it hydrogen bonds to residues 246 - 248 (part of β-strand j). The presence of this extra strand changes the nature of the connections between the two domains. Instead of two β-strands and a helix, as in Lf, the domains are now connected by three β-strands. The contacts made and implications of the changes will be discussed more fully in section B3.2.9.3

Table 20 Summary of the α-helices and β-sheets in FeLf$_N$.

α-helices		β-sheets	
Name	Residue range	Name	Residue range
1	12 - 30	a	5 - 10
2	41 - 53	b	32 - 39
3	60 - 69	c	54 - 59
5	121 - 136	d	74 - 82
6	144 - 153	e	91 - 100
7	191 - 203	f	112 - 116
8	212 - 218	g	153 - 158
8a	222 - 226	h	204 - 211
9	263 - 279	i	226 - 231
10	315 - 322	j	246 - 258
		k	304 - 310
		l	323 - 326

Fig. 58 Conformation of residues 316 to 327 in Fe$_2$Lf and FeLf$_N$.
The FeLf$_N$ chain is shown in pink and the Fe$_2$Lf chain in blue. Residues 315 - 320 have a similar conformation in both structures while residues 320 - 327 have quite different configurations.
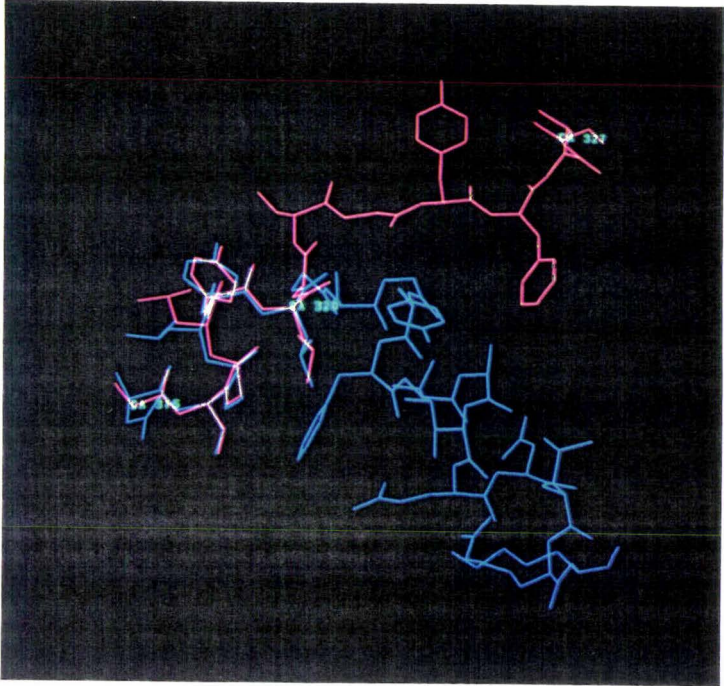
The overall similarity of the FeLf$_N$ and Fe$_2$Lf (N-lobe) models was further analysed by superimposing the two structures and then calculating the *rms* deviation of the atoms (Table 21).

Table 21 Summary of the *rms* deviation between FeLf$_N$ and Fe$_2$Lf.

In each case the residue ranges listed were used to superimpose FeLf$_N$ and Fe$_2$Lf and then the *rms* deviation was calculated. * Atoms which differed in position by more than 1.0 Å between the two structures were omitted (~20).

Residue range	Atoms included	*rms* deviation
4 - 312	Mainchain	0.45 Å (0.40)*
4 - 312	Cα only	0.41 Å
92 - 250	Cα only	0.28 Å
4 - 91 and 251 - 312	Cα only	0.39 Å

In general the models agree quite well, the *rms* deviation of the two mainchains being only 0.45 Å and 0.40 Å if the twenty atoms which differ in position by more than 1.0 Å are removed. If only the individual domains are superimposed and then the *rms* deviation is calculated it is clear that domain 2, with a deviation of only 0.28 Å, is the most conserved part of the structure. Although the deviation is greater for domain 1 the *rms* deviation of domain 1 is still less than the *rms* deviation when the intact lobes are compared. The lower *rms* deviation of the individual domains, compared to the whole lobe, is accounted for by the fact that the orientations of the two domains are slightly different in FeLf$_N$. A rotation of 2.7° is required to correctly orientate domain 1 after superposition of domain 2. If the two models are superimposed using only domain 2 it is clear that the binding cleft in FeLf$_N$ is actually slightly more open than in the N-lobe of Fe$_2$Lf (Fig. 59).

Although the rotation is only small the orientation of the two domains is clearly different and it is quite possible that this slight change in domain orientation may contribute to the more facile release of iron from FeLf$_N$ than from Fe$_2$Lf (Section A3.2.2.5). Analysis of the contacts made between the two domains reveals that while the interactions are very similar there are subtle differences. In particular, in Fe$_2$Lf the N1 and N2 domains are linked by two salt bridges, 216 OE1...NZ 301 and 217 OD2...NZ 296, which are both considerably weakened in FeLf$_N$. Both contacts are longer, being 2.6 Å compared with 3.0 Å and 3.3 Å compared with 3.7 Å, respectively. While the altered orientation of the domains may be a consequence of the conformational change at the back of the iron binding site and may contribute to the different iron release properties of FeLf$_N$ it is equally likely that the difference is a function of crystal packing. At this stage it is not possible to

choose between these two possibilities therefore the functional relevance of the weakened salt bridges is not known.

Fig. 59 Comparison of FeLf$_N$ and Fe$_2$Lf after superposition using only domain 2. The Cα plot of FeLf$_N$ (pink) and Fe$_2$Lf (N-lobe) (blue) is shown after the two models have been superimposed using the Cα atoms in only domain 2.
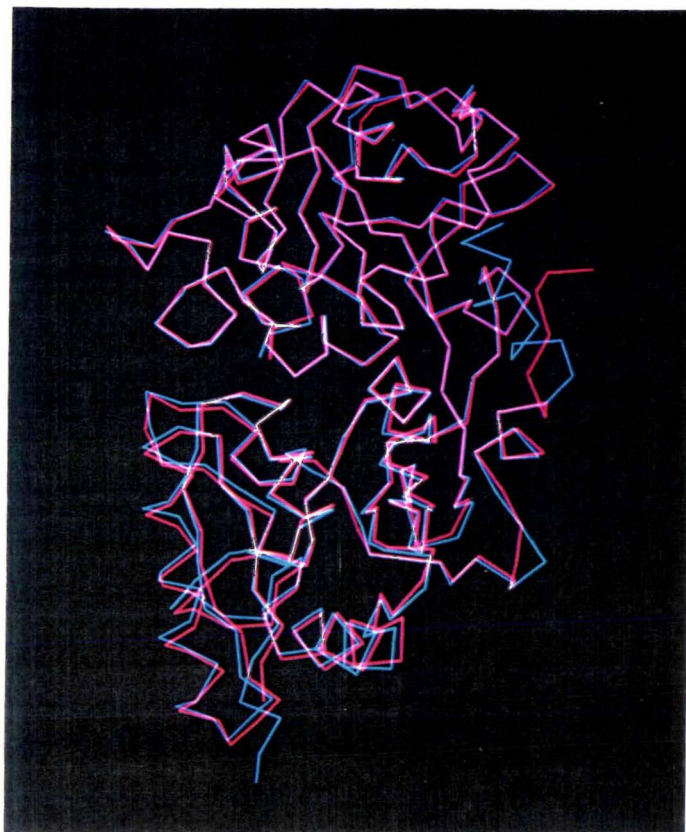
$.2.9.2   Analysis of the metal and anion binding sites in FeLf$_N$.

The iron binding site in FeLf$_N$ is very well defined and is essentially identical to that found in the N-lobe of Fe$_2$Lf. In both structures the iron atom is coordinated to four protein ligands through the sidechains of Asp 60, Tyr 92, Tyr 192 and His 253. The two remaining ligands are provided by the anion which is bound in a bidentate manner. The details of the metal/ligand bond lengths and angles are summarised in Table 22. In all cases the bond distances approach the ideal value of ~2.0 Å. However, while the iron atom is bound by six ligands the octahedral geometry is somewhat distorted by the presence of the anion which has a small (63°) chelate bite angle. All the other angles between the bonds are close to 90° or 180°.

Table 22 Geometry of the iron site.

Ligand	Ligand/Fe distance (Å)	Angle °	
Asp 60 OD1	1.99	60 OD1-Fe-OH 92	90
		60 OD1-Fe-OH 192	164
		60 OD1-Fe-NE2 253	87
		60 OD1-Fe-O1	87
		60 OD1-Fe-O2	94
Tyr 92 OH	2.00	92 OH-Fe-O1	160
		92 OH-Fe-O2	97
Tyr 192 OH	1.92	192 OH-Fe-NE2 253	86
		192 OH-Fe-OH 92	106
His 253 NE2	2.15	253 NE2-Fe-O2	167
		253 NE2-Fe-O1	104
Carbonate O1	2.04	O1-Fe-OH 192	80
		O1-Fe-O2	63
Carbonate O2	2.10	O2-Fe-OH 192	97

Table 23 Anion/protein hydrogen bonds.

Hydrogen bond	Length (Å)	Angle °	
O1....NE 121	2.90	121 NE-H....O1	132
O1....NH2 121	2.59	121 NH2-H....O1	143
O2....N 123	2.81	123 N-H....O2	151
O3....OG1 117	2.64	117 CB-OG1....O3	118
O3....N 124	3.31	124 N-H....O3	148

In Table 23 the hydrogen bonds made between the protein structure and the carbonate are summarised. The full bonding potential of the anion is satisfied by the interactions it makes with the iron atom, the sidechains of Thr 117 and Arg 121, and the N-terminus of helix 5, residues 123 and 124 (Fig. 60). All of the hydrogen bonds are close to linear and have lengths of between 2.6 and 2.9 Å except for the rather longer O3....NH 124 bond (Table 23). The similarity of the iron binding sites in $FeLf_N$ and Fe_2Lf (N-lobe) is indicated by the fact that the *rms* deviation of the six residues (117, 121, 60, 92, 192, 253) involved in metal and anion binding is only 0.41 Å and the iron/ligand distances differ by no more than 0.1 Å between the two structures, the probable error in the coordinates at the iron site. The two sites are superimposed in Fig. 61.

Fig. 60 The anion binding site: electron density and contacts.

The carbonate and the contacts made by this moiety are indicated. The $2F_{obs} - F_{calc}$ map shown here was contoured at 1.0σ.
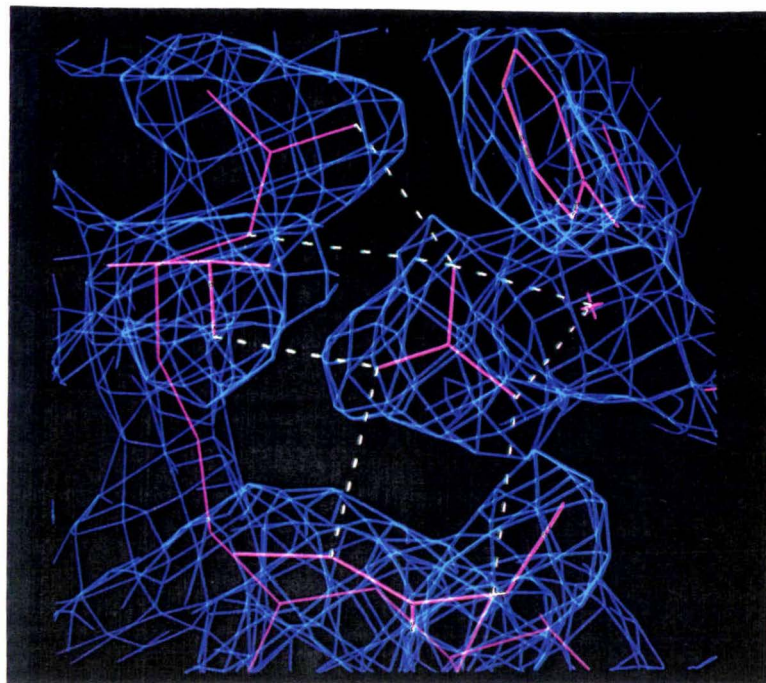
Fig. 61 Superposition of the $FeLf_N$ and Fe_2Lf iron binding sites.

$FeLf_N$ is shown in pink while Fe_2Lf is shown in blue. Where the two structures superimpose they appear white.
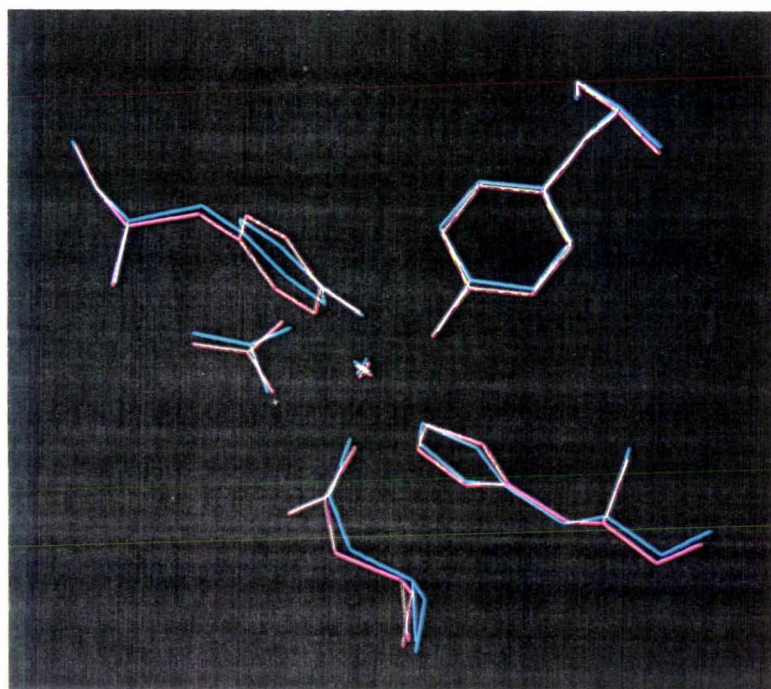
As well as the high level of similarity in the immediate vicinity of the iron atom the hydrogen bonds made by the ligands are also identical in both structures. These hydrogen bonds, shown schematically in Fig. 62, provide the majority of the interactions between the two lobes and help maintain the geometry of the ligands. Aspartic acid 60 appears to have an especially important role as it binds to the iron atom through one carboxylate oxygen, OD1, and through its other oxygen, OD2, receives hydrogen bonds from the peptide nitrogen and OG1 of Thr 122, in domain 2, and from the peptide nitrogen of Gly 62 in domain 1. Contacts between the two domains are also made by Arg 210 which hydrogen bonds to both Tyr 82 in domain 1, and Tyr 192 in domain 2 (Fig. 50b). Tyr 92 and His 253 do not make any hydrogen bonds with protein groups but they are both hydrogen bonded to well defined water molecules which are highly conserved in both structures (Section B3.2.9.4).

The environment of the metal binding site in Lf_N is also very similar to that found in the N-lobe of sTf (Sarra *et al.*, 1990). The *rms* deviation for residues 117, 121, 253, 60, 92, and 192 is only 0.43 Å. The similarity of the metal binding site in all three structures, Lf_N, sTf_N and Fe_2Lf, suggests that the iron binding and release properties of the 'transferrins' are not determined by the immediate environment of the iron atom. Instead it is more likely that changes at positions distant to the metal binding site are responsible for the different properties of the fragments and the intact molecules. However, slight differences in the environment of the iron atom are likely to be responsible for the shift in λ_{max} (Section A3.2.2.1). The aromatic ring of Tyr 192 is rotated slightly in $FeLf_N$ compared with Fe_2Lf. This could affect the position of λ_{max} since the visible absorption band arises from a Tyr-metal charge transfer interaction (Harris and Aisen, 1989). The difference in orientation is small, however, in relation to the likely error level in the two structures. Other smaller differences do occur but a higher resolution structure of both $FeLf_N$ and Fe_2Lf is required before the changes responsible for the shift in λ_{max} can be correctly identified.

Fig. 62 Schematic diagram of the hydrogen bonding network in and around the metal and anion binding sites of Lf$_N$.

The lengths of the hydrogen bonds are shown in Å. Residues belonging to domain 1 are enclosed by a circle and those belonging to domain 2 are enclosed by a square.
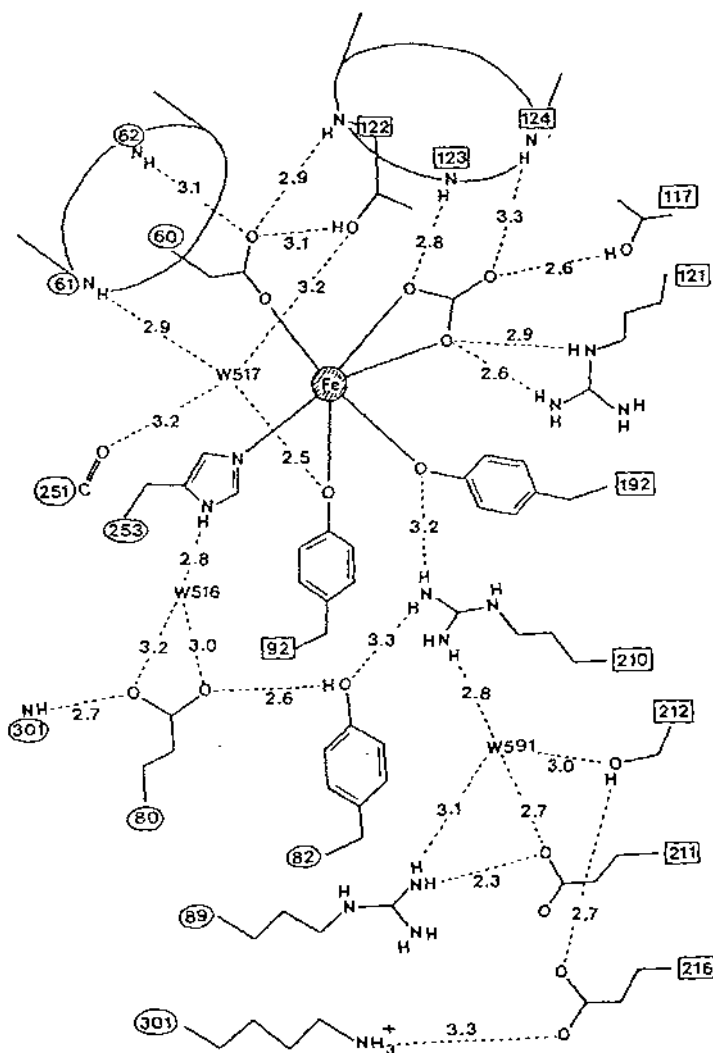
B3.2.9.3   Conformational changes in FeLf$_N$

Residues 312 - 333

As discussed in section B3.2.7 the position of residues 313 - 333 was not clear in the initial electron density maps, but after several rounds of refinement it became possible to position residues 313 - 327. The first residues included in the model were 313 - 320 which were added using, as a model, the coordinates for these

residues from Fe$_2$Lf. Residues 313 - 320 have been refined and rebuilt 5 times and the *rms* deviation of the mainchain atoms between the two structures is 0.7 Å indicating that while there are some differences the overall conformation is similar. Residues 321 - 333 proved more difficult to position and at this stage of refinement and rebuilding only residues 321 - 327 have been included in the model. As already discussed (Section B3.2.9.1) the assignment of helices and sheets in this region has changed. The difference in the conformation appears to originate from Gly 321 which is at the C-terminus of helix 10 in Fe$_2$Lf. In Fe$_2$Lf this residue has mainchain torsion angles of (90°, 152°) and makes an abrupt turn such that the angle between helix 10 and 11 is 90°. In FeLf$_N$ Gly 321 has approximately helical torsion angles (-62°, -8°) allowing helix 10 to continue as far as Gly 323, the last hydrogen bond being 320 O....N 323. From Gly 323 the polypeptide chain is extended and the residues have torsional angles characteristic of a β-strand. The torsion angles of residues 320 - 327 in FeLf$_N$ and Fe$_2$Lf are summarised in Table 24.

Table 24 Comparison of the mainchain torsion angles for residues 315 to 326. The residues involved in helices are shown in bold.

Residue Number	Fe$_2$Lf Dihedral angles		FeLf$_N$ Dihedral angles	
	$\phi°$	$\psi°$	$\phi°$	$\psi°$
315	-118	175	-78	169
316	**-54**	**-53**	**-53**	**-48**
317	**-49**	**-53**	**-40**	**-60**
318	**-66**	**-27**	**-72**	**-28**
319	**-74**	**-40**	**-62**	**-52**
320	**-70**	**-30**	**-71**	**-32**
321	95	152	**-62**	**-8**
322	**-65**	**-41**	**-63**	**-41**
323	**-55**	**-65**	-68	161
324	**-63**	**-54**	-153	139
325	**-54**	**-40**	-79	135
326	**-82**	**-30**	-39	106

Refinement of residues 321 - 327 is not complete but they do appear to fit the density quite well (Fig. 63) and there seems to be no doubt that they have been correctly placed. The correctness of their position is also supported by the B values and correlation coefficients for residues 321 to 326 which, although higher than

average, are still quite reasonable. Residue 327 is the only one which may have been incorrectly positioned.

Fig. 63 Residues 320 - 326; atoms and density.

The position of residues 320 to 326 and the corresponding electron desity, contoured at 0.75σ, is shown.
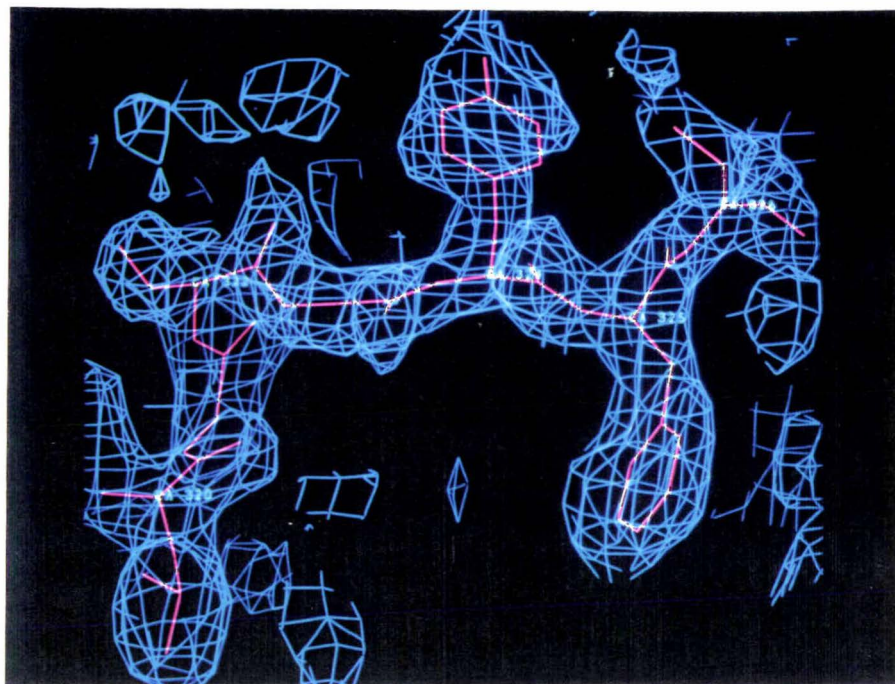
The change in the configuration of these residues is most interesting as it raises questions about the driving forces which allow these residues to form a helix in one structure and an extended strand in another structure. The apparent ease with which helix 11 unravels and becomes an extended chain suggests that this helix is partially unstable in Fe_2Lf, and the presence of the C-lobe stabilises the helical form. Secondly, and perhaps of more importance in this study, the position and contacts made by these residues in $FeLf_N$ may partially account for the different iron binding and release properties of $FeLf_N$. The changes seen in the structure of $FeLf_N$ are primarily due to two features (i) the absence of the C-lobe and (ii) repositioning of residues 322 - 327.

Removal of the C-lobe has several effects. Firstly the two salt bridges between the two lobes (Arg 313 - Asp 379 and Asp 315 - Lys 386) are no longer formed. Instead Arg 313 is reoriented so that both NH1 and NH2 can hydrogen bond with the mainchain oxygen of Ser 259. As the density for the sidechain of Asp 315 is poorly defined this residue has not been included in the model and the interactions made by this group are not known. The environment of the residues in the N-lobe which contributed to the hydrophobic interactions between the two lobes in Fe_2Lf

have also changed. Previously 311, 314 and 318 interacted with residues 688 - 689 from the C-lobe but now they are packed against residues 133 and 134 from a symmetry related molecule. In addition the environment of Phe 325 is different as it now points inwards, towards the N-lobe (see below) instead of packing against residues 322 - 325 from the C-lobe.

The unravelling of helix 11 and the repositioning of residues 321 - 333 has had several affects. Firstly analysis of the hydrogen bonds made by residues 320 - 333 in Fe$_2$Lf shows that although these residues form many hydrogen bonds most of them are 'internal' due to the helical nature of this peptide and only five hydrogen bonds (330 O....NH1 133, 333 O....NH2 133, 332 NH2....OE2 66, 324 OH....ND2 126 and 324 OH....OG1 122) are made with the rest of the N-lobe. In FeLf$_N$ the hydrogen bonds made by residues 320 - 327 are quite different. Residues 320 to 323 are now part of helix 10 and thus make hydrogen bonds which characterise helices (ie. 320 O....N 323). Residues 323 - 327 are extended and form three hydrogen bonds with the j strand of the N-lobe at the back of the iron site, 247 O....N 326, 249 O....N 324 and 249 N....O 324. Effectively these interactions result in the formation of a short piece of 3 stranded antiparallel β-sheet at the back of the iron site. In addition, in the crystal structure solved here the Lf$_N$ molecules pack in a 'back to back' manner and each molecule interacts with the back of a neighbour resulting in the presence of a six stranded antiparallel β-sheet. Two hydrogen bonds between 321 O....N 325 and 325 N....O 321 link the two molecules. The interactions made by residues 321 - 326 with β-strand j and β-strand l, from a symmatry related molecule, are shown in Fig. 64. The two hydrogen bonds formed with the symmetry related molecule probably help to stabilise the structure but it is unlikely that they are responsible for the unwinding of helix 11.

The loss of the four hydrogen bonds between 320 - 333 and the N-lobe in Fe$_2$Lf has resulted in several localised rearrangements in the structure, especially in helix 5 (residues 121 - 136) which forms one wall of the N2 domain and is covered by helix 11 in Fe$_2$Lf. Firstly the sidechain of Arg 133, which previously interacted with 330 O, has a new orientation and now NH1 hydrogen bonds to residue 130 O while NH2 interacts with a water molecule. Secondly removal of Tyr 324, and the loss of the two hydrogen bonds between Thr 122 and Asn 126 and the OH group of Tyr 324, has increased the access of solvent (see below) to this region of the structure. Both Thr 122 OG1 and Asn 126 ND2 now hydrogen bond to a water molecule which occupies the position previously filled by Tyr 324 OH (Fig. 65). Lastly the loss of the hydrogen bond between Arg 332 NH2 and Glu 66 OE2 has not affected the position of Glu 66 but the sidechain of Arg 120

Fig. 64 Hydrogen bonding pattern involving β-strand 1.

In this figure the 2F$_{obs}$ - F$_{calc}$ has been contoured at 0.75σ. The atoms from the molecule in question are shown in pink and those from a symmetry related molecule are shown in yellow. All of the hydrogen bonds made by β-strand 1 are shown as dotted lines.
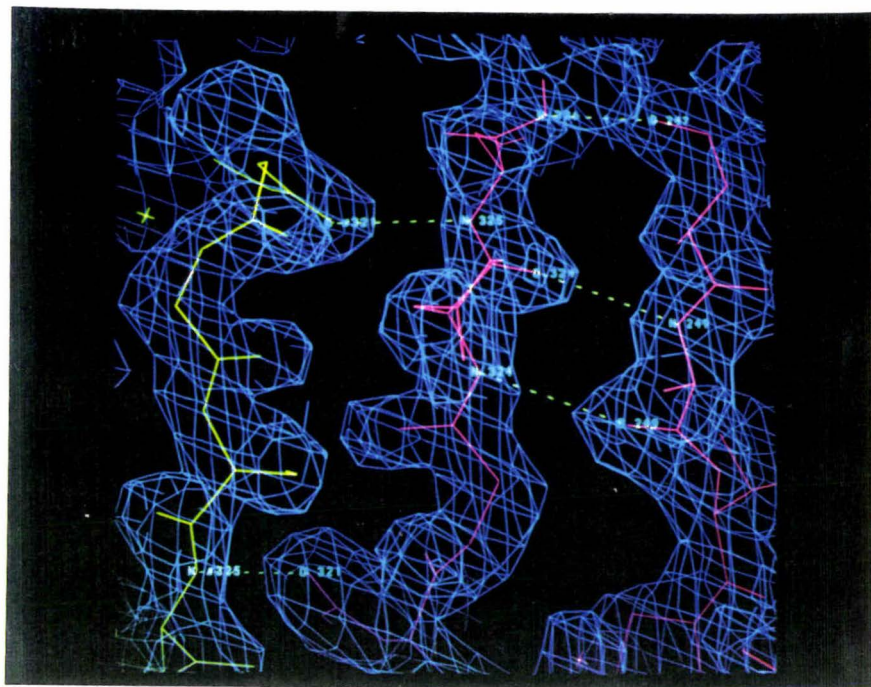
Fig. 65 Helix 5: contacts due to the absence of Tyr 324 and presence of Phe 325.

Residues 120 to 132 and 320 to 327 in FeLf$_N$ are shown in pink. Residues 320 - 333 from Fe$_2$Lf are shown in blue, residues 120 to 132 have a very similar position in both structures. The water molecules that replace Tyr 324 are shown in pale blue and the hydrogen bonds made by these are indicated.
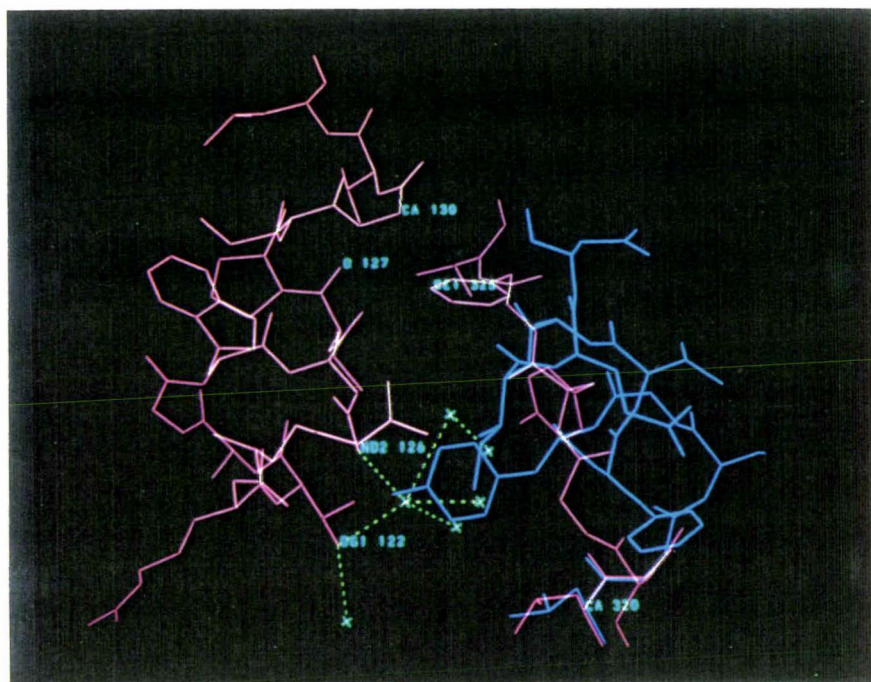
has moved and now OE2, of Glu 66, interacts with only NH2 of Arg 120 instead of both NH1 and NH2. In reality the sidechain of Arg 120 is probably quite mobile, as the density for this residue is poorly defined, and it may occupy both configurations.

As a consequence of the altered conformation of residues 320 - 327 the solvent accessibility of some residues has changed. The changes in solvent accessibility, calculated using the method of Lee and Richards (1971), of the binding site, the hinge region, helix 5 and the C-terminal peptide are summarised in Table 25. While the greatest changes involve residues in the C-terminal peptide the increased accessibility of Thr 90, His 91 and Val 250 may be of most significance. The hinge region in the interdomain strands is located at these residues and their increased accessibility may have an important role in determining the equilibrium between the open and closed forms. The increased accessibility of Thr 122 may also be important as it is adjacent to the carbonate anion and the other iron ligands. Finally the environments of Val 127, Gly 130 and Thr 131 are also altered. In Fe_2Lf several water molecules fill the cavity between these residues and helix 11 (residue 327 in particular), but in $FeLf_N$ Phe 325 packs quite closely against helix 5 (3.2 Å between 325 CE1 and 127 O) decreasing the solvent access to this part of helix 5 (Fig. 65 above). This may also affect the mechanism of domain opening because Gly 130 has been shown to have a different environment in Fe_2Lf as compared to ApoLf (Gerstein *et al.*, 1992).

Intermolecular interactions

The other major change in sidechain conformation involve residues on the surface of the protein which interact with symmetry related molecules. Besides the two hydrogen bonds between the β-strands at the back of the iron site there are 9 contacts to both the N1 and the N2 domains. The intermolecular interactions are summarised in Table 26. Half of these interaction involve electrostatic interactions between basic residues in domain 1 (Arg 24, Arg 27, Lys 28 and Arg 30) and acidic residues in domain 2 (Glu 178, Asp 205, Asp 220, Glu 221). The interactions involving residues 24 - 30 mean that these residues are more ordered than in Fe_2Lf and that they have different conformations. The prevalence of electrostatic interactions is quite possibly related to the crystallisation conditions, which involve low ionic strength buffers and may force these residues to 'seek each other out'.

Table 25 Changes in solvent accessibility.

Solvent accessibilities for residues in Fe_2Lf (Lf) and $FeLf_N$ (Lf_N) were calculated using the method of Lee and Richards (1971). $\Delta = Lf_N - Lf$ and is in $Å^2$.

C-terminal peptide				Helix 5				Hinge region			
Residue	Lf	Lf_N	Δ	Residue	Lf	Lf_N	Δ	Residue	Lf	Lf_N	Δ
Arg 309	92	99	7	Thr 122	0	14	14	Arg 89	47	47	0
Val 310	0	12	12	Ala 123	2	2	0	Thr 90	43	58	15
Pro 311	19	82	63	Gly 124	0	0	0	His 91	23	43	20
Pro 312	97	98	1	Trp 125	8	9	1	Tyr 92	0	0	0
Arg 313	66	154	88	Asn 126	10	58	48	Arg 249	135	71	-64
Ile 314	1	44	43	Val 127	12	8	-4	Val 250	3	11	8
Asp 315	0	29	29	Pro 128	0	0	0	Pro 251	17	14	-3
Ser 316	12	16	4	Ile 129	1	7	6	Ser 252	0	0	0
Gly 317	3	51	47	Gly 130	40	24	-16				
Leu 318	3	135	132	Thr 131	48	33	-15				
Tyr 319	14	24	10	Leu 132	2	1	-1	Binding site			
Leu 320	9	49	40	Arg 133	26	119	93	Residue	Lf	Lf_N	Δ
Gly 321	26	57	31	Pro 134	79	107	28	Asp 60	10	11	1
Ser 322	30	81	51	Phe 135	76	88	12	Tyr 92	0	0	0
Gly 323	50	35	-15					Tyr 192	6	5	-1
Tyr 324	25	91	66					His 253	16	21	5
Phe 325	18	110	92					Thr 117	0	2	2
Thr 326	37	56	19					Arg 121	23	20	-3

Table 26 Summary of the interactions between symmetry related molecules.

[a] All direct contacts less than 3.3 Å in distance are listed. [b] Atom in molecule (1/2-x, 1/2+y, 1-z); [c] Atom in molecule (1/2-x, -1/2+y, 1-z); [d] Atom in molecule (x, y, 1+z); [e] Atom in molecule (x, y, -1+z); [f] Atom in molecule (1-x, y, 1-z).

Contact[a]	Distance (A)	Contact[a]	Distance (A)
24 NE....OE2 221[b]	3.12	205 OD2....NE 30[e]	2.97
24 NH2....OE2 221[b]	2.91	220 OD1....NH1 27[c]	2.95
27 NH1....OD1 220[b]	2.95	220 OD2....NH2 27[c]	3.18
27 NH2....OD2 220[b]	3.18	221 OE2....NE 24[c]	3.12
28 NZ....OE2 178[c]	3.09	221 OE2....NH2 24[c]	2.91
30 NE....OD2 205[d]	2.97	281 O....ND2 179[c]	3.22
178 OE2....NZ 28[b]	3.09	282 O....ND2 179[c]	3.26
178 OE2....OG 283[b]	2.45	283 OG....OE2 178[c]	2.45
179 ND2....O 281[b]	3.22	321 O....N 325[f]	3.00
179 ND2....O 282[b]	3.26	325 N....O 321[f]	3.00

B3.2.9.4 Solvent structure

As described earlier (Section B3.2.7) solvent molecules were added to the model as density became visible and all of the solvent molecules included were treated as water molecules. In total 186 water molecules have been included in the model structure. Twenty of these are so-called 'second shell waters' as they do not hydrogen bond with any protein atoms while the remainder are first shell waters and have at least one protein hydrogen bonding partner. A rigourous analysis of the water molecules has not been carried out and it is quite possible that some of the water molecules added at the end of round 5 may have been incorrectly included. However, most of the solvent molecules included so far are believed to be 'real'. Comparison of the positions of the water molecules in FeLf$_N$ with those in Fe$_2$Lf has identified 86 water molecules in common between the two structures. Water molecules were recorded as 'common' if the position of two water molecules differed by less than 2.0 Å when the FeLf$_N$ and Fe$_2$Lf structures were superimposed. The distribution of the 'unique' and 'common' waters is shown in Fig. 66. Although it is not very clear in Fig. 66 many of the unique waters are on the surface of the molecule while a greater proportion of the common waters are in the iron binding cleft or in pockets on the surface of the protein.

Fig. 66 Position of the common and unique waters in FeLf$_N$.

In the following stereo diagram the Cα plot of FeLf$_N$ is shown in pink, the water molecules in common with Fe$_2$Lf are white and the water molecules unique to Lf$_N$ are blue.
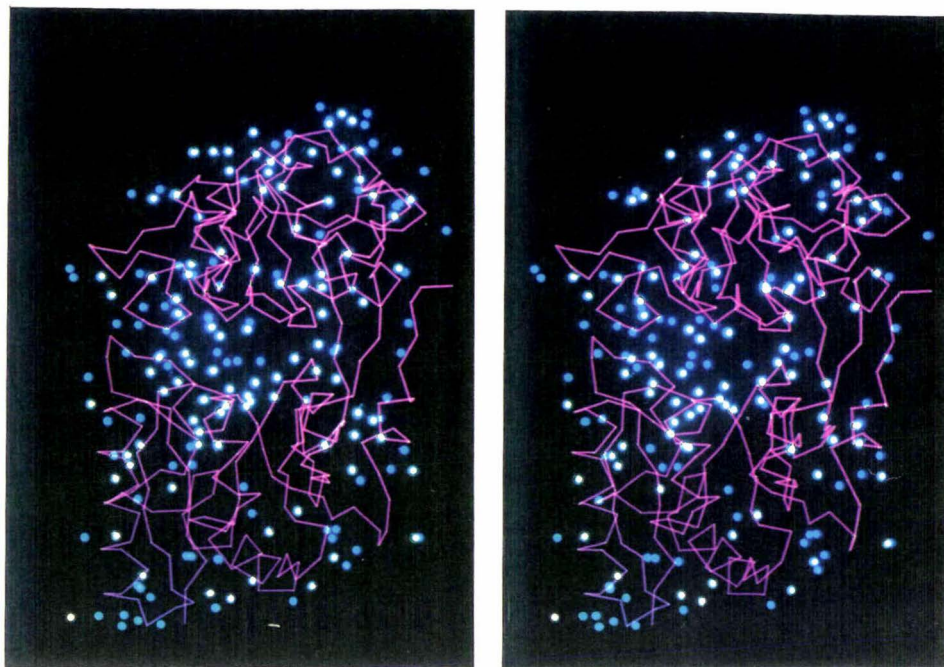
Analysis of the contacts made by each class of water molecule has shown that the average 'common' first shell water has 1.92 protein hydrogen bonding partners while the average unique water has only 1.4 hydrogen bonding partners. This demonstrates that the water molecules which are conserved between the two structures make, on average, more hydrogen bonds with the protein structure than do those which are unique to FeLf$_N$. This trend is even more clearly demonstrated within the 'common' waters where the water molecules with the most conserved position (deviation of between 0 and 0.5 Å) make a significantly greater number of hydrogen bonds (Table 27) to the protein structure than do those with less conserved positions (deviation of between 0.5 and 1.0 Å).

There is also a correlation between the class of water molecule ('common' and 'unique') and the average B value of each class. The average B value for all of the waters is 46.4 Å^2 while the average B value of the conserved waters is ~4 Å^2 lower at 42.5 Å^2 and the average B value of the unique waters is ~4 Å^2 higher at 50.8 Å^2. This trend has been reported previously (Baker, 1988) and reflects the tighter binding of the common water molecules.

Table 27 Analysis of the contacts made by the common water molecules.

The number in brackets indicates the number of second shell waters. * The number in brackets indicates the average number of hydrogen bonds made by the waters in each class if the second shell waters are included.

Deviation in position	Number of waters #	Average number of hydrogen bonds*
0 - 2.0 Å	86 (5)	1.92 (1.8)
0 - 0.5 Å	34 (2)	2.34 (2.2)
0.5 - 1.0 Å	30	1.6
1.0 - 1.5 Å	20 (3)	1.2 (1.0)
1.5 - 2.0 Å	11	1.2

The contacts made by the water molecules have been further analysed and 31 internal waters have been identified (Table 28). Of these 15 occupy essentially discrete sites and are characterised by a large number of hydrogen bonds, 3.6 per water, almost exclusively to protein atoms (only 5 out of a total of 55 are to other waters). The remainder occupy the large cavity in the interdomain cleft and while on average each water still makes 3.1 hydrogen bonds many of these, 2.0 per water, are to other water molecules. A part of this water network in the interdomain cleft is shown in Fig. 67. Besides making a considerable number of hydrogen bonds the internal water molecules have a lower average B value, 33.9 $Å^2$ compared to 46.4 $Å^2$ for all the waters, and all but three of the internal waters are 'common'. These features demonstrate that the internal water molecules are an integral part of the structure.

Fig. 67 Position and electron density for some waters in the interdomain cleft.
The hydrogen bonds made by the water molecules are indicated and the electron density, contoured at 1.0σ is shown.
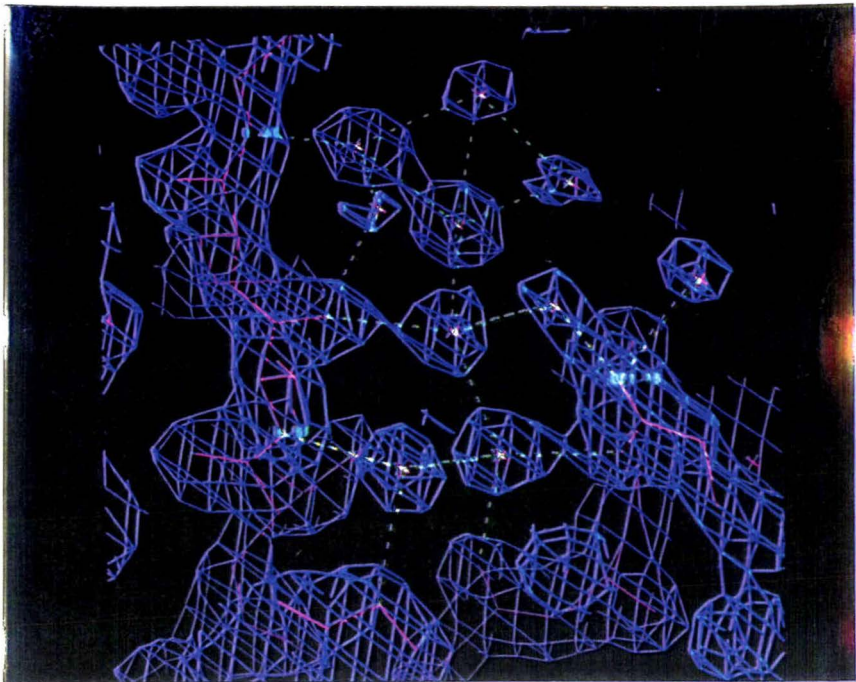
Table 28 Internal water molecules.

Discrete water sites		Interdomain cleft waters	
Water	Hydrogen bond partners	Water	Hydrogen bond partners
W501	8 NE1, 58 N, 58 OG1, 299 O, W500	W500	9 O, 15 OE2, W501, W502
W506	18 NZ, 290 N, 300 O, 302 OD2, 298 O	W502	58 O, W500, W503, W611
W516	80 OE1, 80 OE2, 253 ND1, W610	W503	W 502, W 504, W 505, W621
W517	61 N, 92 OH, 122 OG1, 251 O	W504	60 N, W 621, W 503, W620
W520	64 O, 68 N, 74 O	W505	W503, W621, W555
W525	85 O, 123 O	W555	183 O, W505, W668
W526	96 N, 230 O, 245 O	W559	193 N, W562
W529	105 NE2, 106 N, 231 O, 250 O	W560	190 O, W558, W666
W532	11 O, 189 O	W561	191 OG, 297 OD1
W568	210 O, 211 O, 214 N, W569	W562	193 OG, 301 NZ, W559, W563
W569	211 O, 239 O, W568	W563	213 OG1, 216 OE1, 216 OE2, W562
W575	55 O, 258 N, 258 NH1, 262 O	W611	15 OE1, W502
W580	278 O, 279 O, 283 N, 283 O, 286 O	W620	58 O, W504
W591	89 NH2, 211 OE1, 212 OG, 210 NH2	W621	W503, W504, W505
W610	297 O, 299 O, W516	W656	15 OE1, 297 O, W669
		W669	12 OG, W556, W656

B3.3   **DISCUSSION**

In this chapter the solution and refinement of the FeLf$_N$ structure has been
described in detail. Although refinement is not quite complete the structure is
believed to be accurate (maximum error of ~ 2.5 Å) and the current crystallographic
R factor is 0.184 when data between 8.0 Å and 2.0 Å are included. A considerable
amount of effort was devoted to processing the high resolution images and this is
reflected in the high quality, and completeness, of the data set. The structure has
been solved by molecular replacement using the N-lobe of Fe$_2$Lf as the starting
model. Although essentially straightforward, in that the signals obtained from the
rotation and translation functions were well above the background signal, solution
of the structure was complicated by the uncertainty over which space group the
crystals belonged to. Initially the structure was solved in P1 and then once the
relationship between the molecules in the asymmetric unit was revealed the
coordinates for the model and the data were converted into space group C2, the
correct space group. All of the refinement using the high resolution data was
carried out in space group C2.

Six rounds of refinement and rebuilding have been carried out. This has allowed
residues 4 - 327 to be fitted to the electron density. Comparison of this structure to
the N-lobe of Fe$_2$Lf has shown that the two structures are very similar and the
folding pattern is essentially identical. The deviation in the position of the residues
involved in metal and anion binding is only 0.41 Å. One major conformational
change, involving residues 321 - 327, does however occur. In intact lactoferrin, as
in sTf (Bailey *et al.*, 1988), residues 321 - 333 form a helix which together with β-
strands e and j forms the connection between the two domains. In FeLf$_N$ helix 10
is extended by three residues, to include 315 - 323 while the remaining residues,
323 - 327, form a third β-strand across the back of the iron site. Glycine 321 and
323 appear to play a critical role in the conformational change. In Fe$_2$Lf helix 10
finishes at glycine 321 which provides the break between helices 10 and 11, but in
FeLf$_N$ helix 10 is extended to glycine 323, at which point the chain becomes
extended. This is consistent with a number of other structures where glycines have
been associated with flexible regions of proteins (Huber and Bode, 1978).

The similarity of the metal binding sites in the two structures suggests that changes
to residues other than those directly involved in metal and anion binding are
responsible for the more facile release of iron from Lf$_N$. Besides the main
conformational difference at the C-terminus of FeLf$_N$ changes in the position of
some sidechains and in the solvent accessibility of others also occur. The

differences between the structure of FeLf$_N$ and Fe$_2$Lf are a consequence of (i) crystal packing; (ii) the slight difference in the orientation of the two domains; (iii) the absence of the C-lobe and (iv) the absence of helix 11 and the addition of β-strand 1.

Crystal packing affects residues on the surface of the protein, in particular basic residues in helix 1 and acidic residues between residues 178 and 221. Several salt bridges are formed between these residues and consequentially these sidechains are better defined in FeLf$_N$ than in Fe$_2$Lf. Two hydrogen bonds are made at the back of the iron site between β-strands 1 and 1'. These bonds may help stabilise the structure seen in FeLf$_N$ but are unlikely to be responsible for the change in conformation of residues 321 - 333. In general the contacts made between neighbouring molecules are involved in stabilising the crystal packing but probably do not make a significant impact on the spectral and iron binding properties of Lf$_N$. Crystal packing could also be responsible for the difference in the orientation of the two domains. The change in domain closure is similar to the variation in domain orientation seen between hLf and bLf, and between hLf and sTf (Baker *et al.*, 1991; Baker and Lindley, 1992). Similar differences in domain closure have also been reported for the bacterial periplasmic proteins which share the same folding pattern and the same mechanism of ligand-induced conformational change (Baker *et al.*, 1987; Quiocho, 1991). In all cases the significance of the altered orientation of the domains is not known; it could reflect specific differences in interdomain interactions but is quite likely to involve no more than crystal packing effects (Dobson, 1990; Faber and Matthews, 1990). Two interdomain salt bridges have been considerably weakened by the 2.7° rotation of the domains but the position occupied by most of the other residues in the interdomain cleft is very similar. In addition the positions of the water molecules in the iron binding cleft are highly conserved, with all but three of them corresponding to within 1.0 Å between the two structures. This suggests that the solvent access to the iron site, via the cleft, is essentially identical.

Removal of the C-lobe has allowed helix 11 to unravel and changed the environment at the back of the iron binding site. Iron binding and release by members of the transferrin family involves a hinge movement of β-strands e and j which run across the back of the iron site and link the two domains. The hinge movement is essentially parallel to a line through the Cα atoms of Thr 90 and Pro 251 (Gerstein *et al.*, 1992). In addition to the hinge movement helices 5 and 11 rotate against one another during domain opening and closure; helix 5 moves with N2 while helix 11 stays fixed with N1 (Anderson *et al.*, 1990). This movement is associated with changes in the packing environment of residues at the interface

between the two helices; in particular Ala 327 and Gly 130 are closely packed in the open form but move further apart in the closed form.

The absence of the C-lobe increases solvent access to the β-strands (residues 90, 91 and 250 in particular) at the back of the iron site. The significance of the differences in solvent accessibility are difficult to estimate as they are associated with changes within the N-lobe, notably the loss of helix 11 which has been replaced by β-strand 1 in FeLf$_N$. This change in conformation means that in FeLf$_N$ domain opening will involve the flexing of three β-strands and the environment of helix 5 in both the open and closed states will be altered. This may be important as the environment of Gly 130 is different and this is one of the residues which may be important in determining the mechanism of domain opening and closing seen in Lf. In addition removal of Tyr 324 increases the solvent accessibility of Thr 122, adjacent to the bound anion and Asp 60, one of the metal ligands.

These two affects, the loss of helix 11 and the increased solvent accessibility of some residues, would seem to be the changes most likely to contribute to the more facile release of iron from Lf$_N$. This theory is supported by the properties of Lf$_{N30}$, a 30 kDa proteolytically-derived N-terminal fragment of human lactoferrin (Legrand *et al.*, 1990). This fragment released iron at an even higher pH than Lf$_N$ and was found to be 50% desaturated at pH 6.3 compared to pH 4.8 for Lf$_N$. Lf$_{N30}$ only extends as far as residue 281 therefore it is missing many residues which contribute to the N1 wall of the binding cleft (notably 293 - 302) as well as helices 10 and 11 (see Fig 4 in intro). The absence of these residues would be expected to further increase the solvent accessibility and cause the higher pH of iron release.

In total there are quite a few changes due to the loss of the C-lobe and repositioning of residues 312 - 333 and although these are almost certainly responsible for the properties of Lf$_N$ it is difficult to conclude which make the main contribution. In all likelihood it is a combination of these changes which causes FeLf$_N$ to release iron at a higher pH.

Besides providing information about the possible structural reasons for the altered properties of FeLf$_N$ this structure has provided the highest resolution structure of a 'transferrin' metal binding site. This has clearly shown the bidentate nature of the anion coordination and the exquisite design of the anion binding site which fulfils the complete bonding potential of the preferred anion, carbonate. The accuracy of this structure means that a detailed comparison with other structures will be permitted and even quite small changes can be evaluated. Crystals of a fragment

representing the N-terminal half of human transferrin have been reported (Wang *et al.*, 1992) and once this structure has been solved comparison of the two N-terminal structures may allow a structural basis for the different properties of the two fragments to be determined. It is likely that the position of the last ~20 residues in the C-terminal half of transferrin may also differ in Tf$_N$. In the structure of the proteolytically derived N-terminal half-molecule of transferrin (Sarra *et al.*, 1990), which was thought to represent residues 1 - 332, the polypeptide chain could not be traced beyond residue 305. The authors suggested that this was either due to proteolysis at an earlier position than expected or due to a conformational change. In light of the current structure a conformational change is most likely but the change will probably differ in detail from that of Lf$_N$ given the presence of a disulphide bridge (number 10) which ties the connecting peptide to the N-lobe (Bailey *et al.*, 1988); no such bridge occurs in Lf$_N$.

Part B : Chapter Four

ApoLf$_N$ DATA COLLECTION AND STRUCTURE SOLUTION

B4.1   **INTRODUCTION**

The growth of ApoLf$_N$ crystals, which diffracted to high resolution, allowed the structural analysis of Lf$_N$ to be extended to include the structure of ApoLf$_N$. It was hoped that this study would allow the changes in the structure associated with iron binding and release to be examined. In the structure of intact ApoLf, solved in 1990 (Anderson *et al.*, 1990), the N-lobe was found to be open and, although iron free, the C-lobe was closed (Fig. 68). This finding suggested that there was an equilibrium between the open and the closed forms. It was hoped that solution of another apo structure would expand our understanding of the mechanism of iron binding and release. As for FeLf$_N$ the structure has been solved by molecular replacement using, as models, fragments of the intact structure. Although refinement of ApoLf$_N$ is incomplete the structure to date has revealed several interesting features which will be discussed in some detail.

Fig. 68 Ribbon diagram of ApoLf

This diagram was prepared using the program Ribbon (Richardson, 1985; Priestle, 1988). In this representation α-helices are shown as pink spirals and β-strands as blue arrows.

N-lobe   C-lobe
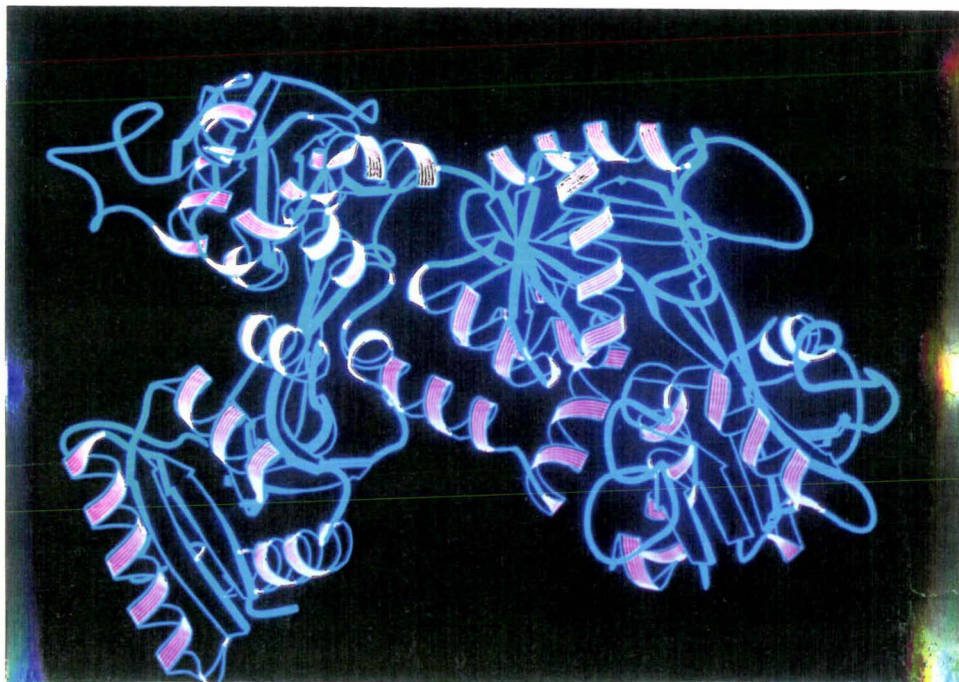
B4.2   RESULTS

B4.2.1   Data collection and processing

Data collection

At the same time that the high resolution data set was collected from the DGFeLf$_N$ crystal a data set was also collected from a small crystal of ApoLf$_N$ in Tsukuba, Japan using synchrotron radiation and Weissenberg photography (Section B1.2). The setting parameters used for data collection are summarised in Table 29.

Table 29 Setting parameters used for collection of the ApoLf$_N$ data in Japan.

Source	Beam-line 6A2, Photon Factory
Method	Sceenless Weissenberg photography
Wavelength	1.0 Å
Crystal to IP distance	286 mm
Coupling constant	Variable
Oscillation range ($\omega°$/IP)	Variable
Number of osc^n / IP	13
ω scan speed °/sec	2
Exposure per image (secs)	Variable
Number of IP's	13 split horizontal plates

Initially the small crystal (0.5 mm x 0.5 mm x 0.02 mm) was aligned with the a axis parallel to the spindle and the b axis parallel to the incident X-rays. Thirteen pairs of horizontal image plates were used to collect 92° of data. Because these crystals have one long ~217 Å axis and two short ~58 Å axes the coupling constant and the ω range used for data collection were varied depending on the orientation of the crystal. This reduced the overlap between adjacent spots. The coupling constant and ω range used for each pair of images are given in Table 30.

Data processing

Although discussed second in this thesis the ApoLf$_N$ data set was actually the first to be processed using the programs described in section B1.3.2. This may have affected the quality of the processing, as much depends on the accuracy of the setting parameters determined during processing. The images were aligned using the program *Weis* as described for the FeLf$_N$ images (Section B3.2.5). Once the images had been correctly oriented and the setting parameters refined the intensities

of all the spots, to a maximum resolution of 2.0 Å, were determined (Sections B1.3.2 and B3.2.5). The processing statistics for each image are summarised in Table 30. In most cases no more than about 75% of the expected reflections were observed. This was quite a low level of observed reflections and in hindsight an integration sphere with a radius of only 2.5 Å would have been more appropriate. The number of reflections measured may have also been increased if the measurement box used to determine the intensity of the reflections had been enlarged. These changes would have increased the % of observed reflections and decreased the values of DR and DS which are both considerably higher than obtained when the FeLf$_N$ data was processed (Section B3.2.5). The possibility of reprocessing the images was considered, but, at this stage bigger crystals (0.5 mm x 0.5 mm x 0.075 mm) of ApoLf$_N$ had been grown and sent to the Molecular Structure Corporation (USA) where another data set was collected (see section B4.2.2).

Despite the obvious inadequacies of this data set the files containing the intensity measurements were concatenated and converted using *WeisdataJ* (Section B1.3.1). The data was then scaled and merged as described in section B1.4 except that in this instance all data less then 1σ were discarded. A merging R value of 11.3% was obtained after 9.7% of the reflections had been removed. The final output file contained 14 881 unique reflections which represented 60% of the expected reflections. Of these reflections only 61% had I > 2σ although if only the data to 2.5 Å were included 78% of the expected reflections were present and of these 73% had I > 2σ. The quality of the complete data set is summarised in Fig. 69 where the % of the observed reflections with I > 2σ and 3σ is plotted against resolution. Note that in this instance 1σ data is not included because all the data with intensities less than 1σ had been discarded during merging. Although incomplete and of moderate quality, especially at high resolution, this data set was successfully used to determine the structure of ApoLf$_N$. The average cell dimensions obtained from processing, and used for structure solution were, a = b = 58.35 Å, c = 215.2 Å, α = β = γ = 90.0°.

Table 30 Data collection and processing statistics for the ApoLf$_N$ images.

ω range represents the oscillation range for which data were collected for each image. The coupling constant was varied depending on the orientation of the crystal and the value used for each image is indicated. DR and DS give a measure of missetting of the centre of each reflection, DR is a measure of the horizontal displacement and DS is measure of the vertical displacement. RMS/BGD is the ratio of the standard error to the background level of density.

Image plate	ω range	Coupling Constant	% expected reflections observed	Number of whole reflections	DR	DS	RMS/ BGD
1 top	82.0° - 93.9°	1.8	not processed				
bottom			62.0%	3231	0.51	0.48	
2 top	93.5° - 104.0°°	1.5	67.2%	2013	0.68	-0.72	0.72
bottom			64%	1852	0.72	0.71	0.72
3 top	103.0° - 113.5°	1.3	66.7%	2712	0.72	0.75	0.70
bottom			66.0%	2731	0.70	0.79	0.7
4 top	113.0° - 122.0°	1.2	89.0%	2401	0.22	0.25	0.4
bottom			65.0%	2795	0.65	0.66	0.4
5 top	121.5° - 130.0°	1.1	88.0%	2988	0.21	0.22	0.4
bottom			81.8%	1633	0.52	0.6	0.5
6 top	129.5° - 137.0°	1.0	65.0%	3521	0.75	0.75	0.6
bottom			65.4%	2670	0.72	0.85	0.6
7 top	136.6° - 143.6°	0.9	71.0%	1653	0.75	0.66	0.9
bottom			71.6%	1274	0.69	0.8	0.5
8 top	143.2° - 149.7°	0.9	74.3%	1347	0.51	0.44	0.4
bottom			68.2%	1633	0.68	0.81	0.9
9 top	149.3° - 155.3°	0.9	73.3%	1922	0.68	0.58	0.5
bottom			73.0%	900	0.63	0.76	0.4
10 top	154.9° - 160.9°	0.8	76.0%	1992	0.63	0.54	0.4
bottom			74.0%	1840	0.7	0.66	0.7
11 top	160.5° - 165.5°	0.8	71.0%	2289	0.55	0.54	0.3
bottom			76.3%	1594	0.72	0.66	0.7
12 top	165.1° - 170.1°	0.7	72.0%	1134	0.7	0.62	0.7
bottom			73.0%	2312	0.59	0.64	0.8
13 top	169.7° - 174.3°	0.7	65.7%	2194	0.56	0.66	0.7
bottom			not processed				

Fig. 69 Analysis of the ApoLf_N data set collected in Japan.
In this instance resolution is defined in terms of $\sin\theta/\lambda$. The resolution equivalent to 2.5 Å is indicated by an arrow.
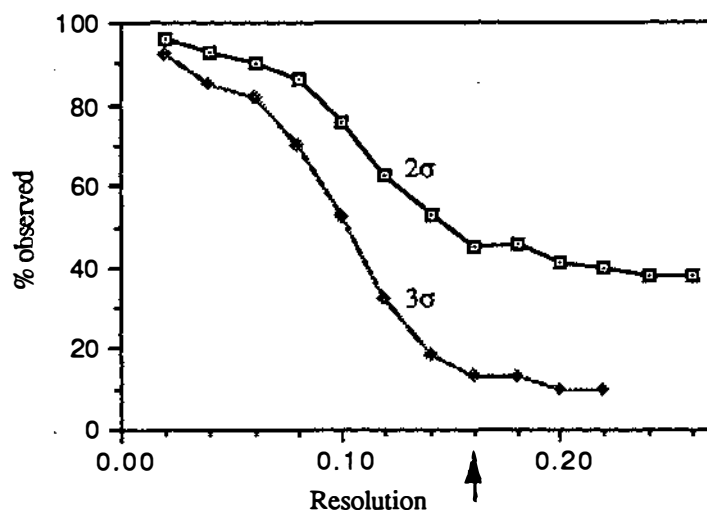
B4.2.2   Solution of the ApoLf_N structure

The Models

Because it has been suggested that an equilibrium exists between the open and closed states of lactoferrin when iron free, it was not known whether ApoLf_N would be open, closed or some state in between. In an attempt to account for all of these possibilities four model structures were used as search models. The four models, representing the N-lobe of both Fe_2Lf and ApoLf, and the two domains of ApoLf are summarised in Table 31.

Table 31 Molecular Replacement search models for ApoLf_N.

Model name	Residues included
Lfnmodel	intact N-lobe of Fe_2Lf, residues 1 - 333
ApoN	intact N-lobe of ApoLf, residues 1 - 333
Domain 1	domain 1 of ApoLf, residues 1 - 91 and 251 - 333
Domain 2	domain 2 of ApoLf, residues 92 - 250

Before the models representing the two domains were created the coordinates for ApoLf were first transformed so that initially they had the same orientation as the Fe_2Lf coordinates. Once transformed the water molecules were removed from all of the models and the iron atom and the bicarbonate ion were excluded from the Fe_2Lf structure. Structure factors were then calculated using a P1 cell large

enough (80 Å x 80 Å x 80 Å) to eliminate the contribution of intermolecular Patterson vectors and an artificial temperature factor of 25.0 $Å^2$ was applied.

The Rotation Function

The rotation matrix required to obtain the best agreement between the model and the data was calculated using ALMN (Section B1.5.1). The two top peaks obtained for each model are given in Table 32.

Table 32 Rotation function solutions for ApoLf$_N$.

α, β and γ are Eulerian angles which were incremented in 2.5° steps through 360° in α and γ, and 5° steps between 0 and 180° in β. The value of rms refers to the root-mean-square deviation of the map from its mean.

Model	Peak	$\alpha°$	$\beta°$	$\gamma°$	peak height	height/rms
Lfnmodel	1	49.6	79.4	165.2	223.1	7.0
rms = 31.7	2	26.5	84.3	336.0	100.8	3.1
ApoN	1	53.5	80.2	164.0	756.5	4.4
rms = 173	2	35.8	90.0	162.0	640.7	3.7
Domain 1	1	51.2	78.2	164.1	1396.2	4.0
rms = 350	2	87.5	50.3	336.0	1181.1	3.4
Domain 2	1	53.5	46.5	136.5	692.3	3.53
rms = 196	2	51.0	22.0	166.5	689.6	3.50

As there is one molecule in the asymmetric unit a single peak was expected. The greatest difference between the top peak and the second highest peak was obtained when the N-lobe of Fe$_2$Lf (Lfnmodel) was used as the model structure. The top peak for this model was four standard deviations higher than the second peak while the other models used did not give a top peak more than one standard deviation higher than the second peak. On the basis of these results it appeared that Lfnmodel was the model which best represented the structure in question. The correct nature of this solution was supported by the fact that although the highest peak obtained for ApoN is not even one standard deviation higher than the next highest peak, and the difference is even less when the two separate domains were used as models, the rotation angles required to position ApoN and Domain 1 are quite similar to those obtained for Lfnmodel.

The translation function

Once the coordinates for Lfnmodel had been rotated by 49.6°, 79.4° and 165.2°

about α, β and γ respectively the position of the molecule in the unit cell was determined. At this stage it was not known whether the real space group was $P4_12_12$ or $P4_32_12$. These two space groups cannot be distinguished by photography. It had been reported (Lattman, 1985) that a translation function solution will only be obtained in the correct space group. Consequently an R factor search (Section B1.5.2), using data between 20.0 Å and 3.5 Å, was carried out in the unique area of space ($x = 0 - 1$, $y = 0 - 0.5$, $z = 0 - 0.5$) using the symmetry operators for both space groups. When a stepsize of 1.0 Å was used an average R value of ~0.49 was obtained for both space groups. A single minimum was also obtained in both space groups although the $P4_12_12$ minimum was lower (Table 33). In an attempt to more clearly distinguish the two space groups a step size of 0.1 Å was used. Using this smaller step size it appeared that $P4_12_12$ was probably the correct space group as the minimum R value decreased to 0.416 in this instance, but a similar decrease was not observed in space group $P4_32_12$ (Table 33).

Table 33 Summary of the translation function results.

Space group	Step size	Ave R value	Minimum R value	Position of the minimum		
				x	y	z
$P4_12_12$	1.0 Å	0.49	0.443	6.25	18.76	7.83
$P4_32_12$	1.0 Å	0.49	0.458	6.25	18.76	61.62
$P4_12_12$	0.1 Å	0.479	0.416	5.63	19.07	7.8
$P4_32_12$	0.1 Å	0.479	0.453	5.94	18.86	61.62

Rigid body refinement of the model in space group $P4_12_12$

After the model structure had been translated by 5.63 Å, 19.07 Å and 7.8 Å on x, y and z respectively, the position of the model was more accurately determined by carrying out a round of rigid body refinement. In an effort to provide an internal control the five tryptophan sidechains were removed from the model. It was anticipated that if the solution was correct well defined density would be visible for these sidechains and if incorrect density would not be visible. In total nine cycles of rigid body refinement were carried out. The models and data used, together with the progress of refinement are summarised in Table 34.

When the two separate domains were used as models rotations of up to 1.8° occurred in cycles 1 - 3 but in cycles 4 - 6 the movements were only about 0.01° - 0.02° indicating that the model had reached its favoured position. At this time the structure of $FeLf_N$ had been partially refined and the position of residues 313 - 333

Table 34 Summary of ApoLf$_N$ rigid body refinement.

Domain 1: residues 1 - 91 and 251 - 333; Domain 2 residues 92 - 250.

Cycles	Data range	Model	Change in R factor
1 - 3	10.0 Å - 5.0 Å	2 domains	0.393 - 0.374
4 - 6	10.0 Å - 3.5 Å	2 domains	0.35 - 0.341
7 - 9	20.0 Å - 3.5 Å	1 - 315,	0.342 - 0.342
		316 - 321, 322 - 333	

was thought to have changed. In an attempt to position these residues in the ApoLf$_N$ structure they were broken into two fragments (316 - 321 and 322 - 333), representing the two helices present in Fe$_2$Lf, and allowed to move as free bodies. Although they rotated by up to 10° the two helices did not reach a position of equilibrium and it was unlikely that they were correctly positioned. After cycle nine a 2F$_{obs}$ - F$_{calc}$ map was calculated and the position of all the atoms was inspected. Density was clearly visible for most of the mainchain atoms and all of the tryptophans, which had been omitted from the structure, giving some confidence that the solution obtained was correct. At this stage the sidechains of 54 residues, for which the density was not well defined, were removed, and the first four amino acids were completely deleted. Although residues 312 - 333 did not appear to be correctly positioned they were left in the model in the hope that they would move to a more favourable position during refinement.

By this time the data set collected and processed at the Molecular Structure Corporation (USA) was available. This data set (called the 'Rigaku data') was collected using a considerably thicker crystal than the one used in Japan (0.5 mm x 0.5 mm x 0.075 mm). The data was provided as a scaled and merged file containing the hkl, F and sigF for 15 947 unique reflections to a maximum resolution of 2.3 Å. This data set had been collected in 110 frames using an oscillation range of 0.8° in most instances although a few were larger. During processing of the data 8 frames agreed badly and were deleted. The final merging R value for all of the data was 6.5%. Analysis of this data showed that 91% of the expected reflections were present and that 81% of these reflections had I > 2σ. As this data set was more complete and of better quality than the data set collected in Japan the structure of ApoLf$_N$ was refined using this data. The cell dimensions were also changed at this time to those which had been determined during collection and processing of the Rigaku data set. The new cell dimensions were a = b = 58.35 Å, c = 217.16 Å, α = β = γ = 90.0°.

B4.2.3   Partial refinement of the ApoLf$_N$ structure

The model which had been correctly positioned by rigid body refinement, and from which all the sidechains with poorly defined density had been removed, was used as the starting model. Four rounds of XYZ refinement combined with B value refinement (Section B1.6) were carried out using the Rigaku data set. The progress of refinement is summarised in Table 35.

Table 35 Progress of ApoLf$_N$ refinement.

The number in parentheses refers to the number of cycles of temperature factor refinement.

* The R factor was computed by the program RFACTOR (TNT refinement package).

Round of refinement	Number of cycles #	Resolution limits (Å)	R factor*	*rms* bond lengths	*rms* bond angles
1 start	10	10.0 - 3.0	0.403	0.03	4.36
finish			0.3	0.037	4.3
2 start	3	8.0 - 2.3	0.342	0.038	4.3
finish			0.34	0.02	3.35
3 start	50 (4)	8.0 - 2.3	0.3	0.028	3.73
finish			0.225	0.027	3.23
4 start	30 (4)	8.0 - 2.3	0.247	0.031	3.4
finish			0.211	0.024	3.09

With the exception of round two, which is really a continuation of round one, each round of refinement involved several cycles of tight weights followed by several cycles of loose weights after which the weights were slowly tightened until the geometry approached ideality. Four cycles of B value refinement were carried out in each of the last two rounds of refinement. In round three no restraints were applied to the B values and they were found to vary considerably but in round four the B values were restricted to lie between 5.0 Å^2 and 90.0 Å^2, and the maximum shift was limited to, at most, 4 times the average shift. When these restraints were applied the distribution of the the B values appeared to improve but it is likely that the B values have not yet converged.

At the end of each round of refinement the position of all the atoms was examined using $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$ maps (Section B1.7). Using the same criteria as used for the FeLf$_N$ structure (Section B3.2.7) sidechains and water molecules were included as density became visible. The models used in each round of refinement are summarised in Table 36.

Table 36 Summary of the model used for refinement of ApoLf$_N$.

Round of refinement	Residue range refined	Number of absent side chains	Number of atoms in the model	Total number of water molecules
1	4 - 333	54	2138	
2	4 - 316	52	2084	
3	4 - 316	19	2297	
4	4 - 312	14	2321	18

At this stage refinement is not complete. The full complement of water molecules have not been included and a conserted effort to position residues 313 - 333 has not been carried out. The final model of ApoLf$_N$ which has been analysed in this study includes residues 4 to 312 with the exception of 14 sidechains. The sidechains which are incomplete in this study include Arg 4, Ser 5, Gln 7, Arg 24, Lys 38, Glu 66, Arg 75, Lys 99, Asn 137, Thr 139, Phe 152, Lys 243, Arg 249 and Arg 272. All of these omitted residues are on the surface of the structure and are probably quite mobile.

B4.2.4 Accuracy of the ApoLf$_N$ structure

The current model of ApoLf$_N$ has an R factor of 0.211 and the *rms* deviations for the bond lengths and angles are 0.024 and 3.09, respectively. The current coordinates of ApoLf$_N$ are provided in Appendix 6. The accuracy of the structure was assessed by several criteria. Firstly the mean error in the coordinates was estimated from the variation in R factor with resolution as described by Luzzati (1952). From the Luzzati plot shown in Fig. 70 the error in the coordinates was estimated to be about 0.25 Å. At low resolution the error is quite high, this is probably due to the fact that only 18 water molecules have been included in the model and consequently the solvent in the crystal has not been fully accounted for.

Fig. 70 Luzzati plot for the partially refined ApoLf$_N$ structure.

The position of the error lines equivalent to errors of 0.1, 0.2, 0.3 and 0.4 Å are indicated.
Resolution is defined in terms of sinθ/λ.
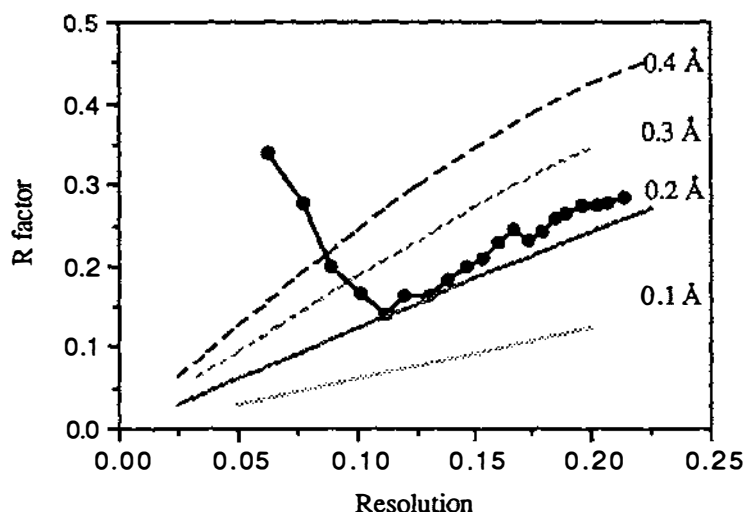
The agreement of the model with the data was quantitated by calculating real space correlation correlation coefficients using the program O (Jones *et al.*, 1991). While most of the correlation coefficients are above 0.85 they are not as good as the correlation coefficients obtained for FeLf$_N$ (Section B3.2.8). There are several poorly defined regions including, residues 4, 38, 86, 134 - 143, 219 - 220, 260 and 294 - 295 where the correlation coefficients fall below 0.75 and rebuilding is required. These are all on the surface of the molecule and the majority of the structure is believed to be correct.

Fig. 71 Mainchain real space correlation coefficients for ApoLf$_N$.
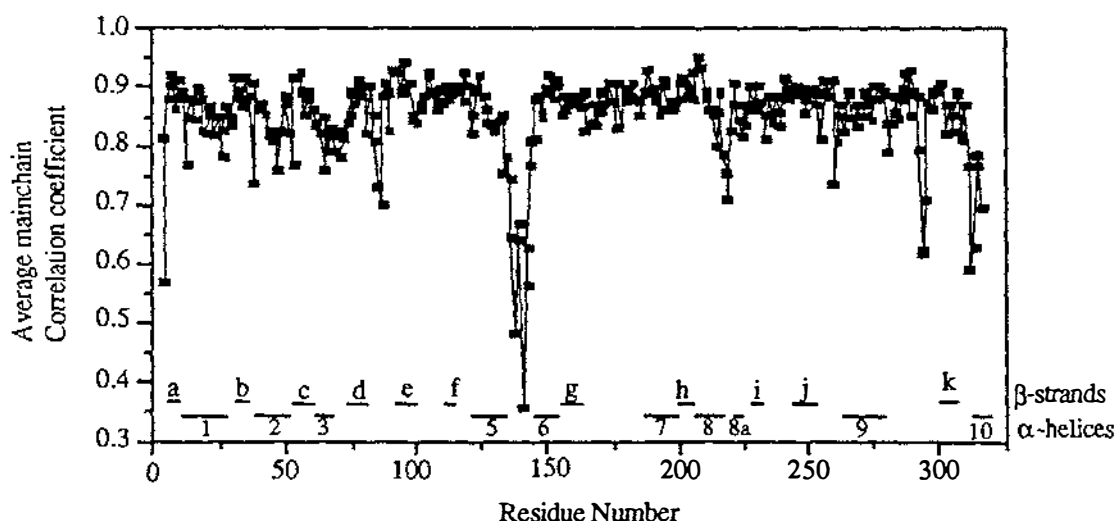
Preliminary analysis of the B values has also shown that although they are quite variable they are distributed in a similar manner to that found in other N-terminal structures (Fig. 72). The average B value for ApoLf$_N$ is 31.8 Å^2 which is actually lower than recorded for the other structures. This low average B value may be partly attributed to the fact that only eight cycles of B value refinement have been carried out and the correct B value for all of the atoms has not been reached. Despite the difference in the overall B value the main point to note is that in general residues with high average B values lie on the surface of the protein, in more mobile and poorly defined regions, while the residues with low average B values are mainly found in the interior of the protein and are involved in secondary structural motifs.

Fig. 72 Plot of average mainchain B values against residue number.
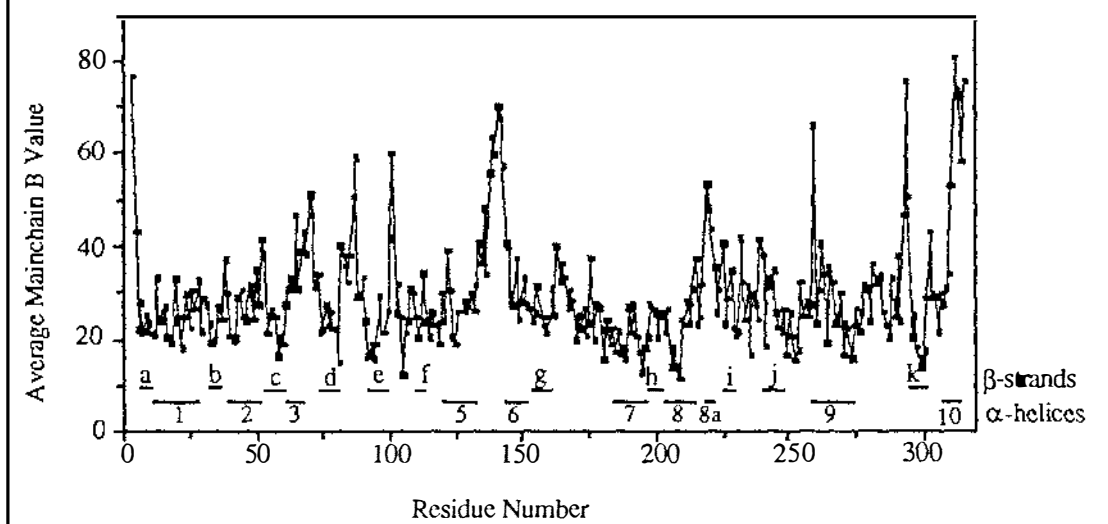
The correctness of the structure was also examined by analysing the distribution of the dihedral angles. The geometry appears very good in that except for Ser 191 and Leu 299, which have been discussed during analysis of the FeLf$_N$ structure (Section B3.2.8), all of the remaining residues lie within acceptable regions on a Ramachandran plot (Fig. 73). This plot is actually better than expected and although the dihedral angles have not been restrained, refinement is incomplete and the dihedral angles may still reflect those of the strarting model.

Fig. 73 Ramachandran plot of mainchain phi/psi angles for ApoLf$_N$.

The allowable regions, as reported by Ramakrishnan and Ramachandran (1965) are indicated by broken lines and glycines are marked by a square rather than a cross.
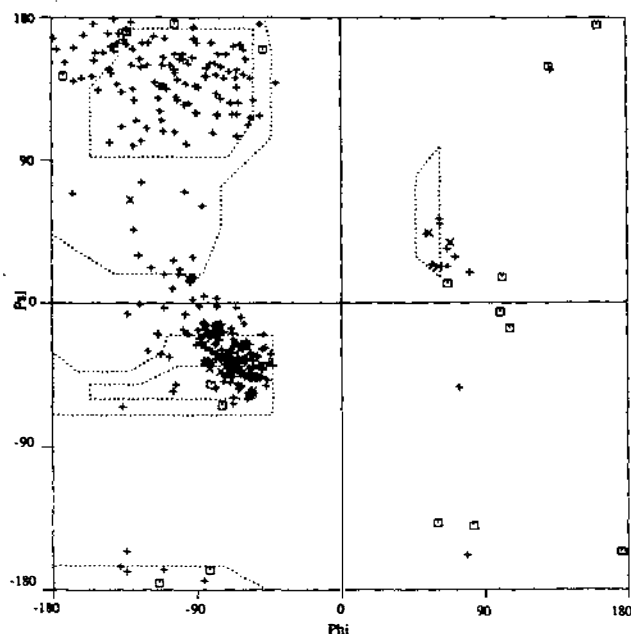
B4.2.5   Analysis of the ApoLf$_N$ structure

As refinement of ApoLf$_N$ is not complete a detailed analysis of this structure is not warranted, however, several general features of the structure will be discussed.

Conformation and crystal packing of ApoLf$_N$

A cartoon diagram of ApoLf$_N$ is shown in Fig. 74. The most notable feature of this structure is that it is essentially identical to the structures of FeLf$_N$ and the N lobe of Fe$_2$Lf, ie. it is **closed**. The *rms* deviation between the two Lf$_N$ structures is only 0.30 Å which is less than the *rms* deviation obtained when either of these structures is compared with the N-lobe of Fe$_2$Lf (Table 37). This is inspite of the fact that the crystal packing is entirely different (see below).

Fig. 74 Cartoon diagram of ApoLf$_N$

This diagram was prepared using the program Ribbon (Richardson, 1985; Priestle, 1988). In this representation α-helices are shown as pink spirals and β-strands as teal arrows.
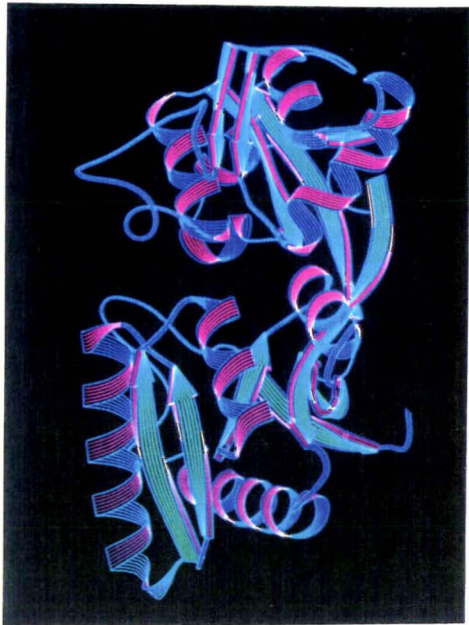
Table 37 Comparison of the deviations between the three closed N-lobe structures.

Structures compared	*rms* deviation
ApoLf$_N$ and FeLf$_N$	0.30 Å
ApoLf$_N$ and Fe$_2$Lf N-lobe	0.34 Å
FeLf$_N$ and Fe$_2$Lf N-lobe	0.36 Å

Like FeLf$_N$, ApoLf$_N$ is also slightly more 'open' than the N-lobe of Fe$_2$Lf, however it is certainly not wide open like the N-lobe of ApoLf. This is the first structure of a supposed iron free lactoferrin where the N-lobe has been found to be closed. Possible reasons for the closed structure may depend on the origin of the density in the iron site which remains unexplained at present (see below). A second intriguing question concerns the crystal packing. If the structures of ApoLf$_N$ and FeLf$_N$ are so similar, as indicated here, why have they crystallised in different space groups and why is the packing of the molecules in the two unit cells so different? As shown in Fig. 75a the molecules in the FeLf$_N$ unit cell are packed in a 'back to back' configuration while the molecules in the ApoLf$_N$ unit cell are packed in a 'front to front' configuration (Fig. 75b) and the back of the each molecule is exposed to the solvent (indicated by an arrow for the purple ApoLf$_N$ molecule). Admittedly the position of residues 313 - 333 has not been determined in ApoLf$_N$ and these will probably occupy some of the space in this channel, but

the orientation of the molecules in the two unit cells is still quite different. It will be interesting to determine the position of the remaining residues in ApoLf$_N$ as these may hold the key to the difference in crystal packing even though almost identical crystallisation conditions were used. Perhaps these residues have yet another orientation in ApoLf$_N$? Initial attempts to build these residues into the structure have not been successful because of the lack of well defined density and the wish to be confident about the position of the residues to avoid incorrectly

Fig. 75 Comaparison of the crystal packing in ApoLf$_N$ and FeLf$_N$.

Fig. 75a shows the packing of the molecules in FeLf$_N$ and Fig. 75b shows the packing in ApoLf$_N$.

Fig. 75a
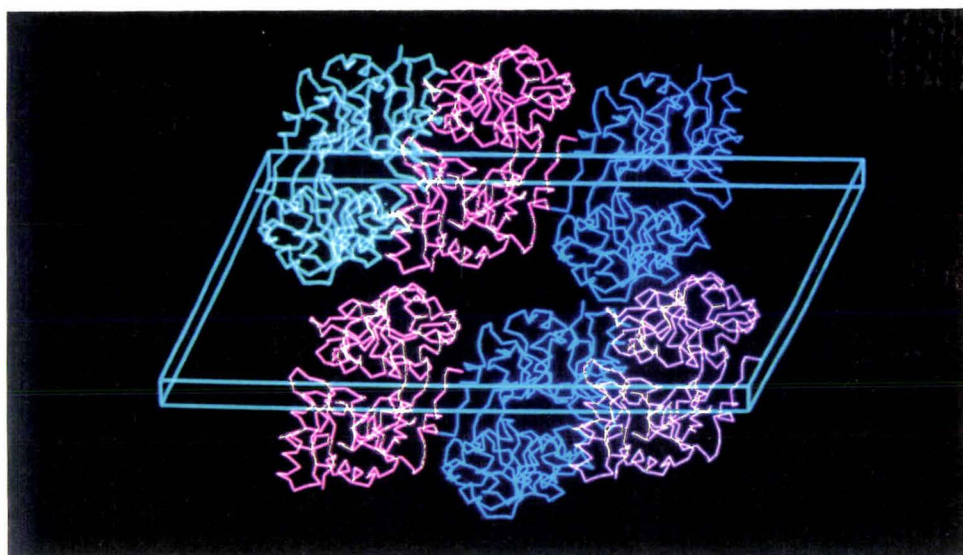
Fig. 75b
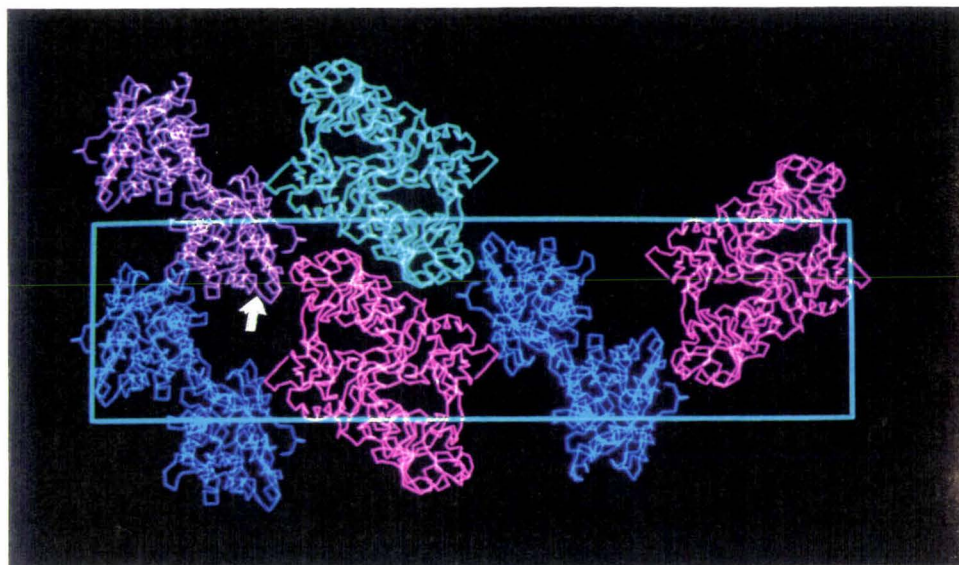
biasing the structure. It is also possible that they are disordered, however, and this question may only be resolved by collection of a higher resolution data set or a derivative data set. Unfortunately limitations on the time available have not permitted this work to be completed in this study.

The binding site in ApoLf$_N$

Besides being closed the ApoLf$_N$ structure is distinguished by the presence of extra density within the metal binding site (Fig. 76). As the protein sample used to prepare these crystals was colourless and supposed to be metal free the presence of this density was most unexpected. At the end of round three of refinement an attempt was made to fit two water molecules into the density. When the position of the two water molecules was not restrained during refinement one of them was moved out of the metal binding site and the B values of both changed rather dramatically. This suggested that the water molecules had been placed in unfavourable positions. Both water molecules were therefore removed from the model and a further eight cycles of refinement were carried out at the end of round four. A new $2F_{obs} - F_{calc}$ map was created at the end of these extra cycles and the density at the metal site analysed. The photograph in Fig. 76 was taken at this time. Clearly 'something' is in the iron site but the nature of the density is unknown. Comparison of the level of density with other parts of the structure showed that it was similar to that of a well defined water molecule, although its shape shows that it is clearly not a single water molecule. To date no viable alternative has been obtained but several possible reasons for this density have been considered. Firstly it is possible that the metal binding site is partially occupied by iron and carbonate and that these moieties are responsible for the density. This however, seems unlikely as the crystals do not appear even faintly pink yet ~25% occupancy would be required to account for the level of density seen in the binding site (estimated by comparison of the level of electron density for a well defined water molecule in ApoLf$_N$ with one in FeLf$_N$). Another alternative is that some other metal ion has bound together with a single water molecule, or even an anion such as azide may be present. It is hoped that further analysis of these crystals may allow this density to be accounted for.

Besides having extra density, the iron site also differs in the orientation of the iron binding ligands. If FeLf$_N$ is reorientated to match ApoLf$_N$ and the iron sites are overlaid (Fig. 77) it is clear that differences exist. Although the extra, unaccounted for density could have distorted the metal binding site the differences in the

Fig. 76 ApoLf_N metal binding site showing the unassigned electron density.
The four ligands involved in metal binding are shown together with the unassigned density. The $2F_{obs} - F_{calc}$ map has been contoured at 1.0σ in this instance.
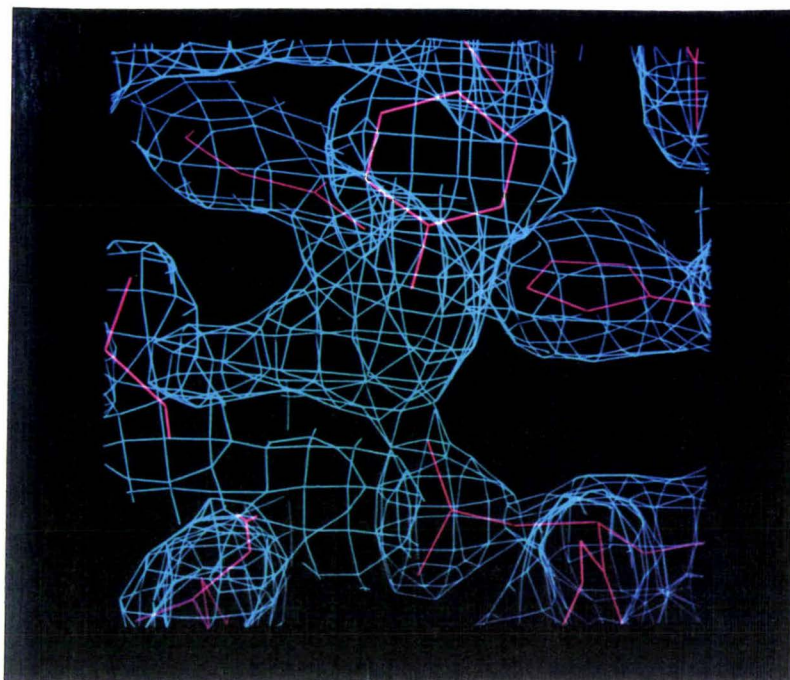
position of at least aspartic acid 60 and histidine 253 appear to be real. The sidechain density for these two ligands is distinct from the unaccounted density and the sidechains are quite well defined. The position of the two tyrosines is more likely to be erroneous as their density merges with that of the extra density and the sidechains are obviously distorted in an attempt to move into the latter. Despite these problems in accurately positioning the ligands there does appear to be a real difference in their position. Once the origin of the density in the metal binding site has been determined it may be possible to account for the differences in geometry seen here.

B4.3   DISCUSSION

Although incomplete analysis of the ApoLf_N structure has revealed several interesting features. The two major points are the fact that the N-lobe is clearly closed, unlike the N-lobe of ApoLf, and that there is obviously some density in the iron site which has not been accounted for. It is likely that the closed orientation of the domains is related to the presence of the electron density in the metal binding site. At present the origin of the density is unknown but it is hoped that eventually it will be accounted for.

Fig. 77 Comparison of the position of the ligands in ApoLf$_N$ and FeLf$_N$.
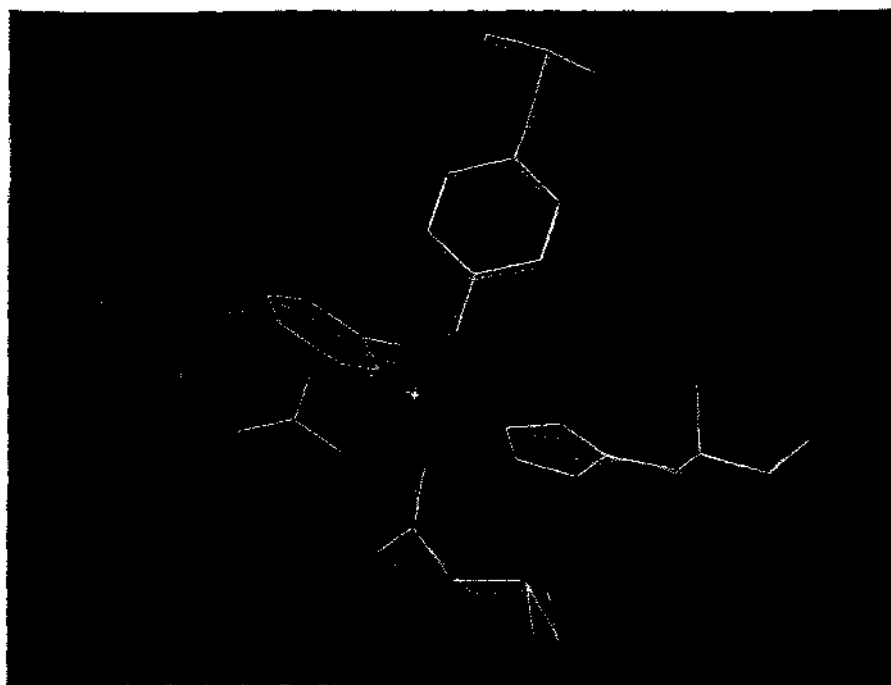
Another interesting feature of this structure is that although it appears very similar to the structure of FeLf$_N$ the crystal packing is quite different in ApoLf$_N$. Perhaps once the orientation of residues 313 - 333 is determined it may be possible to account for this difference. Eventually it would be desirable to determine the structure of an 'open' N-lobe structure so that the changes involved in iron binding and release can be analysed. A structure of this type would also allow a comparison with the N-lobe of ApoLf to be carried out.

Chapter Two

CONCLUSIONS AND FUTURE DIRECTIONS

.1   Conclusions and final discussion

The main aim of this section is to highlight how our understanding of the role of lactoferrin has benefited from a mulitfaceted investigation involving structural, functional, spectroscopic and mutant studies. Many of the important points raised by this investigation have been discussed at the end of each chapter and in an attempt to avoid repetition only the main features will be discussed in this section.

The most important step in this investigation was the construction of an expression system which produced enough protein to allow functional and structural studies to be carried out. If a high level expression system for the amino terminal half of human lactoferrin had not been developed a significant part of this investigation would not have been possible. An equivalent expression system for the C-terminal half of hLf has not so far been developed and this has prevented a similar study on the C-lobe from being carried out in parallel (Sheth, B., Stowell, K.M., Day, C.L. and Tweedie, J.W., unpublished results). More recently human lactoferrin has been produced in fungi (Ward *et al.*, 1992) and human transferrin has been produced in *E. coli* (Ikeda *et al.*, 1992). While both of these systems may be useful for obtaining large amounts of protein more economically, to date the protein produced by these systems has not been fully characterised. It is quite possible that transferrin made in *E. coli* may be incorrectly folded as transferrin, like lactoferrin, has a large number of disulphide bridges that will not be formed in *E. coli*. In contrast intact hLf produced using the expression vector pNUT and BHK cells has been fully characterised. Lactoferrin produced in BHK cells has been shown to be correctly processed and it has iron binding and release properties identical to those of native hLf suggesting that it has been correctly folded (Stowell, 1990; Stowell *et al.*, 1991). On the basis of these results it was reasonable to believe that a fragment representing the amino terminal half of human lactoferrin (Lf$_N$) would also be correctly processed by BHK cells. This has been proved to be the case.

Lf$_N$ produced in BHK cells has been characterised in several ways. The spectroscopic properties of both iron and copper saturated Lf$_N$ were found to be very similar to those of intact hLf. This suggested that the environment of the metal ion in the two proteins was similar and as a consequence it was likely that the amino terminal half of human lactoferrin expressed in BHK cells had been correctly folded. A slight difference in λ_{max} was observed. This could be accounted for

either by small changes in the vicinity of the iron site or by the fact that the λ_{max} for hLf is an average of two sites, one of which could have a lower λ_{max} as found for Lf_N and the other, in this case the C-lobe, could have a higher value. These two possibilities will only be resolved when the C-lobe is expressed on its own or when full-length lactoferrin is produced in which iron binding in the C-lobe has been ablated by mutagenesis, removing the contribution of this site from the λ_{max} of hLf. Although the properties of Lf_N and hLf were similar a striking difference was found in the pH at which iron was released. Lf_N released iron two pH units higher than did hLf suggesting either that structural differences distant from the iron binding site had occurred and these affected iron release, or that interactions between the two lobes contributed to the tight iron binding properties of hLf.

In an attempt to understand the changes responsible for the altered iron release properties of Lf_N the structure of iron saturated Lf_N (FeLf$_N$) was determined. Comparison of the structures of FeLf$_N$ and Fe$_2$Lf has shown that the folding pattern in the two N-lobe structures is highly conserved and that expression of Lf_N in BHK cells and enzymatic deglycosylation of Lf_N has not damaged the protein. The environment of the iron atom in FeLf$_N$ is essentially identical to that seen in Fe$_2$Lf confirming the results form the spectroscopic analysis of the two proteins. The one major difference between the two structures involves residues 315 to 327 at the back of the metal binding site. Instead of forming two helices as in Fe$_2$Lf these residues now form a helix and a β-strand. This changes the connections between the two domains from two β-strands and a helix to three β-strands. The altered environment at the back of the iron site due to the absence of the C-lobe and the conformational change in residues 315 to 327 has been discussed in detail in section B3.3. Perhaps the main point to make from this analysis is that the changes arising from the altered conformation and the different environment at the back of the iron site could quite conceivably affect the mechanism of domain opening and alter the equilibrium between the open and the closed forms of hLf. As a consequence it is believed that in combination these changes are responsible for the altered iron release properties found in Lf_N.

The structure of ApoLf$_N$ has also been solved and partially refined. Unlike ApoLf where the N-lobe was found to be wide open the structure of ApoLf$_N$ solved here is closed. A structure of an open form of Lf_N is required before the effect of the conformational changes, and the absence of the C-lobe, on domain opening can be assessed. It is not possible to model the open form of Lf_N from the structure of ApoLf as the site of the conformational change associated with domain opening in the N-lobe of intact hLf coincides with the position of the structural changes seen in FeLf$_N$.

Characterisation of the mutants made so far, although incomplete, has proved interesting. Firstly removal of the recognition site for the addition of an N-linked carbohydrate group to Lf_N affected both the level of expression and the spectroscopic properties of the protein. This suggests that glycosylation of Lf_N by BHK cells allows the protein to be correctly folded and processed. An important role for the carbohydrate groups attached to hLf in processing is also suggested by the double deglycosylation mutant which is believed to be glycosylated at a third, usually cryptic site, in the absence of the two favoured sites. This work has not been pursued but provides a possible system for defining the factors required for glycosylation and for assessing the role of the carbohydrate groups.

In the second class of mutants changes to residues directly involved in metal and anion binding were made. These studies have shown that single amino acid changes at either the metal (eg. D60S) or the anion binding site (eg. R121D) can have a dramatic affect on the spectroscopic and iron release properties of Lf_N. The difference in the UV-visible and ESR spectra suggests that the coordination of the iron atom is changed, and the increased pH of iron release suggests that the new coordination system is considerably less stable. Other mutations, for instance the conversion of arginine 121 to serine, did not have such a significant effect on the properties of the protein and the changes seen are quite subtle. However, until crystal structures of some of these mutants have been determined it is difficult to identify the exact changes involved.

If the results from the mutant and structural studies are combined they suggest that the tight and specific binding of iron, by lactoferrin, is dependent on both the local environment of the iron atom and the overall conformation of the protein. However, if the results obtained with sTf and sTf_N are combined with those of their lactoferrin counterparts it is possible to suggest that specific interactions between the two lobes of lactoferrin are responsible for the greater pH stability of intact Lf compared to intact sTf. This suggestion is based on the observation that Lf_N, sTf_N and sTf release iron at a similar pH but Lf binds iron much more tightly. The iron binding sites are very similar in sTf and Lf. In combination these results suggest that the coordination of the iron atom and the pattern of folding in each lobe gives a certain level of iron binding, equivalent to that seen in Lf_N, sTf_N and sTf, but that the interactions at positions distant to the metal binding site are responsible for the greater pH stability of Lf. On the basis of this study it is the interactions at the back of the iron site, between the two lobes, which contribute to this effect. Subsequent experiments may allow this theory to be tested.

2.2   Future Directions

In this investigation methods for producing, characterising and mutating the amino terminal half of human lactoferrin have been developed. The establishment of these protocols has greatly expanded the potential for analysing the structure and function of lactoferrin. Using these techniques some questions about the structure and function of lactoferrin have been answered but many more have been raised. In some cases the need for future experiments and the direction that these might take has already been discussed. However, it is my intention in the following sections to briefly highlight some questions which may be answered using other fragments of hLf or chimaeric proteins and questions which may be answered by the construction of more site specific mutants. Further means of analysing the recombinant proteins are also briefly discussed.

2.2.1   Suggestions for future constructs

Construction of an expression system for the C-terminal half of human lactoferrin

At the outset of this project the aim was to express and characterise both the N- and C-terminal halves of human lactoferrin. It was anticipated that an investigation of this type would allow the properties of the two iron binding sites in Lf to be compared. While the N-terminal half has been expressed at high levels, expression of the C-terminal half has not been achieved despite a considerable amount of effort. In the first construct made the cDNA encoding the signal peptide of Lf was joined to the cDNA for the C-terminal half, adjacent to the codon for alanine 340. Although high levels of Lf mRNA were detected in these cells no protein was detected either in the cells or in the media from the cells (B. Sheth, *pers. comm.*). It was postulated that the junction between the signal peptide and the mature protein was not recognised and as a consequence the protein remained in the cells and was rapidly degraded. In an attempt to overcome this possibility another construct was made in which the first seven amino acids at the 5' end of the C-terminal half were changed to match those found at the start of the N-terminal fragment. Unfortunately again no protein was detected in the cells or in the media from cells transfected with this construct (B. Sheth, *pers. comm.*).

The reason for the problems encountered in expressing these constructs is not known. The folding pattern found in the C-lobe of Lf is very similar to that seen in the N-lobe and it is surprising, given the high level expression of Lf$_N$, that the C-lobe is either unable to be processed or is highly unstable. One of the main aims in producing the C-lobe was to allow a comparison between the metal and anion

binding properties of the C-lobe with those found in the N-lobe. While production of the C-lobe would still be desirable perhaps a more feasible means of allowing this investigation to continue would be to construct an intact hLf mutant which can no longer bind iron in the N-lobe. A mutant of this type could probably be produced by site specific mutagenesis and would have the advantage of allowing the properties of the C-lobe to be evaluated in the absence of changes due to lobe separation. A similar strategy in which the iron-binding capacity of the C-lobe is ablated may prove useful for further analysing the properties of the N-lobe as it would allow changes due to lobe separation to be separated from those which occur in hLf but are masked due to the contribution from iron binding in the C-lobe.

Another approach to produce the C-lobe may be to try either a different construct, perhaps using a different signal peptide, or a different expression system. Several new expression vectors have been developed and are commercially available. These may prove useful.

Other iron binding fragments

The production of several other fragments of lactoferrin may allow other questions to be answered. Firstly it may be possible to identify the residues responsible for the different iron binding properties of Lf_N and Lf_{N30}, the proteolytically-derived N-terminal fragment of hLf (Legrand *et al.*, 1990). Lf_{N30} is 50% desaturated at pH 6.3 whereas Lf_N is not 50% desaturated until pH 5.8. The difference between the two fragments is that Lf_N is terminated at position 333 whereas Lf_{N30} only extends as far residue 281. It is therefore missing residues 290 - 305 which form a part of one wall of the cleft and residues 296, 301 and 303 form interdomain contacts. It is likely that the difference in the pH stability arises from the absence of these residues and the presence of residues 321 - 327 in Lf_N which form an additional β-strand across the back of the iron site deceasing the accessibility of some residues to solvent. The contribution from residues 281 - 305 and from residues 320 - 333 in stabilising iron binding could be assessed by constructing a mutant which only extended as faras Leu 320.

In the same study Legrand *et al* reported that an 18 kDa fragment which only included residues 91 - 253 had similar binding properties to Lf_{N30}. An 18 kDa fragment of ovotransferrin is also reported to be capable of binding iron but in this case the spectral properties of this fragment are different and the geometry at the iron site has probably been disrupted (Jhoti *et al.*, 1988). These findings are quite surprising and imply that iron can be bound in the absence of residues 1 - 90 and 253 - 280. A more detailed analysis using deletion mutants or a technique such as

alanine scanning (Cunningham *et al.*, 1990) may prove useful in defining residues/regions which affect the binding properties of lactoferrin. By systematically replacing residues with alanine it would also be possible to define those residues critical for allowing lactoferrin to fold and those critical for iron binding.

As a consequence of these findings it has also been suggested that iron is first bound to domain 2, after which domain 1 closes over providing the full complement of ligands. Although this is a likely scenario it is also possible that iron can bind to domain 1 but as this domain has not been isolated as a separate entity it has not been possible to examine its iron binding ability. A construct could be made in which the two regions of domain 1 are joined together by a short loop. If this protein folded correctly it would be possible to investigate the iron binding capacity of this fragment.

Chimaeric constructs

Many different chimaeric constructs could be made but perhaps one of the more interesting regions to investigate would be the role of residues 1 - 90, and in particular helix 1. This region has been suggested to have a role in the bactericidal properties of lactoferrin (Bellamy *et al.*, 1992), the receptor binding properties of lactoferrin (Rochard *et al.*, 1989; Legrand *et al.*, 1991) and the growth promoting properties of lactoferrin (Diamond *et al.*, 1984). Tomita *et al.* (1992) have made a synthetic peptide containing amino acids 18 to 40 and shown that this will kill bacteria demonstrating that these residues of Lf have bactericidal properties.

The role of this region in receptor binding could be investigated by using chimaeric constructs where regions of lactoferrin are replaced by the equivalent region of lactoferrin from another species or from transferrin. If, as reported, the lactoferrin receptors can distinguish between Lf and Tf; and between Lf from different species (Mazurier *et al.*, 1989), some of these changes would be expected to abolish receptor binding. This may in fact be a good means by which the existence of a specific receptor for lactoferrin can be demonstrated. In a similar manner it may be possible to construct a lactoferrin which binds to the sTf receptor.

Using these same constructs it may be possible to examine the growth promoting properties of lactoferrin and transferrin. While the growth of most cells is stimulated by transferrin a few cell lines are reported to be stimulated by lactoferrin (section 1.2.3); the amino terminal region of both proteins has been implicated in this function (Diamond *et al.*, 1984). If the two regions from Lf and sTf are

exchanged cells which were previously stimulated by sTf may now be stimulated by Lf and conversely those which were previously stimulated by Lf may now be stimulated by sTf. An analysis of this type would allow residues critical for defining specificity and for stimulating growth to be determined.

2.2.2 Suggestions for other site specific mutants

Glycosylation mutants

An important role for the carbohydrate groups attached to lactoferrin has been suggested by this study but a definitive role has not been assigned. In the case of N137A the spectroscopic properties of this protein suggest that there are subtle differences in the folding pattern. One possible reason for this difference is that the change from asparagine to alanine at position 137 may cause helix 5 to be extended as alanine is a helix favouring residue (O'Neil and DeGrado, 1990) and position 137 is at the end of a helix. Alternatively the absence of the carbohydrate group may be sufficient to affect the folding of Lf_N. It would be possible to distinguish between these two possibilities by making other site specific mutants at position 137. If the physical absence of the carbohydrate prevents correct folding changing 137 to another residue would not be expected to affect the properties, but if the presence of alanine is the critical feature the iron binding properties of Lf_N may be restored when Asn 137 is replaced by an amino acid other than alanine.

In the case of the double deglycosylation mutant of hLf the most obvious mutant to construct is that where the three recognition sites for the addition of N-linked carbohydrate groups have been removed. If the protein produced from this construct was not glycosylated it would identify Asn 623 as the site glycosylated in the double mutant (N137A, N478A) made here. A correctly folded but carbohydrate free form of hLf would allow the role of the carbohydrate groups of Lf to be investigated. In particular suggestions that glycosylation is required for iron binding (Legrand *et al.*, 1990) and for receptor binding (Davidson and Lonnerdal, 1988) could be investigated.

Melanotransferrin model

Progress has been made towards modelling the C-lobe binding site of melanotransferrin, which can no longer coordinate iron (Baker *et al.*, 1992), in Lf_N. In this study both Asp 60 and Arg 121 were changed to serine, the amino acid substitutions found in melanotransferrin, but while iron binding was weakened it was not abolished. This suggested that changes other than, or in addition to, the

change from Asp to Ser and from Arg to Ser are necessary to prevent iron binding. Two of the other changes which are most likely to be involved are the conversion of the equivalents of Thr 117 and Thr 122, in the C-lobe of melanotransferrin, to alanine and proline respectively. Thr 122 may be particularly important as it makes two interdomain hydrogen bonds. The affect of these other changes is currently being investigated.

Complete removal of the iron binding capacity from lactoferrin

A mutant of lactoferrin which had no capacity to bind iron would allow several of the other proposed functions of lactoferrin to be investigated. Firstly such a mutant would not be expected to have any bacteriostatic properties if this function is dependent on iron deprivation, but it should retain bactericidal properties if these are independent of iron binding. A mutant which could not bind iron would also allow the role of lactoferrin in myelopoiesis to be investigated. It has been reported (Broxmeyer *et al.*, 1980) that the ability of lactoferrin to control myelopoiesis is dependent on the protein being iron saturated. This has been somewhat controversial as Lf released from neutrophils is iron free and the amount of free iron in circulation is very low. It is possible that the apolactoferrin used in this study had been damaged. If iron saturation is actually required a mutant form of Lf which could not bind iron would not be expected to regulate myelopoiesis.

It is anticipated that a mutant form of hLf with no iron binding ability could be created by removing all of the iron binding ligands but it may only be necessary to remove a subset of the ligands. This work is currently being pursued.

Hinge movement

Several residues at the hinge region of all lactoferrins and transferrins sequenced so far are conserved. Notably Thr 90 is invariant and Val 250 and Pro 251 are almost invariant (Gerstein *et al.*, 1992). These residues may have a critical role in allowing the mechanism of domain opening and closing seen in lactoferrin to occur, and may help define the equilibrium between the open and closed forms of lactoferrin. It would be possible to evaluate the roles of these residues by creating site specific mutants.

Altered metal and anion binding capacity

Lastly but potentially most importantly for commercial applications it may be possible to design lactoferrin that has altered binding properties. If lactoferrin or

transferrin could be designed to specifically bind drugs or other active compounds it may be possible to target these compounds to cells using lactoferrin or transferrin as the carrier molecule. Another potential use of lactoferrin would be as a chelator of other metal ions in the body. If a lactoferrin which has an increased capacity to bind aluminium could be created it may be possible to use it to treat Alzheimers disease as the accumulation of aluminium is associated with the onset of this condition. Replacement of one or more of the oxygen ligands by histidine may produce a lactoferrin which preferentially binds copper. This may also be potentially useful.

2.2.3 Further crystallographic and analytical studies

Crystallographic studies

As already discussed a crystallographic analysis of the recombinant proteins is the only means by which the structural bases of any of the observed spectroscopic and functional changes can be fully assessed. Because of this it will be important to determine the structure of the mutants already created and to crystallise subsequent mutants. Small crystals of some of the mutants have been grown using conditions similar to those used to prepare crystals of Lf_N. It is anticipated that the majority of the mutants are likely to crystallise in the same space group and that the structure of Lf_N can be used to solve the structure of all subsequent mutants.

One of the most important structures now required is that of an open N-terminal fragment. This structure would allow the mechanism of domain opening and closing in Lf_N to be assessed and might help define the altered equilibrium between the open and closed forms of Lf_N. It may be possible to prepare crystals of Lf_N where the domains are open either by using different crystallisation conditions or by preparing the protein in a different way. For instance protein which has lost its iron due to decreased pH rather than due to deglycosylation may behave differently

To allow the properties of the mutants to be completely evaluated it would also be advantageous to crystallise them at low pH. This would be particularly important in the case of R121S where the λ_{max} has been found to differ with pH. This suggests that the coordination of the iron atom changes with pH but only when structures of R121S above pH 7.0 and below pH 6.5 are determined will the structural changes involved be resolved.

Functional studies

Finally to allow the complete characterisation of lactoferrin mutants other methods of assessing the properties of the mutant proteins are required. In particular it would be useful if the strength of iron binding could be measured. Presently the affect of pH on iron binding is monitored but this does not truly reflect the strength of binding. Other methods for assessing the role of lactoferrin in the cell would also be useful because ultimately the goal must be to determine the function of lactoferrin *in vivo*. In this respect an *in vitro* assay for measuring receptor binding, if a receptor does actually exist, is required. A universally accepted method of assessing the bacteriostatic and bactericidal properties of lactoferrin would also be useful.

!!! THE END !!!

REFERENCES

Ainscough, E,W., Brodie, A.M. & Plowman. J.E. (1979) *Inorganica Chimica Acta* 33, 149-153

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

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

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

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

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

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

Akin, D.T., Lu, M.Q., Lu, S.J., Kendall, S. & Arnold, R.R. (1992) in *Proceedings of the International Symposium on Lactoferrin Structure and Function,* in press

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

Ambruso, D.R. & Johnston, R.B. (1981) *J. Clin. Invest.* 67, 352-360

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

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

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

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

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

Baggiolini, M., DeDuve, C., Masson, P.L. & 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. & Watson, J.L. (1988) *Biochemistry* **27**, 5804-5812

Baker, E.N. (1980) *J. Mol. Biol.* **141**, 441-484

Baker, E.N. (1988) *J. Mol. Biol.* **203**, 1071-1095

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

Baker, E.N. & Dodson, E.J. (1980) *Acta Cryst.* **A36**, 559-572

Baker, E.N. & Hubbard, R.E. (1984) *Prog. Biophys. Molec. Biol.* **44**, 97-179

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

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

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

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

Baldwin, D.A., Jenny, E.R. & Aisen, P. (1984) *J. Biol. Chem.* **259**, 13391-13394

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

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

Bali, P.K., Zak, O. & Aisen, P. (1991) *Biochemistry* **30**, 324-328

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

Bannister, J.V., Bannister, W.H., Hill, H.A.O. & Thornalley, P.J. (1982) *Biochim. Biophys. Acta* **715**, 116-120

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

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

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

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

Birgens, H.S., Hansen, N.E., Karle, H. & Kristensen, L.O. (1983) *Scand. J. Haematol.* **54**, 383-391

Bluard-Deconinck, J-M., Williams, J., Evans, R.W., van Snick, J., Osinski, P.A. & Masson, P.L. (1978) *Biochem. J.* **171**, 321-327

Blundell, T.L. & Johnson, L.N. (1976) Protein Crystallography (Horecker, B., Kaplan, N.O., Marmur, J. & Schenaga, H.A., ed.) Academic Press

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

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

Britigan, B.E., Hassett, D.J., Rosen, G.M., Hamill, D.R. & Cohen, M.S. (1989) *Biochem. J.* **264**, 447-455

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

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

Brock, J.H. (1981) *Immunology* **57**, 105-110

Brock, J.H. (1985) in *Metalloproteins* Part II (Harrison, P.M., ed.) pp 183-262 MacMillan Press, London

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

Broxmeyer, H.E. & Ralph, P (1977) *Cancer Research* **37**, 3578-3584

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

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

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

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

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

Chasteen, N.D., (1983) *Adv. Inorg. Biochem.* **5**, 201-233

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.J. (1979) *Biochemistry* **18**, 5294-5299

Crowther, R.A. (1972) in *The Molecular Replacement Method* (Rossman, M.G., ed), pp. 173-178, Gordon and Breach, New York

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

Cunningham, B.C., Henner, D.J. & Wells, J.A. (1990) *Science* **247**, 1461-1465

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

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

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

Diamond, A., Devine, J.M. & Cooper, G.M. (1984) *Science* **225**, 516-519

Dobson, C.M. (1990) *Nature* **348**, 198-199

Dube, S., Fisher, J.W. & Powell, J.S. (1988) *J. Biol. Chem.* **263**, 17516-17521

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

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

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

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

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

Garratt, R.C. & Jhoti, H. (1992) *FEBS Letters* **305**, 55-61

Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M. & Chothia, C. (1992) *J. Mol. Biol. submitted*

Grigera, P.R., Mathieu, M.E. & Wagner, R.R. (1991) *Virology* **180**, 1-9

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

Halliwell, B. & Gutteridge, M.C. (1990) *Arch. Biochem. Biophys.* **280**, 1-8

Hanson, J.C., Watenpaugh, K.D., Sieker, L. & Jensen, L.H. (1979) *Acta Cryst.* A35, 616-621

Harris, W.R. (1983) *Biochemistry* **22**, 3920-3926

Harris, W.R. (1985) *Biochemistry* **24**, 7412-7418

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

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

Hecht, H.J., Szardenings, M., Collins, J. & Schoburg, D. (1992) *J. Mol. Biol.* **225**, 1095-1103

Hekman, A. (1971) *Biochim. Biophys. Acta* **251**, 380-387

Horst, M., Harth, N. & Hasilik, A. (1991) *J. Biol. Chem.* **266**, 13914-13916

Huebers, H., Huebers, E., Csiba, E. & Finch, C.A. (1978) *J. Clin. Invest* **62**, 944-951

Huebers, H., Huebers, E., Csiba, E. & Finch, C.A. (1984) *American J. Physiol.* **247**, R280-R283

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

Huber, R. & Bode, W. (1978) *Acc. Chem. Res.* **11**, 114-122

Ikeda, R.A., Bowman, B.H., Yang, F. & Lokey, L.K. (1992) *Gene* **117**, 265-269

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

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

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

Johnston, J.J., Rintels, P., Chung, J., Sather, J., Benz, E.J. & Berliner, N. (1992) *Blood* **79**, 2998-3006

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

Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) *Acta Cryst.* **A47**, 110-119

Karlinsey, J., Stamatoyannopoulos, G. & Enver, T. (1989) *Anal. Biochem.* **180**, 303-306

Keung, W-M., Azari, P. & Philips, J.L. (1982) *J. Biol. Chem.* **257**, 1177-1183

Kojima, N. & Bates, G.W. (1979) *J. Biol. Chem.* **254**, 8847-8854

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

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

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

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

Lash, J.A., Coates, T.D., Latuze, J., Baehner, R.L. & Boxer, L.A. (1983) *Blood* **61**, 885-888

Lattman, E. (1985) *Methods in Enzymology* **115**, 55-77

Leatherbarrow, R.J. & Fersht, A.R. (1986) *Protein Engineering* **1**, 7-16

Lee, B. & Richards, F.M. (1971) *J. Mol. Biol.* **55**, 379-400

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

Legrand, D., Mazurier, J., Aubert, J-P., Loucheux-Lefebvre, M-H., Montreuil, J. & Spik, G. (1986) *Biochem. J.* **236**, 839-844

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

Legrand, D., Mazurier, J., Maes, P., Rochard, E., Montreuil, J. & Spik, G. (1991) *Biochem. J.* **276**, 733-738

Lineback-Zins, J. & Brew, K. (1980) *J. Biol. Chem.* **255**, 708-713

Liu, Y. & Teng, C.T. (1991) *J. Biol. Chem.* **266**, 21880-21885

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

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

Luzzati, P.V. (1952) *Acta Cryst.* **5**, 802-811

Lydon, J.P., O Malley, B.R., Saucedo, O., Lee, T., Headon, D.R. & Conneely, O.M. (1992) *Biochim. Biophys. Acta*, **1132**, 97-99

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

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

Masamune, Y. & Richardson, C.C. (1971) *J. Biol. Chem.* **246**, 2692-2701

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

Mason, S. (1976) *J. Clin. Path.* **30**, 541-546

Masson, P.L., Heremans, J.F. & Dive, C.H. (1966) *Clin. Chim. Acta*, **14**, 735-739

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

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

Matzuk, M.M. & Boime, I. (1988) *J. Cell Biol.* **106**, 1049-1059

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

Mazurier, J., Lhoste, J-M., Montreuil, J. & Spik, G (1983) *Biochim. Biophys. Acta* **745**, 44-49

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

McPherson, A. (1985) *Methods in Enzymology* **114**, 112-120

McPherson, A. (1985) *Methods in Enzymology* **114**, 120-125

McPherson, A. (1985) *Methods in Enzymology* **114**, 125-127

McPherson, A. (1990) *Eur. J. Biochem* **189** 1-23

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

Mead, P.E. (1992) *Ph.D. thesis*

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

Moguilevsky, N., Retegui, L.A. & Masson, P.L. (1985) *Biochem. J.* **229**, 353-359

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

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

Morris, A-L., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. (1992) *PROTEINS* **12**, 345-364

Moskaitis, J.E., Pastori, R.L. & Schoenberg, D.R. (1990) *Nucleic Acids Research* **18**, 6135

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

Norris, G.E., Anderson, B.F. & Baker, E.N. (1991) *Acta Cryst.* **B47**, 998-1004

North, A.C.T., Phillips, D.C. & Mathews, F.S. (1968) *Acta Cryst.* **A24**, 351-359

O'Neil, K.T. & DeGrado, W.F. (1990) *Science* **250**, 646-651

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

Padda, J.S. & Schryvers, A.B. (1990) *Infection & Immunity* **58**, 2972-2976

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

Panella, T.J., Liu, Y., Huang, A.T. & Teng, C.T. (1991) *Cancer Research* **51**, 3037-3043

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

Parry, R.M. & Brown, E.M. (1974) in *Adv. in Exp. Med. and Biol.* **48**, (Freidman, M., ed.) pp 141-160 Plenum Press, New York

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

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

Powell, M.J. & Ogden, J.E. (1990) *Nucleic Acids Research* **18**, 4013

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

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

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

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

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

Rainard, P. (1986) *Vet. Microbiology* **11**, 103-115

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

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

Rehse, P.H. & Zubak, V. (1992) *High Energy Physics Lab. Manual*

Rey, M.W., Woloshuk, S.L., deBoer, H.A. & Pieper, F.R. (1990) *Nucleic Acids Research* **18**, 5288

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

Riordan, J.F., McElvany, K.D. & Borders, C.L. (1977) *Science* **195**, 884-885

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

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

Roiron, D., Amouric, M., Marvaldi, J. & Figarella, C. (1989) *Eur. J. Biochem.* **186**, 367-373

Roquet, F., Declercq, J-P., Tinant, B., Rambaud, J. & Parello, J. (1992) *J. Mol. Biol.* **223**, 705-720

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

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

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

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

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

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

Samson, R.R., Mirtle, C. & McClelland, D.B.L. (1980) *Acta Paediatr. Scand.* **59**, 517-523

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

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

Schlabach, M.R. & Bates, G.W. (1975) *J. Biol. Chem.* **250** 21182-2188

Schryvers, A.B. & Morris, L.J. (1988) *Infection & Immunity* **56**, 1144-1149

Shepard, W.E.B. (1991) *Ph.D. thesis*

Shirsat, N.V., Bittenbender, S., Kreider, B.L. & Rovera, G. (1992) *Gene* **110**, 229-234

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

Smith, C.A. (1992) *Ph.D. thesis*

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

Roquet, F., Declercq, J-P., Tinant, B., Rambaud, J. & Parello, J. (1992) *J. Mol. Biol.* **223**, 705-720

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

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

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

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

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

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

Samson, R.R., Mirtle, C. & McClelland, D.B.L. (1980) *Acta Paediatr. Scand.* **59**, 517-523

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

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

Schlabach, M.R. & Bates, G.W. (1975) *J. Biol. Chem.* **250** 2 1182-2188

Schryvers, A.B. & Morris, L.J. (1988) *Infection & Immunity* **56**, 1144-1149

Shepard, W.E.B. (1991) *Ph.D. thesis*

Shirsat, N.V., Bittenbender, S., Kreider, B.L. & Rovera, G. (1992) *Gene* **110**, 229-234

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

Smith, C.A. (1992) *Ph.D. thesis*

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

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

Smith, K.L. & Schanbacher, F.L. (1977) *J. Amer. Vet. Med. Assn.* **170**, 1224-1227

Smith, E.L., Hill, R.L., Lehman, I.R., Lefkowitz, R.J., Handler, P. & White, A. (1983) Principles of biochemistry. Mammalian Biochemistry (7th edition) McGraw-Hill International Book Company. Japan

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

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

Stout, G.H. & Jensen, L.H. (1989) *X-ray Structure Determination* Wiley Interscience

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

Stowell, K.M. (1990) *Ph.D. thesis*

Teng, C.T., Pentecost, B.T., Chen, Y.H., Newbold, R.R., Eddy, E.M. & McLachlan, J.A. (1989) *Endocrinology* **124**, 992-999

Teng, C.T. (1992) in *Proceedings of the First International Symposium on Lactoferrin Structure and Functionin,* in press

Tindall, K.R. & Kunkel, T.A. (1988) *Biochemistry* **27**, 6008-6013

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

Tomita, M., Takase, M., Wakabayashi, H. & Bellamy, W. (1992) in *Proceedings of the First International Symposium on Lactoferrin Structure and Function,* in press

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

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

van Snick, J.L. & Masson, P.L. (1976) *J. Exp. Med.* **144**, 1568-1580

Wang, Y., Chen, J. Luo, Y., Funk, W.D., Mason, A.B., Woodworth, R.C., MacGillivray, R.T.A. & Brayer, G.D. (1992) *J. Mol. Biol.* **227**, 575-576

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

Welty, F.K., Smith, K.L. & Schanbacher, F.L. (1976) *J. Dairy Science* **59**, 224-231

Wharton, B.A., Balmer, S.E. & Scott, P.M. (1992) in *Proceedings of the First International Symposium on Lactoferrin Structure and Function,* in press

Williams, J. (1982) *Trends in Biochemical Sciences* **8**, 394-397

Williams, S.C. & Woodworth, R.C. (1973) *J. Biol. Chem.* **248**, 5848-5853

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

Winterbourn, C.C. (1983) *Biochem. J.* **210**, 15-19

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

Woodworth, R.C., Mason, A.B., Funk, W.D. & MacGillivray, R.T.A. (1991) *Biochemistry* **30**, 10824-10829

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

APPENDIX 1

DNA AND PROTEIN SEQUENCE OF hLf

The DNA and protein sequence of human lactoferrin used in ths study is shown below. The amino acid corresponding to each codon is shown under the first base of each triplet. Position 1 of the DNA sequences is given to the A of the ATG start codon. The protein sequence is numbered such that the amino acid numbers of the mature protein, used by Anderson *et al.* (1989), are preserved. This means that it is necessary to call the third arginine of the mature protein 3A and the signal peptide is not numbered (see below). The signal peptide is shown in italics and the last amino acid in Lf$_N$, residue 333, is shown in bold.

```
ATGAAACTTGTCTTCCTCGTCCTGCTGTTCCTCGGGGCCCTCGGACTGTGTCTGGCTGGC 60
M   K   L   V   F   L   V   L   L   F   L   G   A   L   G   L   C   L   A   G   1

CGTAGGAGAAGGAGTGTTCAGTGGTGCGCCGTATCCCAACCCGAGGCCACAAAATGCTTC 120
R   R   R   R   S   V   Q   W   C   A   V   S   Q   P   E   A   T   K   C   F   20
    3   3A  4

CAATGGCAAAGGAATATGAGAAGAGTGCGTGGCCCTCCTGTCAGCTGCATAAAGAGAGAC 180
Q   W   Q   R   N   M   R   R   V   R   G   P   P   V   S   C   I   K   R   D   40

TCCCCCATCCAGTGTATCCAGGCCATTGCGGAAAACAGGGCCGATGCTGTGACCCTTGAT 240
S   P   I   Q   C   I   Q   A   I   A   E   N   R   A   D   A   V   T   L   D   60

GGTGGTTTCATATACGAGGCAGGCCTGGCCCCCTACAAACTGCGACCTGTAGCGGCGGAA 300
G   G   F   I   Y   E   A   G   L   A   P   Y   K   L   R   P   V   A   A   E   80

GTCTACGGGACCGAAAGACAGCCACGAACTCACTATTATGCCGTGGCTGTGGTGAAGAAG 360
V   Y   G   T   E   R   Q   P   R   T   H   Y   Y   A   V   A   V   V   K   K   100

GGCGGCAGCTTTCAGCTGAACGAACTGCAAGGTCTGAAGTCCTGCCACACAGGCCTTCGC 420
G   G   S   F   Q   L   N   E   L   Q   G   L   K   S   C   H   T   G   L   R   120

AGGACCGCTGGATGGAATGTCCCTATAGGGACACTTCGTCCATTCTTGAATTGGACGGGT 480
R   T   A   G   W   N   V   P   I   G   T   L   R   P   F   L   N   W   T   G   140

CCACCTGAGCCCATTGAGGCAGCTGTGGCCAGGTTCTTCTCAGCCAGCTGTGTTCCCGGT 540
P   P   E   P   I   E   A   A   V   A   R   F   F   S   A   S   C   V   P   G   160

GCAGATAAAGGACAGTTCCCCAACCTGTGTCGCCTGTGTGCGGGGACAGGGGAAAACAAA 600
A   D   K   G   Q   F   P   N   L   C   R   L   C   A   G   T   G   E   N   K   180

TGTGCCTTCTCCTCCCAGGAACCGTACTTCAGCTACTCTGGTGCCTTCAAGTGTCTGAGA 660
C   A   F   S   S   Q   E   P   Y   F   S   Y   S   G   A   F   K   C   L   R   200
```

```
GACGGGGCTGGAGACGTGGCTTTTATCAGAGAGAGCACAGTGTTTGAGGACCTGTCAGAC 720
 D  G  A  G  D  V  A  F  I  R  E  S  T  V  F  E  D  L  S  D  220

GAGGCTGAAAGGGACGAGTATGAGTTACTCTGCCCAGACAACACTCGGAAGCCAGTGGAC 780
 E  A  E  R  D  E  Y  E  L  L  C  P  D  N  T  R  K  P  V  D  240

AAGTTCAAAGACTGCCATCTGGCCCGGGTCCCTTCTCATGCCGTTGTGGCACGAAGTGTG 840
 K  F  K  D  C  H  L  A  R  V  P  S  H  A  V  V  A  R  S  V  260

AATGGCAAGGAGGATGCCATCTGGAATCTTCTCCGCCAGGCACAGGAAAAGTTTGGAAAG 900
 N  G  K  E  D  A  I  W  N  L  L  R  Q  A  Q  E  K  F  G  K  280

GACAAGTCACCGAAATTCCAGCTCTTTGGCTCCCCTAGTGGGCAGAAAGATCTGCTGTTC 960
 D  K  S  P  K  F  Q  L  F  G  S  P  S  G  Q  K  D  L  L  F  300

AAGGACTCTGCCATTGGGTTTTCGAGGGTGCCCCCGAGGATAGATTCTGGGCTGTACCTT 1020
 K  D  S  A  I  G  F  S  R  V  P  P  R  I  D  S  G  L  Y  L  320

GGCTCCGGCTACTTCACTGCCATCCAGAACTTGAGGAAAAGTGAGGAGGAAGTGGCTGCC 1080
 G  S  G  Y  F  T  A  I  Q  N  L  R  K  S  E  E  E  V  A  A  340

CGGCGTGCGCGGGTTGTGTGGTGTGCGGTGGGCGAGCAGGAGCTGCGCAAGTGTAACCAG 1140
 R  R  A  R  V  V  W  C  A  V  G  E  Q  E  L  R  K  C  N  Q  360

TGGAGTGGCTTGAGCGAAGGCAGCGTGACCTGCTCCTCGGCCTCCACCACAGAGGACTGC 1200
 W  S  G  L  S  E  G  S  V  T  C  S  S  A  S  T  T  E  D  C  380

ATCGCCCTGGTGCTGAAAGGAGAAGCTGATGCCATGAGTTTGGATGGAGGGTATGTGTAC 1260
 I  A  L  V  L  K  G  E  A  D  A  M  S  L  D  G  G  Y  V  Y  400

ACTGCAGGCAAATGTGGTTTGGTGCCTGTCCTGGCAGAGAACTACAAATCCCAACAAAGC 1320
 T  A  G  K  C  G  L  V  P  V  L  A  E  N  Y  K  S  Q  Q  S  420

AGTGACCCTGATCCTAACTGTGTGGATAGACCTGTGGAAGGATATCTTGCTGTGGCGGTG 1380
 S  D  P  D  P  N  C  V  D  R  P  V  E  G  Y  L  A  V  A  V  440

GTTAGGAGATCAGACACTAGCCTTACCTGGAACTCTGTGAAAGGCAAGAAGTCCTGCCAC 1440
 V  R  R  S  D  T  S  L  T  W  N  S  V  K  G  K  K  S  C  H  460

ACCGCCGTGGACAGGACTGCAGGCTGGAATATCCCCATGGGCCTGCTCTTCAACCAGACG 1500
 T  A  V  D  R  T  A  G  W  N  I  P  M  G  L  L  F  N  Q  T  480

GGCTCCTGCAAATTTGATGAATATTTCAGTCAAAGCTGTGCCCCTGGGTCTGACCCGAGA 1560
 G  S  C  K  F  D  E  Y  F  S  Q  S  C  A  P  G  S  D  P  R  500
```

```
TCTAATCTCTGTGCTCTGTGTATTGGCGACGAGCAGGGTGAGAATAAGTGCGTGCCCAAC 1620
 S   N   L   C   A   L   C   I   G   D   E   Q   G   E   N   K   C   V   P   N   520

AGCAATGAGAGATACTACGGCTACACTGGGGCTTTCCGGTGCCTGGCTGAGGATGCTGGA 1680
 S   N   E   R   Y   Y   G   Y   T   G   A   F   R   C   L   A   E   D   A   G   540

GACGTTGCATTTGTGAAAGATGTCACTGTCTTGCAGAACACTGATGGAAATAACAATGAG 1740
 D   V   A   F   V   K   D   V   T   V   L   Q   N   T   D   G   N   N   E   560

GCATGGGCTAAGGATTTGAAGCTGGCAGACTTTGCGCTGCTGTGCCTCGATGGCAAACGG 1800
 A   W   A   K   D   L   K   L   A   D   F   A   L   L   C   L   D   G   K   R   580

AAGCCTGTGACTGAGGCTAGAAGCTGCCATCTTGCCATGGCCCCGAATCATGCCGTGGTG 1860
 K   P   V   T   E   A   R   S   C   H   L   A   M   A   P   N   H   A   V   V   600

TCTCGGATGGATAAGGTGGAACGCCTGAAACAGGTGCTGCTCCACCAACAGGCTAAATTT 1920
 S   R   M   D   K   V   E   R   L   K   Q   V   L   L   H   Q   Q   A   K   F   620

GGGAGAAATGGATCTGACTGCCCGGACAAGTTTTGCTTATTCCAGTCTGAAACCAAGAAC 1980
 G   R   N   G   S   D   C   P   D   K   F   C   L   F   Q   S   E   T   K   N   640

CTTCTGTTCAATGACAACACTGAGTGTCTGGCCAGACTCCATGGCAAAACAACATATGAA 2040
 L   L   F   N   D   N   T   E   C   L   A   R   L   H   G   K   T   T   Y   E   660

AAATATTTGGGACCACAGTATGTCGCAGGCATTACTAATCTGAAAAAGTGCTCAACCTCC 2100
 K   Y   L   G   P   Q   Y   V   A   G   I   T   N   L   K   K   C   S   T   S   680

CCCCTCCTGGAAGCCTGTGAATTCCTCAGGAAGTAA 2136
 P   L   L   E   A   C   E   F   L   R   K   *   691
```

APPENDIX   2

INTERNAL   hLf   SEQUENCING   PRIMERS

Sequence and position of the 7 oligonucleotide primers used to determine the sequence of the amino terminal half of human lactoferrin. The bases are numbered using the sequence of hLf given in appendix 1. Oligonucleotides CLD3 - CLD6 are complementary to the bottom strand and were used to obtain the sequence of the top strand while CLD11 - CLD13 are complementary to the top strand and were used to obatin the sequence of the bottom strand.

| Oligo name	Position	Sequence
CLD3	196 - 213	ATCCAGGCCATTGCGGAA
CLD4	416 - 433	TTCGCAGGACCGCTGGAT
CLD5	635 - 653	CTCTGGTGCCTTCAAGTG
CLD6	856 - 873	GCCATCTGGAATCTTCTC
CLD11	801 - 785	CAGATGGCAGTCTTTGA
CLD12	510 - 526	TGGCTGAGAAGAACCTG
CLD13	251 - 235	ATGAAACCACCATCAAG

APPENDIX 3

CONSTRUCTION OF pGEM:hLf3

pGEM:hLf3 has rather a complex family tree!. The coding sequence was produced by PCR and the resultant clone was called pGEM:hLf1 and has been described elsewhere (Stowell, 1990; Stowell *et al.*, 1991). During the sequence analysis of this clone it had been cloned into M13mp18 and it was from this clone that the cDNA was obtained for pGEM:hLf3. The cloning steps involved are shown in the following diagram. The restriction enzyme sites derived from pGEM1 are shown in bold, those from M13 are shown in plain text and those from pGEM:3Zf(+) are shown in italics. Ths cDNA for lactoferrin is represented by a triangle and the orientation of the cDNA is indicated. Note the aim of this construct was to obtain a clone which had the C-terminal end adjacent to the EcoRI site in the multiple cloning site.

pGEM:hLf1

HindIII--PstI....AccI--HincII--XbaI--BamHI....EcoRI

C   N

This cDNA for lactoferrin was excised using BamHI and HindIII and then made
blunt ended before being cloned into SmaI cut M13mp18

HindIII....BamHI--**PstI--SalI--AccI--HincII**--XbaI--KpnI--SstI....EcoRI

C   N

The cDNA for lactoferrin was then excised from this clone using SstI and BamHI. This insert
was cloned into pGEM:3Zf(+) which had been cut with SstI and BamHI.

pGEM:hLf3

EcoRI....*SstI*--KpnI--XbaI--**HincII--AccI--SalI--PstI**--*BamHI*.....*HindIII*

C   N

APPENDIX   4

MUTAGENIC   OLIGONUCLEOTIDES

The sequence and position of the 7 muatagenic primers is given below. The bases coding for the mutated residue are shown in bold and the position of the first base is indicated.

| Mutant made	Position of the first base	Sequence
D60T	227	CTGTGACCCTT**ACT**GGTGGTTTCA
D60S	227	CTGTGACCCTT**AGT**GGTGGTTTCA
R121S	410	CAGGCCTTCGC**AGT**ACCGCTGGAT
R121D	410	CAGGCCTTCGC**GAC**ACCGCTGGAT
R121E	410	CAGGCCTTCGC**GAG**ACCGCTGGAT
N137A	457	CGTCCATTCTTGGC**TTG**GACGGGTCC
N478A	1480	GGCCTGCTCTTCGC**CCA**GACGGGCTC

Once the mutagenic reactions had been carried out the presence of the mutation was determined using one of the internal sequencing primers (Appendix 2). Note another internal sequencing primer CLD8 was used to sequence across the mutation at position 478. The sequence of this primer was 5' GAACTACAAATCCCAACA 3'. The first base of this oligonucleotide was 1299.

APPENDIX 5

CONSTRUCTION OF pGEM:hLf4

pGEM:hLf4 was used to make the full legth double deglycosylation mutant. Like pGEM:hLf3 this clone was derived from pGEM:hLf1 except that in this instance the cDNA for lactoferrin was excised from pGEM:hLf1 with BamHI and HindIII. This insert was than cloned into pGEM:3Zf(+) which had been cut with the same enzymes. The construction of this clone is represented in the following scheme. This plasmid was created to allow production of ssDNA which could be used with the mutagenic primers already obtained.

pGEM:hLf1

HindIII--PstI--SalI--AccI--HincII--XbaI--BamHI--AvaI--SmaI--SacI--EcoRI

C   N

BamHI / HindIII dgested and cloned into pGEM:3Zf(+) which had been digested with BamHI and HindIII.

pGEM:hLf4

HindIII--PstI--SalI--AccI--HincII--XbaI--BamHI--SmaI--AvaI--KpnI--SacI--EcoRI

C   N

APPENDIX 7

PROTEIN SEQUENCE OF Lf$_N$

The amino acid sequence of Lf$_N$, using the single letter code, is shown below. The ligands involved in metal binding, Asp 60, Tyr 92, Tyr 192 and His 253 are shown in bold

```
G  R  R  R  R  S  V  Q  W  C  A  V  S  Q  P  E  A  T  K  C  F   20

Q  W  Q  R  N  M  R  R  V  R  G  P  P  V  S  C  I  K  R  D      40

S  P  I  Q  C  I  Q  A  I  A  E  N  R  A  D  A  V  T  L  D      60

G  G  F  I  Y  E  A  G  L  A  P  Y  K  L  R  P  V  A  A  E      80

V  Y  G  T  E  R  Q  P  R  T  H  Y  Y  A  V  A  V  V  K  K     100

G  G  S  F  Q  L  N  E  L  Q  G  L  K  S  C  H  T  G  L  R     120

R  T  A  G  W  N  V  P  I  G  T  L  R  P  F  L  N  W  T  G     140

P  P  E  P  I  E  A  A  V  A  R  F  F  S  A  S  C  V  P  G     160

A  D  K  G  Q  F  P  N  L  C  R  L  C  A  G  T  G  E  N  K     180

C  A  F  S  S  Q  E  P  Y  F  S  Y  S  G  A  F  K  C  L  R     200

D  G  A  G  D  V  A  F  I  R  E  S  T  V  F  E  D  L  S  D     220

E  A  E  R  D  E  Y  E  L  L  C  P  D  N  T  R  K  P  V  D     240

K  F  K  D  C  H  L  A  R  V  P  S  H  A  V  V  A  R  S  V     260

N  G  K  E  D  A  I  W  N  L  L  R  Q  A  Q  E  K  F  G  K     280

D  K  S  P  K  F  Q  L  F  G  S  P  S  G  Q  K  D  L  L  F     300

K  D  S  A  I  G  F  S  R  V  P  P  R  I  D  S  G  L  Y  L     320

G  S  G  Y  F  T  A  I  Q  N  L  R  K
```